npg

Editorial

DeathBase: a database on structure, evolution and function of proteins involved in apoptosis and other forms of cell death

J Díez[1], D Walter[2], C Muñoz-Pinedo*,[2,3] and T Gabaldón*,[1,3]

Programmed cell death has attracted enormous attention in the past few years. The relevance of this process to human disease has triggered an explosion of articles on cell death, and particularly, on apoptosis, as the main form of cell death in animals. For non-experts, literature searches to find the most relevant information about the key apoptotic molecules become a nightmare: they will find 170 000 citations related to apoptosis in Pubmed. Even for scientists working on cell biology and/or cell death, it is very difficult to identify the key apoptotic proteins, or their roles, among the thousands of proteins that could be related to cell death in one way or another. Finally, experimental information obtained for homologous proteins in different model species is dispersed, and there is a need to integrate it into a clear evolutionary framework. For this reason, we have developed the Death-Base (http://www.deathbase.org).

In DeathBase, users will find lists of proteins involved in cell death in five model species: humans (*Homo sapiens*), mice (*Mus musculus*), fruit flies (*Drosophila melanogaster*), nematodes (*Caenorhabditis elegans*) and zebrafish (*Danio rerio*). Each protein has a separate entry page with basic information such as its sequence and domain structure. Perhaps more importantly, for each protein, users will find manually curated information about its role in cell death: the cell death process and pathway it is involved in, the protein family it belongs to and a few key references to understand the role of this protein in cell death. This information is subjected to manual curation by a group of experts in proteins involved in cell death, whose names and areas of expertise are listed in the 'about' page. In addition, curators can add comments on the role of a specific protein in cell death. These comments can be particularly helpful from a didactic point of view, as curators are encouraged to point out that a specific protein is not involved in cell death. As an example, users will find comments about the true role of some homologs of IAPs, such as survivin, or whether proteins such as AIF participate in non-apoptotic cell death, while its role in apoptosis remains controversial.

Selection of proteins to be included in DeathBase

One of the aims of this database is to eliminate 'noise' in the field by focusing only on the proteins that have direct roles in cell death, rather than including every protein whose overexpression induces apoptosis. We included in the DeathBase proteins that fit in one or more of the following categories: (i) which are part of the core evolutionarily conserved machinery of apoptosis (essentially, caspases, Bcl-2 family members, Apaf-1/FADD-like adapter proteins, IAPs, death receptors and ligands and proteins involved in phagocytosis of apoptotic corpses); (ii) proteins that directly regulate other proteins of the core machinery of apoptosis (e.g., Granzyme B or IAP antagonists); (iii) proteins that contain domains involved in apoptotic interactions such as CARDs or Death Domains; (iv) proteins that participate in non-apoptotic forms of cell death (for instance, RIP or *Drosophila* autophagic proteins); (v) that are homologous to central mediators of apoptosis, even if their role in apoptosis has been discounted or is unproven (for instance, non-apoptotic caspases). In these latter cases, we aim to provide comments and references about their physiological roles. Conversely, for the sake of clarity, DeathBase does not include proteins that are not central to the apoptotic process, even if they may have a role on initiation of a cell death pathway. Among these, we did not include: (i) receptors for survival factors, such as IL-3 or NGF receptors; (ii) proteins that participate in a specific signaling pathway that eventually may end up in induction or activation of apoptotic proteins, such as transcription factors or stress-signaling proteins; (iii) most caspase substrates (for database on caspase substrates, see CASBAH[1]).

Functional data: cell death processes and pathways

We have classified many of the proteins in the DeathBase into a few cell death processes and pathways (Table 1). Apoptotic proteins are classified according to the pathway in which they participate: extrinsic or intrinsic; pre-mitochondrial signaling

[1]Comparative Genomics Group, Centre for Genomics Regulation, Dr. Aiguader, 88, Barcelona 08003, Spain and [2]IDIBELL (Institut d'Investigació Biomèdica de Bellvitge), Gran Via de L'Hospitalet 199, L'Hospitalet, Barcelona 08907, Spain
*Corresponding authors: T Gabaldón, Bioinformatics and Genomics, Comparative Genomics Group, Centre for Genomic Regulation (CRG), Dr. Aiguader, 88, Barcelona 08003, Spain. Tel: + 34 933 160 281; E-mail: tgabaldon@crg.es or C Muñoz-Pinedo, IDIBELL (Institut d'Investigació Biomèdica de Bellvitge), Gran Via de L'Hospitalet 199, L'Hospitalet, Barcelona 08907, Spain; E-mail: cmunoz@idibell.cat
[3]These authors share senior authorship.

Table 1 Examples of classification of proteins in processes, pathways and families

Protein	Species	Cell death process	Cell death pathway	Cell death pathway (secondary)	Protein family
Bax	*Mus musculus*	Apoptosis	Intrinsic pathway	Pre-mitochondrial signaling events	Bcl-2 family, multidomain Bcl-2
Caspase-8	*Homo sapiens*	Apoptosis	Extrinsic pathway	TNF/NF-κB signaling	Caspase
PARP-1	*Mus musculus*	Cell death (undefined)			Other
RIPK1	*Homo sapiens*	Necroptosis, apoptosis	Extrinsic pathway	TNF/NF-κB signaling	Other
Reaper	*D. melanogaster*	Apoptosis			IAP antagonist
Apaf-1	*Homo sapiens*	Apoptosis	Intrinsic pathway	Post-mitochondrial caspase activation	CARD-containing adapter protein

events or post-mitochondrial caspase activation; and apoptotic cell clearance. Proteins that participate in other non-apoptotic forms of cell death have been assigned the descriptor 'cell death (other)'. A few examples of these non-apoptotic cell death processes are autophagic cell death during *Drosophila* development,[2] pyroptosis[3] and cornification.[4] Death-receptor induced necrosis – also called necroptosis – has been included as a separate cell death process descriptor, because at this point there is a large number of proteins that have been shown to participate in this process.[5] In addition, we have created a fourth label that applies to proteins with immune function. The reason is that many caspases, caspase adapter proteins and IAPs are not involved in apoptosis, but they participate in immunity, both in mammals and insects.

Structural data, protein families

Apoptotic proteins in the database are classified into the structural families defined by Reed *et al.*,[6] namely, Bcl-2 (which includes BH3-only proteins), Caspases, Death Ligands, Death Receptors, BIR (Baculovirus IAP repeat)-containing proteins and Death Effector Domain- or CARD (Caspase Recruitment Domain)-containing proteins. Furthermore, the domain composition of a protein is described by mapping its sequence to domains and motifs in PFAM, Interpro Prosite and SMART databases. When available, links to the protein 3D structures in PDB are provided. Possible physical interactions with other proteins can be obtained through IntAct database.

For all reference species, links to ArrayExpress with information on the expression profiles of the genes are provided. Moreover, information on whether the expression of a gene encoding an apoptotic protein might be regulated by small RNAs is provided through links to miRBase.

Evolution and homology-based entries

Setting all apoptotic proteins within an evolutionary framework is a central goal of DeathBase. Therefore, for all curated proteins we performed extensive homology searches in other 23 eukaryotic organisms. These proteins are included in the database by virtue of their homology relationships and are therefore labeled as 'homology based', in opposition to 'manually curated' proteins. Another feature of the DeathBase is the possibility to obtain information about the evolution of every protein in an easy manner. Users can find homologs of a

protein by performing a one-click Blast search on NCBI or EBI through the link provided next to the sequence of each protein. Moreover, multiple sequence alignments of these homologs and maximum likelihood phylogenetic trees, including homologs from the 28 species surveyed, can be easily viewed interactively or downloaded for further processing. Finally, additional phylogenetic information can be retrieved through the links to phylomeDB,[7] a database providing additional phylogenetic trees and alignments as well as phylogeny-based prediction of duplications and orthology relationships.

Querying and browsing the database

DeathBase allows searches for protein entries based on principal IDs or synonyms. In addition, users can browse and see full lists of proteins involved in each cell death process or pathway, and browse by domain composition or protein family for any given species.

Curation of the database and community comments. A team of database curators has been set up with the participation of experts in cell death in a specific model organism or protein family. Users can provide feedback on how to improve the DeathBase by contacting the curation coordinator or the developers. In addition, a wiki-based community annotation page is provided for each entry, where any user can add comments. Similar to the wikipedia, this resource will allow continuous expansion of the information contained in the database. We encourage scientists and students worldwide to use this resource and contribute to our understanding of the fascinating process of cell death.

Conflict of interest

The authors declare no conflict of interest.

1. Luthi AU, Martin SJ. *Cell Death Differ* 2007; **14**: 641–650.
2. Martin DN *et al. Cell Death Differ* 2007; **14**: 916.
3. Labbe K, Saleh M. *Cell Death Differ* 2008; **15**: 1339.
4. Kroemer G *et al.* 2009 *Cell Death Differ* 2008; **16**: 3.
5. Hitomi J *et al. Cell* 2008; **135**: 1311.
6. Reed JC, Doctor KS, Godzik A. *Sci STKE* 2004 re9-.
7. Huerta-Cepas J *et al. Nucleic Acids Res* 2008; **36**: D491–D496.