Prediction and Comparison of Coiled-Coil Proteins in Multiple Genomes

Akiyasu C. Yoshizawa[1] acyshzw@kuicr.kyoto-u.ac.jp
Shujiro Okuda[1] okuda@kuicr.kyoto-u.ac.jp
Masumi Itoh[1] itoh@kuicr.kyoto-u.ac.jp
Toshiaki Katayama[2] ktym@hgc.jp
Shuichi Kawashima[1] shuichi@kuicr.kyoto-u.ac.jp
Minoru Kanehisa[1,2] kanehisa@kuicr.kyoto-u.ac.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University
Gokasho, Uji, Kyoto 611-0011, Japan
[2] Human Genome Center, Institute of Medical Science, University of Tokyo
4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Keywords: coiled-coil, prediction, genome, comparative genomics

1 Introduction

The coiled-coil structure is a typical hyper-secondary structure formed by the mutual intertwining of multiple alpha-helices, and is predicted as the structure of keratin of the intermediate filament by Crick in 1952 [2]. In 1988, the leucine zipper in the structure of transcription factors was found to be a form of the coiled-coil, and in succeeding years, a variety of biological applications of coiled-coils has been discovered: For example, structural proteins such as desmosomal proteins, motor proteins such as myosin, and some proteins related to the dynamic structure of membranes such as SNAREs. In ordinary case, coiled-coil molecules assemble with other coiled-coil molecules, so this may be the simple model of the general case of the receptor-ligand binding.

Before now, several algorithms for prediction of coiled-coils were proposed: The first algorithm was proposed by Parry (1982)[6] and by Lupas (1991)[4], using the periodic occurrence of hydrophobic amino acids in the sequence of a coiled-coil. After that, pairwise correlation algorithm, which is the improvement of the scoring method, was proposed by Berger and Kim (1995)[1], and an another prediction method using hidden Markov model was proposed by Delorenzi and Speed (2002)[3].

Such a structure as to contribute to the broad range of biological phenomena and as to be predicted with fair accuracy is suitable for the comparative genomic analysis, though only few studies were published in which the coiled-coil prediction and the analysis was performed against the whole genome [5]. Hence, we collected complete genomic sequences of 18 eukaryotic and 147 prokaryotic organisms, and partial genomic sequences of 3 eukaryotic organisms, and predicted the coiled-coil molecules.

2 Method and Results

2.1 Prediction Method and Genomic Data

1. For the prediction of coiled-coil domains, we used three programs, COILS2.2 [4], Multicoil [7], and Marcoil [3], as the implementation of Lupas algorithm, of pairwise-correlation algorithm, and of HMM, respectively. We adopted the prediction result of Marcoil (threshold = 0.4), and removing the false positives by using COILS (threshold = 0.3, window size = 28, with/without w option) and MultiCoil (threshold = 0.1, window size = 21 and 28).

2. All the amino acid sequences were obtained from KEGG/GENES and DGENES (Release 29.0).

2.2 Results

Finally, 48,667 sequences from KEGG/GENES and 64,584 sequences from KEGG/DGENES were predicted as molecules containing at least one coiled-coil domain. Figure 1 shows the number of coiled-coil protein in each organism and the ratio to the whole genome.

3 Discussion

In the eukaryotic organisms, both the number and the ratio of coiled-coil molecules show the sudden increase. Especially in *P. falciparum*, more than 30% of its proteome are the coiled-coil proteins, and about a half of them show the sequence similarities each other.

In the prokaryotic organisms, the ratio is different with each taxon, diverging from 4% to more than 10%. The numbers also diverge from about 50 to 400. However, no obvious tendency is observed through the evolution.

The numbers of coiled-coils even in the closely related organisms may be different. For example, each of the strains of non-pathogenic *E.coli* has about 200 coiled-coil molecules, though two strains of O-157 have about 300 molecules.

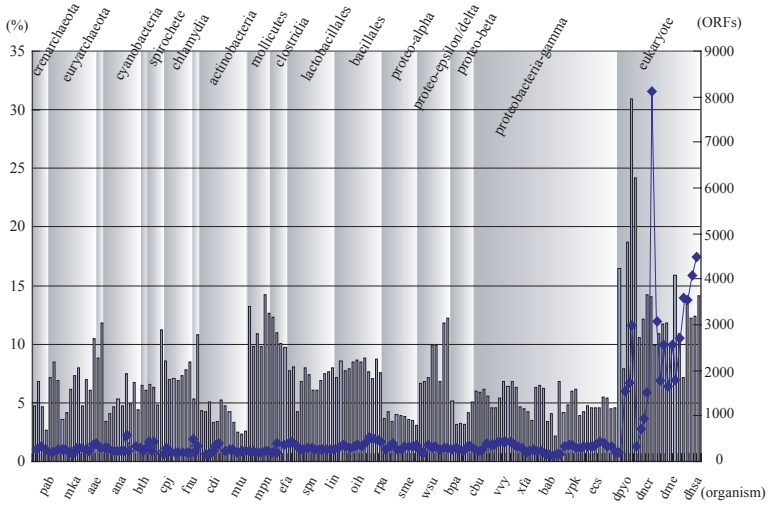

Figure 1: The number (line) and the ratio to the genome (vertical bars) of the coiled-coil molecules. The vertical lines in the background mean the boundaries of the taxons.

Acknowledgements

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

References

[1] Berger, B., Wilson, D.B., Wolf, E., Tonchev, T., Milla, M. and Kim, P.S., Predicting coiled coils by use of pairwise residue correlations, *Proc. Natl. Acad. Sci. USA*, **92**:8259-8263, 1995.

[2] Crick, F.H.C., Is alpha-keratin a coiled coil?, *Nature*, **170**:882-883, 1952.

[3] Delorenzi, M. and Speed, T., An HMM model for coiled-coil domains and a comparison with PSSM-based predictions*, Bioinformatics*, **18**:617-625, 2002.

[4] Lupas, A., Van Dyke, M. and Stock, J., Predicting coiled coils from protein sequences, *Science*, **252**: 1162-1164, 1991.

[5] Newman, J.R.S., Wolf, E. and Kim, P.S., A computationally directed screen identifying interacting coiled coils from *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. USA*, **97**:13203-13208, 2000.

[6] Parry, D.A., 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, 1982.

[7] Wolf, E., Kim, P.S. and Berger, B., MultiCoil: A program for predicting two- and three-stranded coiled coils, *Protein Sci.***, 6**:1179-1189, 1997.