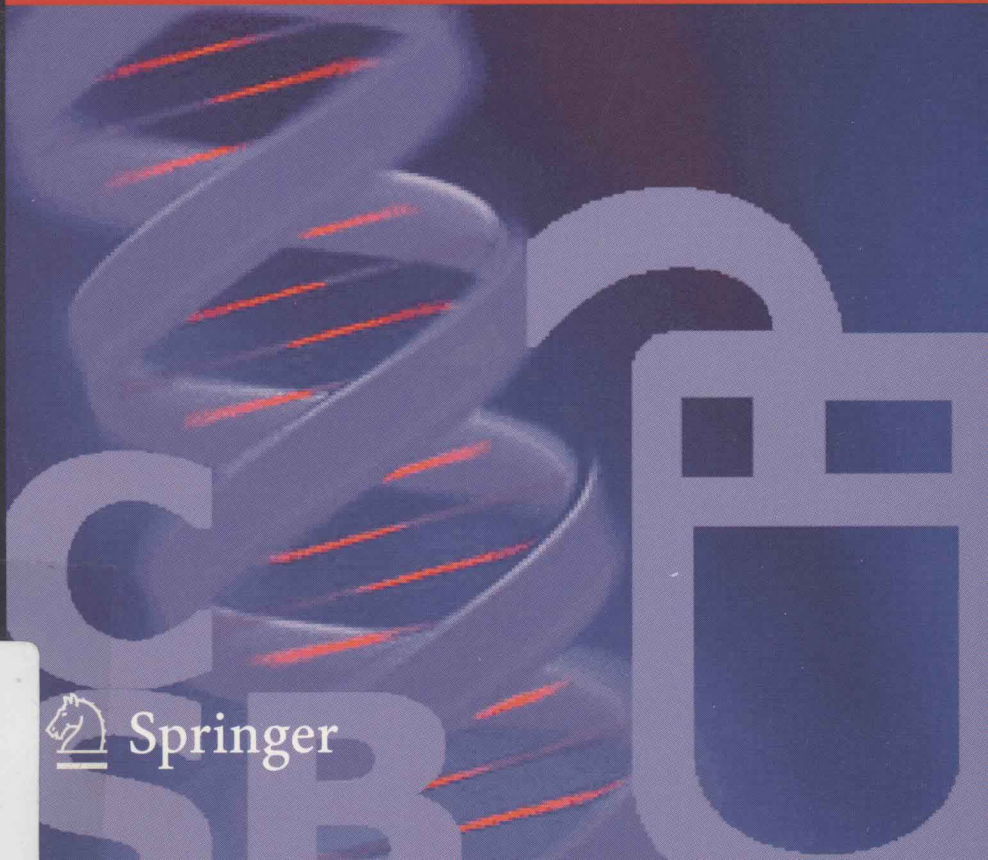
Journal Subline

LNBI 4070

Transactions on Computational Systems Biology V

Corrado Priami
Editor-in-Chief



Springer

Corrado Priami Xiaohua Hu Yi Pan
Tsau Young Lin (Eds.)

Transactions on Computational Systems Biology V

Springer

Volume Editors

Xiaohua Hu
Drexel University, College of Information Science and Technology
3141 Chestnut Street, Philadelphia, PA 19104, USA
E-mail: thu@cis.drexel.edu

Yi Pan
Georgia State University, Department of Computer Science
34 Peachtree Street, Atlanta, GA 30302-4110, USA
E-mail: pan@cs.gsu.edu

Tsau Young Lin
San Jose State University, Department of Computer Science
San Jose, CA 95192, USA
E-mail: tylin@cs.sjsu.edu

Library of Congress Control Number: 2006929800

CR Subject Classification (1998): J.3, H.2.8, F.1

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 1861-2075
ISBN-10 3-540-36048-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-36048-3 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11790105 06/3142 5 4 3 2 1 0

Lecture Notes in Bioinformatics 4070

Edited by S. Istrail, P. Pevzner, and M. Waterman

Editorial Board: A. Apostolico S. Brunak M. Gelfand
T. Lengauer S. Miyano G. Myers M.-F. Sagot D. Sankoff
R. Shamir T. Speed M. Vingron W. Wong

Subseries of Lecture Notes in Computer Science

Lecture Notes in Bioinformatics

Vol. 4075: U. Leser, F. Naumann, B. Eckman (Eds.), Data Integration in the Life Sciences. XI, 298 pages. 2006.

Vol. 4070: C. Priami, X. Hu, Y. Pan, T.Y. Lin (Eds.), Transactions on Computational Systems Biology V. IX, 129 pages. 2006.

Vol. 3939: C. Priami, L. Cardelli, S. Emmott (Eds.), Transactions on Computational Systems Biology IV. VII, 141 pages. 2006.

Vol. 3916: J. Li, Q. Yang, A.-H. Tan (Eds.), Data Mining for Biomedical Applications. VIII, 155 pages. 2006.

Vol. 3909: A. Apostolico, C. Guerra, S. Istrail, P. Pevzner, M. Waterman (Eds.), Research in Computational Molecular Biology. XVII, 612 pages. 2006.

Vol. 3886: E.G. Bremer, J. Hakenberg, E.-H.(S.) Han, D. Berrar, W. Dubitzky (Eds.), Knowledge Discovery in Life Science Literature. XIV, 147 pages. 2006.

Vol. 3745: J.L. Oliveira, V. Maojo, F. Martín-Sánchez, A.S. Pereira (Eds.), Biological and Medical Data Analysis. XII, 422 pages. 2005.

Vol. 3737: C. Priami, E. Merelli, P. Gonzalez, A. Omicini (Eds.), Transactions on Computational Systems Biology III. VII, 169 pages. 2005.

Vol. 3695: M.R. Berthold, R.C. Glen, K. Diederichs, O. Kohlbacher, I. Fischer (Eds.), Computational Life Sciences. XI, 277 pages. 2005.

Vol. 3692: R. Casadio, G. Myers (Eds.), Algorithms in Bioinformatics. X, 436 pages. 2005.

Vol. 3680: C. Priami, A. Zelikovsky (Eds.), Transactions on Computational Systems Biology II. IX, 153 pages. 2005.

Vol. 3678: A. McLysaght, D.H. Huson (Eds.), Comparative Genomics. VIII, 167 pages. 2005.

Vol. 3615: B. Ludäscher, L. Raschid (Eds.), Data Integration in the Life Sciences. XII, 344 pages. 2005.

Vol. 3594: J.C. Setubal, S. Verjovski-Almeida (Eds.), Advances in Bioinformatics and Computational Biology. XIV, 258 pages. 2005.

Vol. 3500: S. Miyano, J. Mesirov, S. Kasif, S. Istrail, P. Pevzner, M. Waterman (Eds.), Research in Computational Molecular Biology. XVII, 632 pages. 2005.

Vol. 3388: J. Lagergren (Ed.), Comparative Genomics. VII, 133 pages. 2005.

Vol. 3380: C. Priami (Ed.), Transactions on Computational Systems Biology I. IX, 111 pages. 2005.

Vol. 3370: A. Konagaya, K. Satou (Eds.), Grid Computing in Life Science. X, 188 pages. 2005.

Vol. 3318: E. Eskin, C. Workman (Eds.), Regulatory Genomics. VII, 115 pages. 2005.

Vol. 3240: I. Jonassen, J. Kim (Eds.), Algorithms in Bioinformatics. IX, 476 pages. 2004.

Vol. 3082: V. Danos, V. Schachter (Eds.), Computational Methods in Systems Biology. IX, 280 pages. 2005.

Vol. 2994: E. Rahm (Ed.), Data Integration in the Life Sciences. X, 221 pages. 2004.

Vol. 2983: S. Istrail, M.S. Waterman, A. Clark (Eds.), Computational Methods for SNPs and Haplotype Inference. IX, 153 pages. 2004.

Vol. 2812: G. Benson, R.D. M. Page (Eds.), Algorithms in Bioinformatics. X, 528 pages. 2003.

Vol. 2666: C. Guerra, S. Istrail (Eds.), Mathematical Methods for Protein Structure Analysis and Design. XI, 157 pages. 2003.

Preface

This issue of *Transactions on Computational Systems Biology* contains a selection of papers presented initially at the 2005 IEEE International Conference on Granular Computing held in Beijing, July 25–27, and a few invited papers. Papers included in this special issue are devoted to various aspects of computational methods, algorithms, and techniques in bioinformatics such as gene expression analysis, biomedical literature mining and natural language processing, protein structure prediction, biological database management and biomedical information retrieval.

Z. Huang, Y. Li and X. Hu present a novel SVM-based method to predict anti-parallel structure from sequence data.

C.H. Liu, I.-J. Chiang, J.-M. Wong, H.-C. Tsai and T.Y. Lin introduce a novel model of concept representation called Latent Semantic Networks using a multilevel geometric structure.

B. Jin and Y.-Q. Zhang propose a new system to evolve the structures of granular kernel trees (GKTs) in the case that we lack knowledge to predefine kernel trees. The new granular kernel tree structure evolving system is used for cyclooxygenase-2 inhibitor activity comparison.

M.K. Ng, S.-Q. Zhang, W.-K. Ching and T. Akutsu study a control model for gene intervention in a genetic regulatory network. At each time step, a finite number of controls are allowed to drive to some target states (i.e., some specific genes are on, and some specific genes are off) of a genetic network.

Z. Peng, Y. Shi and B. Zhai discuss how to manage a large amount of complex biological data by an object deputy database system which can provide rich semantics and enough flexibility. In their system, the flexible inheritance avoids a lot of data redundancy.

R. Satre, H. Sovik, T. Amble and Y. Tsuruoka address the natural language understanding in molecular biology literature. Their prototype system GeneTUC is capable of doing deep reasoning, such as anaphora resolution and question answering, which is not a part of the state-of-the-art parsers.

H.-C. Wang, Y.-S. Lee and T.-H. Huang describe a novel approach to combine microarray data and literature to find the relations among genes. Unlike other techniques, this method not only reduces the comparison complexity but also reveals more mutual interactions among genes.

H.-H. Yu, V.S. Tseng and J.-H. Chuang propose a multi-information-based methodology to score genes based on the microarray expressions. The concept of multi-information here is to combine different scoring functions in different tiers for analyzing gene expressions. The proposed methods can rank the genes according to the degree of relevance to the targeted diseases so as to form a precise prediction base.

X. Zhou, X. Hu, G. Li, X. Lin and X. Zhang explore the use of term relations in information retrieval for precision-focused biomedical literature search. A relation is defined as a pair of two terms which are semantically and syntactically related to each other. Unlike the traditional "bag-of-word" model for documents, their model represents a document by a set of sense-disambiguated terms and their binary relations. A prototyped IR system supporting relation-based search is then built for Medline abstract searches. The experiment shows the expressiveness of relations for the representation of information needs, especially in the area of biomedical literature full of various biological relations.

We would like to thank the authors for contributing their research work to the special issue as well as the Editor-in-Chief of the LNCS Transaction on Computational Systems Biology, Prof. Priami.

The editors of the special issue:
Xiaohua Hu, Drexel University
Yi Pan, Georgia State University
T.Y. Lin, San Jose State University

LNCS Transactions on Computational Systems Biology – Editorial Board

Table of Contents

Anti-parallel Coiled Coils Structure Prediction by Support Vector Machine Classification
Zhong Huang, Yun Li, Xiaohua Hu 1

A Complex Bio-networks of the Function Profile of Genes
Charles C.H. Liu, I-Jen Chiang, Jau-Min Wong, Ginni Hsiang-Chun Tsai, Tsau Young ('T.Y.') Lin 9

Evolutionary Construction of Granular Kernel Trees for Cyclooxygenase-2 Inhibitor Activity Comparison
Bo Jin, Yan-Qing Zhang 25

A Control Model for Markovian Genetic Regulatory Networks
Michael K. Ng, Shu-Qin Zhang, Wai-Ki Ching, Tatsuya Akutsu 36

Realization of Biological Data Management by Object Deputy Database System
Zhiyong Peng, Yuan Shi, Boxuan Zhai 49

GeneTUC, GENIA and Google: Natural Language Understanding in Molecular Biology Literature
Rune Sætre, Harald Søvik, Tore Amble, Yoshimasa Tsuruoka 68

Gene Relation Finding Through Mining Microarray Data and Literature
Hei-Chia Wang, Yi-Shiun Lee, Tian-Hsiang Huang 83

A Multi-information Based Gene Scoring Method for Analysis of Gene Expression Data
Hsieh-Hui Yu, Vincent S. Tseng, Jiin-Haur Chuang 97

Relation-Based Document Retrieval for Biomedical IR
Xiaohua Zhou, Xiaohua Hu, Guangren Li, Xia Lin, Xiaodan Zhang 112

Author Index 129

Anti-parallel Coiled Coils Structure Prediction by Support Vector Machine Classification

Zhong Huang, Yun Li, and Xiaohua Hu

College of Information Science and Technology, Drexel University,
3141 Chestnut Street, Philadelphia, PA, USA, 19104
thu@cis.drexel.edu

Abstract. Coiled coils is an important 3-D protein structure with two or more stranded alpha-helical motif wounded around to form a "knobs-into-holes" structure. In this paper we propose an SVM classification approach to predict the anti-parallel coiled coils structure based on the primary amino acid sequence. The training dataset for the machine learning are collected from SOCKET database which is a SOCKET algorithm predicted coiled coils database. Total 41 sequences of at least two heptad repeats of the anti-parallel coiled coils motif are extracted from 12 proteins as the positive datasets. Total 37 of non coiled coils sequences and parallel coiled coils motif are extracted from 5 proteins as negative datasets. The normalized positional weight matrix on each heptad register a, b, c, d, e, f and g is from SOCKET database and is used to generate the positional weight on each entry. We performed SVM classification using the cross-validated datasets as training and testing groups. Our result shows 73% accuracy on the prediction of anti-parallel coiled coils based on the cross-validated data. The result suggests a useful approach of using SVM to classify the anti-parallel coiled coils based on the primary amino acid sequence.

Keywords: coiled coil, SOCKET algorithm, SVM, protein sequence data.

1 Introduction

Coiled coils structure was first introduced by Crick in 1953 in which he postulated a hallmark structure of "knobs-into-holes" formed by wounded strands of alpha-helices [2]. The coiled coils structure is characterized by a heptad repeats of amino acids (*a-b-c-d-e-f-g*)n. Positions *a* and *d* in one chain are occupied by apolar hydrophobic amino acids to form the core packing structure with the same positions in partner chain (see figure 1). The coiled coils structure is further stabilized by side chain electrostatic interaction of *e-g* between two chains which generally occupied by polar charged amino acids. Recently it has been shown that intrachain interactions between heptad residues also contribute to the stability of the coiled coils structure [6] [12].

Due to its well characterized structure the coiled coils has long been a spotlight of the protein design and prediction study. However, the structure of coiled coils is of great diversity in terms of its interchain orientation and oligomer status. The coiled coils structure can be formed between two, three, four or even five chains and the

C. Priami et al. (Eds.): Trans. on Comput. Syst. Biol. V, LNBI 4070, pp. 1 – 8, 2006.

orientation of each chain can be the same (parallel) or different (antiparallel). The core packing registers *a* and *d* are important for determining the number of strands while e-g interaction seems to be important in choosing the helices partners [3] [4] [7] [11]. Therefore the primary sequence of the heptad repeats may be one of the determining factors on the specificity of the coiled coils but much is still poorly understood so far [12].

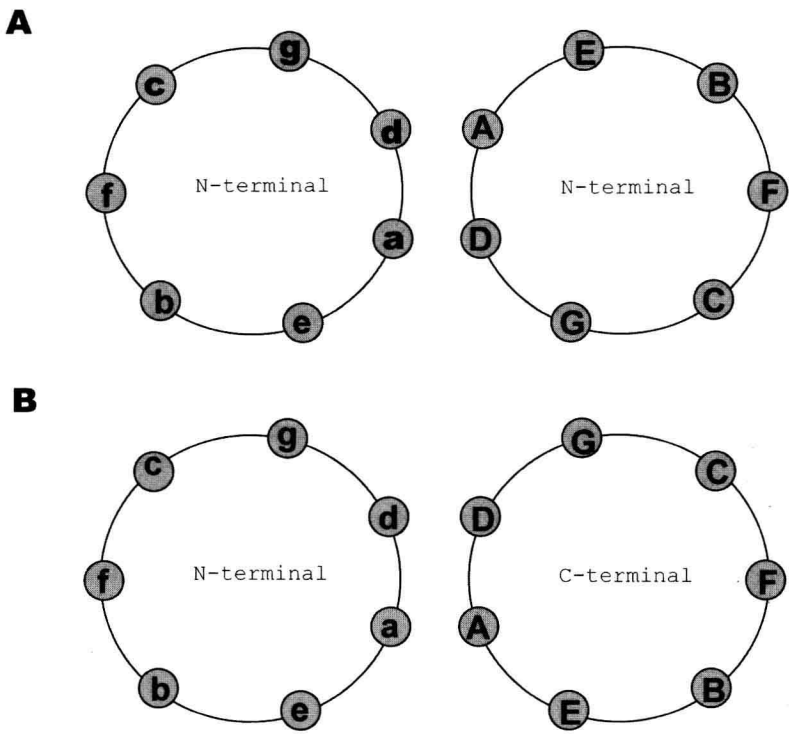


Fig. 1. Illustration of transaction view of two stranded parallel (A) and anti-parallel coiled coils (B)

Two categories of algorithms have been proposed to predict the coiled coils structure based on either the primary amino acid sequence or the atomic 3-D coordinate information. The first category includes COILS [5], PAIRCOIL [1] and MULTICOIL [10]. These algorithms compare the sequence of the target protein with the amino acid sequence database of known two or three stranded parallel coiled coils and give the score of probability. However, none of them are suitable for predicting the anti-parallel coiled coils. The second category of prediction algorithm utilizes the 3-D coordinate information of the polypeptide to predict and define the beginning and ending of coiled coils based on the database of known 3-D structure of coiled coils. SOCKET [9] and TWISTER [8] are two algorithms in this category. The SOCKET algorithm focuses on the core packing structure of the "knobs-into-holes" which is formed by interchain *a-d* interactions. The TWISTER algorithm is designed to identify not only the canonical coiled coils but also the special coiled coils with

discontinuous heptad repeats interrupted by stutters and skips. Both algorithms take the 3-D atomic coordinate PDB file and DSSP file as input and are able to predict two or three stranded parallel and anti-parallel coiled coils. Comparing with the first category of algorithms, using 3-D coordinate as input may seem to be a better choice as the SOCKET algorithm attempts to identify the core packing structure of coiled coils based on the experimentally determined protein 3-D structure. The disadvantage of utilizing SOCKET to predict coiled coils structure is that it requires atomic coordinate information of the polypeptide which may not be readily available. However, with the rapid expanding of the PDB database which currently already collects over 34,000 structures determined by x-ray crystallization and NMR spectroscopy [13], such algorithms can be adopted more widely to predict coiled coils structure.

In this paper we used the machine learning supporting vector machine (SVM) approach to discriminate the anti-parallel coiled coils structure based on the primary amino acid sequence using the normalized amino acid profile of heptad repeat generated by SOCKET [9]. We selected anti-parallel coiled coils proteins with at least two full set of heptad repeats from SOCKET database such that each vector has the same number of features for SVM training and testing. Our preliminary results suggest that SVM is a valuable tool to predict the anti-parallel coiled coils based on the amino acid profile originally determined by atomic 3-D coordinate of the protein.

2 Methods, Results and Discussion

We selected total 78 anti-parallel coiled coils from SOCKET database which currently lists a total of 134 entries [9] based on PDB release #89. The PDB files of 78 anti-parallel coiled coils and 8 parallel coiled coils were downloaded using a perl script available from PDB ftp site. The PDB files from both the anti-parallel and parallel coiled coils were submitted to SOCKET server (http://www.biols.susx.ac.uk/Biochem/Woolfson/html/coiledcoils/socket/server.html) to identify the specific heptad repeats registers. Only the long anti-parallel coiled coils with at least two full sets of the heptad repeats were selected. This is mainly based on our assumption that long coiled coils are more structurally stable and may include more positional information which could contribute to the stability and specificity of coiled coils. Considering the relatively small number of entries currently available, we allow multiple contributions of coiled coils structure from the same protein. Because the SVM only accepts vectors with the same number of features, we chose 2 heptad repeats of total 14 amino acids from each entry. In the case of heptad repeats are more than 2, we allow partial overlap of the heptad repeats assuming each partially overlapped heptad repeats is an independent vector for SVM to avoid loss of any given heptad repeat. Total 41 sequences of at least two heptad repeats of the anti-parallel coiled coils motif are extracted from 12 proteins as the positive datasets (Table 1). Total 37 of non coiled coils sequences and parallel coiled coils motif are extracted from 5 proteins as negative datasets.

We used the normalized frequencies of occurrence at each heptad positions for long anti-parallel coiled coils from Walshaw et al [9] to convert the amino acid of each heptad sequence into amino acid usage frequencies (shown in Table 2).

Table 1. Positive datasets of proteins with anti-parallel coiled coil structures

PDB ID	Protein name
5eau	5-Epi-Aristolochene Synthase
2spc	Spectrin
2ktq	Large Fragment Of DNA Polymerase I
2fha	Human H Chain Ferritin
1ser	seryl-tRNA synthetase complexed with tRNA(Ser)
1ecr	Replication Terminator Protein (Tus) Complexed With DNA
1ecm	Chorismate Mutase
1cnt	Ciliary Neurotrophic Factor
1cii	Colicin Ia
1aqt	F1F0-ATP Synthase
1ab4	59Kda Fragment Of Gyrase A
1a36	Human DNA Topoisomerase I

Table 2. Normalized frequencies of occurrence at each position of the heptad sequence for two stranded long anti-parallel coiled coils from Walshaw et al [9]

Corner	a	b	c	d	e	f	g
A	1.42	1.48	1.28	1.55	0.69	1.04	1.98
C	0	0	0	0.59	0	1.19	0
D	0.09	1.06	1.71	0.56	1.37	1.75	0.25
E	0.54	1.32	1.75	1.31	1.45	1.76	1.23
F	0.84	0.51	0.85	0.96	0.8	0.96	0.64
G	0	0.72	0.81	0.07	0.48	0.38	0.19
H	1.09	1.56	0.31	0.44	0.88	1.47	3.79
I	2.87	0.48	0.24	1.6	1.59	0.79	1.12
K	0.33	2.24	1.52	0.41	1.22	1.55	0.66
L	2.96	1.04	1.33	3.64	1.31	0.35	1.11
M	1.45	1.18	0.59	1.65	0.83	0.83	1.93
N	0.88	1.58	0.94	0.22	0.44	1.78	0.59
P	0	0	0.28	0	0.27	0	0
Q	0.86	1.23	1.75	0.74	1.99	1.16	2.3
R	0.66	1.36	1.48	0.47	2.55	1.78	1.14
S	1.03	0.78	0.58	0.28	0.37	0.83	0.83
T	0.61	1.23	0.73	0.61	0.58	1.39	1.15
V	0.51	0.21	0.53	0.37	0.61	0.41	0.99
W	0	0.56	2.24	0.31	0.53	0.53	0.53
Y	1.23	0.66	0.44	1.69	0.82	0.41	0.2

Table 3. The testing results of SVM using cross-validation approach

Corner	Class Label	Prediction by SVM
1a36-1	1	1
1a36-2	1	-1
1ab4-1	1	1
1ab4-2	1	-1
1aqt-1	1	-1
1cii-1	1	1
1cii-2	1	1
1cii-3	1	1
1cii-4	1	-1
1cii-5	1	1
1cii-6	1	-1
1cii-7	1	-1
1cii-8	1	-1
1cnt-1	1	1
1cnt-2	1	1
1cnt-1	1	1
1cnt-2	1	1
1cnt-3	1	1
1cnt-4	1	1
1cnt-5	1	1
1ecm-1	1	1
1ecm-2	1	1
1ecm-3	1	1
1ecm-4	1	1
1ecr-1	1	1
1ecr-2	1	1
1ser-1	1	1
1ser-2	1	-1
2fha-1	1	1
2fha-2	1	-1
2fha-3	1	-1
2fha-4	1	1
2fha-5	1	1
2ktq	1	1
2spc-1	1	1
2spc-2	1	1
2spc-3	1	1
2spc-4	1	1
2spc-5	1	1

Table 3. *(Continued)*

2spc-6	1	1
5eau	1	1
m6a-1	-1	-1
m6a-2	-1	-1
m6a-3	-1	-1
m6a-4	-1	-1
m6a-5	-1	-1
m6a-6	-1	1
m6a-7	-1	-1
m6a-8	-1	-1
m6a-9	-1	1
m6a-10	-1	-1
m6a-11	-1	1
m6a-12	-1	-1
m6a-13	-1	1
m6a-14	-1	-1
m6a-15	-1	-1
1a93-1	-1	-1
1a93-2	-1	-1
1a93-3	-1	1
1a93-4	-1	1
1fos-1	-1	-1
1fos-2	-1	-1
1fos-3	-1	1
1fos-4	-1	-1
1fos-5	-1	-1
1fos-6	-1	-1
1fos-7	-1	1
1fos-8	-1	1
1fos-9	-1	-1
1fos-10	-1	-1
mbplike-1	-1	1
mbplike-2	-1	-1
mbplike-3	-1	-1
mbplike-4	-1	-1
mbplike-5	-1	-1
mbplike-6	-1	1
mbplike-7	-1	-1
MBP	-1	-1

Two programs written in perl and java were used to convert the list of 14 amino acid sequence into amino acid usage frequency at each position and to generate label file for SVM classification. Due to the small number of samples, we adopted cross-validation approach for SVM training and testing. The total 78 datasets are separated into two groups, with 77 datasets for training and 1 dataset for testing in each cycle. Web interface of SVM (Gist version 2.0.5, http://svm.sdsc.edu) was used in dataset training and testing. The testing results are shown in Table 3.

From total 78 datasets including 41 positive and 37 negative datasets, we identified 31 true positive and 26 true negative respectively using SVM classification. Our results show that the average accuracy for the testing is 73% (summarized in Table 4).

Table 4. Average Accuracy Result

Total Dataset	78
Positive	41
Negative	37
Testing Result	
False Positive	11
False Negative	10
True Positive	31
True Negative	26
Average Accuracy	73%

We demonstrated that SVM classification algorithm can be used to discriminate anti-parallel coiled coils structure from non-parallel coiled coils including the parallel coiled coils structure. Each protein sequence with two heptad repeats was taken as input vector with 14 dimensional features and subjected to SVM training and testing. All 14 features of each vector have same semantics representing the amino acid usage frequencies. Our result indicated that SVM learning algorithm can discriminate two classes using hyperplane with maximum margin between vectors of two classes with relatively high accuracy.

Walshaw et al calculated normalized amino acid frequencies of occurrence at both the parallel and anti-parallel coiled coils and found that anti-parallel coiled coils tends to have broader amino acid usages at each heptad position compared with its parallel counterpart [9]. This characteristic of amino acid profile associated specifically with anti-parallel coiled coils also prompted us to select higher dimensional features which include two repeats of heptad, instead of focusing only on hallmark feature of "knob-into-hole" structure at a/d positions.

3 Conclusion and Future Plan

In this paper we presented a SVM approach for classification of anti-parallel coiled coils structure based on primary sequences. The classification results indicate that the support vector machine learning algorithm is a useful tool in classifying the anti-parallel coiled coils structure, which by far has no direct algorithm to predict its

structure based on its primary sequence. However, more datasets are needed to further validate the approach. With the rapid growing of the PDB database, the SOCKET database has been growing accordingly. We expect the future release of SOCKET database will be helpful in gathering more positive datasets and help to improve the SVM prediction accuracy.

Acknowledgement. This work is supported partially by the NSF Career grant IIS 0448023 and NSF 0514679 and PA Dept of Health Tobacco Formula Grants.

References

1. Berger, B. (1995). "Algorithms for protein structural motif recognition." J Comput Biol 2(1): 125-38.
2. Crick, F. H. C. (1953). "The packing of alpha-helices: simple coiled-coils." Acta. Crystallog. 6: 689-697.
3. Harbury, P. B., T. Zhang, et al. (1993). "A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants." Science 262(5138): 1401-7.
4. Kohn, W. D., C. M. Kay, et al. (1998). "Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils." J Mol Biol 283(5): 993-1012.
5. Lupas, A., M. Van Dyke, et al. (1991). "Predicting coiled coils from protein sequences." Science 252(5010): 1162-4.
6. Oakley, M. G. and J. J. Hollenbeck (2001). "The design of antiparallel coiled coils." Curr Opin Struct Biol 11(4): 450-7.
7. O'Shea, E. K., R. Rutkowski, et al. (1992). "Mechanism of specificity in the Fos-Jun oncoprotein heterodimer." Cell 68(4): 699-708.
8. Strelkov, S. V. and P. Burkhard (2002). "Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation." J Struct Biol 137(1-2): 54-64.
9. Walshaw, J. and D. N. Woolfson (2001). "Socket: a program for identifying and analysing coiled-coil motifs within protein structures." J Mol Biol 307(5): 1427-50
10. Wolf, E., P. S. Kim, et al. (1997). "MultiCoil: a program for predicting two- and three-stranded coiled coils." Protein Sci 6(6): 1179-89
11. Woolfson, D. N. and T. Alber (1995). "Predicting oligomerization states of coiled coils." Protein Sci 4(8): 1596-607
12. Yu, Y. B. (2002). "Coiled-coils: stability, specificity, and drug delivery potential." Adv Drug Deliv Rev 54(8): 1113-29.
13. Berman H.M., Westbrook J. et al (2000). "The Protein Data Bank." Nucleic Acids Res 28 235-242.