Computing Closures of Indirect Similarity Relationships among Structures in Proteins

Kenji Satou[1] Emiko Furuichi[2] Shin'ichi Hashimoto[3] Yukiko Tsukamoto[4]
ken@csce.kyushu-u.ac.jp emiko@grt.kyushu-u.ac.jp gucci@csce.kyushu-u.ac.jp yukiko@grt.kyushu-u.ac.jp

Satoru Kuhara[5] Toshihisa Takagi[6] Kazuo Ushijima[7]
kuhara@grt.kyushu-u.ac.jp takagi@ims.u-tokyo.ac.jp ushijima@csce.kyushu-u.ac.jp

[1,3,7] Department of Computer Science and Communication Engineering, Kyushu University
Hakozaki, Higashi-ku, Fukuoka 812, Japan

[2] Fukuoka Women's Junior College, Gojo, Dazaifu, Fukuoka 818-01, Japan

[4,5] Graduate School of Genetic Resources Technology, Kyushu University
Hakozaki, Higashi-ku, Fukuoka 812, Japan

[6] Human Genome Center, Institute of Medical Science, The University of Tokyo
Shiroganedai, Minato-ku, Tokyo 108, Japan

Abstract

We have developed a deductive database system PACADE for analyzing three dimensional and secondary structures of protein. A function newly introduced to PACADE is described here. It enables to compute a closure of indirect similarity relationships among structure of proteins.

1 Introduction

Because of the exponential increase of these *genome data*, there is a increasing need to develop a sound knowledge base to describe a hypothesis and to check it out with queries. We have developed a deductive database system PACADE for analyzing three dimensional and secondary structures of protein. By describing a set of declarative rules and giving a query to PACADE,

[1]佐藤賢二: 九州大学工学部情報工学科, 〒812 福岡市東区箱崎 6-10-1
[2]古市恵美子: 福岡女子短期大学, 〒818-01 太宰府市五条 4-16-1
[3]橋本真一: 九州大学工学部情報工学科, 〒812 福岡市東区箱崎 6-10-1
[4]塚本由紀子: 九州大学大学院農学研究科遺伝子資源工学専攻, 〒812 福岡市東区箱崎 6-10-1
[5]久原哲: 九州大学大学院農学研究科遺伝子資源工学専攻, 〒812 福岡市東区箱崎 6-10-1
[6]高木利久: 東京大学医科学研究所ヒトゲノム解析センター, 〒108 東京都港区白金台 4-6-1
[7]牛島和夫: 九州大学工学部情報工学科, 〒812 福岡市東区箱崎 6-10-1

a user can flexibly search for structures in proteins. As for the searches for super-secondary structures frequently found in proteins (e.g. Greek Keys and n-stranded meanders,) the deductive inference system of PACADE works well. To make PACADE an even more useful system for molecular biologists, we introduced a novel function which enables automatic searches for similar structures in proteins. Through this function, a user can search for structures of proteins similar to the one being examined.

However, to compute a closure of structures which includes a specific structure of a protein, it is insufficient and inefficient to use the similarity search function of PACADE, combinatorially. We describe here a function newly introduced into PACADE to resolve the problems. This function enables a non-combinatorial computation of a cluster of indirectly similar structures, which includes a specific structure given by a user in the form of a query.

2 Method and Result

Starting from a given structure in a specific protein, one can compute a closure of structures indirectly similar to it by regarding answers as closed queries and iterating direct similarity search in PACADE. Computation of closure is performed by fixpoint computation like the one in ordinary bottom-up evaluator. Thus, it can be optimized by using differential algorithm instead of naive algorithm. On the other hand, the constant bindings in closed queries are available to optimize through Magic Set transformation. A module, added to PACADE to perform closure computation, is implemented using these optimization techniques.

As the result of an example search starting from a specific structure (a0) in interleukin (`"3il8"`), we obtained a closure consisting of 6 structures (a0-a5). In Figure 1, bi-directional arrows represent direct similarity relationships among a0-a5. The graphical representations of a0-a5 reflect the fact that a2, a4 and a5 are similar (Figure 2.)

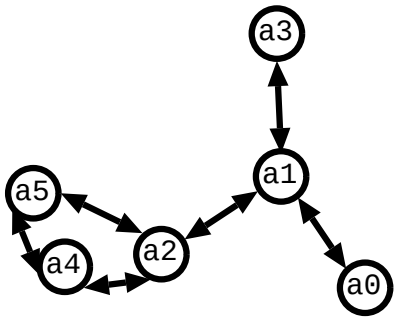

Figure 1: Similarity relationships among a0-a5

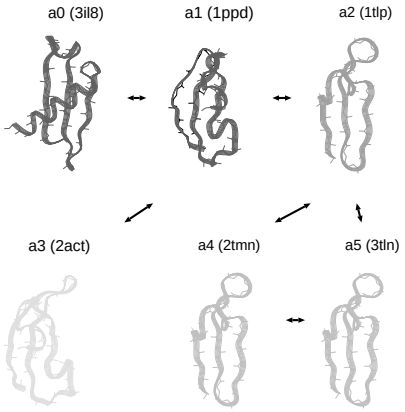

Figure 2: Graphical display of a0-a5

Acknowledgement

This work was supported in part by a Grant-in Aid for Scientific Research on Priority Areas, "Genome Informatics," from the Ministry of Education, Science and Culture of Japan.