J Comput Aided Mol Des (2009) 23:49–61
DOI 10.1007/s10822-008-9237-0

Role of protein flexibility in the design of Bcl-X_L targeting agents: insight from molecular dynamics

William Novak · Hongming Wang · Goran Krilov

Received: 16 June 2008 / Accepted: 12 August 2008 / Published online: 9 September 2008

Abstract Detailed understanding of protein–ligand interactions is crucial to the design of more effective drugs. This is particularly true when targets are protein interfaces which have flexible, shallow binding sites that exhibit substantial structural rearrangement upon ligand binding. In this study, we use molecular dynamics simulations and free energy calculations to explore the role of ligand-induced conformational changes in modulating the activity of three generations of Bcl-X_L inhibitors. We show that the improvement in the binding affinity of each successive ligand design is directly related to a unique and measurable reduction in local flexibility of specific regions of the binding groove, accompanied by the corresponding changes in the secondary structure of the protein. Dynamic analysis of ligand–protein interactions reveals that the latter evolve with each new design consistent with the observed increase in protein stability, and correlate well with the measured binding affinities. Moreover, our free energy calculations predict binding affinities which are in qualitative agreement with experiment, and indicate that hydrogen bonding to Asn100 could play a prominent role in stabilizing the bound conformations of latter generation ligands, which has not been recognized previously. Overall our results suggest that molecular dynamics simulations provide important information on the dynamics of ligand–protein interactions that can be useful in guiding the design of small-molecule inhibitors of protein interfaces.

Keywords Molecular dynamics · Protein flexibility · Protein–protein interfaces · Bcl-X_L · Ligand–protein interactions · ABT-737

Introduction

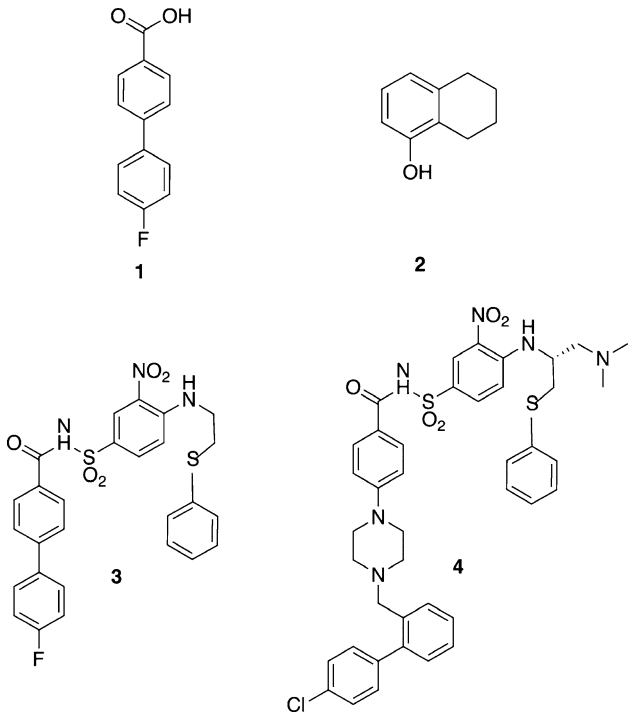
A particular challenge in the design of small molecule inhibitors is understanding how the flexibility of the target protein contributes to ligand binding. This is especially true for small molecule inhibitors targeting the protein interfaces involved in signaling pathways. While such inhibitors hold great promise for therapeutic intervention in cancer, inflammation and other diverse cellular responses [1], protein interfaces are usually large and pliable, and often contain induced or transient binding sites [2]. These properties make rational design of inhibitors challenging [3, 4], since the flexible behavior of the protein is often difficult to predict from static structures typically available from experiments. In this study, we have used molecular dynamics (MD) simulations to investigate the relationship of the receptor flexibility to the binding profile of small molecule inhibitors of protein interfaces. We focus on the Bcl-X_L protein [5], for which a landmark protein-interface inhibitor, ABT-737 [6], is currently being tested as an anti-cancer agent to provide insight into how this dynamic information could be used to enhance the efficiency of the drug design process.

Inhibitors of the B-cell lymphoma 2 (Bcl-2) family of proteins, including Bcl-X_L, are of great interest in cancer treatment. This family is composed of both pro and

Electronic supplementary material The online version of this article (doi:10.1007/s10822-008-9237-0) contains supplementary material, which is available to authorized users.

W. Novak · G. Krilov (✉)
Department of Chemistry, Boston College, 2609 Beacon Street, Chestnut Hill, MA 02135, USA
e-mail: krilov@bc.edu

H. Wang
Computational Science, AstraZeneca R&D Boston, 35 Gate House Dr., Waltham, MA 02451, USA



Scheme 1 Chemical structures of inhibitors 4FC (compound **1**), TN1 (compound **2**), N3B (compound **3**), and ABT-737 (compound **4**)

anti-apoptosis proteins that regulate programmed cell death [7–9]. In some forms of cancer, the relative ratio of protein populations is disrupted, and an over-expression of the pro-survival proteins (Bcl-X_L and Bcl-2) can occur, resulting in increased cellular proliferation, cancerous growth of tumors, and resistance to anti-cancer therapy [10–13]. ABT-737 (compound **4** in this study, Scheme 1), is a pro-survival inhibitor with a remarkable binding affinity for the anti apoptosis Bcl-2 type family of proteins. It has been shown to be effective in increasing the efficiency of chemotherapy treatments and can even suppress tumor growth in some forms of cancer [6, 14]. While the activity of this inhibitor has been well-established, the influence of ligand-induced conformational changes on the binding activity is not well understood. Moreover it has been noted that the binding pocket of Bcl-X_L is quite flexible, suggesting that the dynamic analysis of the protein binding site may be helpful in the design of such inhibitors [15].

The Bcl-X_L protein interface involved in cell signaling consists of several α-helices, α2–α5, shown in Fig. 1, which form a long, flexible hydrophobic binding groove that opens upon substrate binding through secondary structure changes of α3, and motions of helices α2 and α4 [5, 16]. The BH3 domain from a pro-apoptotic partner, such as Bak or Bim binds to this surface during signal transduction [16–18], and alanine scans have revealed [16] that interactions with several hydrophobic side-chains of the BH3 domain are critical for binding. Two of the hydrophobic sites

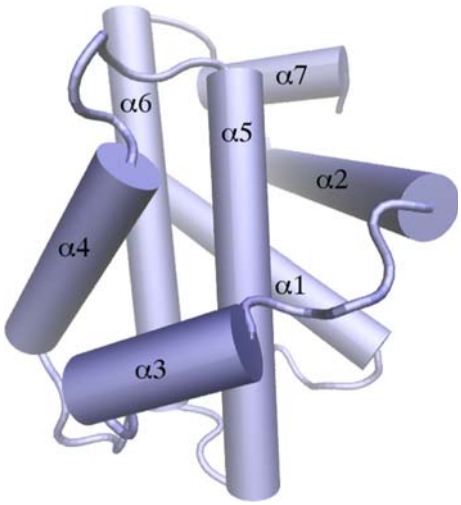


Fig. 1 Schematic depiction of the three-dimensional structure Bcl-X_L with α-helices labeled. The hydrophobic peptide binding groove is composed primarily of helices α3–α5

normally occupied by these side chains are the targets for compounds **1** and **2**, the lead small molecules discovered through 'SAR by NMR' screening [19] found to inhibit BH3 binding [6]. Joining the fragments in an effort to maintain their original interactions with the protein within the two sites led to the design of the compound **3** which displayed nanomolar binding affinity to Bcl-X_L (Scheme 1) [20]. Further structural analysis revealed that the binding of this ligand induces several changes in the binding site topology, resulting in a substantially different binding mode [20, 21]. Surprisingly, inhibitor **4**, which was developed through further optimization of compound **3** aimed at improving the activity against Bcl-2, and selectivity versus human serum albumin (HSA), has shown superior sub-nanomolar activity versus Bcl-X_L [15]. The latter was attributed to favorable interaction of the piperazine-biphenyl extension with a newly induced binding pocket that was not present in the complex with ligand **3** [6]. The observed binding site flexibility and the wealth of available structural data makes Bcl-X_L an attractive system for exploring protein structural flexibility.

In contrast to the large body of experimental work, there have been comparatively few computational studies of the Bcl-2 family and their substrates and inhibitors. Zheng et al. [22] have used the multiple copy simultaneous search method combined with flexible docking to construct an optimized 3D pharmacophore model for Bcl-2 inhibitors, and found that inhibitor potency correlates well with the optimized distance from pharmacophoric points. Based on detailed analysis of interactions obtained from short MD simulations of four pro-apoptotic proteins with Bcl-2 and Bcl-X_L, Pinto et al. [23] have derived a three-point pharmacophore model deemed necessary for binding to these

targets. Mancinelli et al. [24] used a combination of MD and automatic docking to study the binding of the pro-apoptotic Bax with the antiapoptotic Bip peptide, and found that the association with the latter strongly reduces the flexibility of the binding region of Bax, and also stabilizes more distant regions of the protein. More recently, Fu et al. [25] have used normal mode analysis to incorporate protein backbone flexibility into the design of new BH3 peptide sequences that bind to Bcl-X_L, and have shown that a fixed backbone assumption severely restricted the number of BH3 sequences identified as potential Bcl-X_L binders. Eyrisch and Helms [26] have employed MD simulations to study the formation of transient pockets on the surfaces of several proteins participating in protein–protein interactions, including Bcl-X_L, and found that numerous pockets open during the course of the simulation, providing further evidence of the importance of protein flexibility in determining the binding site characteristics.

In this paper, we present the results of MD simulations and free energy calculations for three generations of increasingly potent Bcl-X_L inhibitors leading to ABT-737. We first investigate the stability and flexibility of the Bcl-X_L interface in complexes with different ligands, and how they are modified by ligand binding. We next analyze the dynamic interaction profiles of different inhibitors, and show how the reduction of protein fluctuations leads to more favorable binding interactions. We conclude by discussing how this information may be helpful in guiding the design of small molecule inhibitors of protein–protein interfaces.

Computational methods

Initial structures

Initial configurations for the free Bcl-X_L protein, the Bak peptide complex, and complexes with ligands **1** and **2**, **3**, and **4** were obtained from the respective experimental structures, with PDB codes 1R2D, 1BXL, 1YSG, 1YSI and 2YXJ. The structures of the Bak complex (1BXL) [16], complex **1** and **2** (1YSG) and complex **3** (1YSI) were determined by solution NMR [6], while those of the free protein (1R2D) [27] and the complex **4** (2YXJ) were determined by X-ray crystallography. Hydrogen atoms were added to the proteins as necessary, and the protonation states were assigned according to the pK_a based on pH = 7.0 using the GROMACS package [28–30]. The hydrogen atoms for ligand **4** were assigned manually. Bcl-X_L peptide chain contains a large loop (residues 28–81 in 1YSI, connecting the helices α1 and α2) that was modified in experimental studies to improve the resolution of the NMR structures. In the X-ray structures, the sequence of this loop was not resolved. Hence, in order to match the sequence in 1YSI, an 18 residues loop was built by inserting the corresponding sequence from the 1YSI structure into the appropriate location of the free protein structure (1R2D). The contacts were first optimized through energy minimization, while keeping the protein structure fixed, which was followed by unrestrained minimization of the full protein. The same procedure was used to insert the 17 residue loop into structure of complex **4** (2YXJ). Since the loop is located far from the binding pocket and did not exhibit significant interactions with the bulk of the protein, these modifications are not expected to have significant impact on ligand binding. In addition, four unresolved residues were added to the N terminus and the C terminus of the free protein (1R2D), and seven residues were added the N terminus and three to the C terminus of complex **4** (2YXJ). Finally, 17 residues were removed from the C terminus of complex **1** and **2** (1YSG) and complex **3** (1YSI), including the 8-residue histidine tag which was not part of the native protein, thus providing a consistent 164 residue sequence for Bcl-X_L in all five system. Each complex was placed in dodecahedral cell, with size adjusted to maintain a minimum distance of 10 Å to the cell boundary, and soaked with a pre-equilibrated box of water. All overlapping solvent molecules were removed, and an appropriate number of counterions were added to maintain charge neutrality.

Molecular dynamics simulations

All MD simulations were performed using the GROMACS package. The OPLSAA 2001 force field [31, 32] was used to model all protein and ligand interactions, and the TIP3P model [33] was used for water. The particle-mesh Ewald method (PME) [34] was used to calculate long-range electrostatic interactions with a grid spacing of 0.08 nm combined with a fourth order B-spline interpolation. Van der Waals and short-range electrostatic interactions were smoothly truncated at 9 Å via a potential shift. The atomic partial charges for ligand **3** were assigned by an ESP fit to the quantum mechanical charge distribution for the free ligand in a Poisson–Boltzmann implicit solvent. The latter was obtained from a density functional theory (DFT) calculation with the B3LYP functional and the 6-31G** basis set, using the JAGUAR package (Schrödinger). Nose–Hoover thermostats [35] were utilized to maintain constant simulation temperature and the Berendsen barostat [36] was used to control the pressure. The equations of motion were integrated using the leapfrog algorithm with a 1 fs time step, and full periodic boundary conditions were applied throughout.

In each case, the solvated system was equilibrated via a 2000 step steepest descent energy minimization to remove bad contacts, followed by 200 ps of MD at constant NPT of

1 atm and 300 K, during which all heavy atoms of the solute were restrained. The restraints were then removed, and the solute–solvent system was heated to room temperature (300 K) in 50 K increments over 50 ps. The trajectory was then evolved at constant NVT for a further 10 ns. Configurations were saved in 4 ps intervals, and data from the final 5 ns of each simulation was used for analysis.

MM-GBSA calculations

The relative binding free energies of the inhibitors were determined through the MM-GBSA protocol, where the binding free energy is decomposed into

$$\Delta G_{\text{bind}} = \Delta G_{\text{gas}} + \Delta_{\text{sol}} G_{\text{complex}} - \Delta_{\text{sol}} G_{\text{ligand}} - \Delta_{\text{sol}} G_{\text{receptor}}. \quad (1)$$

Assuming negligible volume change upon ligand binding, the gas phase term is

$$\Delta G_{\text{gas}} = \Delta U_{\text{ele}} - \Delta U_{\text{vdw}} - \Delta U_{\text{conf}}, \quad (2)$$

which incorporates contributions from electrostatic (ΔU_{ele}) and non-polar (ΔU_{vdw}) ligand-receptor interactions, as well as the ligand strain energy (ΔU_{conf}), all of which were obtained from relevant OPLSAA 2001 forcefield terms. ΔU_{conf} was computed as the difference in internal energy of the ligand between the optimized solution conformation and the protein bound state. The solvation contributions

$$\Delta_{\text{sol}} G = \Delta G_{\text{GB}} + \Delta G_{\text{np}}, \quad (3)$$

consist of the electrostatic term (ΔG_{GB}), computed via the surface generalized Born method (SGB) [37], and the non-polar cavity term (ΔG_{np}), which is estimated from the change in the solvent-accessible surface area (SASA) by

$$\Delta G_{\text{np}} = \gamma\, \Delta(\text{SASA}), \quad (4)$$

where $\gamma = 0.005$ kcal/(mol Å^2) is the microscopic surface tension coefficient.

Potentials of mean force

The potentials of mean force $W(\xi)$ were computed via the constraint force method [38], according to:

$$W(\xi) = \int_0^{\xi} \langle f_c \rangle_{\xi'}\, d\xi'. \quad (5)$$

The reaction coordinate ξ specifying the displacement of the ligand from the binding pocket was defined as the distance between the carboxamide oxygen of Asn100 (hydrogen bond acceptor) and the ligand sulfonamide nitrogen (hydrogen bond donor). This distance was constrained over the range from 2.5 to 6.0 Å at equal intervals of 0.1 Å using the SHAKE algorithm [39]. A 1 ns constrained MD simulation was performed at each value of the reaction coordinate to compute the average constraint force $\langle f_c \rangle$. The latter were then integrated numerically to obtain the value of the PMF at each point. All error estimates were computed through block averaging of the simulation data.

Results and discussion

Stability of Bcl-X_L complexes

To analyze the stability of the MD simulations, we considered the root mean squared deviation (RMSD) from the experimental structure for each of the five systems. The latter was computed for the backbone atoms of residues 50–164, comprising helices α2–α7 and connecting loops, which includes all residues constituting the binding groove. Helix α1 and the long loop connecting α1 and α2 were not included due to large uncertainties associated with the generally poor experimental resolution of this region. The RMSD as a function of time for the free protein, the Bak complex, and complexes **1**, **2**, **3**, and **4** are shown in Fig. 2. For most systems, the RMSD profiles were found to plateau within 5 ns, suggesting that a stable equilibrium had been reached. The notable exception was complex **3** (Fig. 2, red line), which showed more pronounced RMSD fluctuations throughout the simulation. Analysis of the trajectory shows that this drift is primarily induced by the destabilization of the ligand in the binding pocket resulting from the unstacking of the nitrophenyl-thiophenyl π-stack system. A similar behavior was observed by Park et al. [21] among some members of a series of substituted *N*,*N*′-diarylureas investigated in efforts to identify substitution patterns promoting the folded bioactive motif. In order to improve stability, the atomic site charges for ligand **3** were

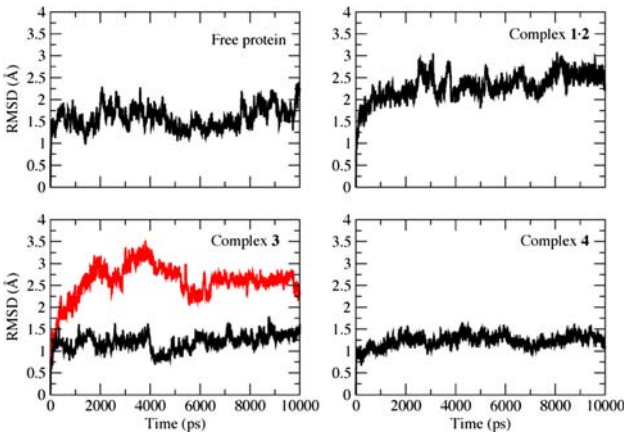


Fig. 2 The protein backbone RMSD from the corresponding experimental structure for the free protein (top left), complex **1**·**2** (top right), complex **3** (bottom left) and complex **4** (bottom right). The red line for complex **3** shows the RMSD prior to re-fitting of the partial charges

re-assigned by fitting to a quantum mechanical charge distribution. The simulation of complex **3** with this new parameterization of charges (Fig. 2, black line) showed a significantly improved stability, with the ligand maintaining the correct orientation in the binding pocket in agreement with the NMR structure. Some deviations from experimental structures were found following equilibration. In particular, the orientation of the Asn100 side chain was altered to facilitate hydrogen bonding with the ligand in all three inhibitor complexes. Similarly, local reorganization of the protein enabling inter-ligand hydrogen bonding was observed in complex **1**·**2**. The implications of these are addressed in the following sections.

Complex structure and flexibility

Initial studies [16, 23, 40] aimed at the design of small molecule Bcl-X_L inhibitors focused on isolating key interactions that facilitated the binding of pro-apoptotic proteins to Bcl-X_L. These studies identified two hydrophobic pockets that were critical to the binding of the BH3 domain of the pro-apoptotic member and the formation of the stable protein–protein interface. The first of these (site 1 in Fig. 3b) is formed by residues Tyr65, Leu72, Val90 and Phe110, and the second (site 2) is framed by Tyr159, Gly102, Phe61 and Ala57. Comparison of the X-ray structure of the free Bcl-X_L protein to the NMR structure of the Bak complex (Fig. 3a) shows that the protein undergoes a major conformational change, including shifts in the positions of helices α2 and α4, as well as partial unfolding of α3, to form the large binding groove containing the above two pockets, that will accommodate the helical Bak peptide substrate. Moreover, the RMS fluctuation profiles suggest that these conformational changes are accompanied by significant changes in the local flexibility of the protein chain. Hence, it behooves us to examine the implications of this extreme plasticity of the Bcl-X_L binding site for the design of small molecule inhibitors. In Fig. 4, we plot the average per-residue RMSD of the protein backbone of the Bak complex and those with the three small-molecule inhibitors obtained from the MD simulations, relative to the free protein crystal structure. Significant deviations can be observed in the active site, in particular residues 50–100 comprising helices α2–α4 which constitute the binding arm forming the walls of the peptide binding groove. Comparison with RMS fluctuation profiles shown in Fig. 5 indicates that variations in local flexibility of the complexes are most pronounced among the binding arm residues, hence it is of interest to investigate this region in more detail. The RMSF data for free protein show a distinct region of high flexibility spanning residues 65–85 at the start of the binding arm, comprising the residues which constitute binding pocket 1 in the Bak complex. The flexibility of this regions is observed to be significantly curtailed in the Bak complex, suggesting that the binding of the peptide leads to local stabilization of the flexible regions of the active site. On the other hand, several sites further down the binding arm are destabilized, including residues 85–90, 95–100 and 126–128. Similar shifts in flexible regions are observable in the RMSF profiles of the small-molecule inhibitor complexes. It is of interest to examine whether this is fortuitous, or if a more general relationship between local flexibility and the ligand binding mode can be established. In Fig. 6 we plot the difference in RMSF per residue for the three complexes relative to each other. An overall increase in RMSF of the binding arm residues of complex **1**·**2** is clearly observable, particularly in the region 56–65 proximal to the binding site of weaker inhibitor **2**, as well as the loop region 76–82 connecting helices α3 and α4, which indicates that the fragments 1 and 2 destabilize the binding pocket. In contrast, complex **3** data shows that the binding of the inhibitor **3** results in stabilization of the originally flexible region of the free protein (residues 65–75), with an overall reduction in fluctuations of ∼50%, suggesting a mode of action more similar to that of the native substrate. This increase in

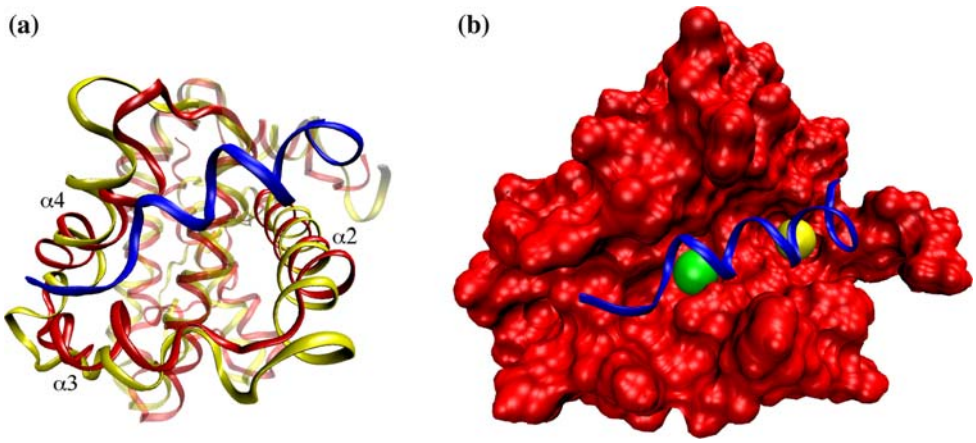


Fig. 3 (**a**) Overlap of the free Bcl-X_L backbone (red) and that of the Bak complex (yellow), with the Bak derived BH3 peptide shown in blue. (**b**) Solvent accessible surface area of the Bcl-X_L in the Bak peptide bound (shown as blue ribbon) conformation. Hydrophobic target sites 1 and 2 are indicated by green and yellow spheres, respectively. Protein helices α2–α4 are noted

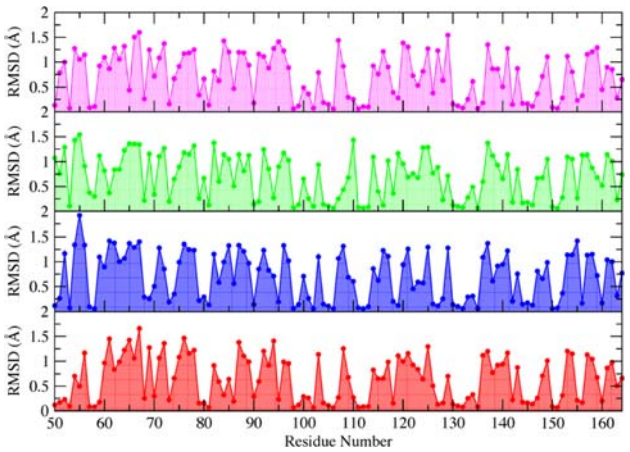


Fig. 4 Average per residue backbone RMSD of Bcl-X_L from the free protein structure for the Bak substrate bound complex (magenta), complex **1·2** (green), complex **3** (blue) and complex **4** (red). Significant differences are observed in the binding arm region, including residues 50–100 and 115–130

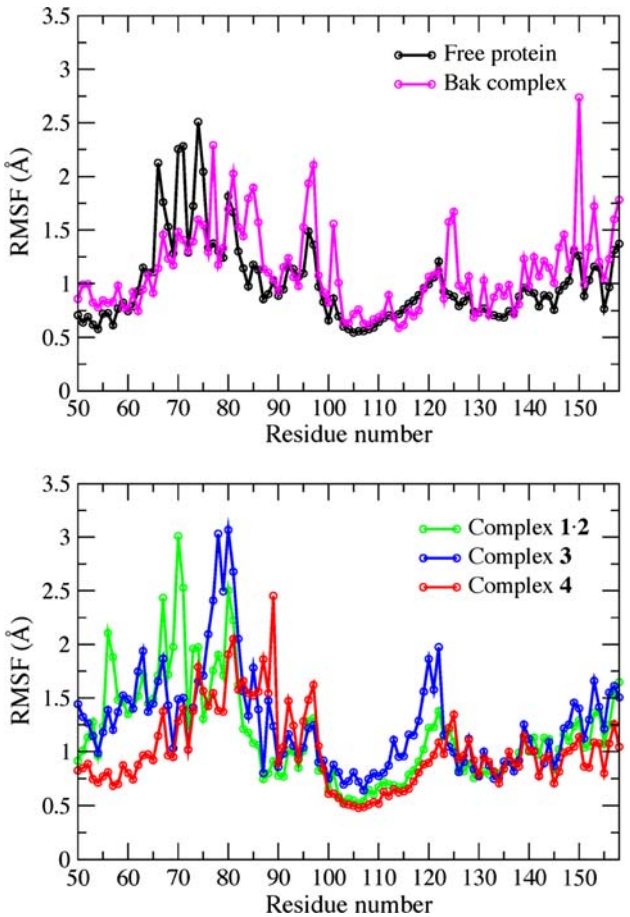


Fig. 5 The RMSF of the Bcl-X_L backbone atoms per residue of the free protein (top, black), the Bak peptide complex (top, magenta), complex **1·2** (bottom, green), complex **3** (bottom, blue) and complex **4** (bottom, red). Shifts in regions of flexibility are clearly visible

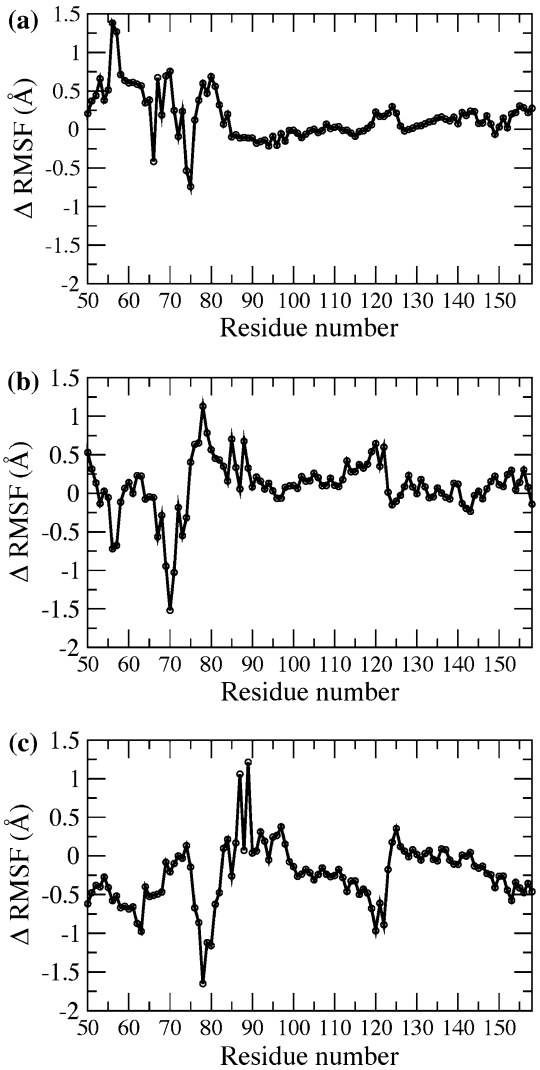


Fig. 6 The difference in the per-residue root mean square fluctuation of the Bcl-X_L backbone between (**a**) complex **1·2** and the free protein; (**b**) complex **3** and complex **1·2**; and (**c**) complex **4** and complex **3**. Progressive stabilization of the protein by new generation ligands is evident

rigidity is most likely due to the propensity of this ligand to form a unique wedge-shaped folded structure upon binding to the protein, which is stabilized by π-stacking interactions between the nitrophenyl and thiophenyl rings. This allows the ligand to anchor the protein through an extended π-stack system formed by interactions of the thiophenyl with Phe61 and Tyr65 of α3 and the nitrophenyl with Tyr159 of α5, which form a large hydrophobic pocket in the binding groove at the intersection of the BH1 and BH3 domains. However, at the same time, increased fluctuations are observed further down the binding arm, where the segment centered at the loop connecting helices α3 and α4 including residues 74–88, has been significantly destabilized. In addition, a new region of instability had formed, spanning residues 115–120 of α5. The latter two are

proximal to the residues that will frame the new binding pocket in complex **4**, which was not present in either the free protein or the Bak complex. Interestingly, a similar binding pocket was observed in the related complex of ligand **3** bound to Bcl-2, and it was the attempt to improve the inhibition properties of **3** against Bcl-2 by accessing this additional hydrophobic binding site that motivated further development of compound **3**. Comparison of the per-residue RMSF difference profiles for complex **3** and complex **4** clearly shows that the fluctuations of these regions are significantly reduced in the latter, indicating the interactions with the newly-added lipophylic biphenyl extension stabilized this region of the protein. The binding arm in complex **4** shows the highest degree of conformational stability, comparable to and even exceeding that of the native Bak complex indicating a correlation between the increased binding affinity of inhibitors and their ability to stabilize the protein fluctuations. Curiously, the segment including residues 86–88 at the base of helix α5 retains a markedly higher flexibility than the rest of the complex **4** binding groove, suggesting a potential for further improvement in the binding affinity through stabilization of this region. We note that neither the crystal structure of the free protein, nor the NMR structures of complex **3** show evidence of the new binding pocket. Instead, the discovery of the most potent inhibitor of Bcl-X_L was a fortuitous consequence of efforts directed at improving the inhibition of Bcl-2. On the other hand, potential for the formation of this additional binding site can be elucidated from the information on dynamic stability of the protein structure available from MD simulations.

The secondary structure of the Bcl-X_L binding arm is of helix-loop-helix topology, hence the conformational changes in the binding arm can also be well-characterized by percent helicity, defined as fraction of time a given residue is observed in a helical configuration. The residue is considered to be helical if the mean square deviation of the associated backbone dihedral angles (ϕ, ψ) is less than 64 deg^2. In Fig. 7 we show the percent helicity for binding arm residues of the free protein, the Bak complex, and complexes **1**·**2**, **3** and **4**. We first note that while in most cases the residues observed to be in helical conformation in experimental structures (light grey bars) are found to be so in at least some of the configurations sampled by MD, the dynamic stability of these structural elements varies significantly across the binding groove region, and among different complexes. This again underscores the importance of dynamic information provided by simulations in interpreting the structural features of flexible binding sites. In comparison with the free protein, binding of the Bak peptide results in the loss of helicity in residues 69 and 70, however this is compensated by the increase in the helicity of the neighboring residues 71 and 72, and significant

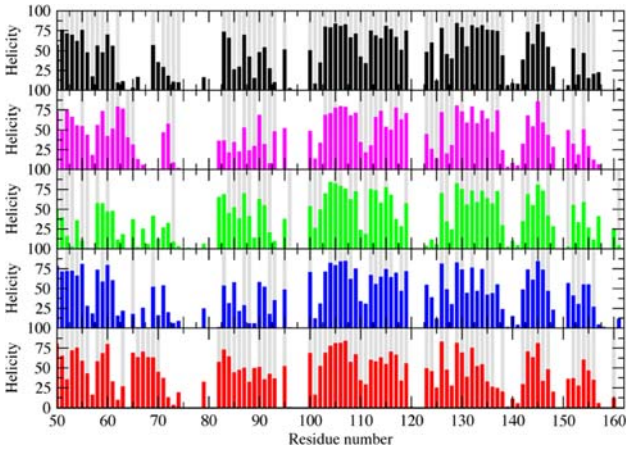


Fig. 7 The secondary structure population of Bcl-X_L shown as the percentage of time each residue was found in a helical conformation. The helicity data is shown for the free protein (black), the Bak peptide complex (magenta), complex **1**·**2** (green), complex **3** (blue) and complex **4** (red). The light gray bars show the helicity of the residues in the corresponding static experimental structures

organization of the previously unstructured segment 62–65. In addition, relatively high degrees of helicity were observed in several regions of α4 and α6 that appear less ordered in the experimental NMR structures. On the other hand, complex **1**·**2** exhibits significant loss of helicity in the region 50–65 due to destabilization of α2 upon binding of ligand **2**. Interestingly, increase in helicity was observed in residues 82–83 at the base of α5, which are distant from binding sites 1 and 2. Such effects were also observed by Manconelli et al. in their studies of the binding of pro-apoptotic peptides to Bax [24]. Complex **3** shows an overall increase in helical stability of α2, but this is accompanied by reduction in the helical stability in region 83–88 at the base of α4, paralleling the RMSF results. This trend is maintained in complex **4**, which shows the highest degree of helical order, exceeding even that of the native Bak complex. In particular, the binding of ligand **4** induces helical ordering of residues 65–70 of α3 and 85–90 of α4, both proximal to the newly formed hydrophobic binding site, that is occupied by the lipophilic extension of ligand **4**. In summary, significant differences were observed in equilibrium populations of helical conformations of binding arm residues for all four complexes relative to the free protein. In most cases, the increase in per residue helicity coincides with the reduction of corresponding RMS fluctuations. The notable exception is the loop region comprising residues 74–81, which remains unstructured in most complexes, although the fluctuations are reduced in complex **4** through interactions with the lipophilic extension. Hence, the helicity data nicely complements the RMSF analysis and further demonstrates the significant conformational changes that the binding arm undergoes upon ligand binding.

Characterization of key ligand–protein interactions

Initial 'SAR by NMR' studies identified compound **1** which showed a fair binding affinity ($K_i = 0.30$ mM) to site 1 of the Bcl-X_L BH3 binding groove, and compound **2** that bound to a proximal site 2 (Fig. 3), though with a lower binding affinity ($K_i = 4.3$ mM), to form complex **1** and **2**. In addition to the high degree of protein flexibility noted in section (ii), our simulations of complex **1**·**2** showed significant fluctuations in the conformations of bound ligands, in particular the more weakly bound ligand **2**, which was observed to undergo full 180° rotation in response to the loosening of the local protein structure. To investigate the binding mechanism in more detail, we computed the total interaction energy of the two ligands with each protein residue. The latter was defined as the sum of Coulomb and Lennard-Jones interactions of all ligand atoms with backbone and side-chain atoms of a given protein residue, and the results are shown in Fig. 8a and b. Significant interaction of ligand **1** with Arg103 is indicative of hydrogen bonding of the Arg103 amine to the ligand carboxylate, which is consistent with the findings of Oltersdorf et al. [6]. However, our simulations show an even stronger interaction of ligand **1** with Asn100, due to hydrogen bonding between the carboxylate of the former to the amide, which had not been observed previously. Moreover, upon reorganization of protein structure, in most conformations, the relative orientation of the ligands was found to be of distinct V-shape, with an angle of ~110° between the major axes bisecting the biphenyl and tetrahydronaphthalene, with the two ligands joined at the apex via an inter-ligand hydrogen bond. The latter was found to be surprisingly stable, being present in 99.5% of configurations. This orientation is remarkably similar to the alignment of nitrophenyl group with the biphenyl and phenylpiperazine in complexes **3** and **4**, respectively, that were achieved by introduction of the acylsulfonamide linkage, and it allows ligand **2** to form favorable stacking interactions with Tyr159, and Phe61 and Tyr65. These interactions are illustrated in Fig. 9a. The efforts to link

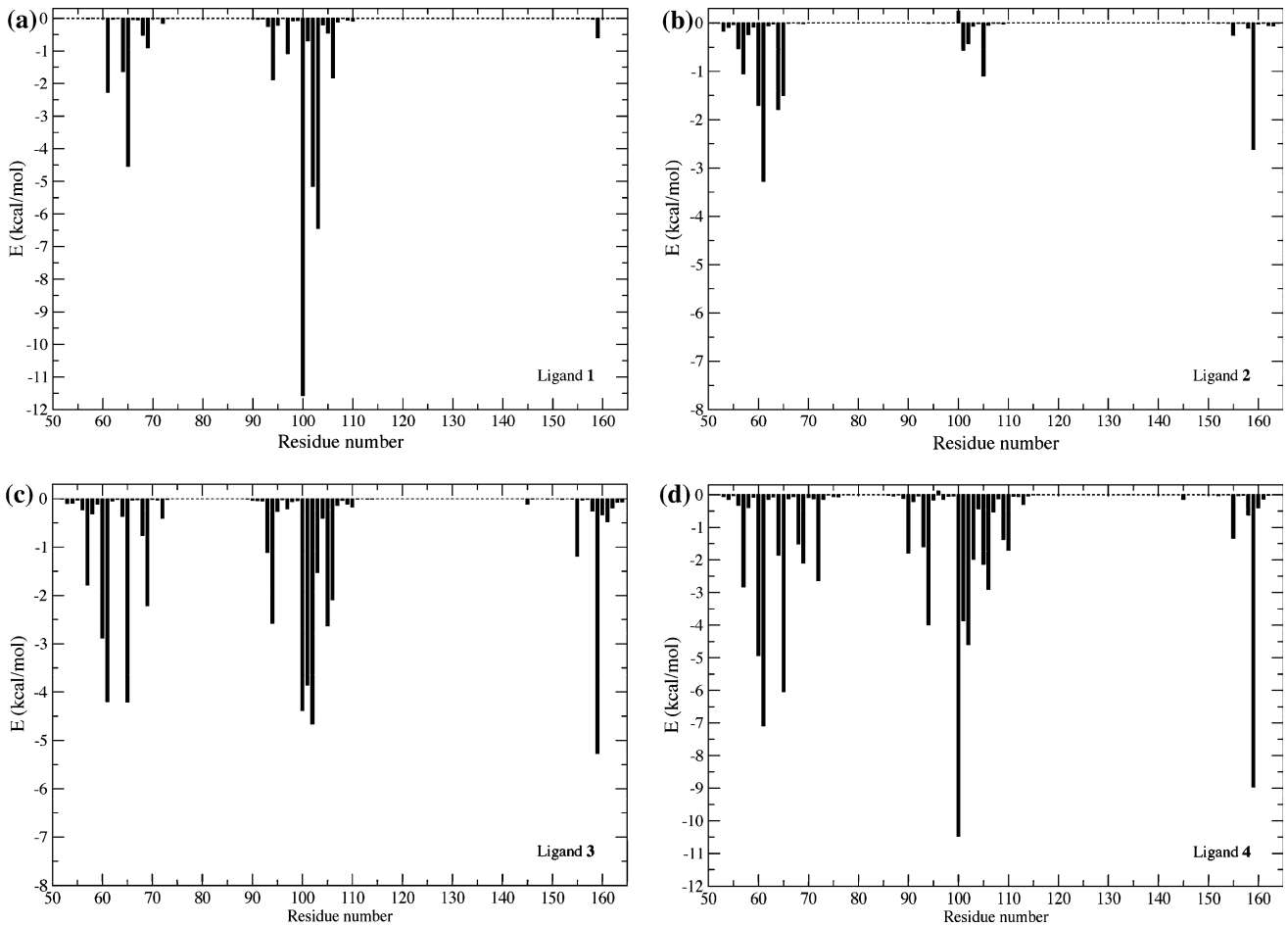


Fig. 8 Average interaction energy of (**a**) ligand **1**, (**b**) ligand **2**, (**c**) ligand **3**, and (**d**) ligand **4** with individual residues of Bcl-X_L, obtained from MD simulations

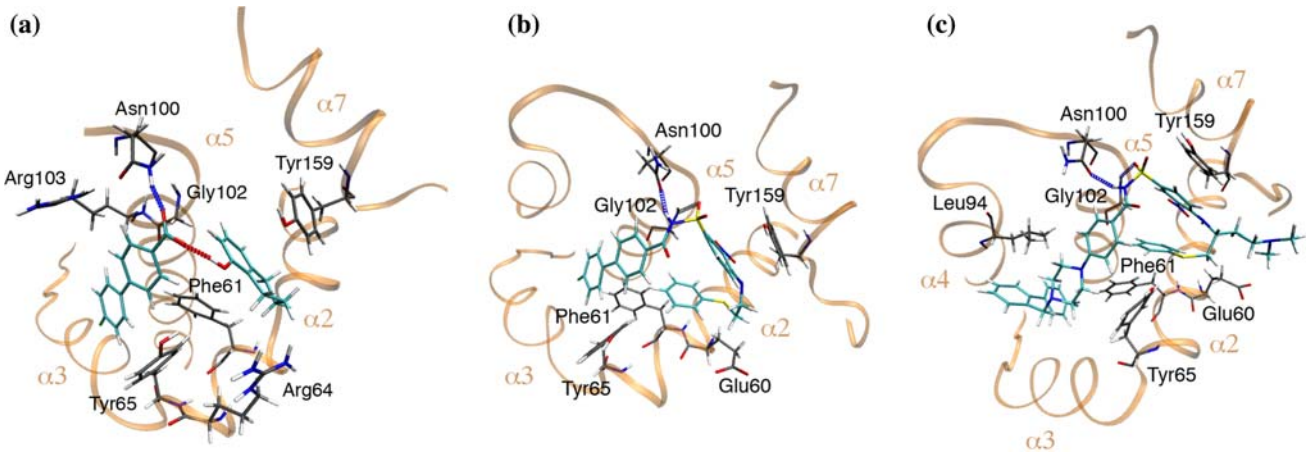


Fig. 9 Schematic representation of Bcl-X_L residues participating in prominent interactions with the ligands in (**a**) complex **1**·**2**, (**b**) complex **3**, and (**c**) complex **4**. Important hydrogen bonds and protein helices are noted

compound **1** to its partner in order to obtain a single molecule binding to both sites led to the discovery of the potent inhibitor **3**, whose high binding affinity for Bcl-X_L has been attributed to a uniquely folded wedge-shaped binding mode. This ligand conformation is believed to impart activity through several favorable ligand–protein interactions illustrated in Fig. 9b: (i) the formation of an extended π-stacking network, made possible by the hydrophobically driven folding of the 3-nitro-4-(2-phenylthioethyl)aminophenyl group which anchors the ligand in what was originally site 2 through interactions of the nitrophenyl of ligand **3** with the aryl side chain of Tyr159, and the thiophenyl with Tyr65 and Phe61, while the intramolecular stacking of nitrophenyl and thiophenyl rings imparts the conformational stability to the ligand; (ii) The characteristic V-shaped geometry of the sulfonamide hinge, placing the axis of fluorobiphenyl group at 110° to the plane of the nitrophenyl, allowing favorable interactions of the former with the hydrophobic binding site framed by Phe69, Ala68, Tyr65 and Leu94; (iii) the favorable hydrogen bonding interactions between the sulfonamide group and the neighboring protein residues. Following a 10 ns simulation of complex **3**, the ligand was found to maintain the stable wedge-shape conformation, despite considerable local fluctuations of protein structure, the most notable changes being the slight displacement of the thiophenyl group outwards leading to a somewhat reduced stacking with the nitrophenyl. The interaction profile of ligand **3** with the protein, given in Fig. 8c, shows some similarity to that of complex **1**·**2**, although there are significant differences. Due to conformational restrictions imposed by the acylsulfonamide linkage, Arg103 side chain proximal to the ligand was no longer able to attain a geometry that would permit hydrogen bonding, which is reflected in significantly reduced interaction of Arg103 with the ligand. On the other hand, strong interactions with Asn100 were maintained in complex **3**, however, in this case through hydrogen bonding with the amide group of the sulfonamide linkage, which was prevalent in 81.8% of conformations. Surprisingly, despite strongly favorable electrostatic interactions, no significant hydrogen bonding was observed between Gly102 and the sulfonamide oxygen, the hydrogen bond being populated in only 0.02% of the conformations sampled. We note that while the importance of Gly102 in the binding of ligand **3** had been proposed before, the participation of Asn100 had not been recognized until now. The other significant set of interactions contributing to strong binding of ligand **3** are the non-polar interactions of the nitrophenyl-thiophenyl system, principally through strong parallel π-stacking of nitrophenyl with Tyr159, and to a lesser extent, the displaced parallel stacking of thiophenyl with Phe61 and Tyr65, as well as a favorable electrostatic interaction with the nearby charged Glu60. Additional stabilization is provided by hydrophobic contacts formed by the fluorobiphenyl with Phe69, Ala68, Tyr65 and Leu94. It is of interest to compare the interaction profile of ligand **3** with that of the best-known inhibitor, compound **4**. We note that the primary goal in developing compound **4** was not to increase the binding affinity to Bcl-X_L, but rather to improve activity against Bcl-2, and selectivity against HSA. The latter was achieved by replacing certain hydrophobic groups of ligand **3** that were solvent exposed in complex **3**, but buried in hydrophobic pockets when bound to HAS, with polar substituents. The interaction profile of ligand **4**, shown in Fig. 8d is quite similar to that of the parent compound **3**, with most of the key interactions found in the latter participating with somewhat more favorable energies, which reflects the overall higher binding affinity of complex **4**. A notable exception is the re-establishment of the hydrogen

bonding interaction with Arg103 which was present in complex **2**, but not in complex **3**, which further explains the higher stability of complex **4**. Furthermore, the chlorobiphenyl extension designed to bind to site 3 of Bcl-2 is found to insert into a newly formed deep binding pocket framed by lipophilic residues Phe61 and Ala68 of α3, Leu72, Val90, and Leu94 of α4, and Ala106 and Phe108 of α5. The key interactions are illustrated in Fig. 9c. It is important to note that this pocket was not present in any of the previous complexes, as well as the free protein, but rather is a product of the binding groove side chain and backbone rearrangement induced by the binding of compound **4**. Moreover, several of the residues that will form this new binding site belong to the highly flexible regions of the binding arm of complex **3** identified in the previous section. These favorable hydrophobic contacts then contribute both to the increased binding affinity of the former as well the reduction of fluctuations in the binding arm, and therefore, increased stability of the protein in complex **4**, as observed in Fig. 5. We also found that the hydrophilic substituents added to compound **4** to improve selectivity versus HAS remain in contact with the solvent throughout the simulation, with the two piperazine nitrogens and the tertiary amine attached to the thioethyl linkage each forming on average $\sim$1.0 hydrogen bonds with water.

We note that the interaction profile for ligand **1** exhibits several interactions that are significantly larger in magnitude than those observed in other profiles in Fig. 8, belying its low binding affinity. This is primarily due to the strong Coulombic interactions between the negatively charged carboxylate and charged protein residues, since inhibitor **1** binds in the ionized form. In spite of several highly favorable interactions, the overall binding affinity of inhibitor **1** is still poor compared to inhibitors **3** and **4**, since attractive interactions are partially cancelled by the large desolvation penalty for removing the ionized inhibitor **1** from water. This is confirmed by the MM-GBSA calculations of the relative binding affinities of ligands given in Table 1, which incorporate contributions from these competing effects. The latter are found to be in good qualitative agreement with the measured K_i and K_d values for the four inhibitors, suggesting that the bulk of the ligand–protein interaction is correctly captured.

Table 1 Relative binding affinities of the ligands computed by MM-GBSA and experimental inhibition constants

Inhibitorr	ΔG_{bind} (kcal/mol)[a]	K_i or K_d[b] (μM)
Ligand **1**	−10.9 ± 4.9	300 ± 30
Ligand **2**	−14.2 ± 2.2	4,300 ± 160
Ligand **3**	−38.2 ± 2.6	0.0036 ± 0.00016
Ligand **4**	−54.1 ± 3.6	0.0008

[a] Averages over six snapshots taken at 1 ns intervals from the last 5 ns of MD simulations. The error bars reported are standard deviations over this set

[b] Experimental data and errors as reported in Ref. [6]. The values are K_d for Ligand **1** and **2**, and K_i for Ligand **3** and **4**

To further explore the origin of the high binding affinity of inhibitors **3** and **4**, we investigate the key hydrogen bonding and π-stacking interactions in more detail. In both complex **3** and **4**, the dominant hydrogen bonding interaction was found to be that between carboxylamide oxygen of Asn100 and the nitrogen of the sulfonamide linkage, populated in 81.8% and 94.2% conformations, respectively. The distributions of donor-acceptor distances of this hydrogen bonding pair for complex **3** and complex **4** are shown in Fig. S1 provided as Supplementary Material. In both cases, the distributions are found to be tightly peaked around 3.0 Å, confirming that this interaction holds the ligands firmly in place. We note that the average NMR structure for complex **3** (PDB code 1YSI) shows the side chain Asn100 oriented away from the sulfonamide resulting in a larger donor–acceptor distance of 5.5 Å. On the other hand, in crystal structure of complex **4** (PDB code 2YXJ) the said side chain is positioned closer to the amide, with donor-acceptor distance of 3.9 Å and orientation very similar to that predicted by our simulations. Since the above distances are somewhat larger than the usual hydrogen bonding range, the importance of latter interaction has not been recognized. However, due to the plasticity of the active site, room temperature dynamic fluctuations are sufficient to bring the donor and acceptor in contact. The strength and persistence of these hydrogen interactions observed throughout our simulations lead us to believe that they are an important contributor to the stability of complexes **3** and **4**.

Gly102 also participates in significant polar interactions with the ligands, through electrostatic attraction and hydrogen bonding with the sulfoxyl oxygen of the sulfonamide linkage. While the latter is observed in 46.7% of sampled conformations of complex **4**, this hydrogen bond is found to be populated in only 0.02% conformations of complex **3**. The lower stability of this hydrogen bond compared to that with Asn100, particularly for complex **3**, is reflected in a relatively broad distribution of donor-acceptor distances, which is shown in Fig. S2 provided as Supplementary Material, with only a small fraction of conformations within hydrogen bonding distance. Moreover, in most of these the peptide amide and sulfonamide oxygen were found not to be in the geometry that would permit hydrogen bond formation. We note that in earlier simulations of complex **3** using the original OPLSAA charge distribution, this hydrogen bond was better populated, even though the overall stability of the complex was poorer, suggesting this interaction is quite sensitive to force field parameterization. Conversely, complex **4** shows a

significantly smaller variation in the distance distribution indicating an overall tighter binding of compound **4** and higher conformational stability of the ligand.

The principal π-stacking interactions are those between the ligand nitrophenyl and Tyr159, and the thiophenyl and Phe61. The distribution of the distances between centers of mass of the participating aromatic rings for the two π-stacking pairs is given in Fig. S3 of the Supplementary Material. In both complexes, the nitrophenyl–tyrosene pair shows a relatively narrow distribution strongly peaked around 3.9 Å, which is consistent with that of parallel-stacked aryl moieties [41]. On the other hand, the thiophenyl-phenylalanine stacks exhibit significantly broader distributions, peaked at 5.5 and 5.0 Å for complex **3** and **4** respectively, which are consistent with a those more weakly bound T-shaped π-interacting aryl dimers.

Finally, to quantify the overall stability of ligands **3** and **4** in their bound state, we computed the potential of mean force (PMF) as a function of the displacement of the ligand from the binding pocket specified by the distance between the donor and the acceptor of the hydrogen bonding partners of Asn100 and the sulfonamide linkage. In Fig. 10, we plot the PMF profiles for complex **3** and **4**. In both cases, a well-defined, deep minimum corresponding to the preferred binding mode is found at 2.9 and 3.0 Å respectively, indicating that the ligand is tightly held in the pocket and maintains a consistent orientation with respect to key protein residues. Another important features is the barrier at 4.4 Å in both complexes, which provides a measure of the free energy cost for displacing a ligand from the binding pocket. The latter is found to be $\sim$2.9 kcal/mol for ligand **3** and $\sim$4.4 kcal/mol for ligand **4**, corresponding to 5–7 k_BT at room temperature, indicating that both ligands are strongly bound. The higher barrier for ligand **4** is consistent with the higher binding affinity of the latter. In order to estimate the overall effect of the principal Asn100 hydrogen bond, we also computed the PMF for complex **4** along the same reaction coordinate with the amide partial charges set to zero (blue dashed line), effectively eliminating the polar contribution of the interactions. Comparison with the original profiles show a shift in the position of the minimum to 3.2 Å, accompanied by significant broadening, and a decrease in the barrier height of $\sim$2.7 kcal/mol, indicating that the removal of the hydrogen bond leads to significant destabilization of the ligand in the binding pocket, and likely enhancement of the escape kinetics.

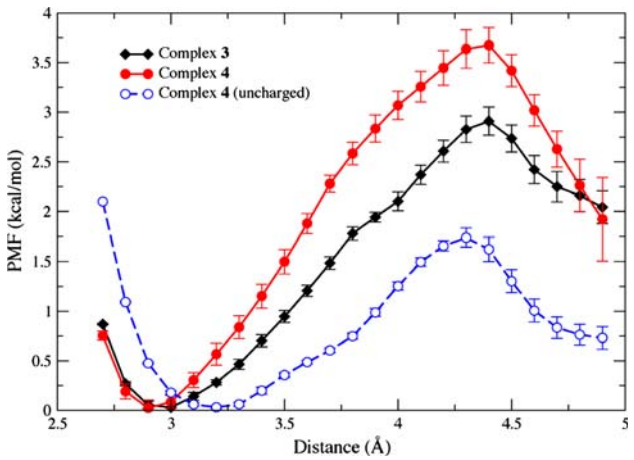


Fig. 10 Potential of mean force for displacing the ligand from the binding site in complex **3** (black line, diamonds) and complex **4** (red line, filled circles) as a function of the distance between carboxylamide oxygen of Asn100 (hydrogen bond acceptor) and the ligand sulfonamide nitrogen (hydrogen bond donor) obtained from constraint force simulations. The dashed blue line (empty circles) shows the PMF for complex **4** with the sulfonamide amide charges set to zero

Conclusions

In this paper, we have used MD simulations to explore the role of structural flexibility in the rational design of small-molecule inhibitors for Bcl-X_L, an important cancer therapy target. The latter has proven to be quite a formidable task for traditional drug design methods which rely on static structural data available from X-ray or NMR studies, since protein–protein interfaces often lack clearly defined deep binding pockets. In addition, the interface regions are usually sufficiently pliable to undergo substantial conformational changes upon binding of a given ligand, resulting in a binding site topography significantly different from those of the free protein or other related ligand complexes, rendering it difficult to make fruitful ligand optimization decisions based on these static structures. This is clearly evident in the design cycle of the potent inhibitor of Bcl-X_L ABT-737 (ligand **4**), which was discovered through a massive multi-stage 'SAR by NMR' study requiring the experimental determination of structures of a large number of ligand–protein complexes.

On the other hand, we have shown that dynamic information obtained through MD simulations can be very helpful in guiding the ligand design by identifying specific regions of the binding site as targets that are not apparent from the analysis of static structures. In particular, through the analysis of local fluctuations of the Bcl-X_L binding arm in complexes with three generations of small molecule inhibitors leading to ABT-737, we have found that the binding of each successive ligand with improved inhibition potency results in stabilization of the region of the binding arm found to be highly flexible in the preceding generation complex. Furthermore, these regions were found to correspond to new binding sites accessed by the modified ligands. This is especially evident in the case of modifications to ligand **3** leading to ABT-737, which were designed

primarily with the aim to improve binding to Bcl-2, by targeting a specific hydrophobic pocket observed in the structure of the ligand **3** complex with that protein. The finding that ABT-737 also exhibited superior binding to Bcl-X_L was surprising, since no evidence of the binding site accessed by the fluorobiphenyl extension was present in complex **3**. It is of interest whether this correlation between local flexibility and potential binding sites made accessible through induced fit effects is specific to Bcl-X_L, or if it is a more general characteristic of pliable protein–protein interfaces. Recent studies [26] uncovering a relationship between transient pockets formed through thermal fluctuation of several protein-binding receptors and corresponding ligand induced binding sites suggest that this may be the case, however a conclusive determination would require the consideration of a much larger set of ligand–protein complexes, which we plan to address this in a future study.

The analysis of ligand–protein interaction profiles has shown that the latter evolve with the dynamic rearrangement of protein structure induced by the binding of different ligands. Several prominent interactions, including hydrogen bonding of the polar glycine and asparagine to the sulfonamide linkage, and the π-stacking of the nitrophenyl-thiophenyl system with the nearby aromatic side chains were identified, and found to be progressively reinforced with each successive inhibitor generation. Somewhat surprisingly, one of the strongest interactions in all three complexes was observed to be the hydrogen bond between the carboxylamide oxygen of Asn100 and the nitrogen of the sulfonamide linkage. The importance of this interaction in stabilizing the bound state conformations of ligands **3** and **4**, which has not been recognized previously, has been confirmed by free energy calculations of the potential of mean force for the displacement of the ligand from the binding site.

In summary, our study has demonstrated the importance of protein dynamics in understanding the interactions of small molecule inhibitors with a flexible receptor, such as the protein–protein interface of Bcl-X_L. Moreover, we have shown how information extracted from MD simulations can help guide the rational design of small molecule inhibitors for targets with pliable binding sites lacking specific features, such as protein–protein interfaces.

Acknowledgments This work was funded by a grant from Boston College to G.K.

References

1. Sharma SK, Ramsey TM, Bair KW (2002) Curr Med Chem. Anti-Cancer Agents 2:311–330. doi:10.2174/1568011023354191
2. Arkin MR, Wells JA (2004) Nat Rev Drug Discov 3:301–317. doi:10.1038/nrd1343
3. Berg T (2003) Angew Chem Int Ed 42:2462–2481. doi:10.1002/anie.200200558
4. Yin H, Hamilton AD (2005) Angew Chem Int Ed 44:4130–4163. doi:10.1002/anie.200461786
5. Muchmore SW, Sattler M, Liang H et al (1996) Nature 381:335–341. doi:10.1038/381335a0
6. Oltersdorf T, Elmore SW, Shoemaker AR et al (2005) Nature 435:677–681. doi:10.1038/nature03579
7. Borner C (2003) Mol Immunol 39:615–647. doi:10.1016/S0161-5890(02)00252-3
8. Cory S, Adams JM (2002) Nat Rev Cancer 2:647–656. doi:10.1038/nrc883
9. van Delft MF, Huang DC (2006) Cell Res 16:203–213. doi:10.1038/sj.cr.7310028
10. Coultas L, Strasser A (2003) Semin Cancer Biol 13:115–123. doi:10.1016/S1044-579X(02)00129-3
11. Kirkin V, Joos S, Zornig M (2004) BBA-Mol Cell Res 1644:229–249
12. Amundson SA, Myers TG, Scudiero D et al (2000) Cancer Res 60:6101–6110
13. Green DR, Evan GI (2002) Cancer Cell 1:19–30. doi:10.1016/S1535-6108(02)00024-7
14. Wendt MD, Shen W, Kunzer A et al (2006) J Med Chem 49:1165–1181. doi:10.1021/jm050754u
15. Bruncko M, Oost TK, Belli BA et al (2007) J Med Chem 50:641–662. doi:10.1021/jm061152t
16. Sattler M, Liang H, Nettesheim D et al (1997) Science 275:983–986. doi:10.1126/science.275.5302.983
17. Petros AM, Nettesheim DG, Wang Y et al (2000) Protein Sci 9:2528–2534. doi:10.1017/S096183680000331X
18. Petros AM, Olejniczak ET, Fesik SW (2004) BBA-Mol Cell Res 1644:83–94
19. Shuker SB, Hajduk PJ, Meadows RP et al (1996) Science 274:1531–1534. doi:10.1126/science.274.5292.1531
20. Petros AM, Dinges J, Augeri DJ et al (2006) J Med Chem 49:656–663. doi:10.1021/jm0507532
21. Park CM, Oie T, Petros AM et al (2006) J Am Chem Soc 128:16206–16212. doi:10.1021/ja0650347
22. Zheng CH, Zhou YJ, Zhu J et al (2007) Bioorg Med Chem 15:6407–6417. doi:10.1016/j.bmc.2007.06.052
23. Pinto M, Perez JJ, Rubio-Martinez J (2004) J Comput Aided Mol Des 18:13–22. doi:10.1023/B:JCAM.0000022559.72848.1c
24. Mancinelli F, Caraglia M, Budillon A et al (2006) J Cell Biochem 99:305–318. doi:10.1002/jcb.20893
25. Fu XR, Apgar JR, Keating AE (2007) J Mol Biol 371:1099–1117. doi:10.1016/j.jmb.2007.04.069
26. Eyrisch S, Helms V (2007) J Med Chem 50:3457–3464. doi:10.1021/jm070095g
27. Manion MK, O'Neill JW, Giedt CD et al (2004) J Biol Chem 279:2159–2165. doi:10.1074/jbc.M306021200
28. Berendsen HJC, Vanderspoel D, Vandrunen R (1995) Comput Phys Commun 91:43–56. doi:10.1016/0010-4655(95)00042-E
29. Lindahl E, Hess B, van der Spoel D (2001) J Mol Model 7:306–317
30. Van der Spoel D, Lindahl E, Hess B et al (2005) J Comput Chem 26:1701–1718. doi:10.1002/jcc.20291
31. Jorgensen WL, Maxwell DS, TiradoRives J (1996) J. Am. Chem. Soc 118:11225–11236. doi:10.1021/ja9621760
32. Kaminski GA, Friesner RA, Tirado-Rives J et al (2001) J Phys Chem B 105:6474–6487. doi:10.1021/jp003919d
33. Jorgensen WL, Chandrasekhar J, Madura JD et al (1983) J Chem Phys 79:926–935. doi:10.1063/1.445869
34. Essmann U, Perera L, Berkowitz ML et al (1995) J Chem Phys 103:8577–8593. doi:10.1063/1.470117
35. Hoover WG (1985) Phys Rev A 31:1695–1697. doi:10.1103/PhysRevA.31.1695

36. Berendsen HJC, Postma JPM, Vangunsteren WF et al (1984) J Chem Phys 81:3684–3690. doi:10.1063/1.448118
37. Ghosh A, Rapp CS, Friesner RA (1998) J Phys Chem B 102:10983–10990. doi:10.1021/jp982533o
38. den Otter WK, Briels WJ (1998) J Chem Phys 109:4139–4146. doi:10.1063/1.477019
39. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327–341. doi:10.1016/0021-9991(77)90098-5
40. Lugovskoy AA, Degterev AI, Fahmy AF et al (2002) J Am Chem Soc 124:1234–1240. doi:10.1021/ja011239y
41. McGaughey GB, Gagne M, Rappe AK (1998) J Biol Chem 273:15458–15463. doi:10.1074/jbc.273.25.15458