HMM Search for Apoptotic Domains

Masahiro Hattori and Minoru Kanehisa

For the purpose of analyzing apoptotic molecular interactions, we have developed a knowledge base, which consists of apoptotic molecular interactions, together with the WWW interface for it. This database and the user interface enabled us to find out entries containing various information about cell death. This information tells us that the apoptotic molecular interactions are likely to be controlled under a series of specific conserved domains. Thus, the viewpoint of domain seems to be more effective than the sequence similarity of the entire protein, when analyzing pathways, molecules, genes and genomes. In this study, we collected several hundred domain sequences of apoptotic interactions from our database, made 14 Hidden Markov Models (HMMs) for the domain groups, and searched against KEGG/GENES database with those HMMs to detect evolutionarily conserved domains.

Keywords: Apoptosis / Database / Hidden Markov Model / Bioinformatics

The characteristic form of cell death, which is accompanied with a cell shrinking, an active proteolysis and a nucleosome level DNA incision, has been observed[1]. Those dying cells show a specific process which is clearly different from the usual cell dying process. They are absorbed into other living cells around them and their biomaterials, for instance amino acids or nucleotides, seem to be recycled within those living cells. This type of cell death is called apoptosis, and is believed to play an important role in tissue and organ development in ontogenesis and homeostasis of living body. It is also known that apoptosis is related to some diseases such as autoimmune disease, virus infection and cancer. Furthermore many drugs such as anticancer agents take effect via induction of apoptosis in the target cells. With those background the studies on apoptosis have become popular and recently many experimental facts have been reported. In these studies the authors have found new apoptotic related genes or molecules and determined the relationships among those specific molecules. They also suggest that the cell death mechanism like apoptosis has been incorporated and diversified into animals during their evolutional history but its main system has been conserved from *C.elegans* to *H.sapiens*.

In this study, in order to analyze apoptotic pathways in the viewpoint of molecular interactions we have devel-

MOLECULAR BIOLOGY AND INFORMATION — Biological Information Science —

Scope of research

This laboratory aims at developing theoretical frameworks for understanding the information flow in biological systems in terms of genes, gene products, other biomolecules, and their interactions. Toward that end a new database is being organized for known molecular and genetic pathways in living organisms, and computational technologies are being developed for retrieval, inference and analysis. Other studies include: functional and structural prediction of proteins from sequence information and development of sequence analysis tools.

Prof
KANEHISA, Minoru
(D Sc)

Assoc Prof
GOTO, Susumu
(D Eng)

Students:
KIHARA, Daisuke (DC)
PARK, Keun-joon (DC)
HATTORI, Masahiro (DC)
BONO, Hidemasa (DC)
IGARASHI, Yoshinobu (DC)
KATAYAMA, Toshiaki (DC)
SATO, Kiminobu (DC)
KINJO, Fumi (DC)
NAKAO, Mitsuteru (DC)
YOSHIZAWA, Akiyasu (DC)
OKUJI, Yoshinori (MC)
KATO, Masaki (MC)
Research Fellow:
SATO, Kazushige (RF)
DANNO, Masaki (RF)
SHIRAISHI, Kotaro (RF)

oped a new knowledge base and the WWW interface for it, which focus on interactions among those apoptotic molecules[2]. We have also tried to analyze several apoptosis-specific domains with a Hidden Markov Model (HMM) technique because of the fact that the molecular interactions concerning with apoptosis are characterized by a set of domains[3].

At first we have collected information on domains or motifs concerning with apoptotic interactions from literature or databases, such as PROSITE or Pfam, and extracted those domain sequences from our database. After that we have obtained 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 (TNFR) superfamily, and the smallest has 2 sequences of C-terminus homologous region of CIDE superfamily. The average size is 29 sequences per 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 domains in KEGG/GENES database[4] by using those HMMs. Currently KEGG/GENES has 7 eukaryotes and 29 prokaryotes. It contains 24 complete genomes of 2 eukaryote and 22 prokaryotes. Our HMM-search procedure was done on each of the 36 species.

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

1) *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.

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

Table 1. HMM search results vs KEGG/GENES

2) *S.pombe* SPAC2C6.16 has two BIR domains and *S.cerevisiae* YJR089W has one. BIR domain is known to be necessary for an inhibition of apoptotic peptidase, 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 inhibitor of apoptotic peptidase (IAP) family. Interestingly, these two sequences don't have an overall similarity to each other.

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 in Japan. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

References

1. J.F. Kerr, A.H. Wyllie and A.R. Currie, Brit. J. Cancer, **26**, 239 - 257 (1972).
2. M. Hattori and M. Kanehisa, Genome Informatics, **8**, 276 - 277 (1997).
3. M. Hattori and M. Kanehisa, Genome Informatics, **10**, 267 - 268 (1999).
4. H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono and M. Kanehisa, Nucleic Acids Research, **27**, 29 - 34 (1999).