



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Evolution of Bcl-2 homology motifs: homology versus homoplasy

Abdel Aouacheria<sup>1</sup>, Valentine Rech de Laval<sup>2</sup>, Christophe Combet<sup>2</sup>, and J. Marie Hardwick<sup>3</sup>

<sup>1</sup> Molecular Biology of the Cell Laboratory, Ecole Normale Supérieure de Lyon, LBMC UMR 5239 CNRS – UCBL – ENS Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

<sup>2</sup> Institut de Biologie et Chimie des Protéines, UMR 5086 BMSSI-CNRS/UCBL, Lyon, France

<sup>3</sup> W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA

**Bcl-2 family proteins regulate apoptosis in animals. This protein family includes several homologous proteins and a collection of other proteins lacking sequence similarity except for a Bcl-2 homology (BH)3 motif. Thus, membership in the Bcl-2 family requires only one of the four BH motifs. On this basis, a growing number of diverse BH3-only proteins are being reported. Although compelling cell biological and biophysical evidence validates many BH3-only proteins, claims of significant BH3 sequence similarity are often unfounded. Computational and phylogenetic analyses suggest that only some BH3 motifs arose by divergent evolution from a common ancestor (homology), whereas others arose by convergent evolution or random coincidence (homoplasy), challenging current assumptions about which proteins constitute the extended Bcl-2 family.**

## Spotting diversity in the Bcl-2 portrait gallery

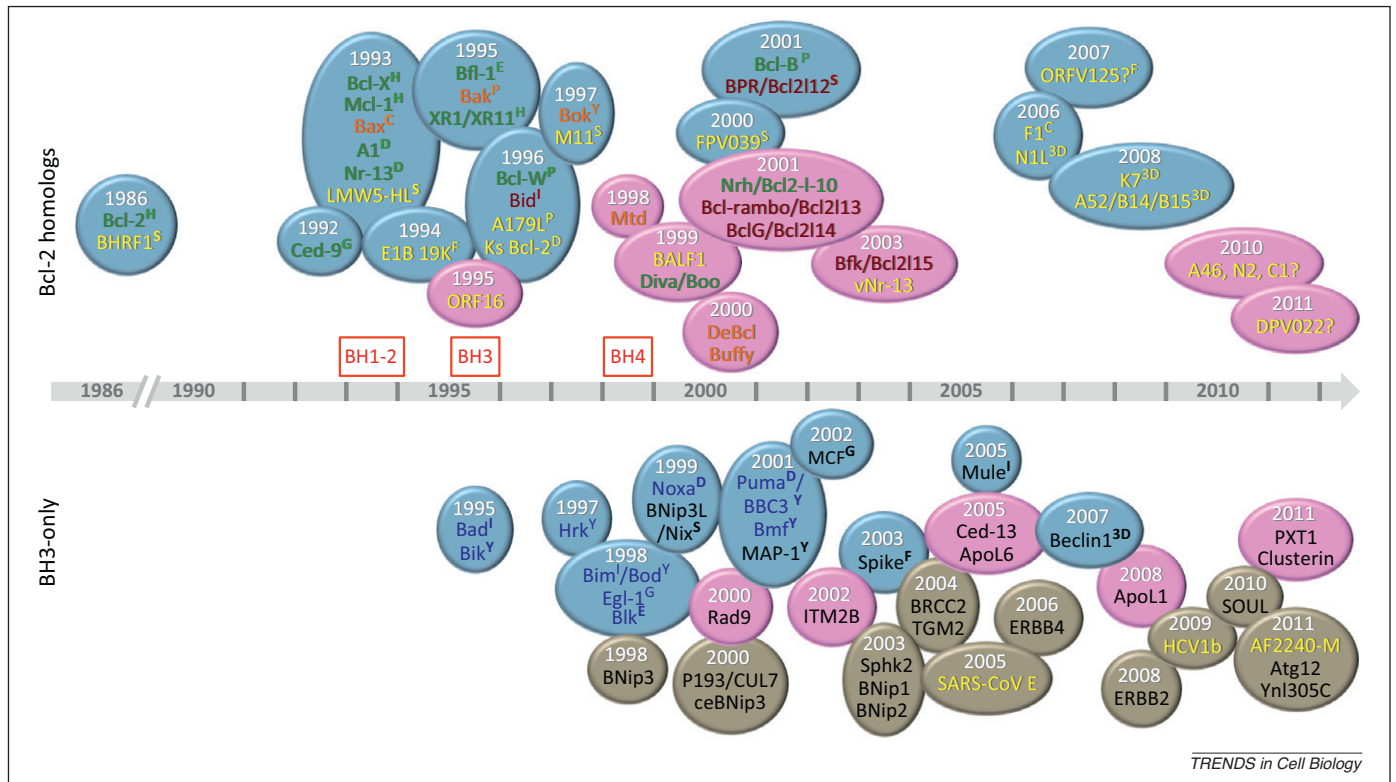
The genome sequence for ancestral species cannot be gleaned from contemporary databases, analogous to the limitations in studying early civilizations that lacked a written language (protohistorical archeology). However, past evolution of a protein family, such as the Bcl-2 family of apoptosis regulators, is protohistorically recorded in the genomes of present-day cells. The lack of ancestral sequences is only one of many obstacles to tracing the evolutionary history of genes and their cellular processes. The chronological order of discovery profoundly influences nomenclature and the interpretation of evidence. Also, there are notorious hurdles in obtaining reliable alignments of divergent (but homologous) or dissimilar (analogous but not homologous) sequences, such as BH3 motifs, for use in searching available databases. Despite limitations, comparisons across species have catapulted forward our understanding of biological processes, exemplified by Bcl-2 family proteins [1]. Thus, cross-species comparisons need to be captured and understood to broaden our biological knowledge of Bcl-2 family members and their central roles in normal development and in cancer.

Bcl-2, the founding member of the eponymous protein family, was discovered more than 20 years ago at the chromosomal breakpoint of the t(14;18) translocation in human follicular B cell lymphomas [2] (Figure 1). This led rapidly to the identification of a family of related proteins [3–5] defined by four unique sequence motifs, termed BH domains (more accurately termed BH motifs) and numbered BH1–4 in their order of discovery. Bcl-2 homologs are evolutionarily conserved throughout metazoans, though it is less clear whether their apoptotic mechanisms (versus other mechanisms) are conserved beyond mammals [6].

In principle, the Bcl-2 family has been subdivided into several subfamilies based on their composition of BH motifs (Figure 2). We took a closer look at the BH nomenclature. Despite the usefulness of this paradigmatic classification scheme, close inspection of BH motif sequences indicates that subfamily classifications draw heavily on prior knowledge of protein functions (for details, see below). Overextension of computational methods and subjective interpretations of sequence similarities have expanded the Bcl-2 family beyond justifiable limits. To gain a better understanding, we considered the possible origins of BH motifs by applying bioinformatics and phylogenetic concepts to more rigorously identify four distinct subfamilies of the extended Bcl-2 family. Three of these (Bcl2-like, Bax-like, and Bid-like) form a monophyletic group (share a common ancestor) (Figures 3 and 4) and a fourth comprises eight unrelated (or very distantly related) canonical BH3-only proteins (Figure 2). Inclusion of these specific eight in the canonical BH3-only subgroup is also heavily dependent on prior knowledge. Bias in Bcl-2 subfamily nomenclature extends beyond amino acid sequence analysis. Assignment to anti-death or pro-death subgroups is challenging for those family members that lack obvious cell death-related phenotypes (e.g., Bok and Bcl2L13) and for those that can exhibit either anti-death or pro-death activity in different conditions or cell types. For example, *Drosophila* Buffy, mammalian Bcl-x<sub>L</sub>, and worm CED-9 are all traditionally categorized as antiapoptotic, but all three have also been shown to exhibit pro-death activity *in vivo* [7–9]. Here we illustrate the methodological and conceptual issues that have arisen in defining and classifying the various other BH3-containing proteins reported to date.

Corresponding authors: Aouacheria, A. ([abdel.aouacheria@ens-lyon.fr](mailto:abdel.aouacheria@ens-lyon.fr)); Hardwick, J.M. ([hardwick@jhu.edu](mailto:hardwick@jhu.edu)).

Keywords: Bcl-2 family; BH3; sequence motifs; protein homology; apoptosis; evolution



**Figure 1.** Discovery timeline for the conglomerate Bcl-2 family. First publication year for proteins with documented or predicted structural similarity with Bcl-2 (above line) and for proteins reported to contain only the Bcl-2 homology (BH)3 motif (below line). (Note that the Bcl-2 protein coding region was published 2 years after the gene was identified in 1984.) For official and alternative protein names, see Figure 2; antiapoptotic Bcl-2-related proteins (green); Bax-like proapoptotic members (orange); Bid-like divergent Bcl-2 homologs (red); canonical BH3-only proteins (blue); other reported BH3-containing proteins (black); viral proteins (yellow). Techniques employed are indicated with superscripts: S, direct sequencing; H, hybridization; P, PCR; E, expressed sequence tag (EST) screen; C, coimmunoprecipitation; I, protein interactive cloning; Y, yeast two-hybrid; D, differential screening; G, genetic screen; F, functional screen; 3D, structural determination. Experimental identification: blue filled circles. Bioinformatics-based prediction of Bcl-2 family membership: pink filled circles. Visual inspection: beige filled circles.

### BH motif definitions not etched in stone

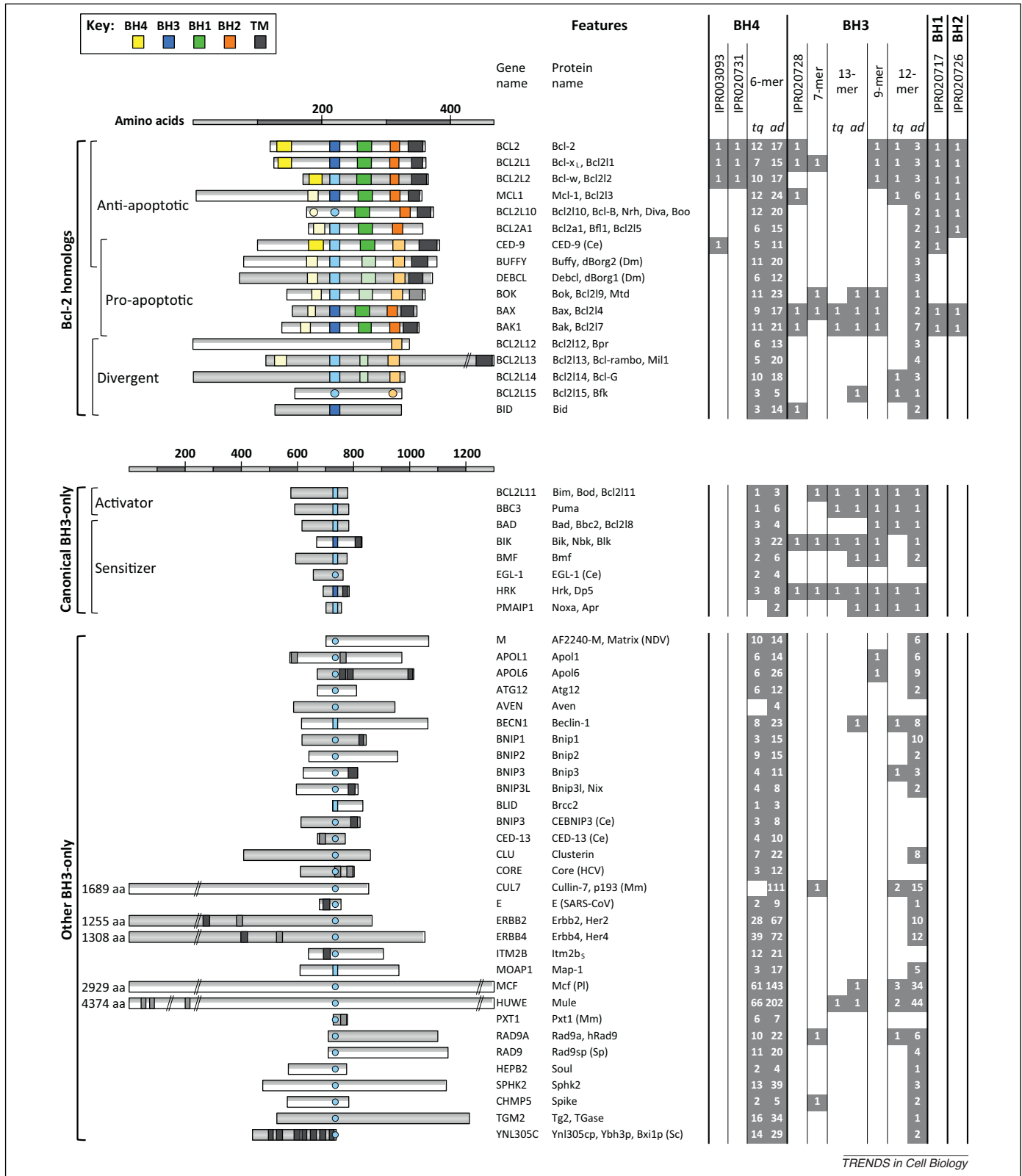
To determine cell fate in response to a wide range of environmental and internal stresses, anti- and pro-death Bcl-2 family members interact in a regulated series of events and conformational changes, the details of which are actively being pursued. 3D structures have been determined for many Bcl-2 family members, revealing a series of approximately nine alpha helices with a central helical hairpin. Although several dimer interfaces have been identified between Bcl-2 family members, the best characterized is the interaction of BH3-only proteins with anti-death Bcl-2 proteins. The BH3-containing helix of BH3-only proteins inserts into a deep elongated binding cleft on anti-death Bcl-2 proteins, resulting in the inactivation of either or both binding partners [10–12]. By inhibiting BH3-only proteins or by directly binding Bax (multi-BH proapoptotic Bcl-2 homolog), Bcl- $x_L$  can prevent homo-oligomerization of Bax that otherwise leads to mitochondrial outer membrane permeability (MOMP). Through MOMP, mitochondrial cytochrome c is released to act as a cytosolic cofactor for assembly of caspase-activating apoptosomes, and cell death occurs within a few minutes [13]. Because antiapoptotic Bcl-2 family proteins promote tumor cell survival, BH3 mimetics have been developed as cancer therapeutics [14].

The N-terminal helix where the BH4 resides in some proteins is present in both anti- and pro-death family proteins and is suggested to mediate interactions with

non-Bcl-2 family proteins [15,16]. BH1 and BH2 motifs are found in both anti- and pro-death Bcl-2 homologs. They flank the central helical hairpin thought to insert into membranes to regulate apoptosis. Together, the BH1, BH2, and BH3 form a binding cleft to capture their own hydrophobic membrane anchor (helix 9). When the tail is displaced, this same binding cleft can be occupied by incoming donor BH3-containing helices from different subclasses of Bcl-2 family proteins or unrelated proteins.

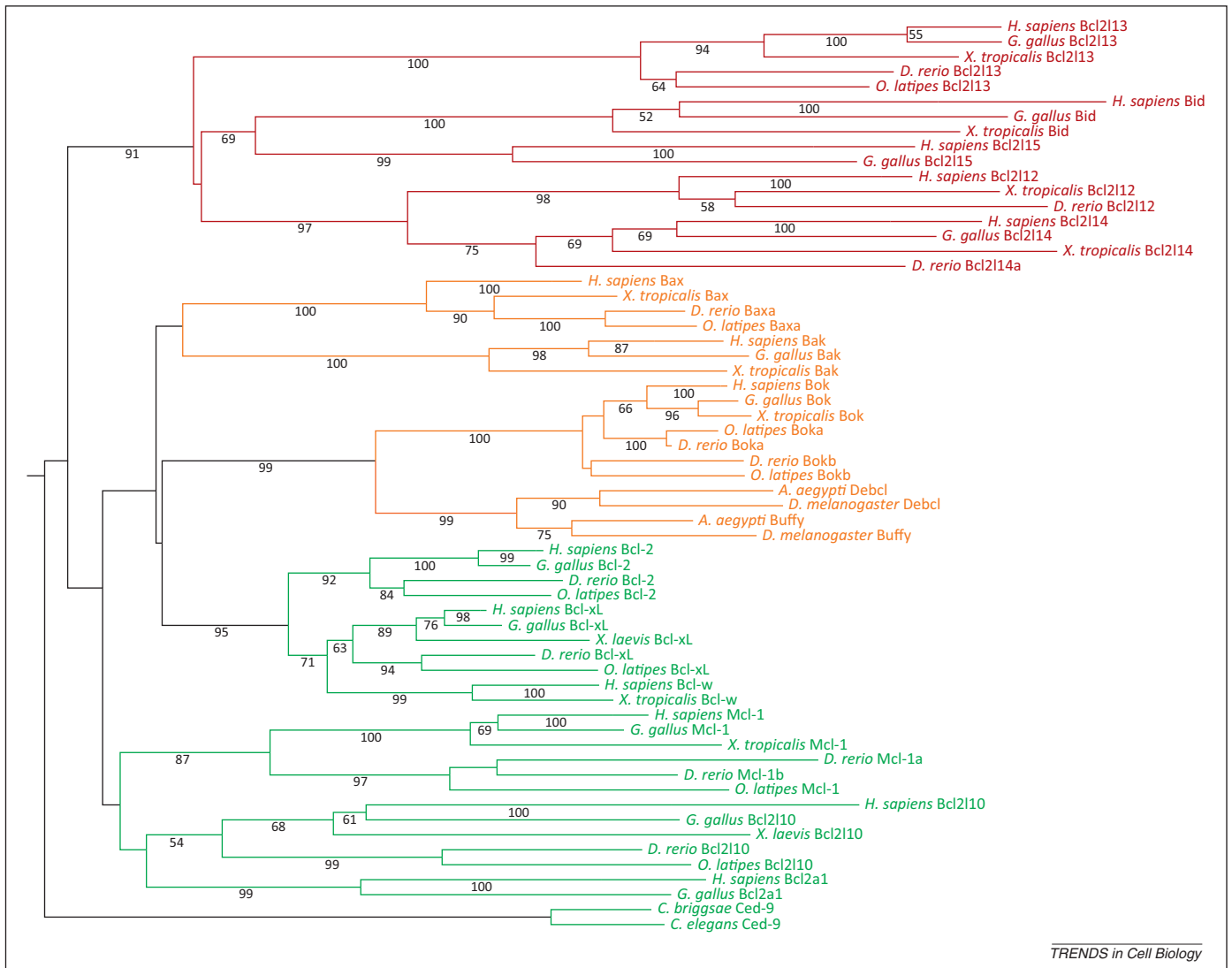
### N-terminal BH4 motifs lack confirmation

The BH4 motif was originally defined using multiple alignments of a subgroup of closely related antiapoptotic proteins (Bcl-2, Bcl- $x_L$ , and Bcl-w) [4]. It is widely assumed that BH4 distinguishes all antiapoptotic Bcl-2 family proteins from pro-death members that lack BH4, but solid evidence for this delineation is lacking. For instance, most multiple alignment programs fail to align the BH4 motifs of the original trio (Bcl-2, Bcl- $x_L$ , and Bcl-w) with the putative BH4 of their antiapoptotic homologs Mcl-1 and Bfl-1/A1 (Figures 2 and S1a in the supplementary information online). Furthermore, when we used InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>), a popular motif-prediction tool that searches a given sequence against a compilation of protein signatures from different databases (e.g., PROSITE, PFAM), BH4 motifs were not identified in these latter proteins or in the other antiapoptotic (Bcl-B/Bcl2L10 and *Drosophila* Buffy) or more distantly related



TRENDS in Cell Biology

**Figure 2.** Evidence for Bcl-2 homology (BH) motif compositions in Bcl-2 family proteins and candidate BH3-containing proteins. Schematic representation of protein size and BH motif composition for Bcl-2 homologous proteins (including Bcl-2-like, Bid-like, and Bax-like subgroups), canonical BH3-only proteins, and a large category of other reported BH3-containing proteins. Note that the various depicted BH3 motifs are supported by heterogeneous experimental evidence (Box 2), while BH1, BH2, and BH4 motifs lack experimental validations in most cases. A three-level color coding system indicates BH type (see color key) and predictability: (i) BH motifs predicted by InterProScan (see color key, meaning BH is predicted); (ii) BH motifs that appear in the NCBI entry for this protein (light shades, meaning BH can be uncertain due to automatic annotation); and (iii) BH motifs not predicted or annotated as above, but nonetheless reported with experimental evidence (colored circles). Transmembrane (TM) segments predicted by one (light gray) or both (dark gray) TMHMM Server v. 2.0 and Phobius web server [51,52]. Total amino acid (aa) number is indicated for proteins not drawn to scale. Abbreviations for non-human proteins: Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; NDV, Newcastle disease virus; Mm, *Mus musculus*; HCV, hepatitis C virus; PI, *Photorhabdus luminescens*; Sp, *Schizosaccharomyces pombe*; SARS-CoV, human severe acute respiratory syndrome (SARS) coronavirus; Sc, *Saccharomyces cerevisiae*. Number of BH motif hits (right): the number of times a BH motif is predicted to occur within each protein based on the following searches (>1 denotes ambiguity). Predefined amino acid sequence signatures from InterPro (IPR) [53], from published patterns *talis qualis* (tq), and from patterns inferred from these



**Figure 3.** Phylogenetic tree of Bcl-2 homologous proteins. Neighbor-joining tree calculated using Poisson-corrected amino acid distances on 74 sites in the core Bcl-2 domain encompassing the Bcl-2 homology (BH)3 motif (or homologous site) through BH2. The tree was rooted with CED-9 sequences from *Caenorhabditis*. Numbers indicate bootstrap percentages after 1000 replications; values below 50% are not reported. Branch lengths are proportional to distances between sequences. Sequence names contain abbreviations of genus and species (see supplementary information 2 online); primarily antiapoptotic Bcl-2-related proteins (green), primarily proapoptotic Bax-like members (orange), and Bid-like divergent homologs (brown).

proapoptotic members (Bax, Bak, and Bok) (Figures 2 and S1b,c in the supplementary information online). These sequence differences are paralleled by the sensitivity of the same three BH4-containing proteins to the BH3 mimetic ABT-737, although ABT-737 does not directly contact the BH4. However, all (uncleaved) Bcl-2 family proteins contain the N-terminal helix that encompasses the BH4 motif [17].

Lack of conformity regarding BH4 nomenclature is understandable given that the conserved domain database (CDD) at the National Center for Biotechnology Information

(NCBI) appears to predict a more widely conserved BH4. This BH4 is annotated in both antiapoptotic (including Mcl-1, BFL-1/A1, and Buffy) and proapoptotic members (Bax, Bak, BOK, and Bcl2L13). However, the CDD identifier (132900) refers to a protein family (PFAM) signature that captures similarities across all the Bcl-2 family (PF00452), leading to the over-prediction of BH4-containing proteins and misannotation. A more recent effort was made to re-define BH4 as a motif present in most globular Bcl-2 family proteins, including both pro- and antiapoptotic members [18]. However, when we used this novel sequence signature

publications *additum* (ad) were used to search the indicated protein and the number of exact matches to the following sequences is reported (shaded cells; unshaded cells had no hits): IPR003093 for BH4; IPR020731 for BH4 is from the PROSITE pattern PS01260 ([DS]-[NT]-R-[AE]-[LI]-V-X-[KD]-[FY]-[LIV]-[GHS]-Y-K-L-[SR]-Q-[RK]-G-[HY]-X-[CW]); 6-mer BH4 tq ([ILVCGMFYWHKT]-[ILVCGMFYWHKT]-X2-[FYWH]-[ILVCGMFYWHKT]) [18]; 6-mer BH3 ad ([ILVCGMFYWHKT]-[ILVCGMFYWHKTE]-X2-[FYWH]-[ILVCGMFYWHKT]); IPR020728 BH3 is from PS01259 ([LIVAT]-X3-L-[KARQ]-X-[IVAL]-G-D-[DESG]-[LIMFV]-[DENSHQ]-[LVSHRQ]-[INSR]); 7-mer BH3 (L-X3-G-D-E); 9-mer BH3 ([LM]-[ERAKQL]-[RAQICEY]-[MILAFSE]-[ASG]-D-[DEQKGR]-[LFMIVD]-[NDVHSEQA]) [54]; 12-mer BH3 tq ([ILVCGMFYWHKT]-X3-[ILVCGMFYWHKT]-[KR]-X2-[AGS]-[DE]-X-[ILVCGMFYWHKT]) [55]; 12-mer BH3 ad ([ILVCGMFYWHKTE]-X3-[ILVCGMFYWHKT]-[KRAQL]-X2-[AGS]-[DE]-X-[ILVCGMFYWHKT]); 13-mer BH3 tq ([ILVCGMFYWHKT]-[AGS]-X2-[ILVCGMFYWHKT]-X2-[ILVCGMFYWHKT]-[AGS]-D-[ED]-[ILVCGMFYWHKT]-[NHDY]) [56]; 13-mer BH3 ad ([ILVCGMFYWHKTE]-[AGSC]-X2-[ILVCGMFYWHKT]-X2-[ILVCGMFYWHKT]-[AGS]-D-[EDKQL]-[ILVCGMFYWHKT]-[NHDEY]); IPR020717 BH1 from PS01080: [LVMENQ]-[FTLS]-X-[GSDECO]-[GLPCKH]-X1/2-[NST]-[YW]-G-[RK]-[LIV]-[LIVC]-[GAT]-[LIVMF](2)-X-F-[GSAEC]-[GSARY]); and IPR020726 BH2 from PS01258 (W-[LIM]-X3-[GR]-G-[WQ]-[DENSAV]-X-[FLGA]-[LIVFTC]) (<http://www.ebi.ac.uk/interpro>).



### Box 1. Limited tools for defining BH3 motif-containing proteins

**Informed visual inspection.** BH3 motifs are typically identified by visual inspection for the characteristic Leu-X3-Gly-Asp/Glu sequence. This strategy has proven successful when searching for newly identified Bcl-2 binding partners, but without additional information this method is difficult to justify.

**BLAST searches.** Some groups have used the pairwise sequence comparison program BLAST to define BH3 motifs [57,58]. However, BLAST was not designed for short peptide sequences.

**Motif searches.** Published BH3 motif searches have relied mainly on deterministic sequence patterns, such as the regular expression PS01259 of the PROSITE database (InterPro signature IPR020728). When we applied this pattern to search the database (UNI60/UniProtKB with sequences  $\leq 60\%$  identity to avoid multiple copies of the same sequence), it recovered only two of the eight canonical BH3-only proteins, BIK and HRK (see Figure 2 in main text and Figure S2 in the supplementary information online). Several groups have defined less stringent consensus sequences in attempt to capture more of the canonical BH3-only proteins [59]. Applying their 7-mer as a search criterion, we identified only one additional BH3-only protein (Bim) and 33,558 additional putative hits (see Figure 4 in main text and Figure S2 in the supplementary information online). Similar problems arise with the reported 9-residue [54], 12-residue [55], and 13-residue regular expressions [56], one of which identified 25% of all proteins. The situation deteriorates if reported consensus patterns are

modified *additum* (ad) to capture all of the BH3 sequences that were included in their respective published alignments (see Figures 2 and 4 in main text and supplementary information 3 online). By extending this search to plants, microorganisms and viruses, we identified putative BH3-only proteins in species that lack Bcl-2 homologs with which to interact (e.g., *Chlamydomonas*, *Arabidopsis*) and that lack true mitochondria (e.g., *Giardia*, *Trichomonas*). Furthermore, there was strikingly little overlap between hit lists, and the number of hits was not significantly different when the proteomes were shuffled or reversed (rearrangements of the same amino acids) (see Figure 4 in main text and supplementary information 3 online). Thus, the number of BH3 motifs identified by these strategies is not significantly different from random chance.

**Hidden Markov models.** Two studies reported using the probabilistic method of profile hidden Markov models (Profile HMMs), which is known to be more sensitive and specific. This method identified a BH3 in PXT1 (Q8K459), but the assigned BH3 PFAM domain is untraceable and may not exist, and the other HMM study provided no details regarding the bioinformatic approaches [60,61]. Our experience with building HMM profiles with BH3-like proteins invariably leads to profile drift in iterated searches, indicating that these sequences collectively contribute noise. Thus, there is a lack of evidence that BH3 motifs of the Bcl-2 family proteins are significantly similar in sequence (Figure S1f,g in the supplementary information online).

for database searching, it produced high false-positive rates, including multiple BH4 motifs within individual Bcl-2 family proteins (Figure 2). Defining BH4 can be further complicated by divergent sequences present on either side of helix 1 (e.g., the long N-terminal extension of Mcl-1 and the large unstructured loop between the BH4 and BH3 of Bcl-x<sub>L</sub>) (Figures 2 and S1c in the supplementary information online). Although these difficulties do not exclude the presence of an evolutionarily conserved BH4 region among all Bcl-2 paralogs (e.g., defined by specific structural contacts), currently available bioinformatics has not clearly justified this classification as a sequence motif.

#### BH1–BH2 motifs define Bcl-2 family subgroups

Most alignment algorithms accurately align the paired BH1 and BH2 motifs that flank the central helical hairpin of both anti- and proapoptotic Bcl-2 family proteins (Figure S1c in the supplementary information online). However, the Bid-like clade of Bcl-2 proteins, which also includes Bcl2L12, Bcl2L13 (Bcl-Rambo), Bcl2L14 (Bcl-G), and Bcl2L15 (Bfk), harbor dissimilar sequences at the equivalent BH1 or BH2 position [19,20] (Figures 2 and 3). Bcl2L13 is encoded in a tail-to-tail orientation with Bid in vertebrate genomes, suggesting that these two paralogs arose from a tandem inverted duplication, followed by functional divergence based on their antisymmetric distribution of BH motifs. Bcl2L13 has recognizable BH1 and BH2 motifs but a less conserved BH3, the antithesis of Bid (Figure S1b in the supplementary information online). Bid has a structurally well-defined and functionally important BH3 motif and has also evolved unique functions in apoptosis [21,22] and in the DNA damage checkpoint [23–25]. By contrast, Bcl2L13 may have lost its role in apoptosis regulation [26].

#### BH3 homology versus homoplasy

The short BH3 motif of Bcl-2 family proteins is a special case because it is present in pro- and anti-death Bcl-2 homologs and in a diverse set of structurally and biologically unrelated

proteins, only eight of which are included in the extended mammalian Bcl-2 family, based on traditional criteria. These eight (plus worm EGL-1) constitute the canonical BH3-only subfamily of the Bcl-2 family and are further subdivided into two groups based on their ability to directly (Bid, Bim, and Puma) or indirectly (Bad, Bik, Bmf, Hrk, Noxa, and in principle *Caenorhabditis elegans* EGL-1) activate the pro-death function of Bax to kill cells (Figure S1d,e in the supplementary information online). In addition, there is an ever-growing list of disparate cellular and viral proteins reported to contain BH3 motifs that regulate apoptosis (Figures 1 and 2, and S1f,g in the supplementary information online).

Although it is widely assumed that all BH3-only proteins are unrelated except for their BH3, conservation of sequence similarity in their BH3 motif has not been rigorously confirmed (or dispelled, as seen in Box 1). Although underappreciated, the best-characterized BH3-only protein Bid (defined as lacking BH1, BH2, and BH4) in fact shares significant sequence similarity with Bcl-2 homologous proteins, consistent with its Bcl-2-like 3D structure [19,21,22]. Thus, in contrast to current nomenclature, Bid does not belong to the subgroup of canonical BH3-only proteins, but instead belongs to a branch of the Bcl-2 phylogenetic tree with Bcl2L12, -13, -14, and -15, which generally lack C-terminal transmembrane domains (Figures 3 and S1b in the supplementary information online). Bid is an example of misclassification by applying the BH motif criteria alone; additional evidence strongly argues for an evolutionary relationship (homology) between Bid and the Bcl-2 family. Yet the criterion for direct activator-type BH3-only proteins is defined by the biochemical properties of Bid (activated/caspase-cleaved tBid or its BH3 peptide). Thus, Bid may have additional functions in apoptosis regulation, beyond its BH3.

In contrast to Bid, Beclin 1 (yeast Atg6) and Atg12 are not Bcl-2 family members because they lack overall amino acid sequence similarity, consistent with their lack of

structural similarity to Bcl-2 [27,28] (Figure 2). However, a large body of compelling functional and structural evidence supports the existence of a BH3-like motif in Beclin 1. Through this BH3-mediated interaction of Beclin 1 with Bcl-2 proteins, autophagy appears to be functionally linked to the control of apoptosis. Although it seems obvious that the BH3 motifs of Bcl-2 and of Beclin 1 have distinct origins, it is also possible that some BH3-only proteins currently classified as Bcl-2 family members have BH3 sequence similarities due to either random coincidence or convergent evolution (selection of the same residues for identical functional reasons). To further clarify, there is no denial that the core BH3 motif sequence Leu-X(3)-Gly-Asp is present in many proteins and can be aligned by hand in a diagram, but this does not imply that they are homologous or are functionally equivalent. Conversely, the lack of verifiable sequence similarity using currently available bioinformatics does not negate the potential importance of BH3-like motifs in biology. There may be other cases where divergence accumulated over evolutionary time-scales, producing proteins that will exhibit similarities only in this short region. Thus, the denomination of BH3-only protein carries some degree of ambiguity because many proteins that logically qualify for this designation belong to different protein families and contain other (non-BH) motifs.

### Tracing the evolution of Bcl-2 family proteins

Several different evolutionary mechanisms may explain the conglomerate Bcl-2 protein family. Numerous viruses, including herpesviruses such as Epstein–Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV), encode compact Bcl-2 homologous proteins (BHRF1 and KS-Bcl-2, respectively) with significant sequence similarity in the BH1–BH2 motifs and high structural similarity [29,30]. The finding that Bcl-2 homologs are encoded in the genomes of several different families of large DNA viruses, probably as a result of independent horizontal gene transfer (HGT) events from their respective host cells, suggest that the Bcl-2 structural fold can be acquired other than by vertical descent.

In contrast to the structurally defined Bcl-2 protein fold, which forms a single protein domain (helical bundle), individual BH motifs do not constitute separate protein domains [17]. Furthermore, sequence similarity of BH3 is not sufficient to automatically infer common ancestral origins [31] (Box 1). This raises the question of how BH3 motifs arose in so many diverse proteins. The most logical assumption to explain the functionally and structurally validated BH3 motifs in unrelated proteins is that they arose independently on multiple occasions (homoplasmy) owing to specific constraints (either biological or physico-chemical), facilitating interactions with Bcl-2 family proteins. The surprising proteome-wide prevalence of potential BH3 motifs (Box 1) suggests that they could correspond to a particular class of short linear motifs [32] used as a binding interface to connect critical components of various biological pathways to the core apoptotic machinery. As in other short linear motifs, the short length of the BH3 signature and the small number of residues involved in functional interaction make BH3 motifs likely

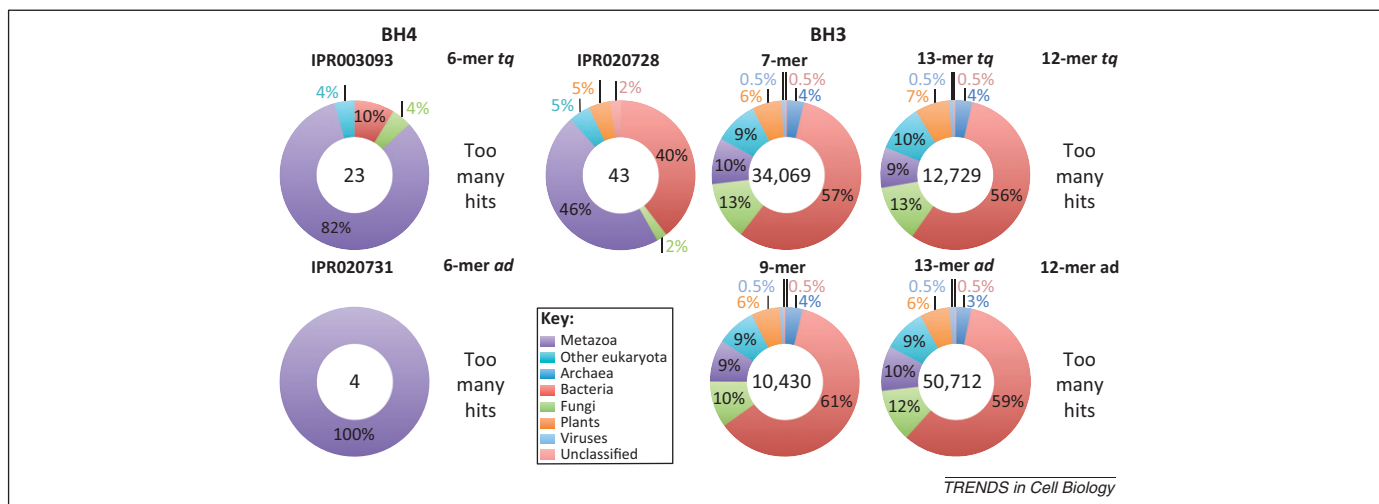
to appear and disappear in unrelated protein sequences by random mutation. This scenario does not exclude the possibility of divergent evolution in some instances, such as between the BH3-coding exons of the proapoptotic Bcl-2 homolog Bak and the BH3-only protein Bik, as we previously hypothesized [19,20]. Moreover, because 3D structure evolves more slowly than sequence, remote homologs may be found among canonical/other BH3-containing proteins as more protein structures are solved.

However, structural similarities do not always reflect homologies. Convergent evolution also could explain the structural similarity between the central helical hairpin of Bcl-2 homologs and that of the needle protein PrgI of *Salmonella typhimurium* [33] or the pore-forming domains of some bacterial toxins (e.g., colicins and diphtheria toxins) [34]. Until additional evidence is generated, it seems reasonable to consider that the existence of this super-secondary structure reflects a common structural requirement that is imposed on membrane-active proteins by thermodynamics.

Despite considerable effort, so far no Bcl-2/Bax/Bid homologs have been found outside the metazoan phyla, apart from animal viruses [19,35], indicating that these proteins are not present in non-metazoan species or that ancestral Bcl-2 homologs in non-metazoan species have become so dissimilar that their common origin cannot be detected from their sequences. Similar difficulties challenge the evolutionary origins of RNA viruses, where high mutation rates obscure ancestral origins [36]. It is hoped that advances in structural genomics and remote homology detection coupled with the growing size of sequence databases will make it possible to capture more ancient evolutionary events and possibly reveal connections between

### Box 2. Criteria for including or rejecting candidate BH3 proteins

Definitive bioinformatic procedures that reliably predict new BH3-containing proteins are currently lacking (Box 1) and new, improved methods are needed. Although previously reported BH3 search strategies potentially narrow the list of candidates somewhat, there is little certainty of this. Similar issues plague our ability to accurately predict microRNA targets [62]. Even with improved bioinformatics, additional criteria are needed to avoid logical contradictions and false positives. Using Bid-like proteins as an example, it is reasonable to consider inclusion of proteins with Bcl-2-like 3D structures even in the absence of recognizable BH motifs [63]. Structural determinations and evidence for direct interactions of putative BH3 peptides with confirmed Bcl-2-homologous proteins using fluorescence polarization, nuclear magnetic resonance spectroscopy, or surface plasmon resonance have been applied to confirm new BH3 motifs, in addition to cell-based evidence to validate biochemical studies [64]. BH3 candidacy would also be supported by functional conservation in closely related species and by the pairing of a BH3 candidate with its target Bcl-2 family member in the same species or perhaps with an invading intracellular pathogen. In the absence of a host–parasite relationship, candidate BH3-only proteins encoded by non-metazoan species lacking Bcl-2 homologues or true mitochondria are difficult to justify (Box 3). There are potentially many clinically relevant peptides and small molecules that fit into the hydrophobic pockets of various Bcl-2 family proteins and mimic natural ligands. Although not derived from a natural protein, the case can be made to include peptides selected from an artificial library as *bona fide* BH3 motifs [65], with the potential appellation of 'BH3-like', 'BH3-mimetic', or 'BH3 functional analog'.



TRENDS in Cell Biology

**Figure 4.** Bcl-2 homology (BH4 and BH3) motif searches identify improbable numbers of hits throughout major phylogenetic groups. InterPro signatures and patterns derived from published alignments (Figure 2) were used to scan UniProt KnowledgeBase with reduced redundancy (no more than 60% identity, UNI60; release 2011\_12) for sequences containing BH3 or BH4 motifs. Pie charts show the percent of proteins in each of the phylogenetic groups that were identified as hits by the indicated signature search. The total number of hits detected is shown within the pie chart.

previously unrecognized Bcl-2-like proteins. In the meantime, we can only attempt to avoid a mammalian-centric perspective. Indeed, interacting proteins are known to evolve and coevolve in a lineage-specific manner [37] and interacting proteins often display similar evolutionary histories [38,39] (Box 2).

It is intriguing that mammalian Bcl-2 family proteins can inhibit (e.g., Bcl-x<sub>L</sub>) or induce (e.g., Bax) cell death when expressed separately in cells from multiple kingdoms, including yeast and other species lacking Bcl-2 homologs [40–43]. This could conceivably reflect conservation of function if mammalian Bcl-2 family proteins

interface with highly conserved cellular machinery. Consistent with this model, Bcl-x<sub>L</sub> and Mcl-1 were recently reported to interact with the highly conserved mitochondrial ATP synthase [44,45]. One potential implication would be that yeast and bacteria encode as yet unrecognized Bcl-2-shaped proteins explaining why they maintain the ability to interact with mammalian Bcl-2 proteins. Pursuit of these answers is seriously compromised until such putative Bcl-2-like proteins are identified in non-metazoans, and there are equally valid alternative explanations that do not evoke conservation of function. Ectopically expressed mammalian Bcl-2 proteins may use different mechanisms to modulate cell survival/death in yeast by acting on distinct effector proteins. Alternatively, Bcl-2 proteins could act on lipid membranes, such as changing membrane curvature in a manner that does not require specific interactions with foreign proteins. Experimental investigation of yeast is likely to provide tremendous insight into many different cell-death mechanisms and may even reveal as yet unknown mechanisms of mammalian Bcl-2 proteins [46,47]. However, the epistemic goal in this case would be different: to uncover the mechanisms of action of Bcl-2 family proteins, which is fundamentally different from establishing phylogenetic inferences (Box 3).

Although the search for genes controlling cell death has benefited extensively from studies using model organisms like *C. elegans* (which possesses only one Bcl-2 homolog, CED-9), our views on cell death evolution mainly result from biased sampling of the tree of life. The decoded genome sequences of the sponge *Amphimedon queenslandica*, the cnidarians *Nematostella vectensis* and *Hydra vulgaris*, the echinoderm *Strongylocentrotus purpuratus*, the urochordate *Ciona intestinalis*, and the cephalochordate *Branchiostoma floridae* revealed that these ‘early-branching’ metazoan species have a large and unique set of Bcl-2 homologs [48–50]. These findings destabilize the idea of CED-9 being a prototypical multi-BH protein and that simpler metazoans are necessarily less equipped than

### Box 3. Do yeasts have BH3-only proteins?

One case in point is the candidate yeast BH3 protein Ybh3/Bxi1 (YNL305C) from *Saccharomyces cerevisiae*, which is a member of the BI-1 protein family predicted with high confidence to be a multimembrane-spanning protein [66–68]. The reported assignment of a BH3 motif to yeast Ybh3 was made by visual inspection and the multiple alignment reveals only two fully conserved residues (Leu and Asp) followed by a less conserved amino acid (Asp or Glu). Using the sequence L-X(4)-D-[DE] to search against the proteome of *S. cerevisiae* yields 2575 hits (in 1834 sequences) and the more stringent pattern [LTYIE]-X-[EQARS]-X-L-[KQRA]-X(3)-D-[EKDS] derived from the reported alignment yields 160 matches (in 156 proteins). This candidate yeast BH3 sequence is also somewhat truncated because it is located at the C terminus of the Ybh3 protein and overlaps one of six transmembrane segments, two unprecedented features among BH3-containing proteins. Furthermore, the putative BH3 sequence lacks conservation in orthologous proteins, potentially challenging its functional importance. Although this BH3 was reported to bind Bcl-x<sub>L</sub> and to promote yeast cell death, two other studies have reported that this yeast gene is pro-survival, not pro-death, consistent with BI-1 function in other species [67–69]. More perplexing is how this yeast protein might ever encounter human Bcl-x<sub>L</sub>. Still it is conceivable that non-metazoan species will be found to encode proteins that closely resemble the 3D structure of Bcl-2 family proteins to serve as a potential target of this yeast BH3 [70]. Nevertheless, currently available tools to define BH3 motifs in non-metazoans suggest that Ybh3 is just as likely to reflect a fortuitous event, although this does not negate its utility in research if it binds tightly and specifically to the cleft of Bcl-2.



complex ones. The apparent conservation of apoptotic functions in *Hydra* Bcl-2 family proteins, but not in *Drosophila* Bcl-2 proteins, is even more puzzling [6].

Furthermore, all of the organisms listed above are modern representatives of early-branching lineages that have evolved along their own lines and that have developed their own repertoire of Bcl-2 family genes, and do not themselves represent 'ancestral' or 'primitive' metazoan species.

### Concluding remarks

Sequence analyses indicate that Bcl-2 family proteins comprise three clades (Bcl-2-like, Bax-like, and Bid-like) that share 3D structural folds and common ancestry despite lacking BH motifs, and a fourth fast-growing group of phylogenetically unrelated proteins with limited sequence similarity in the BH3 motif. The examples presented here illustrate the difficulty in interpreting functional versus phylogenetic similarities of the diverse BH3 sequence motifs reported to date, and in handling the potentially huge number of predicted BH3-containing proteins. More reliable strategies are therefore needed to identify and classify functional BH3 motif-containing proteins by defined and generally acceptable standards. Distinguishing evolutionary similarities (homology) from other similarities (analogy) will be important to achieve a stable classification scheme for the Bcl-2 family.

### Acknowledgments

We thank Dr Manolo Gouy and Lucian Soane for helpful discussions. This work was supported by National Institutes of Health grants RO1 NS37402 and RO1 GM077875 (J.M.H.). V.R.L. is the recipient of a doctoral fellowship from La Ligue Contre le Cancer (Comité de Saône-et-Loire).

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, at <http://dx.doi.org/10.1016/j.tcb.2012.10.010>.

### References

- Hengartner, M.O. and Horvitz, H.R. (1994) *C. elegans* cell survival gene *ced-9* encodes a functional homolog of the mammalian proto-oncogene *bcl-2*. *Cell* 76, 665–676
- Tsujimoto, Y. *et al.* (1984) Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science* 226, 1097–1099
- Chittenden, T. *et al.* (1995) Induction of apoptosis by the Bcl-2 homologue Bak. *Nature* 374, 733–736
- Huang, D.C. *et al.* (1998) The conserved N-terminal BH4 domain of Bcl-2 homologues is essential for inhibition of apoptosis and interaction with CED-4. *EMBO J.* 17, 1029–1039
- Oltvai, Z.N. *et al.* (1993) Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. *Cell* 74, 609–619
- Hardwick, J.M. *et al.* (2012) Multipolar functions of BCL-2 proteins link energetics to apoptosis. *Trends Cell Biol.* 22, 318–328
- Tanner, E.A. *et al.* (2011) Bcl-2 proteins and autophagy regulate mitochondrial dynamics during programmed cell death in the *Drosophila* ovary. *Development* 138, 327–338
- Ofengeim, D. *et al.* (2012) N-terminally cleaved Bcl-xL mediates ischemia-induced neuronal death. *Nat. Neurosci.* 15, 574–580
- Jagasia, R. *et al.* (2005) DRP-1-mediated mitochondrial fragmentation during EGL-1-induced cell death in *C. elegans*. *Nature* 433, 754–760
- Dai, H. *et al.* (2011) Transient binding of an activator BH3 domain to the Bak BH3-binding groove initiates Bak oligomerization. *J. Cell Biol.* 194, 39–48
- Dewson, G. *et al.* (2008) To trigger apoptosis, Bak exposes its BH3 domain and homodimerizes via BH3:groove interactions. *Mol. Cell* 30, 369–380
- Dewson, G. *et al.* (2012) Bax dimerizes via a symmetric BH3:groove interface during apoptosis. *Cell Death Differ.* 19, 661–670
- Albeck, J.G. *et al.* (2008) Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. *Mol. Cell* 30, 11–25
- Rudin, C.M. *et al.* (2012) Phase II study of single-agent navitoclax (ABT-263) and biomarker correlates in patients with relapsed small cell lung cancer. *Clin. Cancer Res.* 18, 3163–3169
- Denis, G.V. *et al.* (2003) Bcl-2, via its BH4 domain, blocks apoptotic signaling mediated by mitochondrial Ras. *J. Biol. Chem.* 278, 5775–5785
- Shibasaki, F. *et al.* (1997) Suppression of signalling through transcription factor NF-AT by interactions between calcineurin and Bcl-2. *Nature* 386, 728–731
- Petros, A.M. *et al.* (2004) Structural biology of the Bcl-2 family of proteins. *Biochim. Biophys. Acta* 1644, 83–94
- Kvansakul, M. *et al.* (2008) Vaccinia virus anti-apoptotic F1L is a novel Bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands. *Cell Death Differ.* 15, 1564–1571
- Aouacheria, A. *et al.* (2005) Phylogenomics of life-or-death switches in multicellular animals: Bcl-2, BH3-Only, and BNip families of apoptotic regulators. *Mol. Biol. Evol.* 22, 2395–2416
- Hardwick, J.M. and Youle, R.J. (2009) SnapShot: BCL-2 proteins. *Cell* 138, 404.e1–404.e2
- Chou, J.J. *et al.* (1999) Solution structure of BID, an intracellular amplifier of apoptotic signaling. *Cell* 96, 615–624
- McDonnell, J.M. *et al.* (1999) Solution structure of the proapoptotic molecule BID: a structural basis for apoptotic agonists and antagonists. *Cell* 96, 625–634
- Kamer, I. *et al.* (2005) Proapoptotic BID is an ATM effector in the DNA-damage response. *Cell* 122, 593–603
- Liu, Y. *et al.* (2011) Proapoptotic Bid mediates the Atr-directed DNA damage response to replicative stress. *Cell Death Differ.* 18, 841–852
- Zinkel, S.S. *et al.* (2005) A role for proapoptotic BID in the DNA-damage response. *Cell* 122, 579–591
- Kataoka, T. *et al.* (2001) Bcl-rambo, a novel Bcl-2 homologue that induces apoptosis via its unique C-terminal extension. *J. Biol. Chem.* 276, 19548–19554
- Huang, W. *et al.* (2012) Crystal structure and biochemical analyses reveal Beclin 1 as a novel membrane binding protein. *Cell Res.* 22, 473–489
- Suzuki, N.N. *et al.* (2005) The crystal structure of plant ATG12 and its biological implication in autophagy. *Autophagy* 1, 119–126
- Huang, Q. *et al.* (2002) Solution structure of a Bcl-2 homolog from Kaposi sarcoma virus. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3428–3433
- Huang, Q. *et al.* (2003) Solution structure of the BHRF1 protein from Epstein-Barr virus, a homolog of human Bcl-2. *J. Mol. Biol.* 332, 1123–1130
- Oettinghaus, B. *et al.* (2011) Tonight, the same old, deadly programme: BH3-only proteins, mitochondria and yeast. *EMBO J.* 30, 2754–2756
- Diella, F. *et al.* (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* 13, 6580–6603
- Shortridge, M.D. and Powers, R. (2009) Structural and functional similarity between the bacterial type III secretion system needle protein PrgI and the eukaryotic apoptosis Bcl-2 proteins. *PLoS ONE* 4, e7442
- Muchmore, S.W. *et al.* (1996) X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* 381, 335–341
- Reed, J.C. *et al.* (2004) The domains of apoptosis: a genomics perspective. *Sci. STKE* 2004, re9
- Duffy, S. *et al.* (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9, 267–276
- Guillemin, Y. *et al.* (2011) Characterization of unique signature sequences in the divergent maternal protein Bcl2l10. *Mol. Biol. Evol.* 28, 3271–3283
- Goh, C.S. *et al.* (2000) Co-evolution of proteins with their interaction partners. *J. Mol. Biol.* 299, 283–293
- Pazos, F. *et al.* (2005) Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. *J. Mol. Biol.* 352, 1002–1015

- 40 Kane, D.J. *et al.* (1993) Bcl-2 inhibition of neural death: decreased generation of reactive oxygen species. *Science* 262, 1274–1277
- 41 Fannjiang, Y. *et al.* (2004) Mitochondrial fission proteins regulate programmed cell death in yeast. *Genes Dev.* 18, 2785–2797
- 42 Silva, R.D. *et al.* (2011) Modulation of Bax mitochondrial insertion and induced cell death in yeast by mammalian protein kinase Calpha. *Exp. Cell Res.* 317, 781–790
- 43 Vander Heiden, M.G. *et al.* (2002) Bcl-x(L) complements *Saccharomyces cerevisiae* genes that facilitate the switch from glycolytic to oxidative metabolism. *J. Biol. Chem.* 277, 44870–44876
- 44 Chen, Y.B. *et al.* (2011) Bcl-xL regulates mitochondrial energetics by stabilizing the inner membrane potential. *J. Cell Biol.* 195, 263–276
- 45 Perciavalle, R.M. *et al.* (2012) Anti-apoptotic MCL-1 localizes to the mitochondrial matrix and couples mitochondrial fusion to respiration. *Nat. Cell Biol.* 14, 575–583
- 46 Polcic, P. and Forte, M. (2003) Response of yeast to the regulated expression of proteins in the Bcl-2 family. *Biochem. J.* 374, 393–402
- 47 Silva, R.D. *et al.* (2011) The importance of humanized yeast to better understand the role of bcl-2 family in apoptosis: finding of novel therapeutic opportunities. *Curr. Pharm. Des.* 17, 246–255
- 48 Lasi, M. *et al.* (2010) The molecular cell death machinery in the simple cnidarian *Hydra* includes an expanded caspase family and pro- and anti-apoptotic Bcl-2 proteins. *Cell Res.* 20, 812–825
- 49 Robertson, A.J. *et al.* (2006) The genomic underpinnings of apoptosis in *Strongylocentrotus purpuratus*. *Dev. Biol.* 300, 321–334
- 50 Zmasek, C.M. *et al.* (2007) Surprising complexity of the ancestral apoptosis network. *Genome Biol.* 8, R226
- 51 Kall, L. *et al.* (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35, W429–W432
- 52 Krogh, A. *et al.* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580
- 53 Hunter, S. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.* 37, D211–D215
- 54 Liu, Z. *et al.* (2005) Apolipoprotein I6, a novel proapoptotic Bcl-2 homology 3-only protein, induces mitochondria-mediated apoptosis in cancer cells. *Mol. Cancer Res.* 3, 21–31
- 55 Sinha, S. and Levine, B. (2008) The autophagy effector Beclin 1: a novel BH3-only protein. *Oncogene* 27 (Suppl. 1), S137–S148
- 56 Day, C.L. *et al.* (2008) Structure of the BH3 domains from the p53-inducible BH3-only proteins Noxa and Puma in complex with Mcl-1. *J. Mol. Biol.* 380, 958–971
- 57 Fleischer, A. *et al.* (2002) Proapoptotic activity of ITM2B(s), a BH3-only protein induced upon IL-2-deprivation which interacts with Bcl-2. *Oncogene* 21, 3181–3189
- 58 Schumacher, B. *et al.* (2005) *C. elegans* ced-13 can promote apoptosis and is induced in response to DNA damage. *Cell Death Differ.* 12, 153–161
- 59 Lanave, C. *et al.* (2004) Comparative genomics: the evolutionary history of the Bcl-2 family. *Gene* 333, 71–79
- 60 Kaczmarek, K. *et al.* (2011) Overexpression of peroxisomal testis-specific 1 protein induces germ cell apoptosis and leads to infertility in male mice. *Mol. Biol. Cell* 22, 1766–1779
- 61 Lee, D.H. *et al.* (2011) Interaction of a putative BH3 domain of clusterin with anti-apoptotic Bcl-2 family proteins as revealed by NMR spectroscopy. *Biochem. Biophys. Res. Commun.* 408, 541–547
- 62 Li, J. *et al.* (2009) A probabilistic framework to improve microRNA target prediction by incorporating proteomics data. *J. Bioinform. Comput. Biol.* 7, 955–972
- 63 Placzek, W.J. *et al.* (2011) Identification of a novel Mcl-1 protein binding motif. *J. Biol. Chem.* 286, 39829–39835
- 64 Aranovich, A. *et al.* (2012) Differences in the mechanisms of proapoptotic BH3 proteins binding to Bcl-XL and Bcl-2 quantified in live MCF-7 cells. *Mol. Cell* 45, 754–763
- 65 Lee, E.F. *et al.* (2009) Novel Bcl-2 homology-3 domain-like sequences identified from screening randomized peptide libraries for inhibitors of the pro-survival Bcl-2 proteins. *J. Biol. Chem.* 284, 31315–31326
- 66 Buttner, S. *et al.* (2011) A yeast BH3-only protein mediates the mitochondrial pathway of apoptosis. *EMBO J.* 30, 2779–2792
- 67 Cebulski, J. *et al.* (2011) Yeast Bax inhibitor, Bxi1p, is an ER-localized protein that links the unfolded protein response and programmed cell death in *Saccharomyces cerevisiae*. *PLoS ONE* 6, e20882
- 68 Chae, H.J. *et al.* (2003) Evolutionarily conserved cytoprotection provided by Bax Inhibitor-1 homologs from animals, plants, and yeast. *Gene* 323, 101–113
- 69 Teng, X. *et al.* (2011) Gene-dependent cell death in yeast. *Cell Death Dis.* 2, e188
- 70 Holm, L. and Sander, C. (1996) Mapping the protein universe. *Science* 273, 595–603