J Mol Evol (1995) 41:775–783

JOURNAL OF MOLECULAR EVOLUTION

Molecular Evolution and Secondary Structural Conservation in the B-Cell Lymphoma Leukemia 2 (*bcl-2*) Family of Proto-oncogene Products

Dyfed Lloyd Evans, Robert E. Mansel

Department of Surgery, University of Wales College of Medicine, Heath Park, Cardiff, CF4 4NX, United Kingdom

Received: 14 January 1995 / Accepted: 10 July 1995

Abstract. The nature of the *bcl-2* family of proto-oncogenes was analyzed by sequence alignment, secondary structure prediction, and phylogenetic techniques. Phylogenies were inferred from both the nucleic acid and amino acid sequences of the human, murine, rat, and chicken sequences for BCL-2 and BCL-X, human MCL1, murine A1, the nematode *Caenorhabditis elegans* and *Caenorhabditis briggsiae* ced-9 proteins, and the sequences BHRF1 from Epstein-Barr and LMW5-HL from African swine fever viruses. Both sequence alignment and secondary structure prediction techniques supported the conservation of both the overall secondary structure and the carboxy-terminal transmembrane domain in all members of the family. All the treeing methods employed (distance matrix, maximum likelihood, and parsimony) supported a tree in which the pro-apoptotic proteins BCL-2 and BCL-X represent the most recent additions to the group. All the trees also indicated that the viral proteins BHRF1 and LMW-HL arose from a common ancestor, an ancestor they shared in common with the pro-apoptotic control protein BAX, indicating that this function of BAX evolved only recently. The most ancient branches are represented by the nematode ced-9 protein and by the control genes MCL1 and A1, which in the treeing methods employed represent separate lineages within the most ancient grouping. These results demonstrate the evolution of a highly conserved family of developmental control genes from nematode to man—genes that encode proteins essential for normal development but which are highly conserved in terms of predicted structure and possible cellular localization. The evolutionary analysis also indicates that the family may be even larger than originally predicted and that other members are waiting to be discovered.

Key words: *bcl-2* family — Proto-oncogene products — Molecular evolution

Correspondence to: D. Lloyd Evans

Introduction

The control of cell number in multicellular eukaryotes represents a balance between cell proliferation and cell death. One of the most well understood apoptotic mechanisms exists in the *bcl-2*-regulated system, where over-expression of the protein BCL-2 protects cultured cells against several types of apoptotic stimuli (see Reed 1994; and Hoffman and Liebermann 1994; for recent reviews). If, in contrast, the metastasis-inducing gene translocation leads to up-regulation of p53 oncogene expression, the cells die via an apoptotic mechanism. p53 seems to function by up-regulating the expression of BAX, a pro-apoptotic member of the BCL-2 family. BAX dimerizes with BCL-2, thus negating its anti-apoptotic function (Oltvai et al. 1993). A third BCL-2 homolog is expressed in two forms—the long form, $BCLX_1$, is as effective an anti-apoptotic agent as BCL-2, but the short form, $BCLX_s$, seems to act in a pro-apoptotic manner. However, in contrast with BAX, no $BCLX_s$–BCL-2 dimers have been detected to date, and thus its mechanism of action remains unclear. In addition, expression of BCLX seems to be highly tissue-specific, with high levels of message only being detected in developing lymphocytes and the central nervous system (Reed 1994).

Other members of this gene family include human *MCL1* and murine *A1*. Identified in differentiating human myeloid and murine homeopoetic cells, respectively, these seem to be early response genes, switched on early in development, before the cellular machinery for peptide processing is in place. The function of both these gene products remains unknown, but it is surmised that they may act in an anti-apoptotic manner. The nematode BCL-2 homolog, ced-9, has been cloned from both *Cænarhabditis elegans* and *Cænarhabditis briggsiae* (Hengartner and Horvitz 1994). Both gene products have been shown to be anti-apoptotic, functioning in a manner that is almost indistinguishable from BCL-2. Indeed, they are essential in determining the cell lineages that will ultimately lead to the development of the adult nematode. Two further members of the family, BHRF1 and LMW5-HL, were identified and cloned from the Epstein-Barr and African swine fever viruses, respectively (Henderson et al. 1993; Neilan et al. 1993). Both are early transcription products, and BHRF1 has been positively shown to be anti-apoptotic. As such, both genes may be important for viral survival in their respective hosts.

Structurally, little is known about either BCL-2, or its homologs. The carboxy-terminal 22 amino acids have been shown to encode a transmembrane domain that targets the protein to the mitochondrial membrane. This domain has significant homology to the archetypal yeast transmembrane protein Mas70p (Nguyen et al. 1993). BCL-2 also seems to function as an antioxidant, a mode of action associated with its anti-apoptotic nature (Kane et al. 1993), but the domain with antioxidant function remains unknown. The current investigation used computational techniques to probe the implications of the evolution of the *BCL-2* family of proteins for their structure and function.

Methods

Sequences were obtained from the GenBank database using the BLAST server to search against the BCL-2 amino acid sequence back-translated to nucleic acid code (TBLASTN). Accession numbers for the sequences identified by this search method were Z23115, D11382, Z23110, M13994, U10101, U10597, L14680, L09548, L31532, L22473, L16462, L22472, L26546, L26545, and J02070.

The CLUSTAL V program of Higgins and Sharp (1988) was used to align the sequences. This program first aligns all the sequences progressively to produce an approximate score matrix. On the basis of this matrix, the alignment is created by the sequential addition of most related sequences. The subsequent alignment was optimized by hand to remove obvious defects before truncation to ensure that the sequences began and ended at conserved positions—thus avoiding biasing the scores by the presence of unrelated sequences. Gaps of more than two amino acids were replaced by an initial dash to represent a deletion, followed by a row of question marks, thus encoding large deletions as a potential single evolutionary event. This was done to avoid biasing the output trees on the basis of gaps in the sequences.

Secondary Structure Determination. Transmembrane domains were analyzed with the TopPred program (Sipos and von Heijne 1993), with sequences input one at a time. Other secondary structural features were determined using the optimal alignment as input for the PHD neural network program of Rost and Sander (1994). This program calculates secondary structures for soluble proteins and was thus only used on the cytosolic portion of the sequences. Output from this program was combined with that from TopPred to produce the overall secondary structural prediction.

Phylogenetic Techniques. Phylogenetic trees were built with the programs FITCH, SEQBOOT, DNAML, DNAMLK, DNAPARS, DNAPENNY, NEIGHBOR, and PROTPARS in J. Felsenstein's (1993) 3.5 version of the phylogenetic inference package (PHYLIP). Where appropriate, each program was run at least ten times, with the input sequences in different orders to eliminate input bias as a source of error.

Matrices of pairwise distances between DNA sequences were generated by the program DNADIST (PHYLIP) using either the Kimura two-parameter model (Kimura 1990) or the model of Jin and Nei (1980), with the rate of substitution varying according to a gamma distribution. In the former case the ratio of the transition rate (α) to transversion rate (β) was used at the default setting ($\alpha = 2\beta$) whereas in the latter, this value was set at unity ($\alpha = \beta$) and the rate of substitution was set at 0.1. Sets of randomly generated distance matrices were obtained from the DNA sequence alignment by using SEQBOOT to generate 500 resampled replicates.

Results

Amino Acid Sequence Alignment

The alignment of sequences homologous to BCL-2, as shown in Fig. 1, was inferred by the hierarchical clustering method of the CLUSTAL V package (Higgins and Sharp 1988). The alignment shows a remarkable degree of similarity between the various sequences–similarity that is distributed evenly over the whole sequence (Fig. 1). The major exception to this is the extended amino termini of the ced-9 and MCL1 proteins, extensions that have subsequently been lost during the course of evolution.

In the alignment, three highly conserved regions appear between residues 277–286, 294–306, and 344–357. These are highly conserved, with the basic motifs clearly present in all the sequences. Both viral sequences (LMW5-HL and BHRF1) lack the "classic" carboxy-terminal domain as seen in the other members of the family, though the program TopPred strongly predicts a membrane-spanning domain in this region (Fig. 1B). A striking feature is the similarity of this region in the BCL-2 family to the yeast Mas70p signal anchor, explaining the mitochondrial targeting of BCL-2 and suggesting that BCLX, MmA1, and HsMCL1, where the motif is most highly conserved, might also be mitochondrial-associated proteins. A feature to note is the amino-terminal truncation of the chicken BCLX protein, suggesting that this sequence might represent that of the pro-apoptotic alternate splice variant cognate with human $BCLX_S$.

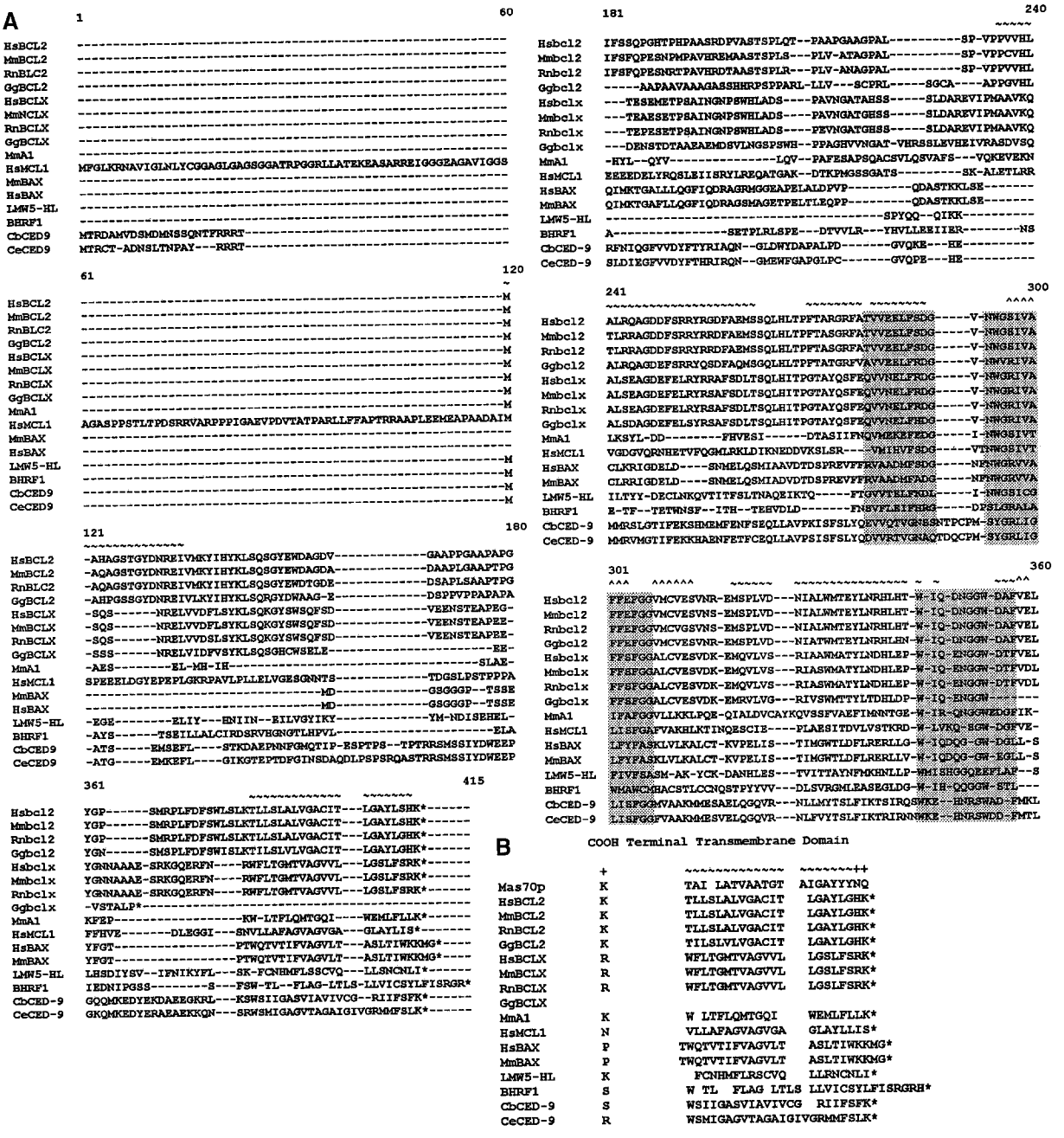

Fig. 1. **A** CLUSTALV alignment of the BCL-2 homologs: BCL-2 from *Homo sapiens* (*HsBCL2*), *Mus musculus* (*MmBCL2*), *Rattus norvegicus* (*RnBCL2*), and *Gallus gallus* (*GgBCL2*); BCL-X from *Homo sapiens* (*HsBCLX*), *Mus musculus* (*MmBCLX*), *Rattus norvegicus* (*RnBCLX*), and *Gallus gallus* (*GgBCLX*); *Homo sapiens* early response gene MCL1 (*HsMCL1*) and *Mus musculus* early response gene (*A1*); BAX from *Homo sapiens* (*HsBAX*), *Mus musculus* (*MmBAX*); the viral proteins *LMW5-HL* from African swine fever virus and *BHRF1* from Epstein-Barr virus; and the cell survival protein *CED-9* from *Caenorhabditis elegans* (*CeCED-9*) and *Caenorhabditis briggsiae* (*CbCED-9*). *Hyphens* represent gaps, and *boxed areas* delineate highly conserved regions. A secondary structure prediction is overlaid on the alignment, with ~~ representing α-helices and ^^ β-strands. Regions omitted from the phylogenetic analyses were the gapped sites 156–168, 287–291, and 324–326, as the gaps were solely caused by either ced-9 or MmA1. **B** Alignment of the hydrophobic amino-terminal domains of the family against the mitochondrial-targetted transmembrane domain of yeast Mas70p. ~~ represent the transmembrane helical domain and + the anchoring charged residues.

A DNA alignment corresponding to the amino acid alignment in Fig. 1 was created by inserting gaps at appropriate positions in the nucleotide sequences and by subsequently removing every third-position nucleic acid from the aligned sequences. The resulting amino acid alignment (sites 120–408, excluding the gapped regions 156–168, 287–291, and 324–326, solely caused by either ced-9 or MmA1 sequences) and the corresponding DNA alignment (536 sites, comprising first and second codon positions produced from the amino acid alignment) were

used for phylogeny inference by distance, maximum likelihood, and parsimony methods.

Secondary Structure Prediction

Aligned sequences, excluding the 22-amino-acid transmembrane tail, were converted to PIR format and used as input for the PHDsec secondary structure prediction analysis program of Rost and Sander (1994). This program implicitly utilizes evolutionary relationships, in that the input consists of blocked alignments of related sequences. PHD relies on a neural network trained with sequences from soluble, globular proteins. As such, it will not function on transmembrane domains. The alignment from 120–382 was therefore fed into PHD for structural prediction.

The output from PHD is overlaid on the sequence alignment (Fig. 1A) and shows that the proteins are of mixed type, containing predicted regions of alpha helix, sheet structure, and turns, though alpha helices form the predominant element.

The transmembrane regions (383–415) were individually analyzed with the TopPred program (Sipos and von Heijne 1993). Averaged output from this program was added to the output from PHD to yield an overall secondary structure prediction. The regions most strongly predicted to be membrane spanning by TopPred are shown in Fig. 1B aligned against the yeast Mas70p signal anchor.

Distance-Matrix Treeing from the Polypeptide Sequences

A distance matrix was created from the aligned amino acid sequences using the PROTDIST program from the PHYLIP package, utilizing the Dayhoff PAM001 matrix (Dayhoff 1979) to create a transition probability matrix, which allows prediction of the probability of changing from any given amino acid to any other. An unrooted phylogenetic tree inferred from the matrix by the method of Fitch and Margoliash (1967) showed three major lineages—the first containing the ced-9 sequences together with MCL1 and A1, the second the pro-apoptotic protein BAX along the viral members BHRF1 and LMW5-HL. The third lineage is comprised of the anti-apoptotic proteins BCL-2 and BCL-X. The BCL-2 and BCL-X sequences are clustered closely together, whereas there is a far greater evolutionary distance between the other sequences. A dendrogram identical to that in Fig. 2A (not shown) was obtained by searching the tree that fitted the matrix data using the neighbor-joining method of Saitou and Nei (1987).

Distance-Matrix Treeing from DNA Sequence Data

Topologies identical to that in Fig. 2A were inferred by the method of Fitch and Margoliash (1967), from distances calculated from the DNA sequence alignment (first and second codon positions). The results were the same whether using the Kimura two-parameter model (Kimura 1980) under the assumption that the transition rate is twice the transversion rate (Fig. 2B), or the Jin and Nei (1990) model, which uses Kimura's model of base substitution but assumes that the rate of substitution varies from site to site according to a gamma distribution. For this model the rates of transition and transversion were set to equality and the rate of substitution was set to 0.1, reflecting the structural similarity between the pro-

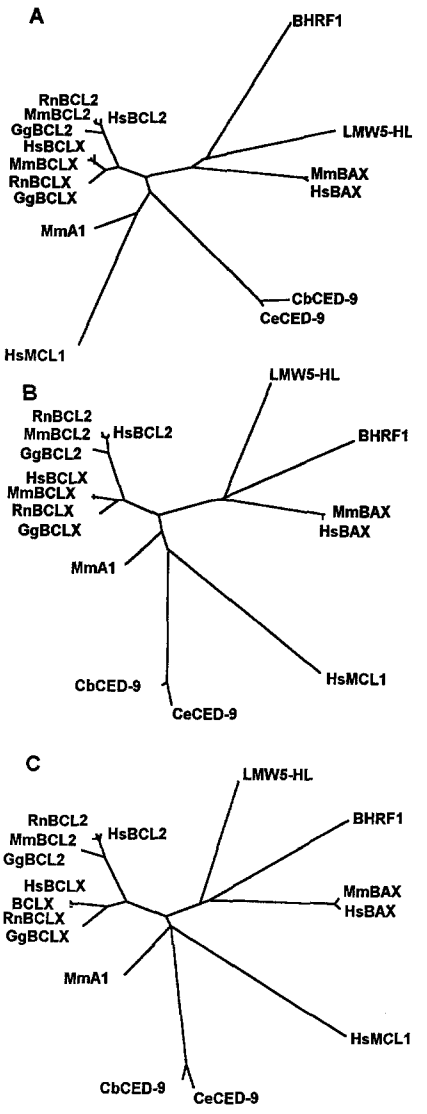

Fig. 2. **A** An unrooted dendrogram calculated by the method of Fitch and Margoliash (1967) from the amino acid alignment of Fig. 1A (sites 120–409, excluding the regions 156–168, 287–291, and 324–326) from pairwise distances determined on the basis on a PAM001 score matrix. **B,C** Unrooted trees calculated according to Fitch and Margoliash from the DNA sequence alignment (536 sites first and second codon positions) using Kimura distances **(B)** and Jin and Nei distances **(C)**. Branch lengths obtained by these methods are proportional to the average numbers of inferred substitutions at each site. Tree topologies as in Fig. 2B and C were the same whether the complete alignment (536 sites) or gap-free blocks (270 sites) were used. Species designations are the same as in Fig. 1.

teins, yielding the tree in Fig. 2C, but an identical tree was obtained when this value was set to 1. Moreover, the topologies remained the same when the data were analyzed using the stepwise neighbor-joining method of Saitou and Nei (1987).

Bootstrap Confidence on the Distance-Matrix Trees

Any conclusions drawn from the data presented in Fig. 2 will be dependent on the robustness of the stems separating the three major groupings. The reality of the topology presented here and the nature of the inferred topologies were therefore examined by the statistical bootstrapping technique (Diaconis and Efron 1983; Felsenstein 1985).

A set of bootstrapped distance matrices was generated from the DNA alignment by random resampling of sites with replacement (Felsenstein 1993). Phylogenetic trees were then inferred by the method of Fitch and Margoliash. The results, shown in Fig. 3, allow a number of conclusions. The division of the BCL-2 groups into three distinct groupings was supported by 100% of the bootstrapped matrices, as were the proteins included in the BCL2/BAX branch; 98% of the trees also supported the coherence of the other two branches, with *Mus musculus* A1 occurring in the BAX/viral branch in only 4.7% of the trees. The common root for the viral proteins is confirmed in 66.4% of the Kimura distance trees and 55.9% of the Jin and Nei distance trees. These numbers are less significant, but other possible tree conformations, such as LMW5-HL arising before BHRF1 (16.8% and 25.7%), or BHRF1 arising before LMW5-HL (9.4% and 10.3%), are very poorly supported. The least-certain branching in the tree relates to the ced-9/early response lineage. The placing of MmA1 as the latest ramification in this branch is supported by 58% of the Kimura trees, but by only 43.4% of Jin and Nei trees. Alternate conformations are HsMCL1 arising after MmA1 (14.7% and 22.6%), or HsMCL1 and MmA1 sharing a common ancestor (8.3% and 15.7%), but none of these alternate placements are more significant than the branch conformations in the most likely trees (Fig. 3A and B). Surprisingly, the trees based on Kimura distances put only a 61.8% confidence limit on the placement of the GgBCLX branch. This may be as a result of the possibility that this sequence represents $BCLX_S$, and therefore contains a carboxy-terminal truncation compared to other members of the family.

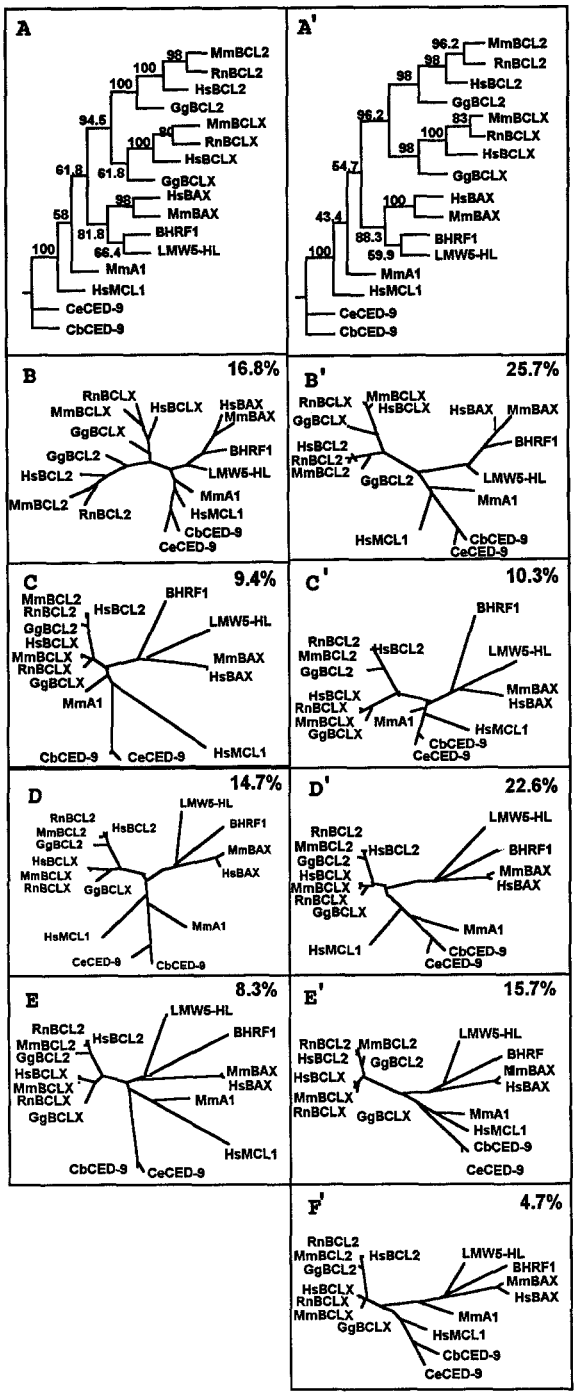

Fig. 3. Bootstrap analysis of the Fitch-Margoliash topologies derived from the DNA sequence alignment (536 sites) using either the Kimura distances **(left panels)** or Jin and Nei distances **(right panels).** Branch lengths represent the average distances calculated from 150 replicate trees. Replicate distance matrices were generated with the SEQBOOT program of PHYLIP 3.5. Species designations are the same as in Fig. 1.

Phylogeny Treeing by the Maximum-Likelihood Method

Under the maximum-likelihood criterion, each branch of the phylogenetic tree is tested for statistical significance (Felsenstein 1993). As shown in Fig. 4A, an unrooted tree inferred from the aligned DNA sequences (268 sites) by the maximum-likelihood method was topologically identical to the trees generated by a least-squares fit to a distance matrix. More importantly, all the internodal distances supporting the branching order of the various lineages were highly significantly positive ($P < 0.01$).

A second, rooted, dendrogram was also inferred from the DNA alignment using the maximum likelihood criterion under the constraint that the branch lengths must

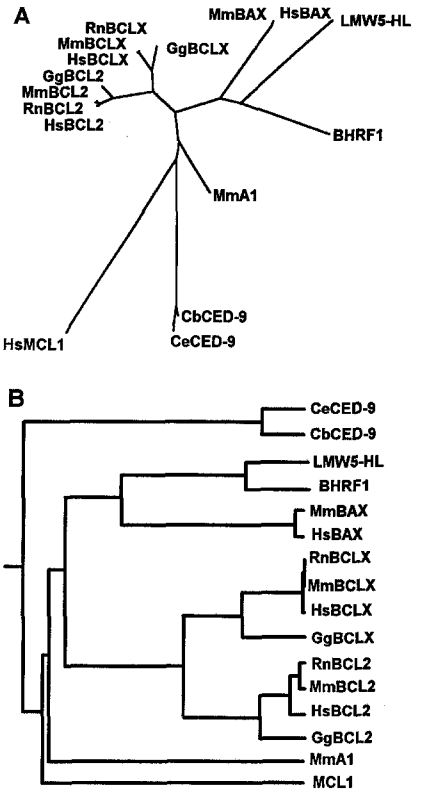

Fig. 4. **A** Maximum likelihood analysis of the DNA sequence alignment (536 sites, first and second codon positions, DNAML program). **B** Maximum likelihood tree with contemporary tips assuming a molecular clock (DNAMLK program). The ratio of transition (α) to transversion (β) was taken as $\alpha = 2\beta$, the default value for both programs. Symbols have the same meaning as in Fig. 1.

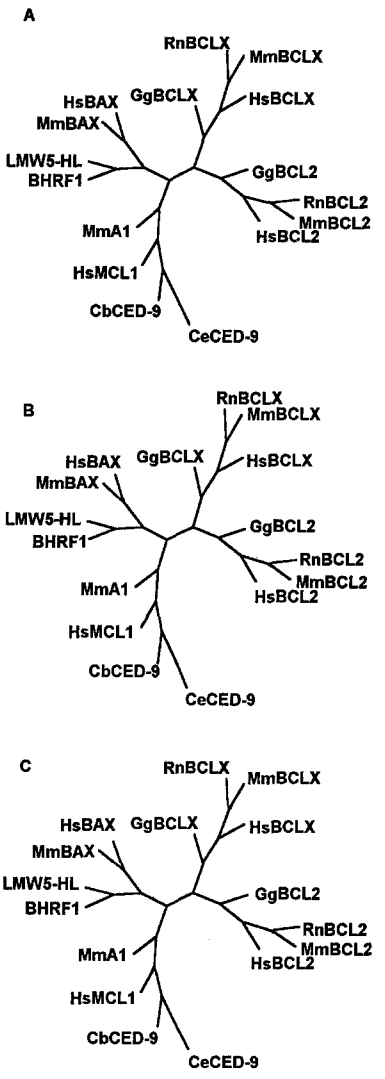

Fig. 5. **A** Protein parsimony tree inferred from the amino acid sequence alignment (268 sites). **B,C** DNA parsimony trees inferred from the nucleotide sequence alignment comprising first and second positions (536 sites, A) and from second positions only (268 sites, B). The tree derived from first codon positions only (268 sites) had the same topology as the tree shown in B. Symbols have the same meaning as in Fig. 1.

be consistent with a molecular clock, thus yielding a tree whose tips are all equidistant from an inferred root (Felsenstein 1993). This method (Fig. 4B) produced a tree topology identical to that shown in Fig. 4A, with the root placed between the nematode and MCL1 sequences.

Maximum Parsimony Treeing

The maximum parsimony method represents the variability in a sequence (amino, or nucleic acid) with the smallest number of postulated changes (Felsenstein 1993). The tree based on the amino acid alignment (Fig. 5A, 268 sites) agreed with the branching pattern inferred from both the distance and maximum likelihood methods. This tree definitely shows the ced-9 and hemopoietic members of the family to be a coherent grouping between the BCL and BAX/viral branches.

Figure 5B and C show maximum-parsimony trees constructed from the nucleotide sequence alignment (536 sites, first and second codon positions) and from alignments comprising the second codon positions (268 sites) only. The trees inferred from second positions are likely to be the least affected by regional or genomic evolutionary changes in GC content (Zuckerkandl 1987).

These trees are remarkably consistent with the previous dendrograms, and both DNA trees agree with the tree derived by protein parsimony (Fig. 5A). Indeed, a tree generated from first codon positions only (not shown) gave a topology identical to that seen in Fig. 5B and C. The striking concurrence of tree topology when inferred solely on the basis of either first or second codon positions argues against the possibility that the degeneracy of the silent first positions of arginine, serine, and leucine codons might artifactually have biased the assessment of tree topology.

The branch-and-bound algorithm (Felsenstein 1993) acts as the obverse of DNA parsimony in that it assesses whether all of the most parsimonious trees implied by a given nucleotide sequence alignment have indeed been found. Analysis of all 526 first and second codon sites yielded a single most parsimonious tree (2,744 steps)

with exactly the same topology as the trees yielded by DNA parsimony (Fig. 5A,B).

Bootstrap Confidence on Parsimony Trees

Confidence limits on the tree topologies in Fig. 5 were placed by bootstrapping. The results of 500 maximum parsimony trees on the amino acid alignment (Fig. 6, left panels) and nucleic acid alignment (Fig. 6, right panels) show that the tree conformation obtained by distance matrix, maximum likelihood, and maximum parsimony is supported by 54.6% of the bootstrapped replicates based on the amino acid alignment and 48.4% of the bootstrapped replicates based on the nucleic acid alignments. The structure of the BCL-2/BCL-X branch is supported by 100% of the replicates, with variations in both the ced-9/early response and BAX/viral branches. The conformation of the BAX/viral branch where human and murine BAX form one lineage and the viral proteins LMW5-HL and BHRF1 form a second lineage is supported by 73% of the amino acid replicates and 70.3% of the nucleic acid replicates. Alternate conformations have LMW5-HL diverging before BHRF1 (18.7% and 22.2%) or BHRF1 diverging before LMW5-HL (8.3% and 7.5%). The predominant conformation of the ced-9/early response branch has MmA1 diverging before HsMCL1 and is supported by 81.6% of the amino acid replicates and 78.1% of the nucleic acid replicates. Alternate conformations have HsMCL1 diverging before MmA1 (13.9% and 14.7%), or ced-9, HsMCL1, and MmA1 forming independent lineages with a common ancestor (4.5% and 7.2%).

Such a high degree of bootstrap congruence concurs with the consistency of trees inferred from either first or second position DNA alignments—the congruence of both bootstrap and selection of codon positions argue against significant variability among nucleotide positions.

Discussion

Secondary Structure Prediction

Recent analyses indicate that BCL-2 may function as a GTP/GDP binding protein (G-protein) (Haldar et al. 1990) and that its main role may occur during development. As such, it may act in a similar manner to ced-9 by protecting cell lineages expressing it from undergoing premature cell death (Veis Novack and Korsmeyer 1994). The BCL-2 proteins contain none of the classical G-protein domains, but there are three closely related regions from amino acids 280–286, 294–306, and 344–357, each in defined regions of predicted secondary structure (helix, strand, helix, respectively) that could fold into a structure resembling the canonical G-protein

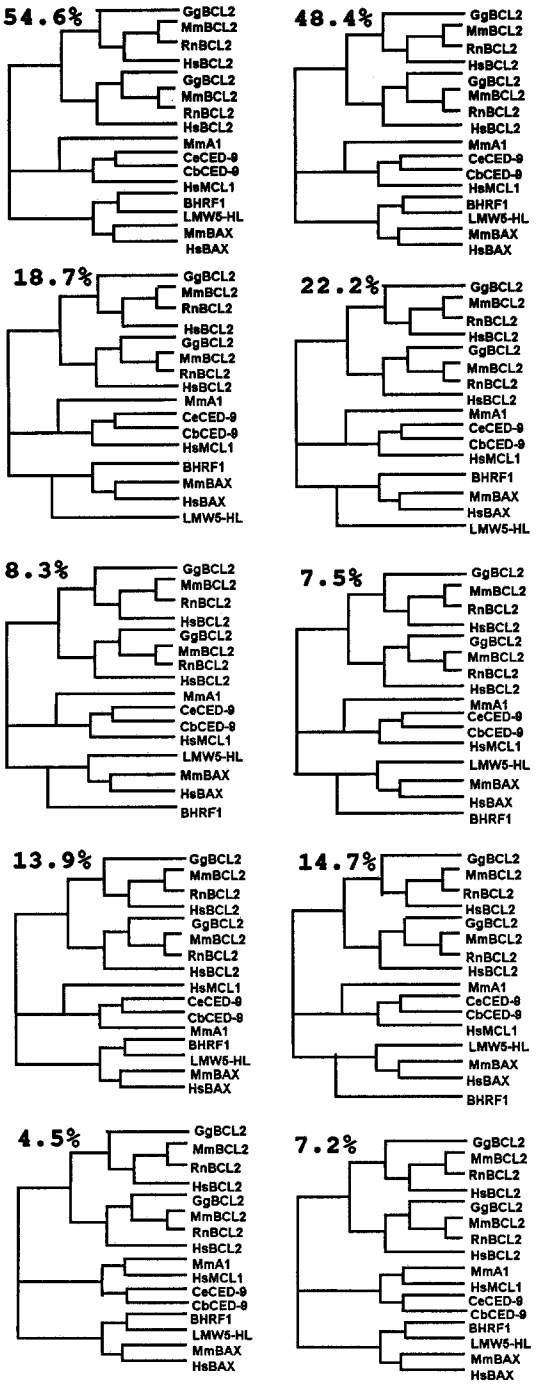

Fig. 6. Bootstrapped parsimony for the amino acid alignment **(left panels)** and the nucleic acid alignment **(right panels).** Percentages refer to 500 bootstrapped replicates. Symbols have the same meaning as in Fig. 1.

binding pocket (Pai et al. 1989). In addition, the sequence in the region 294–306 might represent a novel P-loop motif, as suggested by the sequence similarity of this region with other P-loop families (Savaste et al. 1990). These domains have recently been shown to be essential for the anti-apoptotic function of BCL-2 (Borner et al. 1994).

Overall, the sequences are remarkably similar, showing a high degree of structure-dependent conservation,

despite the differing rates of evolution of the organisms in which they are found. Secondary structure analysis predicts the presence of six α-helices, including the terminal transmembrane helix, and three strands that could form a short β-sheet at the core of the molecule.

The high degree of conservation of the carboxy-terminal transmembrane domain in the family suggests that membrane attachment may be important for the function of these molecules. This may relate to the intracellular antioxidant nature of many of these proteins (mitochondria, nuclear membranes, and the endoplasmic reticulum being major oxidative sites). The only exception to this is the chicken BCLX sequence, which with its carboxy-terminal truncation may represent the sequence of $BCLX_s$, the shorter, pro-apoptotic form of the protein. Such structural and functional conservation in this family makes this an ideal subject for evolutionary analysis to examine the functional divergence and the interdependence between sequence and structure.

Molecular Evolution Analyses

The trees presented here permit the determination of relationships between various members of the BCL-2 family of proteins. First, all the trees, regardless of treeing procedure, support the monophyly of the BAX/viral branch. That is, all these proteins are derived from a common ancestor with branching occurring when the ancestor of LMW5-HL and BHRF1 splits off from the BAX branch. The implication of this is that the event occurred when the viral ancestor of Epstein-Barr virus and African swine fever virus gained its BCL-2 homolog by recombination with a host genome. This points to a common ancestor for both EBV, the most distantly related member of the herpesviridae, and ASTV, the most distantly related member of the poxviridae. Distance methods point to this being an early event, which is compatible with our knowledge of the evolution of the herpesviridae and poxviridae (Koonin and Senkevich 1992).

The most distant members of the lineage are the nematode ced-9 proteins, as supported by Fig. 5B. Other early genes in the lineage are HsMCL1 and MmA1; both are early response genes that are expressed in differentiating myeloid cells and in hemopoietic cell lines, respectively (Kozopas et al 1993; Lin et al 1993).

The BAX lineage also seems to be an ancient lineage. This is surprising, as BAX functions as an anti-apoptotic agent by binding to and thus negating the function of BCL-2. This might suggest that BAX gained its anti-apoptotic function fairly recently, as might be seen from the fact that these proteins still possess a putative transmembrane domain and the conserved motifs at regions 277–286, 294–306, and 344–357. A mutational analysis of BCL-2 in comparison to BAX might yield a number of clues as to the function of the latter.

Of all the lineages examined, BCL-2 and BCL-X are the most recent. Both act in an anti-apoptotic manner, though there is a short form of BCL-X, $BCL\text{-}X_s$ that acts as a pro-apoptotic agent (Boise et al. 1993). BCL-2 acts in embryogenesis (Baer 1994), but it is also expressed in differentiating lymphoid cells (Lin et al. 1993).

The BCL-2 family of proteins is therefore a coherent group, split into three main groups. Each of these shares highly conserved primary structures, as shown by the high degree of conservation in secondary structure predictions and the high congruency between the various treeing methods employed to examine the lineage. Functionally BAX stands out as the only true anti-apoptotic member, though this role may have been gained relatively recently, possibly in response to the evolution of BCL-2. All other genes mediate cell survival and are generally expressed early during development, or early in the maturation of a cellular lineage. Even the two viral members of the family are expressed early during infection and probably promote host cell survival (Henderson et al. 1993; Neilan et al. 1993). Thus, the essential function of these proteins has changed little from nematode to man, though the precise site of expression has altered considerably. The nature of these proteins therefore makes it highly likely that further members remain to be isolated in other species, in that similar genes probably occur in all multicellular organisms as control of cell numbers is essential for the development of complex organisms. There may also be novel members of the family that control cell numbers in other differentiating cell lineages. In this respect, therefore, the members of the BCL-2 family of proteins described to date represent only a subset of the total.

The data presented here support the BCL-2 family of proteins being a coherent and evolutionarily ancient grouping. The function of these proteins is crucial to the survival of the host organism, thus explaining the high degree of both primary and predicted secondary structural conservation. Not only that, but there is also a high degree of conservation at the codon level, as demonstrated by the conserved first and second codon usage. In addition, conservation of overall sequence and structural elements in the viral members of the family argues for the dependence of function on primary sequence integrity.

Acknowledgments. We thank J. Felsenstein for freely contributing the PHYLIP package, B. Rost and C. Sander for use of the PHD server, J. Melhuish for help with the figures, and I. Hopkinson for advice and encouragement. This work was supported by Scotia Pharmaceuticals and benefited from use of the Seqnet facility.

References

Baer R (1994) *Bcl-2* breathes life into embryogenesis (1994) Am J Pathol 145:7–10

Boise LH, González-Garcia M, Postema CE, Ding L, Lindsten T, Turka

LA, Mao X, Nuñez G, Thompson CB (1993) *bcl-x,* a *bcl-2*-related gene that functions as a dominant regulator of apoptotic cell death. Cell 74:597–608

Borner C, Martinou I, Mattmann C, Irmler M, Schaerer E, Martinou J-C, Tschopp J (1994) The protein bcl-2α does not require membrane attachment, but two conserved domains to suppress apoptosis. J Cell Biol 126:1059–1068

Dayhoff MO (1979) Atlas of protein sequence and structure, Vol 5, Suppl 3, 1978. National Biomedical Research Foundation, Washington, DC

Diaconis P, Efron B (1983) Computer-intensive methods in statistics. Sci Am 248:116–130

Felsenstein J (1993) Phylogenetic inference program (PHYLIP) manual, version 3.5. University of Washington, Seattle.

Felsenstein J (1985) Confidence limits in phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 15:279–284

Haldar S, Beatty C, Croce CM (1990) Bcl-2 alpha encodes a novel small molecular weight GTP binding protein. Advances in Enzyme Regulation 30:145–153

Henderson SA, Huen D, Rowe M, Dawson C, Johnson G, Rickinson A (1993) Epstein-Barr virus-coded BHRF1 protein, a viral homologue of Bcl-2, protects human B cells from programmed cell death. Proc Natl Acad Sci USA 90:8479–8483

Hengartner MO, Horovitz HR (1994) C elegans cell survival gene ced-9 encodes a functional homolog of the mammalian proto-oncogene bcl-2. Cell 76:665–676

Higgins DA, Sharp PM (1988) Clustal: a package for performing multiple sequence alignments on a microcomputer. Gene 73:237–244

Hoffman B, Lieberman DA (1994) Molecular controls of apoptosis: differentiation/growth arrest primary response genes, proto-oncogenes and tumor suppressor genes as positive & negative modulators. Oncogene 9:1807–1812

Jin L, Nei M (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol Biol Evol 7:82–102

Kane DJ, Sarafian TA, Anton R, Hahn H, Butler Gralla E, Selverstone Valentine J, Örd T, Bredesen DE (1993) Bcl-2 inhibition of neural death: decreased generation of reactive oxygen species. Science 262:1274–1277

Kimura M (1980) A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Koonin EV, Senkevich TG (1992) Evolution of thymidine and thymidylate kinases: the possibility of independent capture of TK genes by different groups of viruses. Virus Genes 6:187–196

Kozpaz KM, Yang T, Buchan HL, Zhou P, Craig RW (1993) *MCL1,* a gene expressed in programmed myeloid cell differentiation, has sequence similarity to *BCL2.* Proc Natl Acad Sci USA 90:3516–3520

Lin EY, Orlofsky A, Berger MS, Prystowsky MB (1993) Characterization of A1, a novel hemopoietic-specific early-response gene with sequence similarity to *bcl-2.* J Immunol 151:19791988

Neilan JG, Lu Z, Afonso CL, Kutish GF, Sussman MD, Rock DL (1993) An African swine fever virus gene with similarity to the proto-oncogene *bcl-2* and the Epstein-Barr virus gene *BHRF1.* J Virol 67:4391–4394

Nguyen M, Millar DG, Young VW, Korsmeyer SJ, Shore GC (1993) Targeting of Bcl-2 to the mitochondrial outer membrane by a COOH-terminal signal anchor sequence. J Biol Chem 268:25265–25268

Oltvai ZN, Milliman CL, Korsmeyer SJ (1993) Bcl-2 heterodimerizes in vivo with a conserved homolog BAX that accelerates programmed cell death. Cell 74:609–619

Pai EF, Kabsch W, Krengel U, Holmes KG, John J, Wittinghofer A (1989) Structure of the guanine-nucleotide-binding domain of the Ha-*ras* oncogene product p21 in the triphosphate conformation. Nature 341:209–214

Reed JC (1994) Bcl-2 and the regulation of programmed cell death. J Cell Biol 124:1–6

Rost B, Sander C (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19:55–72 press).

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Savaste M, Sibbald PR, Wittinghofer A (1990) The P-loop—a common motif in ATP-and GTP-binding proteins. Trends Biochem Sci 15: 430–434

Sipos L, von Heijne G (1993) Predicting the topology of eukaryotic membrane protein. Eur J Biochem 213:1333–1340

Veis Novack D, Kursmeyer SJ (1994) *Bcl-2* protein expression during murine development. Am J Pathol 145:61–73

Zuckerkandl E (1987) On the molecular evolutionary clock. J Mol Evol 26:36–46