J Mol Model (2012) 18:1885–1906
DOI 10.1007/s00894-011-1201-6

ORIGINAL PAPER

Molecular dynamics simulations of the Bcl-2 protein to predict the structure of its unordered flexible loop domain

Pawan Kumar Raghav · Yogesh Kumar Verma · Gurudutta U. Gangenahalli

Received: 2 October 2010 / Accepted: 24 July 2011 / Published online: 25 August 2011

Abstract B-cell lymphoma (Bcl-2) protein is an anti-apoptotic member of the Bcl-2 family. It is functionally demarcated into four Bcl-2 homology (BH) domains: BH1, BH2, BH3, BH4, one flexible loop domain (FLD), a transmembrane domain (TM), and an X domain. Bcl-2's BH domains have clearly been elucidated from a structural perspective, whereas the conformation of FLD has not yet been predicted, despite its important role in regulating apoptosis through its interactions with JNK-1, PKC, PP2A phosphatase, caspase 3, MAP kinase, ubiquitin, PS1, and FKBP38. Many important residues that regulate Bcl-2 anti-apoptotic activity are present in this domain, for example Asp34, Thr56, Thr69, Ser70, Thr74, and Ser87. The structural elucidation of the FLD would likely help in attempts to accurately predict the effect of mutating these residues on the overall structure of the protein and the interactions of other proteins in this domain. Therefore, we have generated an increased quality model of the Bcl-2 protein including the FLD through modeling. Further, molecular dynamics (MD) simulations were used for FLD optimization, to predict the flexibility, and to determine the stability of the folded FLD. In addition, essential dynamics (ED) was used to predict the collective motions and the essential subspace relevant to Bcl-2 protein function. The predicted average structure and ensemble of MD-simulated structures were submitted to the Protein Model Database (PMDB), and the Bcl-2 structures obtained exhibited enhanced quality. This study should help to elucidate the structural basis for Bcl-2 anti-apoptotic activity regulation through its binding to other proteins via the FLD.

P. K. Raghav · Y. K. Verma · G. U. Gangenahalli (✉)
Stem Cell and Gene Therapy Research Group,
Institute of Nuclear Medicine and Allied Sciences (INMAS),
Lucknow Road,
Timarpur, Delhi 110054, India
e-mail: gugdutta@rediffmail.com

Keywords Bcl-2 · Loop modeling · MD simulations · Apoptosis · Flexible loop domain · Essential dynamics

Abbreviations

1GJH	Bcl-2 isoform 2 structure (from NMR)
2XA0	Model of complex of Bcl-2 and Bax peptide (from X-ray diffraction)
1G5J	Model of complex of Bcl-X_L and Bad peptide (from NMR)
SM	Swiss-Model generated model of Bcl-2 based on the 1GJH template
CW	ClustalW alignment based Bcl-2 model generated by MODELLER
MOD	MODELLER-generated model based on the 1GJH template and obtained using the same alignment as used by SM
MODLOPT	MODELLER loop-optimized model
PM0076467	PMDB ID of the MD-simulated average structure
PM0077081-PM0077103	23-Structure ensemble resulting from clustering
MD	Molecular dynamics
ED	Essential dynamics
PMDB	Protein Model Database
FLD	Flexible loop domain

Introduction

The FLD region (amino acids 34–92) of Bcl-2 (239 amino acids) lies between the BH4 and BH3 domains, lacks a defined structure, and is not structurally conserved among Bcl-2 family members (Fig. 1) [1]. The BH domains are responsible for Bcl-2's anti-apoptotic function, which is

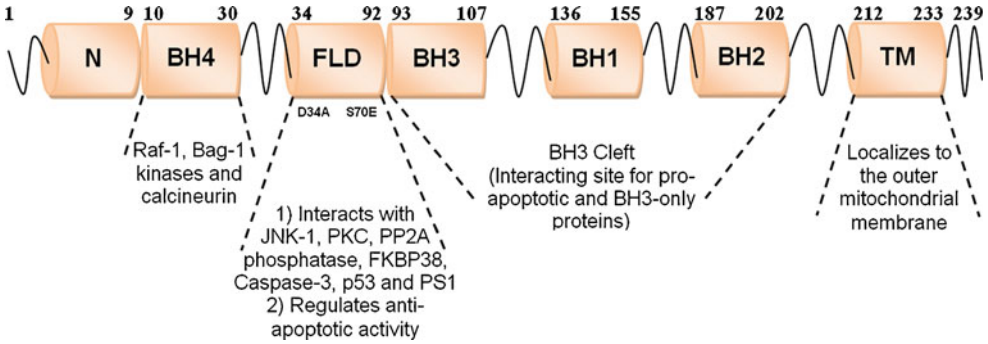


Fig. 1 Functional division of the Bcl-2 protein into domains (represented as *tubes*), *N*- N-terminal, *BH*- Bcl-2 homology, *FLD*- flexible loop domain and *TM*- transmembrane, connected through the flexible region represented as *wires*. The interacting proteins and functions of the domains of Bcl-2 are shown by *dotted lines*. Important residues such as Asp34 and Ser70 lie in the FLD region; the substitution of these residues (i.e., Asp34Ala and Ser70Glu) upregulates the anti-apoptotic activity of Bcl-2

linked with its homodimerization and its heterodimerization with its own family members and those of other families, whereas FLD regulates the activity of Bcl-2. The FLD contains phosphorylation sites—Thr56, Thr69, Ser70, Thr74, and Ser87—that are essential for the regulation of Bcl-2 activity. By using 2-D peptide mapping and sequencing, the residues Thr56, Thr74, and Ser87 were found to be phosphorylated in response to a microtubule damaging agent (paclitaxel) that also arrest T cells at G2/M phase of cell cycle. Changing these sites to Ala led to enhanced survival following death signal as well as paclitaxel treatment [2]. These residues also constitute a ubiquitin-dependent cleavage/MAP kinase site. It was discovered that there was cleavage of Bcl-2 at this site upon TNF-α induced cell death in endothelial cells due to a ubiquitin-dependent proteosome complex. The phosphorylation of these residues has been shown to abolish Bcl-2 degradation [3]. On the other hand, single phosphorylation of Bcl-2 at Ser70 is essential for full and potent cell survival activity. The residue Ser70 was shown to be phosphorylated by IL-3, PKC and EPO to maintain normal cellular homeostasis [4], whereas the Ser70Ala mutant is not phosphorylated after IL-3/Bryo stimulation, and is unable to prolong cell survival upon either IL-3 deprivation or etoposide treatment. Ser70Glu substitution also suppresses etoposide-induced apoptosis more potently than wild-type Bcl-2 [5]. Furthermore, the DNA damage induced by p53–Bcl-2 binding was shown to be associated with the weaker Bcl-2–Bax interaction and increased apoptotic cell death in a mechanism regulated by FLD. This demonstrates that FLD is also involved in regulating the binding of Bcl-2 with Bax and indirectly regulating such interactions with pro-apoptotic proteins [6]. Additionally, FLD contains a caspase 3 (apoptosis effecter protease) recognition and cleavage site at Asp34. Cleavage at Asp34 by caspase 3 renders Bcl-2 unable to inhibit apoptosis. The mutation of Asp34Ala makes Bcl-2 resistant to caspase 3 mediated cleavage and hence enhances anti-apoptotic activity [7]. The FLD-deleted mutant of Bcl-2 displays enhanced ability to inhibit apoptosis without impairing Bcl-2 hetero-dimerization with pro-apoptotic proteins. Full-length Bcl-2 was found to be ineffective at preventing anti-IgM induced cell death of an immature B cell line (WEHI-231). In contrast, this mutant protected WEHI-231 cells from death [8]. Moreover, a molecular interaction between FKBP38 and Bcl-2 has been shown to occur through the unstructured loop of Bcl-2, and this appears to regulate phosphorylation in the loop [9]. Normally proteins are degraded by cellular proteases; nonetheless, the flexible loop in Bcl-2 has been shown to shield or protect it from rapid degradation by cellular proteases due to the presence of a random coil structure that exhibits a long half-life [8, 10, 11].

All these studies indicate that the FLD plays an important role in providing stability to the Bcl-2 protein and regulation of its anti-apoptotic activity. Understanding the interactions of various proteins with FLD would likely help us to tweak apoptosis signaling and the treatment of various malignancies [12] in which Bcl-2 expression is not under control, such as acute myeloid leukemia (AML) [13]. Therefore, the FLD represents an attractive target for drugs that modulate protein–protein interactions, and studies of the FLD would likely facilitate drug discovery and improve our understanding of human diseases.

No data on the 3D conformation of the FLD are currently available; nevertheless, the NMR structure of Bcl-2, in which the FLD is replaced with residues of the Bcl-X_L protein, is available [14]. In our study, sequence-based results showed that the FLD is an extensively disordered region. In order to predict the conformation of the FLD, we elucidated the structure of the Bcl-2 protein using Swiss-Model [15] and MODELLER software [16]. Based on the stereochemical parameters, we observed that MODELLER generated a refined model. Further, the FLD (as

generated by MODELLER) was optimized by the loop modeling method of MODELLER, and then energy minimization was performed. Since the accurate prediction of the secondary and tertiary stable structure and dynamics of the FLD cannot be achieved through experimental measurements [17], we used MD simulation to predict a putative conformation and the flexibility of the FLD. The MD simulation of the energy-minimized Bcl-2 model was carried out using GROMACS (Groningen Machine for Chemical Simulations) [18]. The average structure of the Bcl-2 protein was obtained from the 15 ns MD-simulated trajectory. Further, the predicted average model was energy minimized and validated by stereochemical and overall quality checks. Subsequently, the average model was checked for its ability to bind with pro-apoptotic peptides by docking. Clustering was performed to ensemble the structures based on the RMSD along the MD simulation trajectory. The electrostatic behavior (dipole moment) was predicted for the entire MD simulation trajectory, and was found to be in accordance with the X-ray structure. ED was used to reduce the dimensionality and to predict the essential subspaces for large collective motions that are relevant when generating a biologically functional Bcl-2 model. Our results show that the model obtained has enhanced quality and may be useful for studies focusing on, for example, mutagenesis and protein–protein docking. The average model and ensemble were submitted to the PMDB [19], and are available for further analysis.

Computational methods

Sequence analysis of Bcl-2 protein

The human Bcl-2 sequence (Genbank I.D. 231632; Swissprot I. D. P104152) containing 239 amino acids was used for sequence analysis and modeling. BLAST [20] identified the sequences homologous to Bcl-2 in different organisms. PDBblast was used to search for Bcl-2 homologs with solved 3D structures. The BLAST parameters of an expect value of 10, a hitlist size of 100, a threshold of 11, and a word size of 2 were used, and the BLOSUM62 matrix was employed. Unordered regions in the protein were predicted by the DISOPRED server [21]. The IUPred method [22] was used to obtain the specific amino acid composition of the disordered region of the FLD, which does not form a stable, well-defined structure.

Predicting the structure of Bcl-2 and FLD optimization

The Swiss-Model server (http://swissmodel.expasy.org) and the MODELLER program were used to generate the 3D model of Bcl-2. The Swiss-Model automated modeling mode constructed the Bcl-2 model (*SM*) based on the 1GJH template (73.171% sequence identity). MODELLER 9v7 generated the 3D structure of the Bcl-2 protein based on the 1GJH template by satisfying spatial restraints for the aligned regions.

Two models were generated using two different alignments between the Bcl-2 sequence and the 1GJH template by MODELLER. The first model (MOD) was obtained by manual alignment (similar to the alignment used to generate the SM model). The second model (CW) was obtained using ClustalW [23] PIR format alignment between the Bcl-2 and 1GJH template. This alignment was generated using the EBLOSUM 62 matrix [24] with a gap penalty of 10 and an extend penalty of 0.5. The first model (MOD) was subjected to refinement of the FLD using a loop optimization protocol (the loopmodel class of the DOPE-based method) followed by MD simulations (using the conjugate gradient optimization method) at the temperatures 150, 300, 400, 800 and 1000 K using MODELLER. An initial loop conformation (FLD) for residues 34–92 of the MOD model was generated using a fast loop refinement method by simply positioning the atoms of the FLD, uniformly spaced, on the line that connects the main-chain carbonyl oxygen and the amide nitrogen atoms of the N- and C-terminal anchor regions. This MODELLER loop-optimized (MODLOPT) structure was further energy minimized (globally) using the GROMOS96 43B1 force field [25] by Swiss-Pdb viewer [26]. Pymol [27] was employed to calculate structural alignment differences between the energy-minimized MODLOPT and SM models. The modeled structures were validated by ProSA-web [28] and SAVes server (http://nihserver.mbi.ucla.edu/SAVES/) using the PROCHECK [29], ERRAT [30] and WHAT CHECK [31] programs.

MD simulations

MD simulation of the energy-minimized MODLOPT model was performed in order to predict the native conformation of Bcl-2 and the structural changes induced by the FLD in its domains. We used GROMACS package 4.0 [32], which is a versatile collection of programs and libraries for simulating molecular dynamics and subsequently analyzing trajectory data. The simulations were carried out on a single PC (3.40 GHz Core 2 Duo processor, Pentium IV, 4 GB RAM; Hewlett Packard) running the Windows Vista operating system (using Cygwin). The GROMOS 96 53a6 [33, 34] force field including all hydrogens, along with a simple point-charge (SPC) water model [35], was used for energy minimization. The pre-equilibrated [36] SPC water was added to an octahedral box, and the protein was then placed in the center of the box. 16809 solvent molecules were embedded into the box, which extended at least 9 Å from the Bcl-2 protein to the edge of the box. Two Na^+ ions were added to the solvent to neutralize the charges on the Bcl-2 protein. The protein and nonprotein groups were energy minimized with a tolerance of 2000 kJ mol^{-1} nm^{-1} using the steepest descent method for 500 steps. All bonds

were constrained using the LINCS algorithm [37], and the simulation was performed under NPT conditions, using the v-rescale coupling algorithm [38] and the Parrinello–Rahman coupling algorithm [39], which stabilized the temperature and pressure (P=1 bar, τ_P=0.1 ps; T=300 K, τ_T=0.1 ps). A smooth particle mesh Ewald (PME) method [40] was used with a cut-off of 1.4 nm for electrostatic [41] and van der Waals (vdW) [41–43] interactions. The electrostatic interactions were calculated with PME using a grid spacing of 0.12 nm. Periodic boundary conditions (PBC) were employed to eliminate surface effects [44]. The final MD simulations were carried out with a time step of 3 fs [45, 46] and without any position restraints [47]; 5,000,000 steps were performed for a total of 15 ns.

Analysis of the MD simulations

All analyses were carried out using programs included in GROMACS (version 4.0.7), VMD [48], and Pymol. Trajectories were subjected to energy analysis, global structural analysis (measuring the radius of gyration, R_g), and analyses of the RMSD (root mean square deviation) after least-squares fitting to the protein atoms except for hydrogens, secondary structure content, solvent accessibility, and intramolecular hydrogen bonding. The flexible regions and the stability of the Bcl-2 protein were predicted via the RMSF (root mean square fluctuation). The contact map was calculated using the minimum distance matrix in order to identify the native contacts. The solvent-accessible surface area (SASA) was calculated for the FLD residues. Salt bridges between oppositely charged residues in the FLD within a minimum cutoff distance of 0.5 nm were investigated. Secondary structure analysis was performed using DSSP [49]. The average structure was generated, energy minimized (using the steepest descent method), and validated using SAVes and ProSA-web before submitting it to the PMDB (id: PM0076467). ProSA-web was used to evaluate the stereochemical errors in and the quality of the model. The Z-score was measured to check the compatibility between the model's sequence and structure. The linkage method was used by the clustering tool (g_cluster) of GROMACS to generate an ensemble of structures. The secondary structures of the experimental and the PM0076467-superimposed models were predicted by the STRIDE [50] program. The electrostatic behavior between charged residues was calculated by g_dipoles.

Autodock 4.2 [51] was employed to generate the Bax (peptide)–Bcl-2 and Bad (peptide)–Bcl-2 complexes. AutoDockTools (ADT) was used to add hydrogen bonds (using AD4 atom types) to the peptides/proteins and to assign Gasteiger charges to 32 active torsions of the peptides. As our aim was to assess the binding efficacy of the BH3 receptor cleft, we made the search space large enough to include the whole BH3 cleft, and increased the exhaustiveness using affinity grids of 78×76×58 points and a spacing of 0.375 Å around the protein (via Autogrid). The Lamarckian genetic algorithm [52] was used for the conformational search. Each Lamarckian job consisted of 50 runs. The initial population consisted of 150 structures, and the maximum number of energy evaluations and generations was 2,500,000. The default values were used for the remaining parameters. The docking poses and hydrogen bonds were visualized via Accelrys Discovery Studio Visualizer 2.0 (http://accelrys.com/).

Essential dynamics analysis

The essential degrees of freedom (essential subspace) of Bcl-2 were extracted from the trajectories according to the ED method used (covariance analysis or principal component analysis) [53–56]. The ED method involves constructing the covariance matrix in order to observe the fluctuations in the coordinates of Bcl-2. Correlated motions were observed during the MD trajectories through the eigenvectors of the non-mass-weighted covariance matrix (C) for atomic position fluctuations. Before constructing C, the overall rotation and translation was removed to allow the visualization of internal motion. This was achieved by performing least squares fitting to the average structure based on the C_α coordinates. After the fitting procedure, the internal motions described by the trajectory $x(t)$ and the covariance matrix C were constructed from the coordinates of the positions of the C_α atoms:

$$C_{ij} = 1/S \sum\nolimits_t \{x_i(t) - < x_i >\}\{x_j(t) - < x_j >\} \quad (1)$$

where S is the total number of configurations, t=1, 2, … S; $x_i(t)$ are the position coordinates, with i=1, 2, ….$3N$; N is the number of atoms from which C is constructed, and $<x_i>$ is the average for coordinate i over all configurations [53].

The covariance matrix (621×621) was diagonalized to obtain the eigenvectors and eigenvalues that provide information about the correlated motions and overall flexibility throughout the Bcl-2 protein. The eigenvectors were then sorted according to their eigenvalues in descending order. Usually, the first ten eigenvectors were sufficient to describe almost all of the conformational subspace accessible to the protein. Principal component analysis was used to identify the essential subspaces explored by the simulations of the Bcl-2 protein that likely indicate changes in the BH3 cleft. The dimensionality of the essential subspace was monitored by noting the fraction of total motion described by the reduced subspace, and this was computed as the sum of the eigenvalues relative to the included eigenvectors.

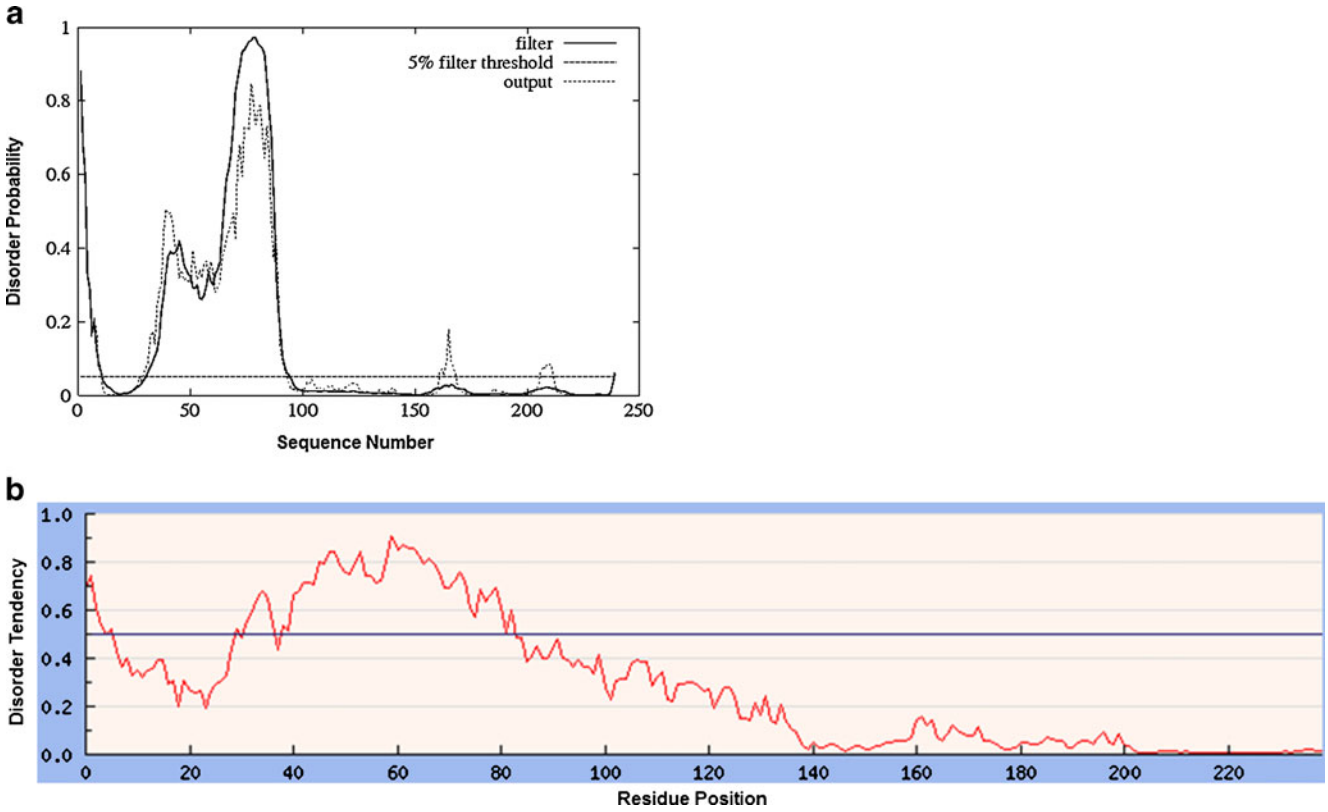


Fig. 2 **a** Plot of the disorder profile, which shows the probability of a disordered region in each residue, (*black line*). **b** The predicted tendency to be disordered (*red line*) shows large peaks for the FLD and N-terminal regions

Results and discussion

FLD disorder prediction

PDBblast gave 55 hits, among which the 1GJH (human Bcl-2, isoform 2, chain_A, NMR structure) sequence showed 73% sequence identity, 74% positives, and 19% gaps, with a score of 301 bits, an e-value of 2×10^{-82}, and a query coverage of 86%. We observed maximum sequence similarity to FLD only with the 1GJH (166 residues) template. No hits were obtained with individual pBLAST and PSI BLAST searches for the FLD (59 residues). In order to find a match for the FLD, we used the PS^2 [57] and HHpred server [58]. However, neither of these servers found a suitable template in their structure databases. These results revealed that the FLD is an unordered region. In addition, the FLD region was found to be largely unordered using two specific structural tests (DISOPRED and IUPred). The DISOPRED software predicted unordered regions, specifically in FLD (residues 34–92) and at the N-

Table 1 PROCHECK Ramachandran distributions, ERRAT overall quality factors, Verify3D scores, WHAT CHECK Z-scores, and ProSA-web Z-scores for the 1GJH, SW, CW, MOD, MODLOPT, and PM0076467 models

3D model	Ramachandran statistics				ERRAT (%)	Verify3D (%)	WHAT CHECK Z-score	ProSA-web Z-score
	Core (%)	Allowed (%)	Generous (%)	Disallowed (%)				
1GJH	73.6	24.3	1.4	0.7	79.355	96.36	−7.098	−6.61
SM	72.9	23.5	2.4	1.2	55.330	91.26	−5.909	−2.07
CW	84.5	10.1	4.2	1.2	42.347	97.34	−1.182	−5.97
MOD	88.7	9.5	1.2	0.6	36.788	99.04	−1.171	−6.52
MODLOPT	85.7	12.5	1.8	0.0	46.821	81.25	−2.732	−7.10
PM0076467*	81.5	16.7	1.2	0.6	80.233	98.08	−2.404	−6.63

* MD-simulated average structure

a

```
>P1;Bcl-2
MAHAGRTGYDNREIVMKYIHYKLSQRGYEWDAGDVGAAPPGAAPAPGIFSSQPGHTPHPA
ASRDPVARTSPLQTPAAPGAAAGPALSPVPPVVHLTLRQAGDDFSRRYRRDFAEMSSQLH
LTPFTARGRFATVVEELFRDGVNWGRIVAFFEFGGVMCVESVNREMSPLVDNIALWMTEY
LNRHLHTWIQDNGGWDAFVELYGPSMR

>P1;1GJH
--HAGRTGYDNREIVMKYIHYKLSQRGYEWDAGD---------------DVEENRTEAPE
GTESEV-------------------------VHLTLRQAGDDFSRRYRRDFAEMSSQLH
LTPFTARGRFATVVEELFRDGVNWGRIVAFFEFGGVMCVESVNREMSPLVDNIALWMTEY
LNRHLHTWIQDNGGWDAFVELYGPSMR
```

b

```
>P1;Bcl-2
MAHAGRTGYDNREIVMKYIHYKLSQRGYEWDAGDVGAAPPGAAPAPGIFSSQPGHTPHPA
ASRDPVARTSPLQTPAAPGAAAGPALSPVPPVVHLTLRQAGDDFSRRYRRDFAEMSSQLH
LTPFTARGRFATVVEELFRDGVNWGRIVAFFEFGGVMCVESVNREMSPLVDNIALWMTEY
LNRHLHTWIQDNGGWDAFVELYGPSMR

>P1;1GJH
--HAGRTGYDNREIVMKYIHYKLSQRGYEWDAG--------------------------
---DDVEEN---RTEAPEGTES--------EVVHLTLRQAGDDFSRRYRRDFAEMSSQLH
LTPFTARGRFATVVEELFRDGVNWGRIVAFFEFGGVMCVESVNREMSPLVDNIALWMTEY
LNRHLHTWIQDNGGWDAFVELYGPSMR
```

Fig. 3 **a** Manually aligned Bcl-2 (207 residues) and 1GJH (166 residues) sequences in PIR format. This alignment was used by MODELLER to generate MOD. **b** PIR-format alignment of Bcl-2 (207 residues) and 1GJH (166 residues), as generated using the Gonnet matrix in ClustalW

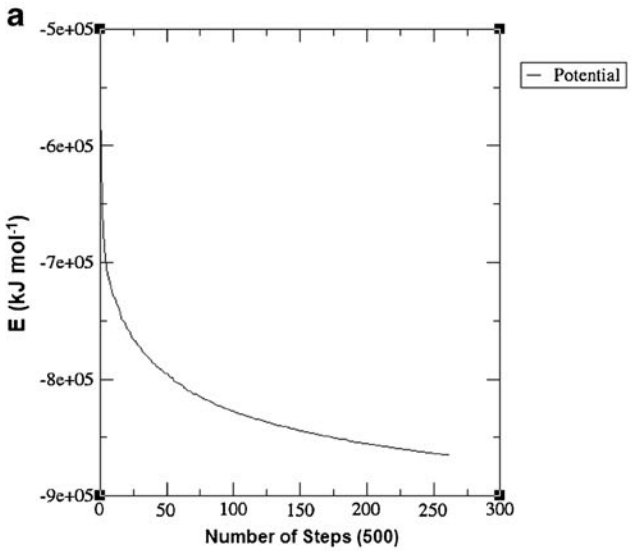
terminal domain (residues 1–9) (Fig. 2a). The DISOPRED program predicted that the Bcl-2 protein is unable to yield a good consensus for the FLD segment. The scattered output curve exhibits the presence of non-alpha and non-beta regions in the FLD zone. IUPred predicted the tendency for disorder in the Bcl-2 protein, and indicated that all four BH domains were well structured. The FLD residues show a high tendency to be disordered, with a probabilistic score of >0.5 (for residues 32–83) (Fig. 2b), but as this score is <1.0, the region does not appear to be completely disordered [59].

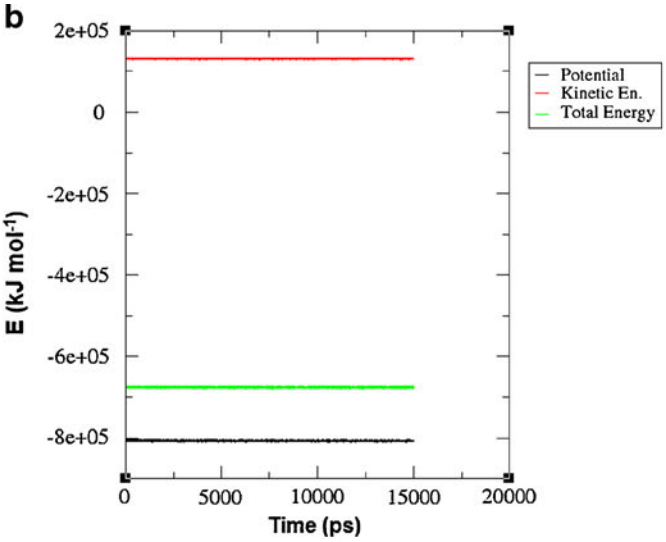
Molecular modeling and optimization

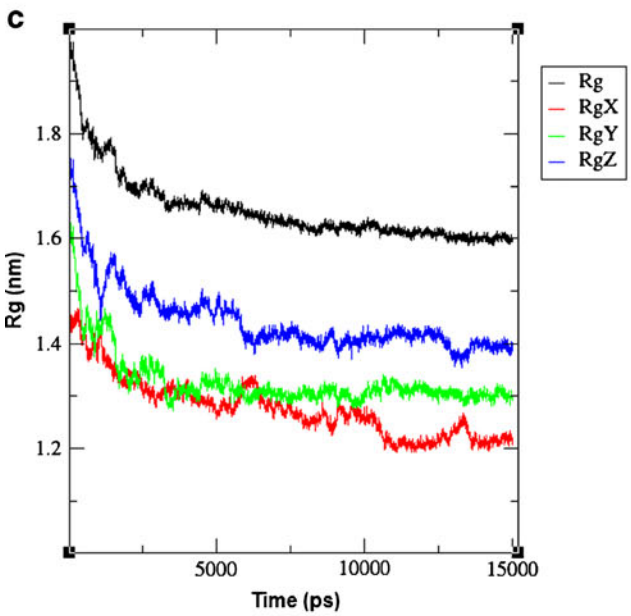
Upon aligning the 1GJH sequence with its PDB file, we noted differences in the FASTA format and coordinate files for the first two residues (i.e., Met1 and Ala2). So, we manually replaced these residues with gaps in the PIR-format alignment file. After comparing the two different models (MOD and CW) (Table 1), we observed that manual alignment (Fig. 3a) generates a better model (MOD) than that (CW) generated using default alignment with the ClustalW program (Fig. 3b). MOD showed fewer conformational strains as compared to CW, so we chose this model for further analysis. MOD (residues 1-207) showed N-terminal, BH4, BH3, BH1, BH2 and FLD (regulatory domain) regions. In pair-wise sequence alignment, we did not observe identity when comparing the complete FLD region of Bcl-2 with 1GJH, so it is likely that MOD would be constrained in the FLD. The accurate conformation of FLD was predicted by removing a bump, using the loop optimization protocol in the loop modeling method of MODELLER. Three loop models were outputted, among which the last model (MODLOPT) was chosen because this model had the best-refined loop. MODLOPT was then subjected to energy minimization for 22 energy iterations (total energy range: E=7888150.500 kJ mol^{-1} to −1671.042 kJ mol^{-1}). An RMSD of 4.189 Å for 1374 atoms was observed between the energy-minimized MODLOPT and SM models. We also identified two beta-strands in the FLD of SM, between residues 35 and 37 and residues 48 and 50.

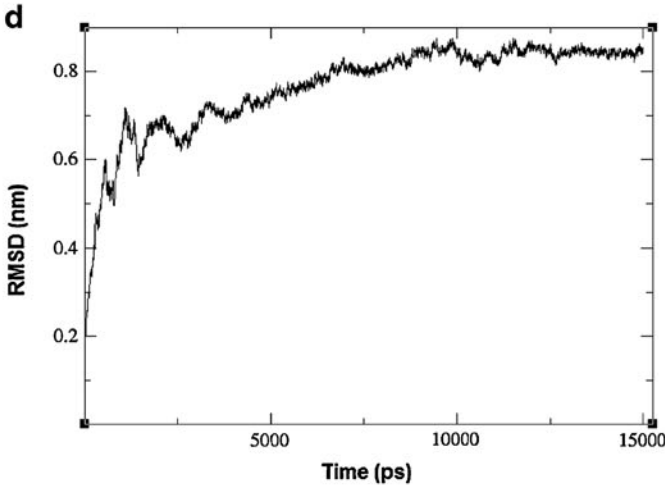
A comparison of all of the structures (Table 1) showed that MODLOPT was the best. ERRAT was used to calculate the overall quality of each model, and this showed that MODLOPT has the highest overall quality (42.857) of any of the models (Table 1). The 1GJH model had an overall quality of 79.355, but when the FLD was added to this model, its quality was observed to decrease [14]. WHAT CHECK was employed to evaluate the

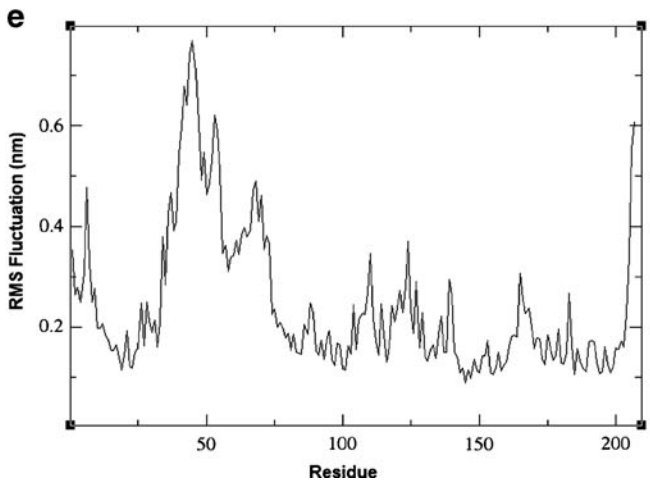
Fig. 4 **a** Potential energy of energy-minimized MODLOPT, which converged after 261 steps. **b** Stable potential, kinetic and total energy plot for the MD simulation. **c** Time evolution of the radius of gyration shows the compactness of structures with respect to time. The radius of gyration of a group of atoms was computed about the *x*- (*RgX*), *y*- (*RgY*) and *z*- (*RgZ*) axes, as shown by the three colored lines, which indicate the global shape of the molecule along the *x*-, *y*-, and *z*-coordinates. **d** Root mean square deviations (RMSD) of structures with respect to simulation time. **e** RMS fluctuations (in nm) of residues; greater fluctuation was observed in the loop region


a
Potential
E (kJ mol-1)
Number of Steps (500)


b
Potential
Kinetic En.
Total Energy
E (kJ mol-1)
Time (ps)


c
Rg
RgX
RgY
RgZ
Rg (nm)
Time (ps)


d
RMSD (nm)
Time (ps)


e
RMS Fluctuation (nm)
Residue

geometries of the models, and it showed that the Ramachandran Z-score (−7.098) was very low for 1GJH but normal (i.e., between −4 and +4) for the other 3D models. Furthermore, a very low chi-1/chi-2 correlation Z-score was observed for the 1GJH model (SD of Z-score= −6.113), but the corresponding values for the other structures were normal.

MD simulations

After validation, we realized that the energy-minimized structure of MODLOPT needs to be improved in terms of stereochemical and overall quality through optimization. MD simulations can improve the FLD structure by using a conformational search approach to study this type of unordered region [60, 61]. Thus, we employed MD to predict the folding of the FLD (i.e., the equilibrium between the folded and unfolded states of Bcl-2 containing the unstructured region). We observed conformational changes at 300 K on the minimization and optimization of the FLD domain. This observation encouraged us to attempt an extensive study of the conformational behavior of the Bcl-2 protein at 300 K over a time scale of 15 ns. The energy-minimized MODLOPT was solvated in an aqueous environment and then energy minimized. After that, the GROMACS force field was used for MD simulations. We chose several structural features along the MD trajectories in order to predict the near-native conformation [42], since discriminating models on the basis of free energy alone is not a reliable approach, due to the very small difference between the free energies of native and decoy structures [62–69].

Predicting Bcl-2 protein flexibility

We performed MD simulation analysis to predict the flexibility of FLD and its effects on the Bcl-2 protein. The energy coordinates were observed to converge after 261 steps, at which point the lowest potential energy (−865117.25 kJ mol^{-1}) was calculated (Fig. 4a). The average potential energy was observed to be −653040 kJ mol^{-1}. Finally, the potential, kinetic and total energies of the system throughout the 15,000 ps MD simulation run were calculated (Fig. 4b). An average potential energy of −806634 kJ mol^{-1}, an average kinetic energy of 130843 kJ mol^{-1}, and an average total energy of −675791 kJ mol^{-1}

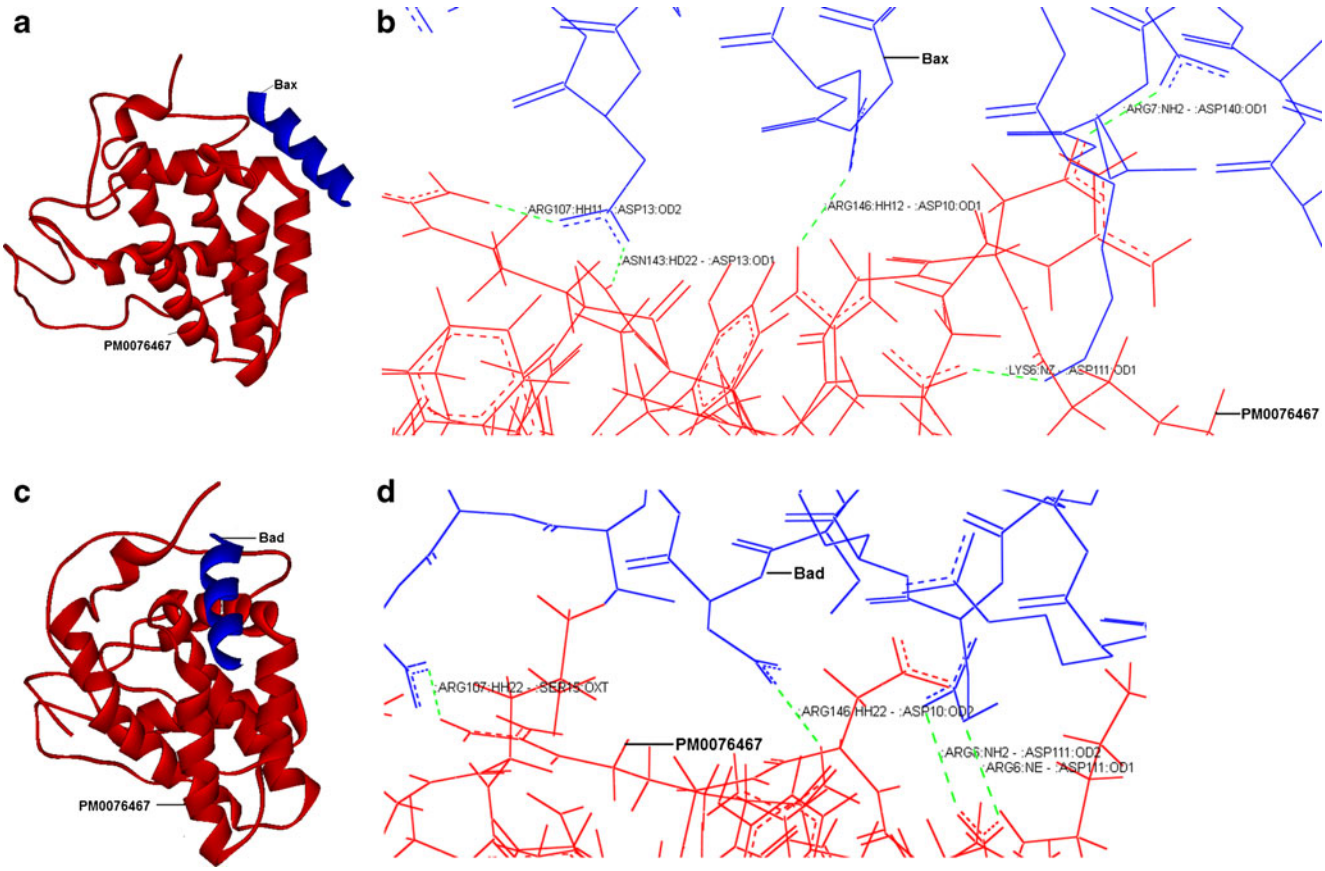


Fig. 5a–d Docking poses of PM0076467 (*red*) with **a** Bax (*blue*) peptide and with **c** Bad peptide; the hydrogen bonds observed between the complexes are shown in **b** and **d**, respectively

Table 2 Members of the 23 clusters, obtained on the basis of the RMSD; each cluster's middle member was submitted to the PMDB

Cluster	No. of members of the cluster	Mean RMSD (nm)	Middle structure	RMSD of middle structure (nm)	Members of the cluster (ps)	PMDB id
1	1	0	–	–	0	PM0077081
2	2	0.095	3	0.095	3, 6	PM0077082
3	2	0.097	9	0.097	9, 12	PM0077083
4	1	–	15	–	15	PM0077084
5	17	0.146	42	0.132	18–66	PM0077085
6	4	0.104	72	0.096	69–78	PM0077086
7	7	0.119	90	0.112	81–99	PM0077087
8	8	0.114	114	0.106	102–123	PM0077088
9	4	0.104	132	0.098	126–135	PM0077089
10	1	–	138	–	138	PM0077090
11	4	0.101	147	0.094	141–150	PM0077091
12	27	0.153	195	0.141	153–231	PM0077092
13	32	0.169	279	0.153	234–327	PM0077093
14	26	0.152	366	0.140	330–405	PM0077094
15	4	0.103	414	0.099	408–417	PM0077095
16	28	0.153	456	0.137	420–501	PM0077096
17	53	0.181	615	0.166	504–660	PM0077097
18	26	0.141	693	0.130	663–738	PM0077098
19	50	0.169	831	0.154	741–888	PM0077099
20	68	0.182	1002	0.161	891–1092	PM0077100
21	91	0.190	1257	0.173	1095–1365	PM0077101
22	148	0.198	1635	0.176	1368–1809	PM0077102
23	4397	0.291	7656	0.248	1812–15,000	PM0077103

were obtained. These three energies were observed to remain stable during the whole run.

The compactness, shape, and folding of the overall Bcl-2 structure at different time points during the trajectory can be seen in the plot of Rg (Fig. 4c). Rg was computed for atoms that were explicitly mass-weighted. 1.60 nm was the average converged value for Rg for 13,000–15,000 ps simulations. This converged and equilibrated Rg value showed that stability was obtained by the FLD. The evaluation of the Rg revealed that there was a difference between the initial and final value for the structure (it decreases from 1.9 nm to 1.8 nm). These results indicate that the final structure is more compact than the initial structure. The RMSD value of 0.84 nm was obtained after a least squares fit to the structure (except for hydrogens) of the protein at the equilibrium plateau from 12,714 ps to 15,000 ps (Fig. 4d). A decrease in Rg indicates FLD folding, whereas the attainment of an equilibrium state indicates stability of structures with similar RMSDs. The RMSF of the residues shows the overall quality of the model, the flexibility of the FLD region, and the displacement of the residues about the average position of the model. The RMSF plot (Fig. 4e) shows that the fluctuations of the FLD peak at residues Gly36 (0.4236 nm of fluctuation), Ala45 (0.7697 nm of fluctuation) and Arg68 (0.4904 nm of fluctuation). The RMSF plot is uniform, especially in the BH3 receptor cleft, indicating the importance of this region in Bcl-2 protein activity.

The energy-minimized average structure extracted from MD trajectory can be used to design new inhibitors or to identify a new binding site that could be useful in drug discovery [70]. Therefore, we computed the average structure from the MD simulations, which was then further energy minimized and validated by ProSA-web (Z-score of −6.63) (Table 1). The average structure obtained after the MD simulations displayed an overall quality factor of 80.233%, as predicted by the ERRAT program (Table 1); this shows improved model quality. This model was observed to have 81.5% of its residues in the core, 16.7% in the allowed, 1.2% in the generously allowed, and 0.6% in the disallowed regions. Verify3D analysis showed that 98.08% of the residues had average 3D-1D scores of >0.2 and G-factor dihedral values of −0.26 (acceptable values of the PROCHECK G-factor lie within the range of 0 to −0.5). Thus, all of the stereochemical and overall quality parameters were better than those of the other models (Table 1). The validated model was subsequently submitted to the PMDB (id: PM0076467).

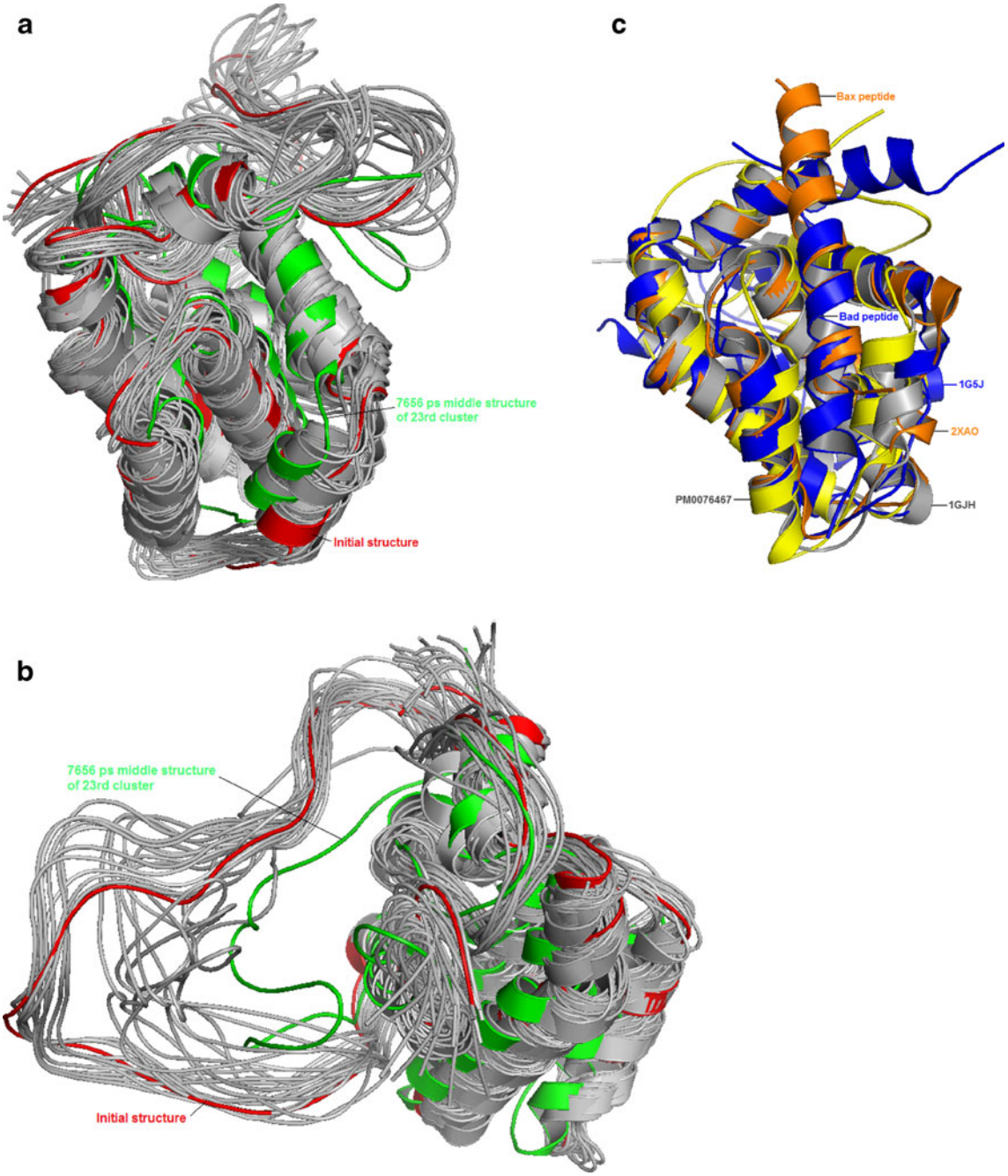


Fig. 6 **a** Top view of the ensemble of 23 structures, showing the flexibility of the BH3 cleft. The initial structure is shown in *red*, and the structure at 7656 ps (middle structure of the 23rd cluster) is shown in *green*. **b** Side view of the ensemble of 23 structures, showing the structural changes in the FLD. The initial structure is shown in *red*, and the structure at 7656 ps (the middle structure of the 23rd cluster) is shown in *green*. **c** Superimposition of the models PM0076467 (*yellow*), 1GJH (*gray*), 2XA0 containing Bax peptide (*orange*) and 1G5J containing Bad peptide (*blue*), showing the variability of the BH3 cleft

In order to analyze the accuracy of the predicted average model, we docked it with pro-apoptotic peptides. The binding mode in the PM0076467 structure was predicted by docking this structure with the Bax peptide (residues 59–73, BH3 motif) (Fig. 5a) and the Bad peptide (residues 110–124, BH3 motif) (Fig. 5c). The docked complexes show that both of the peptides bind efficiently into the BH3 cleft. We identified the same hydrogen bonds in these complexes as those found in previously solved complex structures [71, 72]. Arg 107: HH11–Asp 13:OD2, Asn 143:HD22–Asp13:OD1, Arg146: HH12–Asp10:OD1, Asp111:OD1–Lys6:NZ and Asp140: OD1–Arg7:NH2 hydrogen bonds were observed between PM0076467 and Bax (Fig. 5b), and Arg107:HH22–Ser15: OXT, Arg146:HH22–Asp10:OD2, Arg6:NH2–Asp111:OD2

Table 3 Secondary structure contents of 1GJH, 2XA0, and PM0076467 for the most flexible region (residues 108–125) designated as H (alpha helix), G (310 helix), T (turn) and C (coil)

Residue	Number	Secondary structure contents of models		
		1GJH	2XA0	PM0076467
Tyr	108	T	H	H
Arg	109	T	H	H
Arg	110	T	H	H
Asp	111	T	H	H
Phe	112	H	C	H
Ala	113	H	C	H
Glu	114	H	C	H
Met	115	H	G	H
Ser	116	H	G	H
Ser	117	H	G	H
Gln	118	C	G	H
Leu	119	C	C	C
His	120	C	C	C
Leu	121	C	C	C
Thr	122	C	T	C
Pro	123	T	T	H
Phe	124	T	T	H
Thr	125	T	H	H

and Arg6:NE–Asp111:OD1 hydrogen bonds were noted between PM0076467 and Bad (Fig. 5d). These results indicate that FLD folding does not hinder or affect the binding of antagonistic peptides to the BH3 cleft of Bcl-2, and that the BH3 receptor cleft may play an important role in FLD folding.

Clustering of conformations

Several computationally designed and redesigned proteins have led to significant breakthroughs in biotechnological and biomedical applications, such as designing new biocatalysts [73–76], enhancing protein binding affinity [77] and redesigning protein binding specificity [78–80], and redesigning a protein in order to bind to cofactors [81]. MD simulation has an important role to play in elucidating the functional mechanisms attained by protein structures [82]. We have computed an average structure from a 15 ns simulation, but this does not necessarily mean that it is a physically meaningful structure. To characterize the behavior of the models along the trajectories, we computed an ensemble of structures based on the RMSD using the clustering tool of GROMACS. Here, we built 23 clusters using the linkage method, with all of the clusters corresponding to RMSD values ranging from 0 to 0.291 nm (Table 2). In this way, we obtained a coarse estimate for the conformational space along the trajectory explored by the same or different models. This indicates that all of the conformations that are in the same cluster have the same potential energy, because two clusters are formed when the system crosses a sufficiently high energy barrier to reach a new potential energy valley. The results shown in Table 2 allow improved discrimination between different types of models than the results shown in Fig. 4d in the form of a graph. The models obtained from 1812 ps to 15,000 ps belonged to the same cluster, with similar RMSDs. This implies that models in this cluster have the same energy and do not deviate from one another much.

To predict the flexibility of the clusters, we superimposed the middle structures of all 23 clusters. Superimposition of the 23 clusters (Fig. 6a and b) showed that the FLD folds due to changes that occur in the BH3 receptor cleft. We also observed that these changes occur mainly due to the highly flexible behavior and variability of the secondary structure content of residues 108–135, in the region connecting BH3 and BH1. These observations suggest a folded binding site for the FKBP38 protein, which is reported to bind with the FLD [9]. The changes in the BH3 cleft are also likely to regulate the binding of other proteins to the FLD. Subsequently, each cluster's middle structure was energy minimized by the Swiss-Pdb viewer program using the default parameters, and then submitted to the PMDB (Table 2). This ensemble of closely related structures should help us to predict the backbone flexibility based on computationally designed and experimentally obtained structures [83–85].

We observed the presence of the flexible region by superimposing the MD-simulated average structure (PM0076467) onto two experimental Bcl-2 structures (1GJH and 2XA0, the model of the complex of Bcl-2 with the Bax peptide) and one structure of Bcl-X_L (i.e., 1G5J, the structure of the complex of Bcl-X_L with the Bad peptide) (Fig. 6c). Stability was achieved by reducing the length of the FLD, thus stabilizing the secondary structure [86–88]. The BH3 cleft was observed to be flexible—which is essential if Bcl-2 is to optimally heterodimerize with pro-apoptotic proteins; the flexible nature of the native protein is essential to its function [89]. Substantial differences in the BH3 cleft were observed in the dynamics of the structural features of the backbones of the simulated and template structures. However, we identified the stabilized conformations (including that of the FLD) from the MD simulations.

Superimposing the models (2XA0, IGJH, 1G5J, and PM0076467) helped us to calculate the differences between them on the basis of the RMSD (Fig 6c). The observed RMSD between 2XA0 and 1GJH was 1.105 Å, that between 1G5J and 1GJH was 1.644 Å, that between 1G5J and 2XA0 was 1.234 Å, that between PM0076467 and

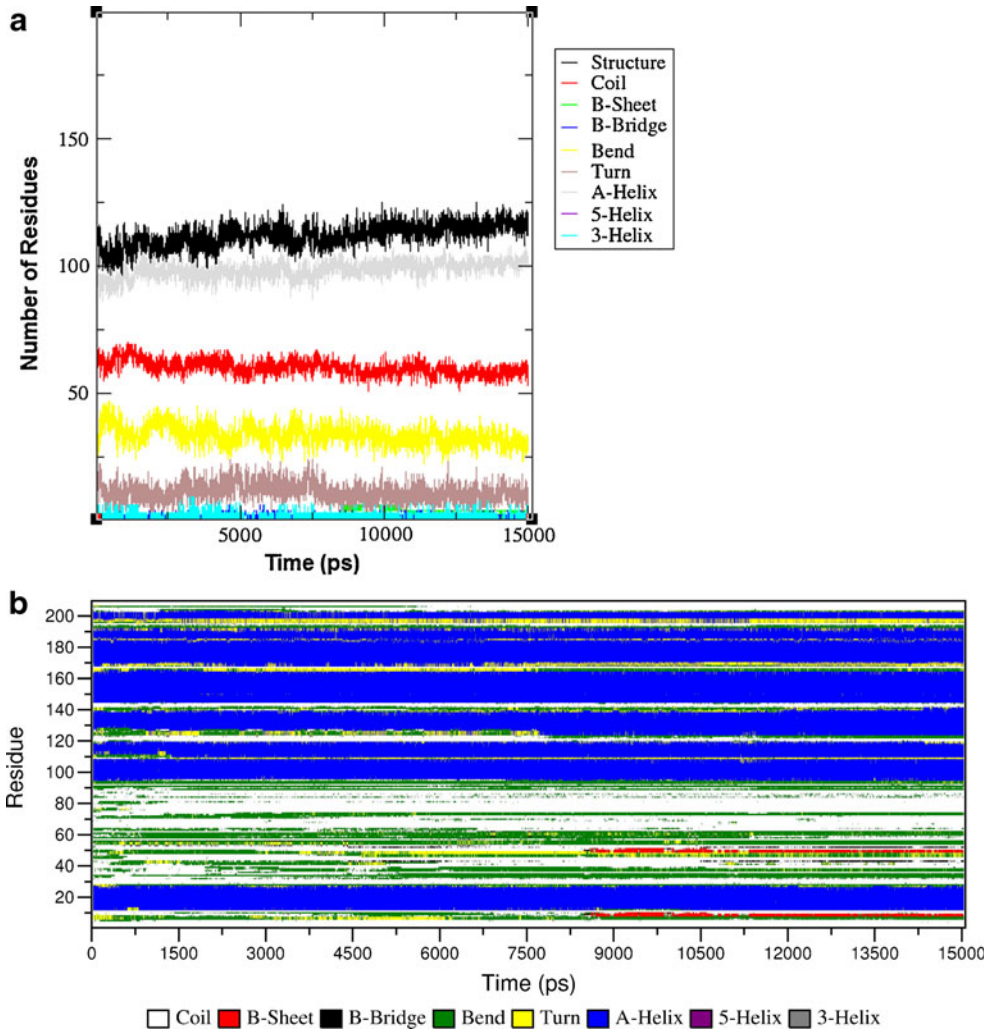


Fig. 7 **a** Average secondary structure contents along the trajectory of Bcl-2. Note that they increase from the contents of the corresponding starting structure. **b** Predicted secondary structure assignment for Bcl-2 residues along the 15,000 ps trajectory

1GJH was 2.084 Å, and that between PM0076467 and 2XA0 was 2.001 Å. We have observed that 2XA0 BH3 cleft is more inclined towards the Bcl-X_L structure. These results revealed that the 2XA0 and 1GJH models were associated with a transition state, as shown by their differences in the BH3 cleft. These BH3 cleft differences may be due to the large dynamic motion found in the region connecting BH3 and BH1 (residues 108–135). This region was found to be the most flexible one; it affects or constricts the BH3 cleft.

Since secondary structure content predicts the stability of a protein, we calculated the secondary structure content and observed transitions in this flexible region (Table 3). The region covering residues 108–111 was observed as a turn in 1GJH or a helix in 2XA0 and PM0076467, whereas the region covering residues 112–114 was found to be a helix in 1GJH and a coil in 2XA0. The structural features of the PM0076467 model show more of a resemblance to the structural features of 2XA0 than 1GJH. Residues 126–135 correspond to a helix in 1GJH, PM0076467, and 2XA0, indicating the stability of this region. Residues 108–116 represent the most important and flexible region, and the one that is likely to be responsible for most of the transitional behavior in the BH3 cleft. We did not perform any mutational studies, because in some cases it appeared that a residue mutation near or at the active site increases the flexibility, thereby decreasing the activity of the interacting site by disrupting its rigid active-site geometry [90].

Structure and stability analysis

The average numbers of residues involved in secondary structures were calculated by the DSSP program. The average structure was observed to contain an alpha-helix (94–104 residues), coils (57–68 residues), bends (24–44 residues), turns (2–19 residues), a beta-sheet (4–6 residues), a beta-bridge (2 residues) and a 3-helix (3–8 residues) (Fig. 7a). It is worth mentioning that the percentage of bends in the simulated conformation deviates greatly from the starting structure, but becomes stable after 12,000 ps and is maintained up to 15,000 ps (Fig. 7b). This may be

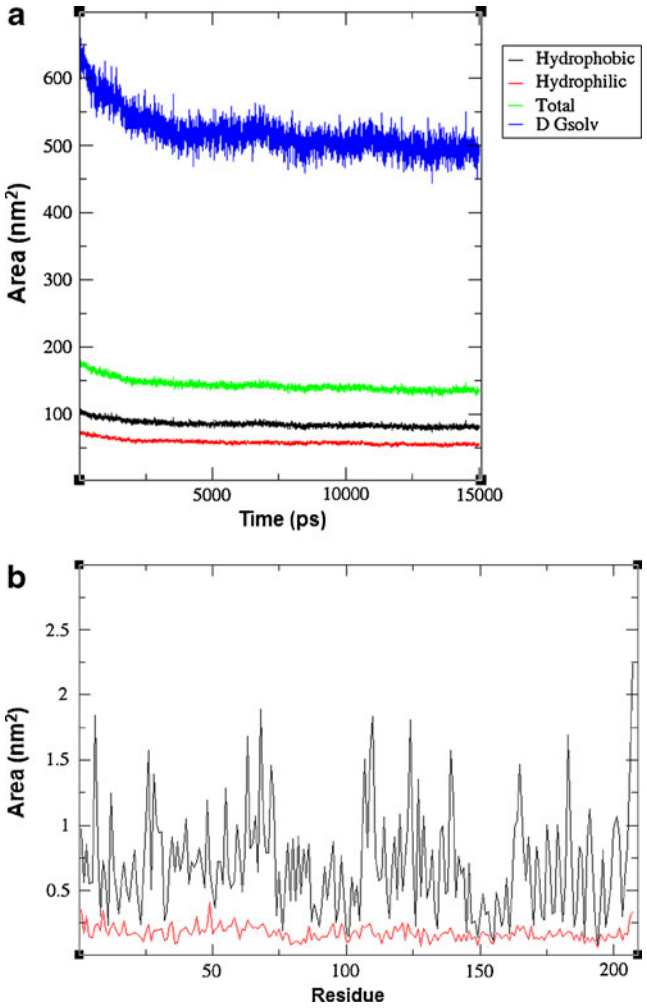


Fig. 8 **a** Calculated solvent-accessible surface area (SASA) during the whole simulation, showing the hydrophobic, hydrophilic, and total areas equilibrated at 15,000 ps. **b** The SASA per residue, displayed as *black peaks*; the *red line* indicates the standard deviations

due to the fact that several coils are converted into bends and beta-sheets in the FLD and the N-terminal region during simulations, but no alpha-helix was observed. This indicated that the FLD would probably only contain beta-sheets, as shown by the whole simulation. The structure assignment prediction for residues Ala45 to Ala60 in the FLD region shows stable beta-strands (in red) and bends (in green). This observation is in accord with the beta-strands observed in the SM model, and indicates that beta-strands are likely to exist in the native structure of the Bcl-2 FLD region.

The solvent-accessible surface area (SASA) was calculated (Fig. 8a and b), and this showed that both hydrophilicity and hydrophobicity decreased and then equilibrium was attained after 12,500 ps of simulation. The hydrophobic SASA was found to be 77–83 nm^2 and the hydrophilic SASA 52–56 nm^2. In addition, the free energy of solvation was found to cover a larger area than the total SASA (Fig. 8a). The calculation of the average SASA (over time, per residue) indicated that Pro40, His55, Pro59, Arg63, Arg68 and Gln73 had the greatest exposure to the solvent (Fig. 8b). The residues Arg63 and Arg68 covered areas of 1.67 nm^2 and 1.88 nm^2 of the solvent and showed the highest hydrophilicity. The hydrophobic residues Ala61, Ala76, Ala81 and Val89 covered areas of 0.489265, 0.193879, 0.423293, and 0.314359 nm^2; among these, Ala76 showed the highest hydrophobicity (Fig. 8b). These hydrophobic residues likely comprise a potent active site in the FLD for a ligand or another protein due to its hydrophobicity.

The secondary and tertiary structural changes that occurred during the whole simulation were predicted by the mean smallest minimum distance matrix (Fig. 9a), which accurately describes the native state of the protein. This can be obtained from the matrix of native contacts. The maps derived from these contacts contain essential geometric topological information, and describe structural interaction patterns in the protein [91, 92]. We observed the formation of contact maps between the backbone residues, and large changes in the FLD at different time scales (Fig. 9b, c and d). The formation of new contacts between residues 35 and 80 and residues 75 and 110 (in the average/mean distance matrix) was observed to stabilize the protein structure, and it helped to produce the secondary structure conformation of the FLD. This also suggests that the binding of FKBP38 and other interacting proteins to the FLD probably takes place at this region.

It appears that the hydrogen bonds stabilize the protein and have crucial role to play in protein folding [93–101]. The average number of intramolecular hydrogen bonds was calculated in different simulations. The intramolecular hydrogen bonds within the whole protein were found to involve 296 donors and 584 acceptors in total (within a cutoff distance of 0.35 nm and angle of 30°). The intramolecular hydrogen bonds that formed within the FLD (Table 4) show that Ala residues participate most frequently in hydrogen bonding. The intramolecular hydrogen bonds (within the FLD) were categorized according to whether they were between hydrophobic and hydrophilic, hydrophobic and neutral group, or hydrophilic and neutral group residues. Hydrogen bonding between hydrophobic and hydrophilic group residues was observed for the Val92N–Pro90O, Thr69N–Ala67O, Arg68N–Val66O, Ala67N–Pro65O, Val66N–Asp64O, Asp64N–Ala61O, Arg63N–Ala61O, Phe49N–Pro44O, Phe49N–Pro46O, Ile48N–Pro46O, Ala43N–Pro39O, Ala43N–Pro40O, Ala42N–Pro39O, and Ala42N–Pro40O bonds. Hydrogen bonds between hydrophobic and hydrophilic group residues were present inside the core and on the surface of the protein, and so they would probably participate in the stabilization of the FLD.

Besides intramolecular hydrogen bonding, salt bridges (Table 5) were observed between oppositely charged

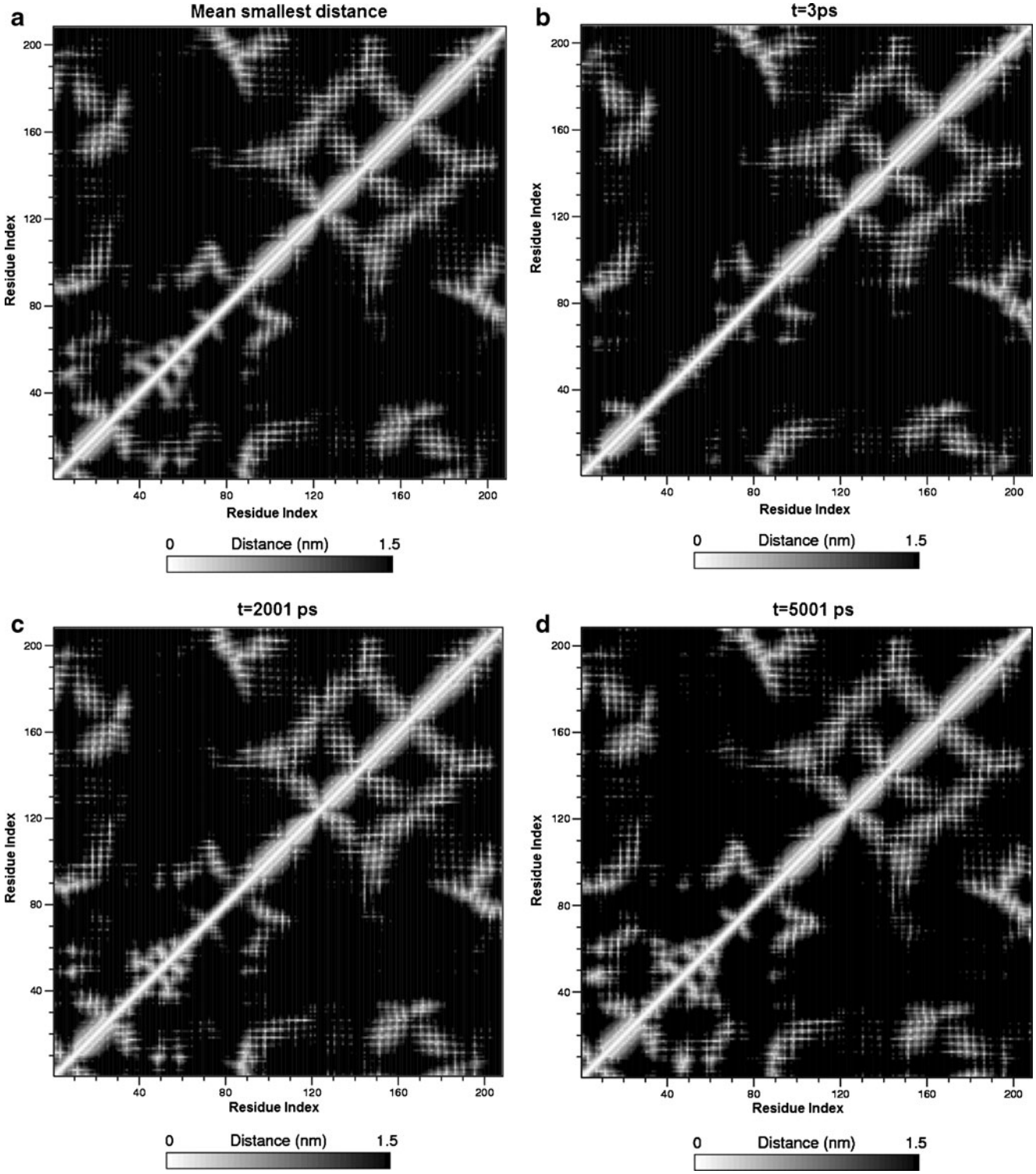


Fig. 9a–d Contact maps formed between the residues of Bcl-2, as calculated from the minimum distance matrix: **a** mean changes; **b** at 3 ps; **c** at 2001 ps; **d** at 5001 ps, for the whole simulation

residues. These salt bridges, may have a stabilizing effect on the overall structure of the Bcl-2 protein. This stabilizing effect is mainly attributed to the Arg and Asp residues present within the FLD. The residues Asp34, Asp64, Arg63, and Arg68 were noted to be particularly important, and are likely to be responsible for FLD stability. The mutation of these residues would understandably decrease the stability of the protein.

Table 4 Intramolecular hydrogen bonds within the FLD

Donor	Acceptors
Thr74N	Pro71O, Leu72O
Asp64N	His55O, Ala61O, Ser62O
Arg63N	His55O, Ala61O
Ser62N	Pro59O, Ala60O
Ala61N	His58O, His59O
Ala60N	Ile48O, His58O
His58N	Ser50O, Thr56O, Ala61O, Ser62O
Thr56N	Gln52O, Pro53O, Gly54O, Ser62O
His55N	Gln52O, Pro53O
Gly54N	Ala38O, Pro39O, Pro40O, Gln52O
Gln52N	Ser50O, His55O, Thr56O, Pro57O
Ser51N	Ala42O, Pro44O, Phe49O
Ser50N	Gly48O, Pro57O
Phe49N	Pro44O, Ala45O, Pro46O, Gly47O
IlE48N	Ala45O, Pro46O
Ala43N	Pro39O, Pro40O, Gly41O, Ser51O
Ala42N	Pro39O, Pro40O, Ser51O
Gly41N	Pro39O, Ser51O
Ala38N	Val35O, Gly36O

These results should help to enhance our knowledge of the interplay of forces that lead to the stability of the FLD through salt bridges and hydrogen bonding. All these results suggest that the FLD naturally adopts a folded state rather than a random coil structure.

Calculating the dipole moment in order to predict the electrostatic behavior

Permanent electric dipole moments contribute to the electrostatic forces (long-range forces) in biomolecules, and play an important role in determining biomolecule folding, structure, and properties. Alpha-helix conformations of proteins lead to large macro-dipoles that induce strong electric fields [102]. The fluctuations of a protein's polar groups in response to conformational changes play a key role in the folding of its structure and its binding properties [103]. The total dipole moment of Bcl-2 was evaluated (Fig. 10), and the result was found to be similar to that observed for 2XA0 (221.10 D). We observed that the differences in dipole moment direction (*x*, *y* and *z*) and the average fluctuations were in accordance with the X-ray structure (2XA0). Moreover, the 1GJH template displayed the highest dipole moment (635.91 D), which indicates that this structure has high overall rigidity and displays low flexible strength and kinetics in its associations with small molecules or proteins. This would also affect its anti-apoptotic activity, which is reported to decrease with structural rigidity [104]. This implies that 1GJH shows completely different electrostatic behavior to 2XA0 due to its large dipole moment from its charged residues, which in turn indicates unfavorable charge–macro-dipole interactions in the 1GJH template, although the interactions are favorable for simulated structures. The dipole moment analysis accurately characterized the active Bcl-2 models along the trajectory, which further validated the use of our MD simulations to obtain the near-native or native conformation.

Table 5 Residues that form salt bridges in the FLD region

Positive	Negative
Arg12	Asp34
Arg63	Asp102
Arg63	Asp64
Arg68	Asp102
Arg63	Asp34
Arg98	Asp34
Arg207	Asp64
Arg68	Asp64
Arg98	Asp64
Arg63	Glu29
Arg68	Glu29
Lys17	Asp34
Lys17	Asp64

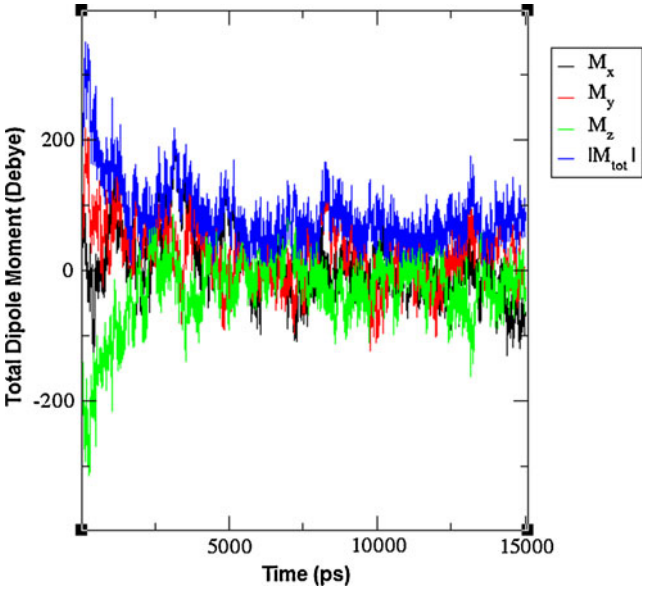


Fig. 10 Total dipole moment (*blue*) versus time, showing the electrostatic behavior during simulations

Essential dynamics

Molecular dynamics focuses on internal protein motions, and the correlation of the measured flexibility based on ordered parameters with the configuration entropy [44].

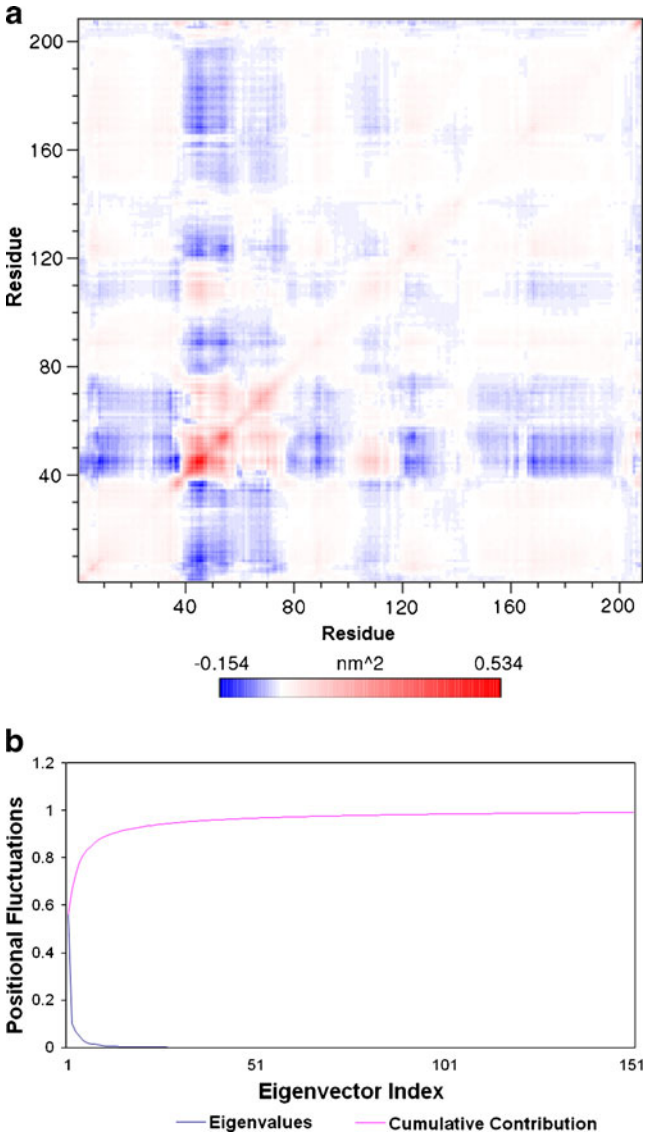


Fig. 11 **a** The covariance matrix shows anticorrelated and correlated motions between atoms. *Red* indicates that two atoms move together (correlated motions); *blue* indicates movement in opposite directions (anticorrelated motions). The *color intensity* indicates the amplitude of the RMS fluctuations (see the *color bar*). **b** Positional fluctuations, indicating the eigenvalues and the cumulative contributions along 149 eigenvectors. Eigenvalues fluctuations shown in decreasing order of magnitude and cumulative contribution fluctuation by increasing order of magnitude, as obtained from the C_α coordinate covariance matrix from a solvent simulation

Therefore, we applied principal component analysis to the Bcl-2 trajectory in order to identify large-scale collective motions of atoms and predict the flexible behavior of the FLD and the BH3 cleft.

The collective (correlated) motions of the atoms in the protein are key to its biological function [90, 105]. This study revealed that the structures underlying the atomic fluctuations (corresponding to B-factors) occur due to correlated interactions between Bcl-2 atoms. Correlated motions at the atomic level helped us to predict if the overall fluctuations of Bcl-2's C-α atoms in the system have functional or biophysical relevance. In addition, these atomic fluctuations describe the flexible behavior of the Bcl-2 protein. The covariance matrix captured the degree of collinearity in atomic motions for each pair among 207 residues. Cross-correlations between residue fluctuations helped us to identify highly correlated, moderately correlated, and anticorrelated regions (Fig. 11a). The covariance 621×621 symmetric matrix (Fig. 11a) shows that the large group of atoms in the FLD moves in an anticorrelated manner in relation to other domains. Using this matrix, we detected that the residues 40–75 (in the FLD) are highly correlated (red color), and that there is a weak atomic correlation (light red) between residues 105–118 (in-between the BH3 and BH1 domains) and residues 40–55 (in the FLD). This proved that the FLD and the BH3 cleft may play important roles in the regulation of Bcl-2 activity. These results indicate that any change in the FLD would likely have some effect on the BH3 cleft, as observed in the binding of p53 to FLD and the consequent decreased interaction of Bcl-2 with Bax through its BH3 cleft [6]. These results are in accordance with our superposition results (Table 3 and Fig. 6c). Hence, the results indicate that most of the internal motions of the C_α atoms of Bcl-2 protein are confined within a subspace with very small dimensions.

The diagonalization of this matrix leads to the generation of eigenvectors and eigenvalues (Fig. 11b). Each eigenvector describes a collective motion performed by particles, whereas the eigenvalues indicate how much a particular atom participated in the motion [106]. The calculated eigenvalues and cumulative contribution in the collective motion were calculated for the first 149 eigenvectors. 80% of the motion of the system was described by the first five eigenvectors, which explained the overall positional fluctuations contributing to the largest motions (Fig. 11b). Similarly, the C_α atom displacements showed that the largest motions are confined to the first five eigenvectors and producing large motions in the FLD and the BH3 cleft.

The eigenvectors describe the essential degrees of freedom (containing large-scale global motions), which are vital for protein function, and represent the part of conformational space called essential subspace, whereas near-constrained subspace refers to less interesting local fluctuations or Gaussian distributions [56, 105, 107–111]. The two-dimensional graphs of eigenvector 1 versus eigenvector 2, eigenvector 2 versus eigenvector 3, eigenvector 9 versus eigenvector 10, and eigenvector 20 versus eigenvector 25 show the protein motion in conformational space (Fig. 12a, b, c and d). No fluctuations or motions were observed along the 25th eigenvector subspace for the

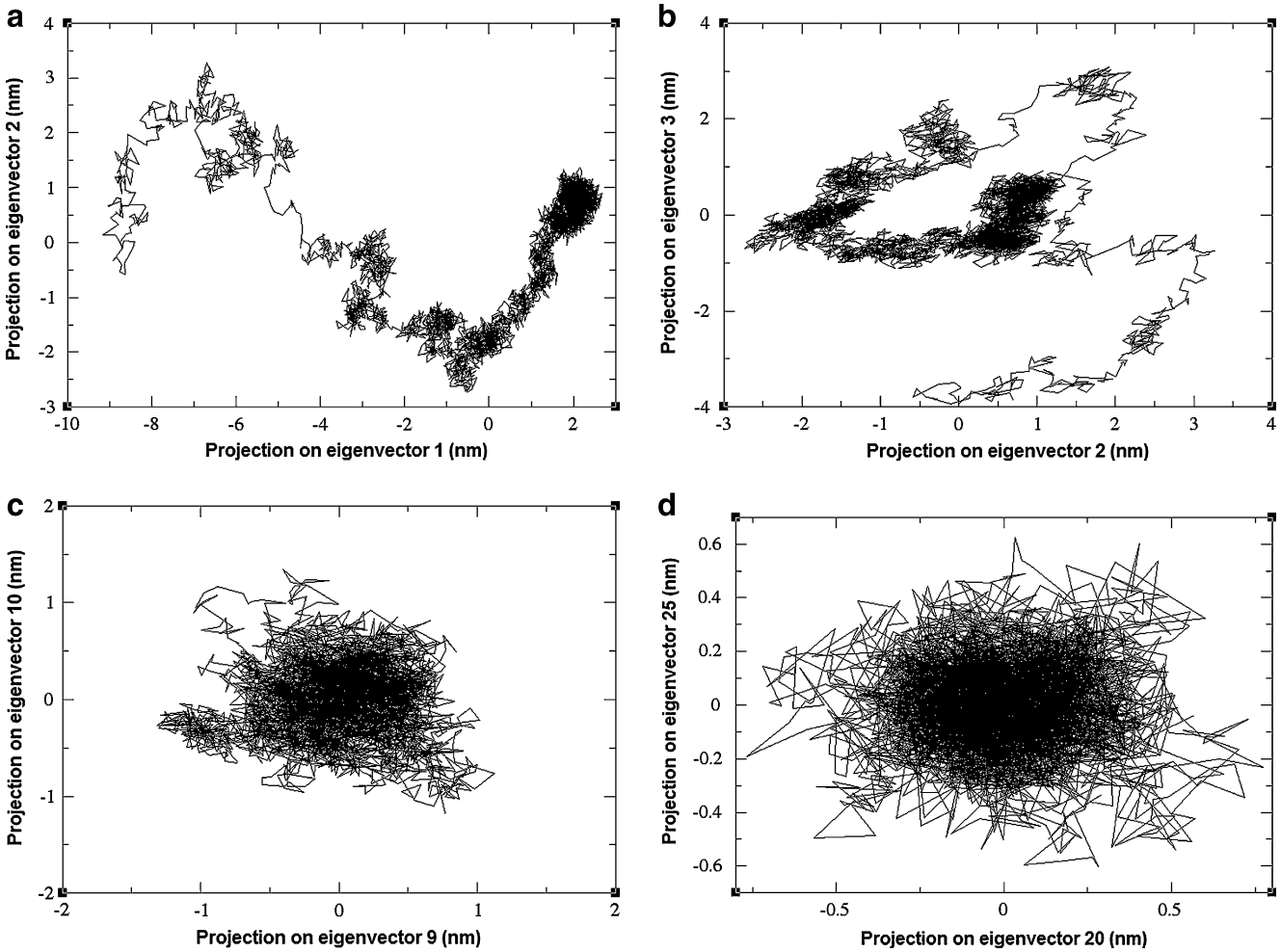


Fig. 12a–d Two-dimensional projections of the Bcl-2 trajectory on eigenvectors **a** 1 and 2, **b** 2 and 3, **c** 9 and 10, and **d** 20 and 25

Bcl-2 protein. The projections on eigenvectors 1 and 2 (Fig. 12a) show random walk and large collective motions in this essential subspace, which are observed to be relevant to protein function. The shapes of the projections were observed to be mutually independent (oval distribution) for projections on eigenvectors 9 and 10 (Fig. 12c) and 20 and 25 (Fig. 12d). These physically constrained subspaces are much less important for protein function, and are referred to as Gaussian fluctuations (irrelevant local fluctuations).

Table 6 Minimum and maximum energy scores of the projections along the ten eigenvectors, as calculated from the covariance matrix

Eigenvectors	Minimum		Maximum	
	Scores	Time (ps)	Scores	Time (ps)
1	−26.898708	99.0	8.427352	9450.0
2	−8.904247	4695.0	11.013029	558.0
3	−11.996984	51.0	7.409557	1824.0
4	−6.242445	2666.0	5.714304	1068.0
5	−10.428877	6.0	8.632933	1101.0
6	−4.668922	531.0	5.037412	78.0
7	−7.668024	45.0	5.500106	528.0
8	−4.987044	11406.0	3.829373	9309.0
9	−5.176743	465.0	5.057445	1131.0
10	−3.973459	8358.0	3.590667	582.0

The motions in the essential subspace are anharmonic diffusional motions of the protein, which randomly moves from one local minimum on the potential energy surface to the other. These motions were identified by the scores attained from extreme structure projections for eigenvectors 1–10 along the 15 ns MD trajectory (Table 6). The dynamic behavior of the protein is predicted by essential dynamics projections along the trajectory [112]. The projection of a trajectory onto the eigenvectors gives an indication of the sampling of conformational space. These spaces were obtained by the average local minimum and maximum scores describing the largest motions along the first eigenvector corresponding to extreme structure with respect to time. These scores, referred to as projections, show the motion along the axis, the total extent of the motion, and they predict the stability of the protein. The first eigenvec-

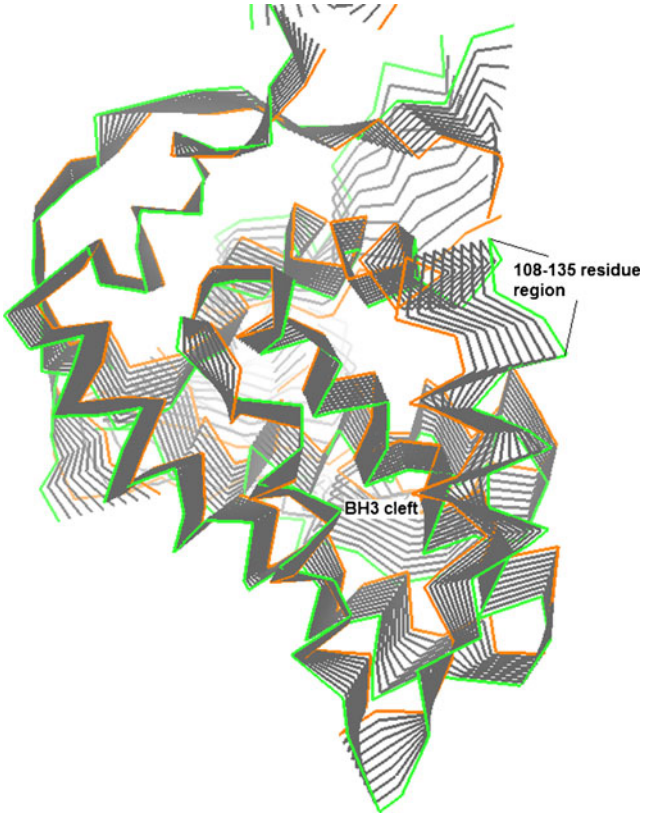


Fig. 13 BH3 cleft motion shown by ten extreme frames along the first eigenvector. All extreme structures were depicted using Pymol. The motions of atoms were observed to cause structural changes from the first frame (*green*) to the last frame (*orange*). *Arrows* indicate the direction of motion in the BH3 cleft; the region containing residues 108–135 shows the largest atomic motions

tor describes the largest motions, which take the longest time to converge during sampling. Our aim was to analyze the essential subspace motions and determine the residues that contribute to such motions, especially in the BH3 cleft. Therefore, the ten extreme structures extracted along the first eigenvector allow us to visualize the motion along the axis and the total extent of the motion, which are largely provided by the BH3 cleft (residues 108–135) (Fig. 13). These results provide a plausible reason for the difference in the BH3 clefts of the 1GJH and 2XA0 models.

Summary

In contrast to other members of the Bcl-2 family, the Bcl-2 and Bcl-X_L proteins include a large region known as the FLD. Our sequence and structural analysis results indicate that these proteins do not have the same FLD regions. The whole structure of the Bcl-2 protein has not yet been determined experimentally; only the isoform NMR structure (1GJH) has been solved (excluding the FLD). The structural features of the binding sites in the FLD are unknown, and no information is available on the structural changes to the Bcl-2 protein induced by the binding of other proteins to the FLD. Mutational and protein interaction studies have identified the importance of the FLD in the Bcl-2 protein, but the structural changes induced by FLD are still unknown. In view of the biological importance of the FLD, we focused our attention on creating a model of the complete Bcl-2 protein that would likely help us to get a better understanding of the regulation of Bcl-2 anti-apoptotic activity, which is mediated through FLD.

The structure of the complete Bcl-2 protein was elucidated by homology modeling. MODELLER 9v7 produced a better model (MOD) using the manually aligned 1GJH template than Swiss-Model. Obtaining an accurate Bcl-2 homology model is expected to reduce the errors that occur when performing protein–protein interaction (docking) and mutagenesis studies. Further, the energy-minimized 3D model of MODLOPT was observed to be an improved-quality model, and to have a better overall quality factor than the other models (CW and MOD). It could therefore be used for further computational studies. The presence of many coils in the FLD structure is accounted for by the presence of 19 Pro residues, which act as alpha-helix and beta-sheet breakers [113]. We also observed that the presence of a high number of Pro residues in the FLD disrupts the continuity of its structure, leading to an unstructured conformation. Thus, in order to predict the FLD folding, MD simulation of the energy-minimized model of MODLOPT was performed, which revealed that the Bcl-2 loop is very flexible, which may play a role in the structural diversity of FLD models. However, we observed that FLD stability was largely due by the presence of the large number of Pro residues in the FLD. These take part in hydrogen bonding between hydrophilic and hydrophobic residues, and were observed to form more stable bonds than those between hydrophilic and neutral group residues, hydrophobic and neutral group residues, or vice versa (Table 4). The residues Arg63 and Arg68 contributed the highest hydrophilicities to the SASA, and are also likely to play an important role in providing stability to the FLD by forming salt bridges. The average structure of Bcl-2 was extracted from the trajectories of MD simulations, and further validated by PROCHECK, WHAT CHECK, ERRAT and ProSA-web software. Our study indicates that the average structure can be considered a near-native conformation, as it shows good overall quality in comparison with other models. The MD-simulated average model was validated by several programs and was found to have the best 3D conformation, with improved folding of the unstructured flexible loop—a prerequisite for understanding the mechanism of mutation and the interaction of proteins with the FLD. The average structure resulting from the trajectory could vary greatly, so conformation clusters

with the same RMSD were identified by g_cluster. The ensemble of structures of Bcl-2 can also be used to study the structural basis for its activation and regulation upon binding with other proteins at the FLD, considering its improved quality as compared to structures already available in the database. The electrostatic behavior of the whole MD-simulated trajectory was found to agree with the X-ray structure (2XA0), indicating the authenticity of the simulations. The quality of the model produced should aid attempts to make reliable predictions of the bioactive FLD's antagonists/agonists via in silico analysis.

Further, PCA analysis was used as a tool to obtain information about essential subspaces. PCA is very useful for predicting the directions of motion of MD-simulated structures in the essential subspace. This helped us to predict the unfolding mechanism behind the conformational and dynamic properties of the BH3 cleft, for which great conformational variability was observed. PCA also helped us to identify motions that are crucial to protein function (Fig. 13) in the essential subspace shown by the first eigenvector (Table 6). This study also indicates that the identification of representative subspaces from the PCA is useful for elucidating the structure–function relationship for the FLD.

Acknowledgments We are thankful to Director R. P. Tripathi, INMAS, Delhi, India for his support. This work is funded by the Defence Research and Development Organization (DRDO) (Govt. of India) and the Department of Science and Technology, India. We are also thankful to Meenakshi Anurag of IGIB, CSIR, Delhi, India for help with using DSSP in GROMACS, and Tsjerk A. Wassenaar of the University of Groningen, the Netherlands, for providing the second chapter on PCA from his Ph.D. thesis.

References

1. Gurudutta GU, Verma YK, Singh VK, Gupta P, Raj HG, Sharma RK, Chandra R (2005) Structural conservation of residues in BH1 and BH2 domains of Bcl-2 family proteins. FEBS Lett 579:3503–3507
2. Antonsson B, Martinou JC (2000) The Bcl-2 protein family. Exp Cell Res 256:50–57
3. Dimmler S, Breitschopf K, Haendeler J, Zeiher AM (1999) Dephosphorylation targets Bcl-2 for ubiquitin-dependent degradation: a link between the apoptosome and the proteosome pathway. J Exp Med 189:1815–1822
4. Ruvolo PP, Deng X, May WS Jr (2001) Phosphorylation of Bcl-2 and regulation of apoptosis. Leukemia 15:515–522
5. Meier P, Finch A, Evan G (2000) Apoptosis in development. Nature 407:796–801
6. Xiangming D, Fengquin G, Tammy F, Jessica A, May WS (2006) Bcl-2's flexible loop domain regulates p53 binding and survival. Mol Cell Biol 26:4421–4434
7. Nicholson DW, Ali A, Thornberry NA, Vaillancourt JP, Ding CK, Gallant M, Gareau Y, Griffin PR, Labelle M, Lazebnik YA, Munday NA, Raju SM, Smulson ME, Yamin TT, Yu VL, Miller DK (1995) Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis. Nature 376:37–43
8. Chang BS, Minn AJ, Muchmore SW, Fesik SW, Thompson CB (1997) Identification of a novel regulatory domain in Bcl-X_L and Bcl-2. EMBO J 16:968–977
9. Kang CB, Tai J, Chia J, Yoon HS (2005) The flexible loop of Bcl-2 is required for molecular interaction with immunosuppressant FK-506 binding protein 38 (FKBP38). FEBS Lett 579:1469–1476
10. Ciechanover A (1994) The ubiquitin-proteosome proteolytic pathway. Cell 79:13–21
11. Hartl FU, Hayer-Hartl M (2002) Molecular cheparones in the cytosol: from nascent chain to folded protein. Science 295:1852–1858
12. Verma YK, Gangenahalli GU, Singh VK, Gupta P, Chandra R, Sharma RK, Raj HG (2006) Cell death regulation by B-cell lymphoma protein. Apoptosis 11:459–471
13. Deng X, Kornblau SM, Ruvolo PP, May WS Jr (2001) Regulation of Bcl2 phosphorylation and potential significance for leukemic cell chemoresistance. J Natl Cancer Inst Monogr 28:30–37
14. Petros AM, Medek A, Nettesheim DG, Kim DH, Yoon HS, Swift K, Matayoshi ED, Oltersdorf T, Fesik SW (2001) Solution structure of the antiapoptotic protein bcl-2. Proc Natl Acad Sci USA 98:3012–3017
15. Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL repository and associated resources. Nucleic Acids Res 37:D387–D392
16. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2006) Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics 15:5.6.1–5.6.30. doi:10.1002/0471250953.bi0506s15
17. Wang J, Cao Z, Shuqiang L (2009) Molecular dynamics simulations of intrinsically disordered proteins in human diseases. Curr Comput Aided Drug Des 5:280–287
18. Lindahl E, Hess B, van der Spoel D (2001) Gromacs 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7:306–317
19. Castrignano T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A (2006) The PMDB Protein Model Database. Nucleic Acids Res 34:D306–D309
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
21. Ward JJ, Sodi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in protein. J Mol Biol 337:635–645
22. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol 347:827–839
23. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
24. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89:10915–10919
25. Van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M (2006) Biomolecular modeling: goals, problems, perspectives. Angew Chem Int Edn Engl 45:4064–4092
26. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723
27. DeLano WL (2002) The PyMOL molecular graphics system. http://www.pymol.org

28. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410
29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291. doi:10.1107/S0021889892009944
30. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519
31. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272–272
32. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theor Comput 4:435–447
33. Oostenbrink C, Soares TA, van der Veget NFA, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. Eur Biophys J 34:273–284
34. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656–1676
35. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1969) Interaction models for water in relation to protein hydration. Nature 224:175–177
36. Feenstra KA, Hofstetter K, Bosch R, Schmid A, Commandeur JN, Vermeulen NP (2006) Enatioselective substrate binding in a monooxygenase protein model by molecular dynamics and docking. Biophys J 91:3206–3216
37. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472
38. Berendsen HJC, Postma JPM, Van der Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
39. Parrinello M, Rahman A (1980) Crystal structure and pair potentials: a molecular dynamics study. Phys Rev Lett 45:1196–1199
40. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593
41. Marlino A, Mazzarella L, Carannante A, Difiore A, Di Donata A, Notomista E, Sica F (2005) The importance of dynamic effects on the enzyme activity. J Biol Chem 280:17953–17960
42. Taly JF, Marine A, Gibrat JF (2008) Can molecular dynamics simulations help in discriminating correct from errorneous protein 3D models? BMC Bioinforma 9:6
43. Malek K, Odijk T, Coppens MC (2005) Diffusion of water and sodium counter-ions in nanopores of a β-lactoglobulin crystal: a molecular dynamics study. Nanotechnology 16:S522–S530
44. Brooks CL, Karplus M, Pettitt BM (1988) Proteins. A theoretical perspective of dynamics, structure and thermodynamics (Advances in Chemical Physics). Wiley, New York
45. Seshasayee AS, Raghunathan K, Sivaraman K, Pennathur G (2006) Role of hydrophobic interactions and salt-bridges in β-hairpin folding. J Mol Model 12:197–204. doi:10.1007/S00894-005-0018-6
46. Zheng H, Wang S, Zhang Y (2009) Increasing the time step with mass scaling in Born–Oppenheimer ab initio QM/MM molecular dynamics simulations. J Comput Chem 30:2706–2711
47. Feenstra KA, Hess B, Berendsen HJC (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem 20:786–798
48. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual Molecular Dynamics. J Mol Graph 14:33–38
49. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637
50. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res 32:W500–W502
51. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791
52. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem 19:1639–1662
53. Amadei A, Linssen AB, Berendsen HJC (1993) Essential dynamics of proteins. Proteins 17:412–425
54. Amadei A, Ceruso MA, Di Nola A (1999) On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. Proteins 36:419–424
55. Kitao A, Hirata F, Go N (1991) The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. J Chem Phys 158:447–472
56. Garcia AE (1992) Large-amplitude nonlinear motions in proteins. Phys Rev Lett 68:2696–2699
57. Chen CC, Hwang JK, Yang JM (2006) $(PS)^2$: protein structure prediction server. Nucleic Acids Res 34:W152–W157
58. Ding J, So BA, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction detection and structure prediction. Nucleic Acids Res 33:W244–W248
59. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21:3433–3434
60. Madhusudhan MS, Marti-Renom MA, Eswar N, John B, Pieper U, Karchin R, Yi Shen M, Sali A (2005) Comparative protein structure modeling. In: Walker JM (ed) The proteomics protocols handbook. Humana, Totowa, pp 831–860
61. Zhang H (2002) Protein tertiary structures: prediction from amino acid sequences. In: Encyclopedia of life sciences. Nature, London, pp 1–7. doi:10.1038/npg.els.0003040
62. Novotny J, Rashin AA, Bruccoleri RE (1988) Criteria that discriminate between native proteins and incorrectly folded models. Proteins 4:19–30
63. Vorobjev YN, Almagro JC, Hermans J (1998) Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. Proteins 32:339–413
64. Janardhan A, Vajda S (1998) Selecting near-native conformations in homology modeling: the role of molecular mechanics and solvation terms. Protein Sci 7:1772–1780
65. Lazaridis T, Karplus M (1999) Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J Mol Biol 288:447–487
66. Gatchell DW, Dennis S, Vajda S (2000) Discrimination of near-native protein structure from misfolded models by empirical free energy functions. Proteins 41:518–534
67. Kinjo AR, Kidera A, Nakamura H, Nishikawa K (2001) Physicochemical evaluation of protein folds predicted by threading. Eur Biophys J 30:1–10
68. Dominy BN, Brooks CL III (2002) Identifying native-like protein structure using physics-based potentials. J Comput Chem 23:147–160

69. Ak F, Gallicchio E, Wallquist A, Levy RM (2002) Distinguishing native conformations of proteins from decoys with an effective free energy estimation based on the OPLS all-atoms force field and the surface generalized born solvent model. Protein 48:404–422
70. Azizian H, Rahrami H, Pasalar P, Amanlou M (2010) Molecular modeling of *Helicobactor pylori* arginase and the inhibitor coordination interactions. J Mol Graph Model 28:626–635
71. Pinto M, Perez JJ, Martinez JR (2004) Molecular dynamics study of peptide segment of the BH3 domain of the proapoptotic proteins Bak, Bax, Bid and Hrk bound to the Bcl-X_L and Bcl-2 proteins. J Comput Aided Mol Des 18:13–22
72. Ku B, Liang C, Jung JU, Oh BH (2010) Evidence that inhibition of Bax activation by Bcl-2 involves its tight and preferential interaction with the BH3 domain of Bax. Cell Res 21:627–641. doi:10.1038/cr.2010.149
73. Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. Proc Natl Acad Sci USA 98:14274–14279
74. Kaplan J, DeGrado WF (2004) De novo design of catalytic proteins. Proc Natl Acad Sci USA 101:11566–11570
75. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De novo computational design of retero-aldol enzymes. Science 319:1387–1391
76. Rothlisberger D (2008) Kemp elimination catalysts by computational enzyme designing. Nature 453:190–195
77. Lazar GA, Dang W, Karki S, Vafa O, Peng JS, Hyun L, Chan C, Chung HS, Eivazi A, Yoder SC, Vielmetter J, Carmichael DF, Hayes RJ, Dahiyat BI (2006) Engineered antibody Fc variants with enhanced effector function. Proc Natl Acad Sci USA 103:4005–4010
78. Ogata K, Jaramillo A, Cohen W, Briand JP, Connan F, Choppin J, Muller S, Wodak SJ (2003) Automatic sequence design of major histocompatibility complex class I binding peptides imparing $CD8^+$ T-cell recognition. J Biol Chem 278:1281–1290
79. Joachimiak L, Koretemme T, Stoddard B, Baker D (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein–protein interface. J Mol Biol 361:195–208
80. Shifman JM, Choi MH, Mihalas S, Mayo SL, Kennedy MB (2006) Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) is activated by calmodulin with two bound calciums. Proc Natl Acad Sci USA 103:13968–13973
81. Cochran FV, Wu SP, Wang W, Nanda V, Saven JG, Therien MJ, DeGrado WF (2005) computational de novo design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor. J Am Chem Soc 127: 1346–1347
82. Daggett V (2006) Protein folding-simulation. Chem Rev 106:1898–1916
83. Desjarlaris JR, Handel TM (1995) De novo design of the hydrophobic cores of proteins. Protein Sci 4:2006–2018
84. Desjarlais JR, Handel TM (1999) Side-chain and backbone flexibility in protein core design. J Mol Biol 290:305–318
85. Krasmer-Pecore CM, LeComte JT, Desjarbis JR (2003) A de novo redesign of the WW domain. Protein Sci 12:2194–2205
86. Vieille C, Zeikus JG (1996) Thermoenzymes: identifying molecular determinants of protein structural and functional stability. Trends Biotechnol 14:183–191
87. Russell RJ, Hough DW, Danson MJ, Taylor GL (1994) The crystal structure of citrate synthase from the thermophilic archaeon, *Thermoplasma acidophilum*. Structure 2:1157–1167
88. Nagi AD, Regan L (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein. Fold Des 2:67–75
89. Carlson HA (2002) Protein flexibility is an important component of structure based drug discovery. Curr Pharm Des 8:1571–1578
90. Beer HD, Wohlfahrt G, McCarthy JE, Schomburg D, Schmid RD (1996) Analysis of the catalytic mechanism of a fungal lipase using computer-aided design and structural mutants. Protein Eng 9:507–517
91. Holm L, Sander C (1996) Mapping the protein universe. Science 273:595–603
92. Vendruscolo M, Kussell E, Domany E (1997) Recovery of protein structure from contact maps. Fold Des 2:295–306
93. Bulaj G, Goldenberg DP (2001) Mutational analysis of hydrogen bonding residues in the BPTI folding pathway. J Mol Biol 313:639–656
94. Chen YW, Fersht AR, Henrick K (1993) Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants. J Mol Biol 234:1158–1170
95. Byrne MP, Manuel RL, Lowe LG, Stites WE (1995) Energetic contribution of side chain hydrogen bonding to the stability of staphylococcal nuclease. Biochemistry 34:13949–13960
96. Pace CN, Shirley BA, McNutt M, Gajiwala K (1996) Forces contributing to the conformational stability of proteins. FASEB J 10:75–83
97. Yamagata Y, Kubota M, Sumikawa Y, Funahashi J, Takano K, Fujii S, Yutani K (1998) Contribution of hydrogen bonds to the conformational stability of human lysozyme: calorimetry and X-ray analysis of six tyrosine → phenylalanine mutants. Biochemistry 37:9355–9362
98. Takano K, Yamagata Y, Kubota M, Funahashi J, Fujii S, Yutani K (1999) Contribution of hydrogen bonds to the conformational stability of human lysozyme: calorimetry and X-ray analysis of six serine → alanine mutants. Biochemistry 38:6623–6629
99. Pace CN (2001) Polar group burial contributes more to protein stability than nonpolar group burial. Biochemistry 40:310–313
100. Grantcharova VP, Riddle DS, Santiago JV, Baker D (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. Nat Struct Biol 5:714–720
101. Krantz BA, Moran LB, Kentsis A, Sosnick TR (2000) D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. Nat Struct Biol 7:62–71
102. Antoine R, Compagnon I, Rayane D, Broyer M, Dugowd Ph, Breaux G, Hogemeister FC, Pippen D, Hudgins RR, Jarrold MF (2002) Electric dipole moments and conformations of isolated peptides. Eur Phys J D 20:583–587
103. Simonson T (1999) Dielectric relaxation in proteins: macroscopic and microscopic models. Int J Quantum Chem 73:45–57
104. Soufian S, Naderi-Manesh H, Alizadeh A, Sarbolouki MN (2009) Molecular dynamics and circular dichoism studies on Aurein 1.2 and retro analog. World Acad Sci Eng Technol 56:858–864
105. Van Aalten DMF, Amadei A, Linssen ABM, Eijsink VGH, Vriend G, Berendsen HJC (1995) The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulation in vacuum and water. Proteins 22:45–54
106. Spoel D van der, Lindahl E, Hess B, Kutzner C, Buuren AR van, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, van Drunen R, Berendsen HJC (2005) Gromacs user manual, version 4.0. http://www.gromacs.org
107. Scheek RM, Van Nuland NAJ, DeGroot BL, Linssen ABM, Amadei A (1995) Structure from NMR and molecular dynamics: distance restraining inhibits motion in the essential subspace. J Biomol NMR 6:106–111
108. Van Aalton DMF, Findley JBC, Amadei A, Berendsen HJC (1995) Essential dynamics of the cellular retinol-binding

protein—evidence for ligand-induced conformational changes. Protein Eng 8:1129–1135
109. Chillemi G, Falconi M, Amadei A, Zimatore G, Desideri A, DiNola A (1997) The essential dynamics of Cu, Zn superoxide dismutase: suggestion of intersubunit communication. Biophys J 73:1007–1018
110. Hayward S, Kitao A, Hirata F, Go N (1993) Effect of solvent on collective motions in globular proteins. J Mol Biol 234:1207–1217
111. Romo TD, Clarage JB, Sorensen DC, Phillips GN Jr (1995) Singular value decomposition analysis of time-average crystallographic refinement. Proteins 22:311–321
112. Hunenberger PH, Mark AE, van Gunsteran WF (1995) Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. J Mol Biol 252:492–503
113. Berg JM, Tymoczko JL, Stryer L (2002) Protein composition and structure. In: Biochemistry, 6th edn. WH Freeman and Co., New York, pp 40-53