Detection of Apoptotic Domains Against KEGG Database by the HMM Search

Masahiro Hattori hattori@kuicr.kyoto-u.ac.jp
Minoru Kanehisa kanehisa@kuicr.kyoto-u.ac.jp
Institute for Chemical Research, Kyoto University
Gokasho, Uji, Kyoto 611-0011, Japan

1 Introduction

We have developed the knowledge base which consists of apoptotic molecular interactions [1], and provided the WWW interface for it. This database and user interfaces enabled us to find out entries containing different information about cell death. This information tells us that the apoptotic molecular interactions are supposed to be controlled under a series of specific conserved domains, like Bcl-2 homology domain, death domain or death effector domain. So the viewpoint of domain seems to be more effective, rather than the entire sequence, when analyzing pathways, molecules, genes and genomes, which are concerned with apoptosis. In this study, we collected a few hundred domain sequences of apoptotic interactions from our database, made 14 hidden Markov models for the domain groups, and searched against KEGG/GENES database [2] with those HMMs (hidden Markov models) to detect conserved domains.

2 Data and Method

At first, we collected information on domains or motifs which are concerned with apoptotic molecular interactions from literature or databases, such as PROSITE [3] or Pfam [4], and extracted those domain sequences from our apoptotic database. After that we got 409 sequences of 14 domain groups, which are originally derived from 200 protein entries of our database. The largest group has 66 sequences of extracellular cysteine repeat domain of tumor necrosis factor receptor superfamily, and the smallest has 2 sequences of C-terminus homologous region of CIDE superfamily. The average size is 29 sequences par domain, which seems enough for HMM learning and further sequence analyses. Secondly, we made the HMMs from those domain sequences by using HMMER2 programs (http://hmmer.wustl.edu/). Thus we got 14 HMMs for the domain groups. These HMMs are statistically tested in order to estimate their ability and efficiency of detection. Finally we looked for any homologous domain in KEGG/GENES database by using those HMMs. Currently KEGG/GENES has 7 eukaryotes and 29 prokaryotes. It contains 24 complete genomes of 1 eukaryote and 23 prokaryotes. Our HMM-search procedure was done on each of the 36 species.

3 Result and Discussion

We could detect potentially interesting protein entries in *S. pombe*, *S. cerevisiae* and *C. elegans* (Table 1).

1) *S. pombe* SPAC2C6.16 has two BIR domains and *S. cerevisiae* YJR089W has one. BIR domain is necessary for an inhibitor of apoptotic peptidase (IAP), and in higher animals three domains are usually found in one protein sequence. Then these two peptides in fungi might be an original type

of IAP family. And interestingly these two sequences do not have an overall similarity to each other. The phylogenetic relationship between them is interesting.

2) *C. elegans* has two BIR homologous peptides, C50B8.2 and T27F2.3, one Caspase like peptidase, Y48E1B.13, and four death domain containing peptides, B0350.2A, B0350.2B, B0350.2C and B0350.2D. Here Y48E1B.13 has both His- and Cys-active sites of Caspase, this seems to be a homologue of Ced-3 apart from its physiological function. Four B0350.2x are thought as the homologue to ankyrin, which is already known as a death domain containing protein. But the function of death domain has not yet been checked in ankyrin.

Table 1: Detected Entries by HMM in Eukaryote Genome

Species	**Gene**	**detected motif**
S. pombe	SPAC2C6.16	two BIRs
S. cerevisiae	YJR089W	one BIR
C. elegans	C50B8.2	two BIRs
C. elegans	T27F2.3	one BIR
C. elegans	Y48E1B.13	Caspase, Cys active site
C. elegans	Y48E1B.13	Caspase, His active site
C. elegans	B0350.2A,B,C,D	death domain

On the other hand, we could get no high scoring domains in prokaryotes but found that several peptides have weak homologies to apoptosis domains. We do not yet know the biological significance of these homologies.

Acknowledgments

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas, 'Genome Science', from the Ministry of Education, Science, Sports and Culture of Japan. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

References

[1] Hattori, M. and Kanehisa, M., The construction of the knowledge base on apoptotic molecular interactions, *Genome Informatics*, 8:276–277, 1997.

[2] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M., KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Research*, 27:29–34, 1999.

[3] Hofmann, K., Bucher, P., Falquet, L., Bairoch, A., The PROSITE database, its status in 1999, *Nucleic Acids Research*, 27:215–219, 1999.

[4] Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Finn, R.D., and Sonnhammer, E.L., Pfam 3.1: 1313 multiple alignments match the majority of proteins, *Nucleic Acids Research*, 27:260–262, 1999.