Database and Bioinformatic Analysis of BCL-2 Family Proteins and BH3-Only Proteins

Abdel Aouacheria, Vincent Navratil, Christophe Combet

Database and bioinformatic analysis of BCL-2 family proteins and BH3-only proteins

Abdel Aouacheria [1,*], Vincent Navratil [2] and Christophe Combet [3]

[1] ISEM, Institut des Sciences de l'Evolution de Montpellier, Université de Montpellier, UMR 5554, CNRS, IRD, EPHE, Place Eugène Bataillon, 34095 Montpellier, France

[2] PRABI, Rhône Alpes Bioinformatics Center, UCBL, Lyon1, Université de Lyon, Lyon, France.

[3] Centre de Recherche en Cancérologie de Lyon, UMR Inserm U1052, CNRS 5286, Université Claude Bernard Lyon 1, Centre Léon Bérard, Lyon, France.

Corresponding author: abdel.aouacheria@umontpellier.fr

Abstract

BCL-2 proteins correspond to a structurally, functionally and phylogenetically heterogeneous group of regulators that play crucial roles in the life and death of animal cells. Some of these regulators also represent therapeutic targets in human diseases including cancer. In the omics era, there is great need for easy data retrieval and fast analysis of the molecular players involved in cell death. In this article, we present generic and specific computational resources (such as the reference database BCL2DB) as well as bioinformatics tools that can be used to investigate BCL-2 homologs and BH3-only proteins.

Key words: databases, bioinformatics, omics, structure–function relationships, protein domains, protein motifs, BCL-2, BH3, apoptosis, cell death

Running Head: *Databases and in silico tools for BCL-2-ology*

1. Introduction

BCL-2 proteins correspond to a still growing group of regulators which play a major role in cell death (by apoptosis) in animals ***(1,2)***. These proteins are among the most studied in cell biology and represent therapeutic targets especially in cancers ***(3-5)***. This protein group is formed by a family of homologs (structurally and evolutionarily) related to BCL-2 and characterized by the presence of one to four BCL-2 Homology motifs (BH1 to BH4), and by a collection of diverse proteins that harbor only the BH3 motif ***(6,7)***. BCL-2 homologous proteins can function as either anti-apoptotic (such as BCL-2 and BCL-xL) or pro-apoptotic factors (like BAX and BAK) based on their influence on the permeabilization of the mitochondrial outer membrane, a key event in apoptosis induction. BH3-only proteins are pro-apoptotic proteins that act upstream of the mitochondrial events, initiating apoptosis in response to developmental cues or intracellular damages. Three decades of research have elucidated the complex interplay of interactions between the various subgroups of BCL-2 proteins ***(8)***. So-

called ‘activator’ BH3-only proteins (e.g. BIM and tBID) can directly activate BAX or BAK by a transient physical interaction leading to their conversion into membrane-associated dimers that further oligomerize into cytochrome-c permeable pores. In contrast to activator BH3-only proteins, ‘sensitizer’ BH3-only proteins (like BAD and NOXA) trigger mitochondrial permeabilization by docking their α-helical BH3 region into a hydrophobic groove present on the surface of anti-apoptotic BCL-2 homologs, thereby releasing BAX/BAK or activator BH3-only proteins. Anti-apoptotic BCL-2 proteins can also prevent mitochondrial cytochrome-c release by interacting with and inhibiting BAX and BAK. Mechanistic aspects of the regulation of the intrinsic (i.e., mitochondrial) pathway of apoptosis by BCL-2 proteins have been extensively addressed elsewhere and the reader is referred to comprehensive reviews on the topic ***(1,9-12,8)***.

Several lines of evidence have progressively complicated the four-digit classification system described above (pro- *versus* anti-apoptotic BCL-2 homologs, sensitizer *versus* direct activator BH3-only proteins) ***(13)***. First, from a functional perspective, anti-apoptotic BCL-2 proteins can be switched (by cleavage or alternative splicing) to death-inducing factors in certain conditions or in certain cells ***(14-17)***, whereas pro-apoptotic BCL-2 proteins were sometimes reported to protect against apoptosis ***(18)***. Moreover, multiple BCL-2 family members have been described as being either pro- or anti-apoptotic at the time of their initial characterization ***(19-26)***. Most importantly, there is a growing realization that both anti- and pro-apoptotic BCL-2 proteins seem to play non-apoptotic roles (partly or completely) distinct from their roles in cell death, like regulation of mitochondrial morphology ***(27)***, calcium homeostasis, cellular metabolism and autophagy ***(28,29)***. As a corollary, BCL-2 proteins were shown to exert functions at subcellular sites other than the outer mitochondrial membrane (such as the endoplasmic reticulum or peroxisomes). Second, at the structural level, BCL-2 homologous proteins (in their soluble forms) were all found to fold as compact α-helical bundles ***(30,31)***. Extensive sequence divergence and insertions and deletions (indels) of amino acid sequences during evolution represent distinctive features of this protein domain ***(32,7)***, which bears resemblance to

globular bacterial toxins and viral regulators ***(33-35)***. In contrast, BH3 motifs occur both in well-studied (or 'classical') BH3-only proteins (BIM, BAD, BMF, HRK, NOXA, PUMA and EGL-1), which are intrinsically disordered proteins ***(36)***, and in globular BCL-2 homologous proteins and a series of unrelated proteins that have (or are predicted to have) a defined tertiary structure. A more nuanced picture have therefore started to emerge, wherein the BH3 motif could represent a novel type of protein-protein interaction module (i.e., a 'short linear motif' or a 'molecular recognition feature') that spreads beyond BCL-2 proteins ***(6)***. Last, phylogenetically, the BCL-2 family of homologs evolved under a dynamic regime during animal history, with lineage-specific diversification events leading to species-specific gene repertoires ***(37,32,7)***. Although some genes (such as BAK or BCL2L1/BCL-xL) appear to be conserved over relatively large evolutionary distances, others are more divergent (such as those forming the BCL-2 gene complement in early-branching metazoans ***(38,39)***) or are found only in restricted taxa (like BFL-1/BCL2A1 in mammals or BCL-WAV in fishes and reptiles ***(37,40)***). A number of BCL-2-related genes were also acquired by viruses from cellular hosts through gene transfer events ***(41)***. Due to their presence in proteins from families with distinct molecular functions and evolutionary histories, BH3 motifs have probably had a more complex evolution than BCL-2 homologs, undergoing duplication-divergence dynamics, but also random/convergent evolution and exon shuffling ***(6)***.

The multiplicity and (structural, functional and phylogenetic) heterogeneity of proteins forming the contemporary BCL-2 clan poses specific challenges to researchers trying to investigate their structure–function relationships, interaction and regulatory networks. Historically, most studies on BCL-2 sequences and structures have been performed through laborious searches and have focused only on a limited set of genes and proteins. However, nowadays, the high-throughput technologies, including Next-Generation Sequencing (NGS), produce a huge volume of raw data ranging from whole-genome, exome, RNA-Seq or targeted NGS, to gene expression levels (DNA microarrays) and three-dimensional structures (e.g. of BCL-2 proteins in complex with specific ligands). This ever-increasing flow of available data needs to be

adequately clustered and processed to extract useful information and to allow 'BCL-2-ologists' to browse them efficiently. Here, we describe (1) general (2) specialized and (3) dedicated computational resources and tools that can be used to investigate BCL-2 and BH3-containing proteins and (4) illustrate how to perform searches with the reference database BCL2DB (https://bcl2db.ibcp.fr/) ***(42)***. Given the many databases created worldwide ***(43)***, a comprehensive coverage of all available tools and resources is not feasible. Only five classes of specialized databases will be considered that are of general interest to 'BCL-2-ologists': (i) signature databases; (ii); molecular interaction and post-translational modification (PTM) databases (iii) structural databases; (iv) (comparative) genomic databases; (v) transcript databases. Part of the knowledge contained within general and specialized databases has been incorporated into databases dedicated to proteins more specifically involved in cell death regulation or execution, including BCL-2 proteins. These databases have been developed for slightly different purposes, but generally without a focus on the BCL-2 group, except BCL2DB.

2. Materials

All the databases and *in silico* tools mentioned in this chapter are listed in **Table 1**.

3. Methods

3.1. General databases

An overview on how to collect information from the databases described in this article is given in **Figure 1**.

1. As a first step, if the sequence of interest has been collected from bibliographic data, use keyword searches in any of the large-spectrum databases of the International Nucleotide Sequence Database Collaboration (INSDC) (see Note **1**) to get general

information. Primary information can be collected rather easily using these databanks as long as the sequence of interest is properly annotated.

2. For the characterization of novel sequences (e.g. issued from sequencing programs) or unannotated protein-coding genes, perform sequence similarity searches using sequence comparison algorithms such as BLAST or FASTA to identify highly similar sequences (see Note **2**).

3.2. Specialized databases

Specialized databases (also called secondary or derived databases) can have multiple characteristics, but they usually draw upon external primary information (deposited in primary databanks) to provide consistent analytical results, with various (often high) levels of curation and utilization of controlled vocabularies. Integrative databases correspond to interconnected resources that function like 'knowledge hubs'. Specialized and integrative databases offer diverse query options (by keyword, by sequence, by BLAST search, etc.) and have become essential everyday tools to the molecular biologist.

1. Pinpoint entries for BCL-2 domains, individual BH motifs (see Note **3**) and specific proteins (see Note **4**) in signature-based repositories (see Note **5**) such as PFAM, PROSITE, ProDom, SMART and PRINTS (see **Table 1** for a list). InterPro and CDD are integrative databases that provide cross-references to all major signature-based databases and numerous useful features such as taxonomic coverage, structural information, etc.

2. As more and more genomes are being sequenced, (comparative) genomic databases flourish that provide functional annotation of the sequences and links to specialized

resources. For studies involving multiple lineages (see Note **6**), comparative genomic databases such as Ensembl, Ensembl Genomes and EggNOG are useful to investigate the composition of BCL-2 or BH3-coding genes in several species, their genomic organization and evolution. Some databases (such as TreeFam) use algorithms to derive orthology-paralogy relationships from automatically reconstructed trees (see Note **7**).

3. As a next step, use molecular interaction databases to study the molecular interaction network of individual BCL-2 proteins (see Note **8**), get a global portray of the 'BCL-2 interactome' and identify novel putative target protein-protein interactions (PPIs). Classical repositories of PPI data (see Note **9**) are HPRD, IntAct, MINT, DIP and BioGRID, whereas specific platforms enable exploration of PPI networks (e.g. STRING, STITCH and Cytoscape). Lists of chemical modulators of BCL-2 proteins (see Note **10**) can be found in classical PPI repositories or dedicated databases (such as DrugBanK, PubChem, iPPI-DB and ProtChemSI). The integrative database dbPTM provides a rich portal to available databases and tools associated with PTM analyses.

4. PDB coordinates of any protein of interest with a solved structure (see Note **11**) can be obtained from the Protein Data Bank or at the NCBI Structure Group and a lot of tools exist that offer solutions for displaying the 3D structures (such as NCBI-ICn3D, the Jmol applet or stand-alone softwares like PyMol). Full-length BH3-only proteins (as well as certain BCL-2 homologous proteins having a significant degree of local disorder) (see Note **12**) have recently been incorporated into databases of intrinsically disordered proteins (such as DisProt or MobiDB) (see Note **13**).

5. Databases of transcript and cDNA sequences represent interesting resources in the following settings: when genomic information is not available, to compare two or more groups of samples in order to identify differentially expressed genes (across species or

according to tissue types, treatments, disease states or developmental stages) or to identify genetic variants (stored in the specialized database dbSNP). Public repositories that can be used for gene expression profiling comprise the NCBI Gene Expression Omnibus (GEO), EBI ArrayExpress (for microarray data) and The Sequence Read Archive SRA (for NGS experiments). Processing of large amount of data usually requires heavy computational power and data storage capabilities but users can also rely on several biologist-accessible resources for their analyses (e.g. GEO2R, Expression Atlas, AltAnalyze, InsilicoDB, GenePattern and Degust) (see Note **14**). Some of this data has been re-packaged to facilitate studies especially in the cancer field (see for instance the Cancer Genome Atlas initiative, the CBio Portal for Cancer Genomics and the Catalogue Of Somatic Mutations In Cancer, COSMIC).

3.3. Dedicated databases

Several dedicated databases have become obsolete (such as the Apoptosis Database ApoDB, DeathBase and AGIS, the Apoptosis Gene Information System), whereas novel databases have been created (from THANATOS, which is basically a catalog of proteins involved in cell death, to more sophisticated or oriented databases as listed below).

1. Use ApoCanD (Database of Human Apoptotic Proteins in the context of cancer) to collect information about cDNA mutations in cell lines and tumor cells and predict their potential impact on protein sequence (for non-synonymous polymorphisms). The website provides links-out to PDB structures, PFAM entries, SUPERFAMILY identifiers, etc.

2. To explore data from proteome studies in oncology, (with an emphasis on the cell death process), visit the Cancer Proteomics Database (see Note **15**). Cross-

comparison of proteome findings is available for a (limited) number of BCL-2 homologous proteins and BH3-only proteins.

3. For research efforts focused on mechanisms of control of gene expression, browse the online database resource ncRDeathDB2.0 that contains an extensive library of non-coding RNAs (including miRNAs but also other classes of ncRNAs) associated with cell death processes in various organisms, or the miRDeathDB database if the focus is exclusively on miRNAs.

3.4. The BCL-2 Database

BCL2DB is a database and web portal giving rapid access to up-to-date knowledge about BCL-2 family proteins and BH3-only proteins. It holds a collection of annotated sequences and structures of BCL-2-related molecules in a standardized format and provides a variety of tools and external links.

1. First, explore the homepage central menu to choose the functionalities and specific tools relevant to your study. Before navigating any further in the website, we highly recommend that the reader familiarizes with the classification used in BCL2DB (Nomenclature Page) (see Note **16**).

2. Use the *Data* menu to access or download (nucleotide or amino acid) (see Note **17**) single sequences or set of sequences in Fasta format (see Note **18**) and to display color-coded multiple alignments (in ClustalW format) (see Note **19**). By clicking on a particular entry, EMBL flat-file format pages are displayed that contain accession numbers, sequences, keywords, bibliographic references and other associated features. Cross-references to ENA, Ensembl, Ensembl Genomes, Gene Ontology, Human Protein Atlas, PDB, RefSeq, NCBI Taxonomy and UniProtKB are also provided.

The navigation system supports queries by species, gene/protein name and BH motif. Users can select a taxonomic subset of sequences for each subfamily and can easily visualize, edit and download sequence alignments.

3. Download or visualize (see Note **20**) the 3D structures available for your protein(s) of interest by clicking on the *Structures* menu. The result tables provides additional information about the experimental method (i.e., X-ray or NMR) used to obtain the atomic coordinates, the X-ray resolution (in Å), the deposition year, source organism and bibliographic reference.

4. BCL2DB offers two generic analytical tools available through the NPS@ server: BLAST and ClustalW, which are accessible through the *Tools* menu or button. Sequences stored in BCL2DB can be searched with BLAST and selected sequences (or previously compiled sequences) can be extracted (or directly aligned) with ClustalW (see Note **21**).

5. To check your protein sequence of interest for being BCL-2-related and for the presence of a BH1, BH2, BH3 or BH4 motif and pinpoint their exact location along the amino acid sequence, use the 'Annotate' tool (see Note **22**). Sequences can be pasted one by one or a file containing multi-Fasta sequences can be uploaded (maximum size allowance is 20 MB) (see Note **23**). The resulting output displays a number of fields including accession number, sequence name and BH motifs. Each result has a link to its detailed page, which contains more information such as available 3D structures or homology models, sequence, etc.

4. Notes

1. The INSDC comprises the NCBI-GenBank, EBI-ENA and CIB-DDBJ databanks. Sequences available in these three databanks are identical.

2. While these programs constitute a great approach to rapidly find pairs of conserved orthologs in many species, BLAST-like methods suffer from several limitations. In our experience with BCL-2 family members, BLAST results are heavily biased towards highly covered taxa (if no taxonomic filter is applied), the closest hit may not always be the nearest phylogenetic neighbor and remote homologs can frequently be missed. Moreover, these techniques are not suited to mine sequence databanks in search of BH motifs, which correspond to relatively small and sometimes degenerate stretches of amino acids (like the BH3 and BH4 motifs ***(7)***).

3. In the case of BCL-2 homologs and BH3-bearing proteins, there are four such motifs: BH1, BH2, BH3 and/or BH4.

4. Note that some tools tend to aggregate BCL-2 proteins with their binding partners (e.g. BAG family chaperones).

5. These repositories should not be considered as redundant as they differ in several aspects: the methodology used to produce the signatures can be different (e.g. regular expressions, pairwise sequence comparison clustering or profiles) as can be the primary source of sequences (e.g. Swiss-Prot, UniProtKB/TrEMBL or the NCBI RefSeq collection). There are also differences in the information used to classify the proteins (e.g. functional conserved motifs, structural data from the CATH and SCOP resources).

6. When the interest is being focused towards a given species or phylogenetic lineage, we recommend to conduct searches directly in the dedicated genomic databases (often

provided by the consortium that generated the data), as they are regularly updated and provide the most accurate information.

7. In most cases and especially for BCL-2 proteins, which form a heterogeneous and divergent group, it is necessary to (i) carefully select the sequences and species of interest; (ii) calculate correct multiple alignments; (iii) use advanced methods for phylogenetic tree inference. Aberrant phylogenetic trees were published over the years that combined sequences of BH3-only proteins with that from BCL-2 homologous proteins or showed highly uncertain nodes (as inferred from their bootstrap values, when available). By experience, automatic phylogenetic tree inference will not give similar results as those obtained manually. Keep in mind that (i) only homologous positions (originating from a common ancestral site) should be aligned together; (ii) the N-terminal halves of BCL-2 homologous proteins are sometimes highly divergent between paralogs and should preferentially be deleted from whole-family alignments; (iii) it is incorrect to draw evolutionary conclusions from phylogenetic trees calculated from multiple sequence alignments (MSA) of unrelated sequences (e.g. BCL-2 homologs and BH3-containing proteins belonging to other protein families); (iv) often, it is imprudent to align divergent sequences from taxonomically distant species without adequate precautions.

8. Numerous proteins from outside the BCL-2 clan have been reported to interact with and modulate the function of the various subgroups of BCL-2 proteins (e.g. ATP synthase – BCL-xL interaction ***(44,45)***). Specifically, post-translational modifications (PTMs) of BCL-2 proteins (through proteolysis, phosphorylation, acetylation, ubiquitylation, etc.) require interaction with regulatory proteins from other protein families.

9. A dedicated database, Bcl-2-Ome, has been developed to explore the interactome of the most studied BCL-2 proteins (albeit divergent BCL-2 homologs and non-classical BH3-only proteins are missing).

10. During the last ten years, several molecules that mimic the effect of BH3-only proteins (termed BH3 mimetics, like the recently approved drug Venetoclax) ***(3-5)*** were also developed to inhibit anti-apoptotic BCL-2 family members through physical association with their hydrophobic groove, promoting BAX/BAK activation and apoptosis induction.

11. Hundreds of structures of BCL-2 proteins, either alone or in complex with ligands, have been solved since the seminal work of Muchmore and co-workers ***(34)***.

12. Like BCL-2 and BCL-xL that have a flexible loop connecting their BH4 and BH3 motifs.

13. Note that entries are only hardly found for BH3 motifs within databases of linear motifs (with the exception of ELM), because these motifs constitute architectural patterns of globular BCL-2 homologous proteins, and due to the relatively long width of the BH3 signature (which lies near the upper limit of ~20-amino acids instead of 10-12 residues for other linear motifs).

14. For a comprehensive list of tools dedicated to the reuse of public genome-wide gene expression data, see ***(46)***.

15. Also referenced as ApoptoProteomics ***(47)***.

16. Proteins of the BCL-2 group fall into several classes: BCL-2 homologous proteins, viral proteins structurally similar to BCL-2 with or without obvious sequence similarity,

classical BH3-only proteins and other BH3-containing proteins (often referred to as ‘BH3-like’ in the literature).

17. These elements are marked as “F” and “C” in the table cells. “R” is a repertoire that gives information to analyze conserved/variable alignment positions, residue frequencies and Shannon entropy.

18. UniProtKB is mined on a regular basis using a set of proprietary profile HMMs that were implemented to specifically recognize the different BCL-2 family orthology groups and the various clusters of BH3-bearing proteins. If a given sequences does not match any of the constituted groups, it is assigned as ‘unclassified’.

19. In MSA, identical residues (*) are in red, strongly similar residues (:) in green, weakly similar residues (.) in blue and unalike residues in black. These alignments can be interactively edited using the provided ‘EditAlignment’ applet.

20. These options are marked as “D” (download) and “V” (view) in the table cells.

21. In main instances, it makes no sense to align full-length BH3-only sequences with that of full-length BCL-2 homologous proteins, especially to drive conclusions about their phylogenetic relationships as these proteins do not share a same ancestor.

22. This tool is based on home-made BH1-BH4 motif profiles with improved sensitivity and specificity compared to signatures available on signature-based databases.

23. Be careful to use a correctly formatted/supported FASTA sequence header (e.g. ref|accession). If you have doubt, simply use >accession as sequence header and be

sure to use a different accession for each sequence. Upload only text files (.txt) and not files in MS Office binary format (such as .doc or .xls files).

Acknowledgement

This work was supported by CNRS and Montpellier University.

5. References

1. Bhola PD, Letai A (2016) Mitochondria-Judges and Executioners of Cell Death Sentences. Molecular cell 61 (5):695-704. doi:10.1016/j.molcel.2016.02.019
2. Moldoveanu T, Follis AV, Kriwacki RW, Green DR (2014) Many players in BCL-2 family affairs. Trends in biochemical sciences 39 (3):101-111. doi:10.1016/j.tibs.2013.12.006
3. Delbridge AR, Strasser A (2015) The BCL-2 protein family, BH3-mimetics and cancer therapy. Cell death and differentiation 22 (7):1071-1080. doi:10.1038/cdd.2015.50
4. Garner TP, Lopez A, Reyna DE, Spitz AZ, Gavathiotis E (2017) Progress in targeting the BCL-2 family of proteins. Curr Opin Chem Biol 39:133-142. doi:10.1016/j.cbpa.2017.06.014
5. Schenk RL, Strasser A, Dewson G (2017) BCL-2: Long and winding path from discovery to therapeutic target. Biochemical and biophysical research communications 482 (3):459-469. doi:10.1016/j.bbrc.2016.10.100
6. Aouacheria A, Combet C, Tompa P, Hardwick JM (2015) Redefining the BH3 Death Domain as a 'Short Linear Motif'. Trends in biochemical sciences 40 (12):736-748. doi:10.1016/j.tibs.2015.09.007
7. Aouacheria A, Rech de Laval V, Combet C, Hardwick JM (2013) Evolution of Bcl-2 homology motifs: homology versus homoplasy. Trends in cell biology 23 (3):103-111. doi:10.1016/j.tcb.2012.10.010
8. Zheng JH, Viacava Follis A, Kriwacki RW, Moldoveanu T (2016) Discoveries and controversies in BCL-2 protein-mediated apoptosis. The FEBS journal 283 (14):2690-2700. doi:10.1111/febs.13527
9. Czabotar PE, Lessene G, Strasser A, Adams JM (2014) Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. Nature reviews Molecular cell biology 15 (1):49-63. doi:10.1038/nrm3722
10. Delbridge AR, Grabow S, Strasser A, Vaux DL (2016) Thirty years of BCL-2: translating cell death discoveries into novel cancer therapies. Nature reviews Cancer 16 (2):99-109. doi:10.1038/nrc.2015.17
11. Tait SW, Green DR (2013) Mitochondrial regulation of cell death. Cold Spring Harbor perspectives in biology 5 (9). doi:10.1101/cshperspect.a008706
12. Uren RT, Iyer S, Kluck RM (2017) Pore formation by dimeric Bak and Bax: an unusual pore? Philosophical transactions of the Royal Society of London Series B, Biological sciences 372 (1726). doi:10.1098/rstb.2016.0218
13. Aouacheria A (2014) The BCL-2 database, Act 2: moving beyond dualism to diversity and pleiotropy. Cell death & disease 5:e981. doi:10.1038/cddis.2013.511

14. Bae J, Leo CP, Hsu SY, Hsueh AJ (2000) MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. The Journal of biological chemistry 275 (33):25255-25261. doi:10.1074/jbc.M909826199
15. Boise LH, Gonzalez-Garcia M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nunez G, Thompson CB (1993) bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell 74 (4):597-608
16. Cheng EH, Kirsch DG, Clem RJ, Ravi R, Kastan MB, Bedi A, Ueno K, Hardwick JM (1997) Conversion of Bcl-2 to a Bax-like death effector by caspases. Science 278 (5345):1966-1968
17. Kucharczak JF, Simmons MJ, Duckett CS, Gelinas C (2005) Constitutive proteasome-mediated turnover of Bfl-1/A1 and its processing in response to TNF receptor activation in FL5.12 pro-B cells convert it into a prodeath factor. Cell death and differentiation 12 (9):1225-1239. doi:10.1038/sj.cdd.4401684
18. Kiefer MC, Brauer MJ, Powers VC, Wu JJ, Umansky SR, Tomei LD, Barr PJ (1995) Modulation of apoptosis by the widely distributed Bcl-2 homologue Bak. Nature 374 (6524):736-739. doi:10.1038/374736a0
19. Aouacheria A, Arnaud E, Venet S, Lalle P, Gouy M, Rigal D, Gillet G (2001) Nrh, a human homologue of Nr-13 associates with Bcl-Xs and is an inhibitor of apoptosis. Oncogene 20 (41):5846-5855. doi:10.1038/sj.onc.1204740
20. Inohara N, Gourley TS, Carrio R, Muniz M, Merino J, Garcia I, Koseki T, Hu Y, Chen S, Nunez G (1998) Diva, a Bcl-2 homologue that binds directly to Apaf-1 and induces BH3-independent cell death. The Journal of biological chemistry 273 (49):32479-32486
21. Jensen SA, Calvert AE, Volpert G, Kouri FM, Hurley LA, Luciano JP, Wu Y, Chalastanis A, Futerman AH, Stegh AH (2014) Bcl2L13 is a ceramide synthase inhibitor in glioblastoma. Proceedings of the National Academy of Sciences of the United States of America 111 (15):5682-5687. doi:10.1073/pnas.1316700111
22. Kataoka T, Holler N, Micheau O, Martinon F, Tinel A, Hofmann K, Tschopp J (2001) Bcl-rambo, a novel Bcl-2 homologue that induces apoptosis via its unique C-terminal extension. The Journal of biological chemistry 276 (22):19548-19554. doi:10.1074/jbc.M010520200
23. Ke N, Godzik A, Reed JC (2001) Bcl-B, a novel Bcl-2 family member that differentially binds and regulates Bax and Bak. The Journal of biological chemistry 276 (16):12481-12484. doi:10.1074/jbc.C000871200
24. Lee R, Chen J, Matthews CP, McDougall JK, Neiman PE (2001) Characterization of NR13-related human cell death regulator, Boo/Diva, in normal and cancer tissues. Biochimica et biophysica acta 1520 (3):187-194
25. Song Q, Kuang Y, Dixit VM, Vincenz C (1999) Boo, a novel negative regulator of cell death, interacts with Apaf-1. The EMBO journal 18 (1):167-178. doi:10.1093/emboj/18.1.167
26. Zhang H, Holzgreve W, De Geyter C (2001) Bcl2-L-10, a novel anti-apoptotic member of the Bcl-2 family, blocks apoptosis in the mitochondria death pathway but not in the death receptor pathway. Human molecular genetics 10 (21):2329-2339
27. Aouacheria A, Baghdiguian S, Lamb HM, Huska JD, Pineda FJ, Hardwick JM (2017) Connecting mitochondrial dynamics and life-or-death events via Bcl-2 family proteins. Neurochemistry international. doi:10.1016/j.neuint.2017.04.009
28. Gross A (2006) BID as a double agent in cell life and death. Cell Cycle 5 (6):582-584
29. Hardwick JM, Soane L (2013) Multiple functions of BCL-2 family proteins. Cold Spring Harbor perspectives in biology 5 (2). doi:10.1101/cshperspect.a008722
30. Kvansakul M, Hinds MG (2015) The Bcl-2 family: structures, interactions and targets for drug discovery. Apoptosis : an international journal on programmed cell death 20 (2):136-150. doi:10.1007/s10495-014-1051-7
31. Petros AM, Olejniczak ET, Fesik SW (2004) Structural biology of the Bcl-2 family of proteins. Biochimica et biophysica acta 1644 (2-3):83-94. doi:10.1016/j.bbamcr.2003.08.012

32. Aouacheria A, Le Goff E, Godefroy N, Baghdiguian S (2016) Evolution of the BCL-2-Regulated Apoptotic Pathway. In: Pontarotti P (ed) Evolutionary Biology. Springer, Switzerland, pp 137-156. doi:DOI 10.1007/978-3-319-41324-2_9
33. Graham SC, Bahar MW, Cooray S, Chen RA, Whalen DM, Abrescia NG, Alderton D, Owens RJ, Stuart DI, Smith GL, Grimes JM (2008) Vaccinia virus proteins A52 and B14 Share a Bcl-2-like fold but have evolved to inhibit NF-kappaB rather than apoptosis. PLoS pathogens 4 (8):e1000128. doi:10.1371/journal.ppat.1000128
34. Muchmore SW, Sattler M, Liang H, Meadows RP, Harlan JE, Yoon HS, Nettesheim D, Chang BS, Thompson CB, Wong SL, Ng SL, Fesik SW (1996) X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. Nature 381 (6580):335-341. doi:10.1038/381335a0
35. Neidel S, Maluquer de Motes C, Mansur DS, Strnadova P, Smith GL, Graham SC (2015) Vaccinia Virus Protein A49 Is an Unexpected Member of the B-cell Lymphoma (Bcl)-2 Protein Family. The Journal of biological chemistry 290 (10):5991-6002. doi:10.1074/jbc.M114.624650
36. Kvansakul M, Hinds MG (2014) The structural biology of BH3-only proteins. Methods in enzymology 544:49-74. doi:10.1016/B978-0-12-417158-9.00003-0
37. Aouacheria A, Brunet F, Gouy M (2005) Phylogenomics of life-or-death switches in multicellular animals: Bcl-2, BH3-Only, and BNip families of apoptotic regulators. Molecular biology and evolution 22 (12):2395-2416. doi:10.1093/molbev/msi234
38. Bender CE, Fitzgerald P, Tait SW, Llambi F, McStay GP, Tupper DO, Pellettieri J, Sanchez Alvarado A, Salvesen GS, Green DR (2012) Mitochondrial pathway of apoptosis is ancestral in metazoans. Proceedings of the National Academy of Sciences of the United States of America 109 (13):4904-4909. doi:10.1073/pnas.1120680109
39. Lee EF, Clarke OB, Evangelista M, Feng Z, Speed TP, Tchoubrieva EB, Strasser A, Kalinna BH, Colman PM, Fairlie WD (2011) Discovery and molecular characterization of a Bcl-2-regulated cell death pathway in schistosomes. Proceedings of the National Academy of Sciences of the United States of America 108 (17):6999-7003. doi:10.1073/pnas.1100652108
40. Prudent J, Popgeorgiev N, Bonneau B, Thibaut J, Gadet R, Lopez J, Gonzalo P, Rimokh R, Manon S, Houart C, Herbomel P, Aouacheria A, Gillet G (2013) Bcl-wav and the mitochondrial calcium uniporter drive gastrula morphogenesis in zebrafish. Nature communications 4:2330. doi:10.1038/ncomms3330
41. Aouacheria A, Banyai M, Rigal D, Schmidt CJ, Gillet G (2003) Characterization of vnr-13, the first alphaherpesvirus gene of the bcl-2 family. Virology 316 (2):256-266
42. Rech de Laval V, Deleage G, Aouacheria A, Combet C (2014) BCL2DB: database of BCL-2 family members and BH3-only proteins. Database : the journal of biological databases and curation 2014:bau013. doi:10.1093/database/bau013
43. Galperin MY, Fernandez-Suarez XM, Rigden DJ (2017) The 24th annual Nucleic Acids Research database issue: a look back and upcoming changes. Nucleic acids research 45 (9):5627. doi:10.1093/nar/gkx021
44. Alavian KN, Li H, Collis L, Bonanni L, Zeng L, Sacchetti S, Lazrove E, Nabili P, Flaherty B, Graham M, Chen Y, Messerli SM, Mariggio MA, Rahner C, McNay E, Shore GC, Smith PJ, Hardwick JM, Jonas EA (2011) Bcl-xL regulates metabolic efficiency of neurons through interaction with the mitochondrial F1FO ATP synthase. Nature cell biology 13 (10):1224-1233. doi:10.1038/ncb2330
45. Chen YB, Aon MA, Hsu YT, Soane L, Teng X, McCaffery JM, Cheng WC, Qi B, Li H, Alavian KN, Dayhoff-Brannigan M, Zou S, Pineda FJ, O'Rourke B, Ko YH, Pedersen PL, Kaczmarek LK, Jonas EA, Hardwick JM (2011) Bcl-xL regulates mitochondrial energetics by stabilizing the inner membrane potential. The Journal of cell biology 195 (2):263-276. doi:10.1083/jcb.201108059
46. Rung J, Brazma A (2013) Reuse of public genome-wide gene expression data. Nature reviews Genetics 14 (2):89-99. doi:10.1038/nrg3394
47. Arntzen MO, Thiede B (2012) ApoptoProteomics, an integrated database for analysis of proteomics data obtained from apoptotic cells. Molecular & cellular proteomics : MCP 11 (2):M111 010447. doi:10.1074/mcp.M111.010447

Figure captions

Figure 1. Bioinformatic pipeline for the analysis of BCL-2 family proteins and BH3-only proteins. A flowchart summarizing the basic bioinformatics analysis of BCL-2 sequences from different input points to general or specific information (as outputs) extracted from primary or secondary databases (listed in Table 1). Results on individual genes or proteins and at the family level (when the various subgroups of BCL-2 homologs and BH3-containing proteins are considered) can vary between databases.

Table captions

Table 1. Web-based bioinformatics tools and resources for the study of BCL-2 family proteins and BH3-only proteins. The name of the primary or derived databases, their URL, PubMed identifier, content source and principle of implementation are given. When several original publications refer to a given database, only the most recent one is indicated. The mention 'Homolog-specific entries' indicates that the web portal of the database allows access to search function by gene, transcript and protein name. In most cases, signatures or entries for individual proteins (e.g. PF06773 for Bim protein N-terminus, PF06393 for BH3 interacting domain, etc.) are omitted and replaced by this mention. For PFAM, PRINTS, PROSITE and SMART, signature identifiers are indicated together with the number of sequences sharing the signature (between brackets).

Bioinformatic tool or resource (acronym and version)	**URL**	**Full name**	**Citation** (PMID)	**Data(base) source**	**Method (if relevant)**	**Content**	
						BCL-2 homologs	BH3
Primary databanks							
DDBJ	http://www.ddbj.nig.ac.jp/	DNA Data Bank of Japan	27924010	Community		Homolog-specific entries	Homolog-specific entries
EBI-ENA	http://www.ebi.ac.uk/ena	European Nucleotide Archive	27899630	Community			
NCBI-GenBank	https://www.ncbi.nlm.nih.gov/genbank/		23193287	Community			
Specialized databases							
Signature databases							

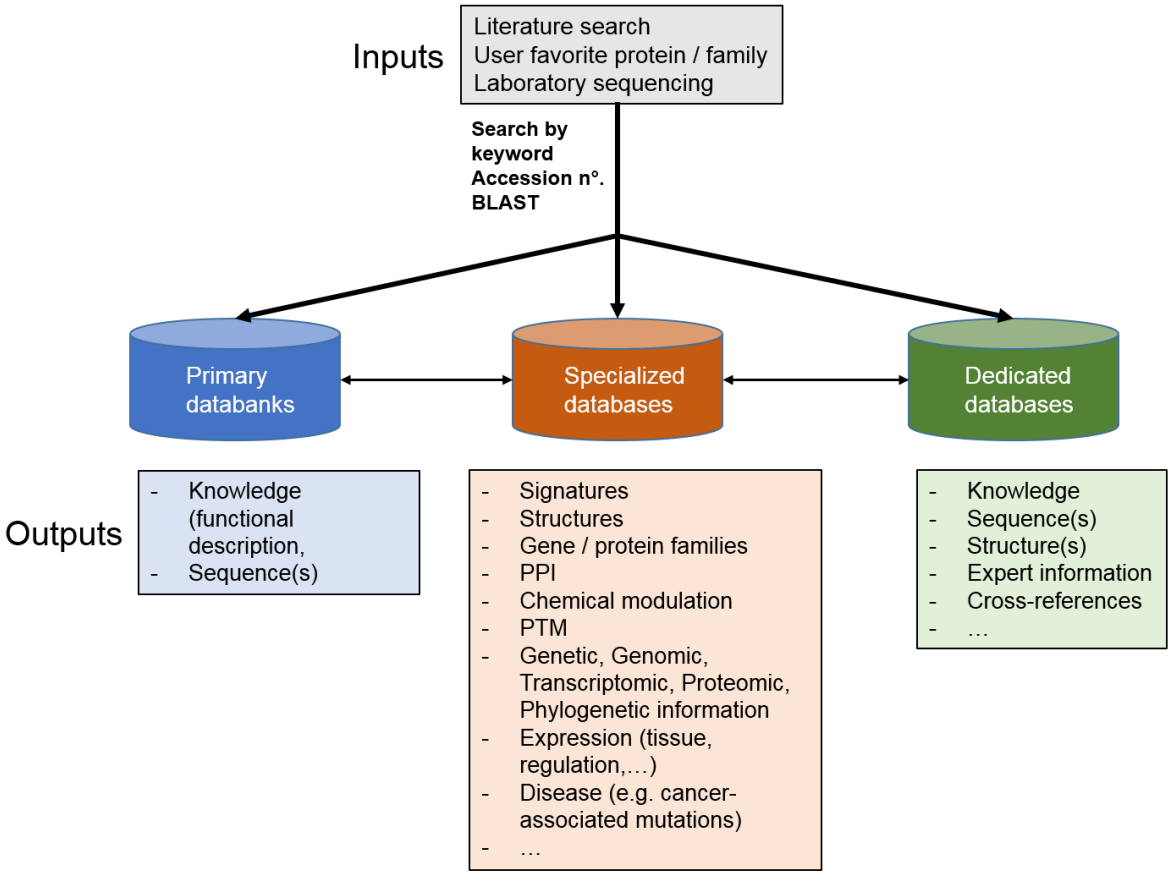
ELM	http://elm.eu.org/	Eukaryotic Linear Motifs	26615199	Expert annotations	Regular expressions		ELME000427
GENE3D	http://gene3d.biochem.ucl.ac.uk/	Database of domain annotations of Ensembl and UniProtKB protein sequences	26578585	UniProtKB/Ensembl	CATH / PFAM assignments, structural clusters and functional families (FunFams)	CATH / PFAM	CATH / PFAM
PANTHER	http://pantherdb.org/	Protein ANalysis THrough Evolutionary Relationships	23193289	Completed genomes	HMMs on functional domains	Homolog-specific entries	Homolog-specific entries
PFAM 31.0	http://xfam.org/	Database of protein families	26673716	UniProtKB	HMMs on functional domains	BCL-2 Family PF00452 (1122) BH4 PF02180 (235) + Homolog-specific entries	Homolog-specific entries
PIRSF	http://pir.georgetown.edu/pirwww/dbinfo/pirsf.shtml	Database of protein families	14681371	UniProtKB	HMMs on functional domains	Homolog-specific entries	Homolog-specific entries
PRINTS 42_0	http://130.88.97.239/PRINTS/index.php	Compendium of protein fingerprints	12520033	UniProtKB	Fingerprints of functional conserved motifs	BCL2 FAMILY PR01862 (24)	
ProDom 2012.1	http://prodom.prabi.fr/	Database of protein domain families	15608179	UniProtKB	Recursive PSI-BLAST searches on functional domains and SCOP information	Homolog-specific entries	Homolog-specific entries
PROSITE Release 2017_07	http://prosite.expasy.org/	Database of protein domains, families and functional sites	23161676	UniProtKB	Patterns, profiles on functional domains	BCL2 FAMILY PS50062 (78) BH4_1 PS01260 (16) BH4_2 PS50063 (19) BH1, PS01080 (44) BH2 PS01258 (45) BH3 PS01259 (30)	BH3 PS01259 (30)
SMART 7.0	http://smart.embl-heidelberg.de/	Simple Modular Architecture Research Tool	25300481	Swiss-Prot, SP-TrEMBL and stable Ensembl proteomes	HMMs on functional domains	BH4 SM00265 (278) BCL BH1-BH2-BH3 SM00337 (1589)	
SUPERFAMILY	http://supfam.org/SUPERFAMILY/	HMM library and genome assignment server	19036790	UniProtKB/SCOP/PDB/InterPro	HMMs on SCOP domains	SCOP entries	
TIGRFAMs	http://www.jcvi.org/cgi-bin/tigrfams/index.cgi	Database of protein families	12520025	UniProtKB	HMMs on functional domains	TIGR00865	
Integrative signature databases							
CDD	https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml	Conserved Domains Database	27899674	NCBI	PSSMs + imported signatures (from Pfam, SMART, COG, PRK and TIGRFAM)	cl02575 cd06845 BH4 cl27450	Homolog-specific entries
InterPro	https://www.ebi.ac.uk/interpro/	Protein sequence analysis & classification	27899635	UniProtKB	Combination of protein signatures (from CDD, Pfam, SUPERFAMILY, PANTHER CATH-Gene3D,	IPR026298 IPR002475 IPR020717 IPR020726	IPR020728 + Homolog-specific entries

					PIRSF and ProDom)	IPR003093 IPR020731 + Homolog-specific entries	
Structural databases							
CATH	http://www.cathdb.info/	Classification of protein structures (Class, Architecture, Topology/fold, Homologous superfamily)	25348408	PDB	Automatic methods and manual curation	1.10.437.10 (Superfamily) + + Homolog-specific entries	
DisProt	http://www.disprot.org/	Database of Disordered Proteins	27899601	Literature	Expert annotations and prediction tools	*BCL-x* DP00449 DP00298 *BCL-2* DP00297	*BAD* DP00563 *BIM* DP00518 *BMF* DP00645
MobiDB	http://mobidb.bio.unipd.it/	Database of protein disorder and mobility annotations	25361972	UniProtKB/DisProt/PDB	Expert annotations and prediction tools	Homolog-specific entries	Homolog-specific entries
NCBI Structure	https://www.ncbi.nlm.nih.gov/Structure/index.shtml	The NCBI Structure Group	22135289	MMDB/CDD/PubChem/NCBI Biosystems database		Homolog-specific entries and small molecules	Homolog-specific entries and small molecules
PDB	https://www.rcsb.org/	Protein Data Bank	27794042	Community		Homolog-specific entries and small molecules	Homolog-specific entries and small molecules
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/	Structural Classification of Proteins	18000004	PDB	Automatic methods and manual curation	56854 56855 + Homolog-specific entries	
SCOPe	http://scop.berkeley.edu/	Structural Classification of Proteins — extended	27914894	PDB	Automatic methods and manual curation	56854 56855 + Homolog-specific entries	
Molecular interaction databases							
BioGRID	https://thebiogrid.org/	Biological General Repository for Interaction Datasets	27980099	Literature	semi-automated text-mining approaches and manual curation	Homolog-specific entries	Homolog-specific entries
DIP	http://dip.mbi.ucla.edu/dip/	Database of Interacting Proteins	11752321	Literature and direct submission	Text mining and manual curation		
DrugBank	https://www.drugbank.ca/	A knowledgebase for drugs, drug actions and drug targets	18048412	GenBank, SwissProt/UniProt, PDB, ChEBI, KEGG, PubChem, PubMed, RxList, PharmGKB and FDA labels	Comprehensive drug and drug target information		BCL-2-like/BAX interaction inhibitors
HPRD	http://www.hprd.org/	Human Protein Reference Database	18988627	Literature	Text mining and manual curation	Homolog-specific entries	Homolog-specific entries
IntAcT	http://www.ebi.ac.uk/intact/	IntAct Molecular Interaction Database	24234451	Literature and direct submission	Text mining and manual curation		
iPPI-DB	http://www.ippidb.cdithem.fr/	Inhibitors of Protein-Protein Interaction Database	26432833	Literature	Data collection and annotation		BCL-2-like/BAX interaction inhibitors
MINT	http://mint.bio.uniroma2.it/	The Molecular INTeraction database	22096227	Literature	Text mining and manual curation	Homolog-specific entries	Homolog-specific entries
ProtChemSI	http://pcidb.russelllab.org/	The database of protein-chemical structural interactions	21573205	PDB	Protein and chemical superimpositions		

PubChem	https://pubchem.ncbi.nlm.nih.gov/	Open Chemistry Database	26400175	545 sources from Chemical and Reagent Vendors, R&D, Governmental Organizations, Curation Efforts and journal publishers	Comprehensive drug and drug target information		BCL-2-like/BAX interaction inhibitors
STITCH	http://stitch.embl.de/	Search tool for interactions of chemicals	19897548	PDSP Ki Database, PDB, KEGG, Reactome, NCI-NPID, DrugBank, MATADOR, GLIDA, PharmGKB, CTD and BindingDB	Text mining and data annotation	Homolog-specific entries	Homolog-specific entries
STRING	https://string-db.org/	Search Tool for the Retrieval of Interacting Genes/Proteins	25352553	BIND, DIP, GRID, HPRD, IntAct, MINT, PID, Biocarta, BioCyc, GO, KEGG, and Reactome	Text-mining and prediction		
Post-translational modification databases							
dbPTM	http://dbptm.mbc.nctu.edu.tw/	An Integrated Resource for Protein Post-Translational Modifications	26578568	Literature, UniProtKB, PDB and dedicated web resources	Data collection, annotation and manual curation	Homolog-specific entries	Homolog-specific entries
(Comparative) genomic databases							
EggNOG	http://eggnogdb.embl.de/#/app/home	Evolutionary genealogy of genes: Non-supervised Orthologous Groups	26582926	Ensembl, UniProtKB, RefSeq, JGI	Data collection, orthology prediction, functional annotation and phylogenetic analysis	Homolog-specific entries	Homolog-specific entries
Ensembl	http://www.ensembl.org/index.html	Automatic annotation on selected eukaryotic genomes	25352552	Public databases, sequencing projects	Species-specific gene sets, comparative, variation and regulatory data		
Ensembl Genomes	http://ensemblgenomes.org/	Extending Ensembl across the taxonomic space	26578574	Genomic data sets, sequencing projects	Genome annotation (genes, variations, sequence conservation)		
TreeFam	http://www.treefam.org/	Database of animal gene trees	24194607	Ensembl, Ensembl Genomes, JGI and Wormbase	HMMs, Multiple Sequence Alignment, phylogenetic analysis		
Transcript databases							
TCGA	https://cancergenome.nih.gov/	The Cancer Genome Atlas initiative	24071849	Genomic data from various types of cancer	From tissue processing to large-scale genomic analysis	Homolog-specific entries	Homolog-specific entries
CBioPortal	http://www.cbioportal.org/	The CBio Portal for Cancer Genomics	22588877	TCGA dataset, literature	Data visualization, analysis and download		
AltAnalyze	http://www.altanalyze.org/	Software for Extended Alternative Splicing Analysis	20513647	User	Analysis and visualization of alternative splicing data in the context of proteins, domains and microRNA binding sites		
COSMIC	http://cancer.sanger.ac.uk/cosmic	Catalogue Of Somatic Mutations In Cancer	27899578	Literature, large-scale genomic screening data	Manual curation, somatic mutation information		
dbSNP	https://www.ncbi.nlm.nih.gov/projects/SNP/	The Single-Nucleotide Polymorphism database	11125122	Public repository of single nucleotide polymorphisms (SNPs) and other variations (small insertions/deletions, microsatellites, short tandem repeats)	Variation retrieval and visualization		
DOR	http://trace.ddbj.nig.ac.jp/dor/	DDBJ Omics Archive	22110025	Public repository of microarray and RNA-seq data, data exchange with EBI ArrayExpress	Archival, retrieval and analytical resources		
Degust	http://degust.erc.monash.edu/	RNA-seq exploration, analysis and visualisation	Unpublished	User	Differential gene expression analysis		
EBI ArrayExpress	https://www.ebi.ac.uk/arrayexpress/	ArrayExpress Archive of Functional Genomics Data	25361974	Public repository of microarray and RNA-seq data, GEO data	Archival, retrieval and analytical resources		

Expression Atlas	http://www.ebi.ac.uk/gxa/home	Expression Atlas	26481351	ArrayExpress	Manual curation and annotation		
GenePattern	https://software.broadinstitute.org/cancer/software/genepattern		16642009	User	Analytical tools for the analysis of gene expression (RNA-seq and microarray), sequence variation and copy number, proteomic, flow cytometry, and network analysis		
ICGC	https://dcc.icgc.org/	International Cancer Genome Consortium Data Portal	20393554	Genomic, epigenomic and transcriptomic data from various types of cancer	Archival, retrieval and analytical resources		
InsilicoDB	https://www.insilicodb.com/	The InSilico DB platform	21937664	User, GEO, TCGA	Data visualisation and analysis tools		
NBCI-GEO	https://www.ncbi.nlm.nih.gov/geo/	NCBI Gene Expression Omnibus	23193258	Public repository of microarray and RNA-seq data	Archival, retrieval and analytical resources		
NCBI-SRA	https://www.ncbi.nlm.nih.gov/sra	Sequence Read Archive	25960871	Public repository of high-throughput sequencing data	Archival, retrieval and analytical resources		
Dedicated databases							
CDP Database	http://apoptoproteomics.uio.no/	The Cancer Proteomics Database	23537399	Literature	Quantitative proteome analyses	Homolog-specific entries	Homolog-specific entries
ApoCanD	http://crdd.osdd.net/raghava/apocand/	Database of human apoptotic proteins in the context of cancer	26861916	COSMIC, CCLE, COLT, PFAM, SUPERFAMILY, 1000 Genomes, dbPTM	Mutation status, copy number variation and gene expression levels, other general information		
BCL2DB	https://bcl2db.ibcp.fr/	The BCL-2 Database	24608034	UniProtKB, Ensembl, ENA, PDB	Profile HMMs, motif prediction, knowledge repository and links to external resources		
Bcl2-Ome	http://for2036.uni-konstanz.de/Bcl2Ome/	A database and interactive web service for dissecting the Bcl-2 interactome	27834951	Literature	Text-mining		
miRDeathDB	http://rna-world.org/mirdeathdb/	A database bridging microRNAs and the programmed cell death	22743998	Literature	Data retrieval, links to external resources		
ncRDeathDB2.0	http://www.rna-society.org/ncrdeathdb/	An all-inclusive information resource on ncRNAs in cell deaths	26431463	Literature	Data retrieval, links to external resources		
THANATOS	http://thanatos.biocuckoo.org/	THe Autophagy, Necrosis, ApopTosis OrchestratorS	Unpublished	Literature	Manual curation, knowledge repository		

Figure 1


Inputs
Literature search
User favorite protein / family
Laboratory sequencing
Search by keyword
Accession n°.
BLAST
Primary databanks
Specialized databases
Dedicated databases
Outputs
- Knowledge (functional description,
- Sequence(s)
- Signatures
- Structures
- Gene / protein families
- PPI
- Chemical modulation
- PTM
- Genetic, Genomic, Transcriptomic, Proteomic, Phylogenetic information
- Expression (tissue, regulation,…)
- Disease (e.g. cancer-associated mutations)
- …
- Knowledge
- Sequence(s)
- Structure(s)
- Expert information
- Cross-references
- …