Energetic Fingerprinting of Ligand Binding to Paralogous Proteins

DOI:
10.1021/acs.jcim.8b00765

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Journal of Chemical Information and Modeling

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Energetic Fingerprinting of Ligand Binding to Paralogous Proteins: The Case of the Apoptotic Pathway

Stefan M. Ivanov\textsuperscript{1,2}, Roland G. Huber\textsuperscript{2}, Irfan Alibay\textsuperscript{3}, Jim Warwicker\textsuperscript{1}, Peter J. Bond\textsuperscript{2,4}\textsuperscript{*}

\textsuperscript{1} Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK
\textsuperscript{2} Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Matrix 07-01, 30 Biopolis Street, Singapore 138671, Singapore
\textsuperscript{3} Division of Pharmacy and Optometry, School of Health Sciences, The University of Manchester, Oxford Road, Manchester, M13 9PT
\textsuperscript{4} Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore

*corresponding author: peterjb@bii.a-star.edu.sg
Abstract

Networks of biological molecules are key to cellular function, governing processes ranging from signal cascade propagation to metabolic pathway regulation. Genetic duplication processes give rise to sets of regulatory proteins that have evolved from a common ancestor, leading to interactomes whose dysregulation is often associated with disease. A better understanding of the determinants of specificity at interfaces shared by functionally related proteins is crucial to the rational design of novel pharmacotherapeutic agents. To this end, a comprehensive dataset of drug and drug-like binders was assembled for the Bcl-xL and Bcl-2 antiapoptotic proteins – archetypal examples of regulatory systems governed by evolutionarily conserved protein-protein interactions. These were first used to derive a two-dimensional quantitative structure-activity relationship (2D QSAR) model, predicting ligand specificity for these homologous proteins. The strengths and weaknesses of high-throughput 2D QSAR were then compared and contrasted to those of theoretically rigorous thermodynamic integration calculations performed on 14 complexes of Bcl-xL-specific, Bcl-2-specific, and potent dual binders bound to the Bcl-xL and Bcl-2 proteins. We demonstrate that free energy calculations provide an added layer of essential information, which traditional QSAR cannot capture. Moreover, we show that protein energetic responses to different ligands, expressed as per-residue energy values, can be used to fingerprint the protein-ligand interaction, extending the framework of four-dimensional molecular dynamics/quantitative structure-activity relationships (4D-MD/QSAR) towards the facilitation of future drug design strategies.
Association between biopolymers is an essential part of physiological processes in all domains of life. Deviations from the genetically predetermined innate interactome are often deleterious, manifesting themselves in a broad spectrum of conditions, ranging from minor symptoms to debilitating syndromes. For example, the antiapoptotic B-cell lymphoma-2 (Bcl-2) family of proteins, including Bcl-xL and Bcl-2, bind to and inactivate BH3-domain proapoptotic proteins. Their overexpression overwhelms the cell's proapoptotic defenses, facilitating malignant proliferation. Conversely, mutations that abrogate binding between pro- and antiapoptotic proteins shift the cellular balance toward premature cell death, and give rise to or are associated with degenerative diseases. Point mutations in leucine zipper transcription factors can lead to altered dimerization and DNA binding, resulting in a great number of documented malignancies. Emulating regulatory protein–protein interactions with small molecule or (stapled) peptide analogues— the so-called Bcl-inhibitors— has been successfully used to rectify imbalances in the apoptosis-regulation pathway, and has produced numerous promising clinical candidates. Encouragingly, one Bcl-2-inhibitor—Venetoclax (ABT-199) —has already obtained FDA-approval for cancer treatment. Targeting other regulatory systems governed by inter- or intrafamily protein–protein recognition and association, such as the ubiquitination pathway, two-component signal transduction, and G protein receptor regulation, is also expected to provide future drug candidates for a wide range of conditions. Thus, elucidating the intricacies of protein–protein and protein–ligand association—the origins of specificity and affinity—is a long-standing goal in the broader scientific field.

Computational strategies for modeling intermolecular interactions include those based on colloid science, bioinformatics or force field-based approaches. Quantitative structure-activity relationship (QSAR) studies are often employed to rationalize observed patterns in binding based on the chemical moieties of the ligands that confer affinity and specificity, with more recent developments in QSAR analysis seeing the explicit inclusion of structural data. Intrafamily protein–protein interactions in key regulatory pathways, such as the Bcl-2 proteins, which regulate apoptosis, present a formidable challenge to drug design and cancer therapy. The high similarity between family members, which have arisen through gene duplication and subsequent divergence, makes off-target effects almost unavoidable and limits the therapeutic efficacy of pharmacological agents. For example, the Bcl-2 protein has been targeted to
treat non-Hodgkin lymphoma and chronic lymphocytic leukaemia, but undesired binding to the highly similar Bcl-xL protein causes dose-limiting thrombocytopenia. To date, all known peptide or small molecule Bcl-2 binders also possess at least some residual affinity for Bcl-xL. Current data for Bcl-2 inhibitors in humans demonstrates that a selectivity of several thousand-fold for the designated target is sufficient to make off-target binding clinically insignificant. The lower bound for selectivity, however, has not been definitively established and may vary in different signaling pathways.

Medicinal chemists typically attempt to enhance binding affinity or achieve selectivity by optimizing polar contacts and hydrogen bonds or salt bridges between a ligand and receptor under the assumption that changes in ligands are small enough to not induce significant conformational change in the receptor. When this is the case, it is straightforward to rationalize the origins of any enhancements to affinity afforded by introducing hydrogen bond donors and acceptors or by enhanced van der Waals contacts. Even under these favorable circumstances, however, it is not easy to anticipate every resulting change in binding energetics, e.g. the free energy cost of reorganizing the receptor from the unbound to the bound ensemble or the change in ligand conformational entropy upon binding. Moreover, ligand – receptor interfaces often involve a great number of interdependent interactions. Optimizing affinity and/or specificity becomes a complex, multidimensional problem, and reducing the dimensionality of this problem is desirable. To this end, we have compiled from literature a comprehensive dataset on binding affinity for Bcl-xL and Bcl-2 inhibitors. We first present a two-dimensional quantitative structure-activity relationship (2D QSAR) analysis, based on low-level molecular descriptors, and compare that to an energetic analysis and 4D QSAR, the fourth dimension being time, based on sampling of conformations obtained from molecular dynamics (MD) simulations for two dual Bcl-2/Bcl-xL binders, two Bcl-xL specific binders, and two Bcl-2-specific binders from a congeneric series of compounds. 2D (low-level) QSAR results are compared and contrasted with 4D (high-level) QSAR from thermodynamic integration calculations totaling 1.65 µs, performed on 12 modeled protein – ligand complexes and 2 template protein – ligand complexes, obtained from the Protein Data Bank (PDB) (Figure 1). We demonstrate how atomic-resolution MD simulations, coupled with free energy calculations, can be used to more fully characterize protein – ligand complexes, and improve upon traditional QSAR.
**Methods**

**2D QSAR**

Only compounds shown to bind to the canonical groove of the antiapoptotic proteins were considered for future analysis; these were converted to SMILES format. For each compound, 165 molecular descriptors were calculated using RDKit.\(^{53,61}\) Moreover, descriptors were additionally normalized by molecular weight and heavy atom count; parameters, scaled to unit standard deviation, with zero or near zero variance, were removed from consideration. Partial least squares regression was then performed on the resulting set of descriptors, fitting the data to the logarithm of the affinity of a compound for Bcl-xL over Bcl-2 \((\log_{10}(\text{affinity for Bcl-xL/affinity for Bcl-2}))\) – a parameter henceforth referred to as specificity.

Ligands with a specificity below -1 are defined as Bcl-xL-selective, ligands with a specificity between -1 and 1 as dual binders, and ligands with a specificity greater than 1 as Bcl-2-specific. The final dataset consisted of 57 Bcl-xL-specific compounds, 112 dual binders, and 19 Bcl-2-specific binders (Figure 2). The compounds were sorted by increasing specificity and split into a training and test set for external validation in a 3:1 ratio – for every three compounds in the training set, one was placed in the test set, moving from Bcl-xL-selective to Bcl-2-selective ligands. The training dataset was used to derive a partial least squares regression model, from which the specificity of the compounds in the test set was predicted. No compound from the training set was present in the test set. To test the robustness and validity of the QSAR model, Monte Carlo cross-validation was performed,\(^{63,64}\) i.e. a random number of random compounds in the test set were swapped with compounds from the training set, after which the QSAR model was derived anew and used to make predictions on the new test set, repeating this procedure 20,000 times. Only swaps within classes were permitted, i.e. Bcl-xL selective compounds were exchanged only for Bcl-xL selective ligands, Bcl-2-selective compounds were exchanged only for Bcl-2-selective ligands, and dual binders were exchanged only for dual binders. Additionally, to remove any bias against the scarce Bcl-2 selective ligands, the dataset was “normalized”, i.e. an analogous analysis was performed with only 19 compounds in each class – 14 in the training set and 5 in the test set. Partial least squares regression was performed with the R \textit{pls} package.\(^{65}\)

**System setup for template structure equilibration simulations**
Six of the compounds (see Figures 2 and 3) were modeled on the 2YXJ and 4LVT crystal structures (ABT-737 (N3C) and ABT-263 (1XJ, Navitoclax) bound to Bcl-xL and Bcl-2, respectively); the models were subjected to molecular dynamics simulation and free energy calculations. Before simulations of the modeled complexes could be initiated, the template structures were simulated first; all simulations were performed with the Amber14 suite. N-termini were capped with an acetyl group, C-termini were capped with a methylamino (-NHCH₃) group. Protein chains were protonated and solvated in a cubic box with TIP3P water with tleap with a minimal wall distance of 13 Å. 0.15 M NaCl was added to approximate a physiological salt concentration whilst ensuring charge neutrality. Protonation states for the ligands were assigned using ChemAxon's Calculator Plugins. Parameters for the ABT-737 and ABT-263 compounds were obtained using the general Amber force field (GAFF 1.7) with AM1-BCC charges using antechamber.

**Simulation protocol for template structure equilibration simulations**

The solvated systems were subjected to a series of 2,000-steps energy minimization with a harmonic restraint of 100 kcal mol⁻¹ Å² first applied to all heavy atoms, followed by all solute heavy atoms, followed by protein heavy atoms. The systems were heated from 100 to 300 K over a period of 1 ns at constant volume with 100 kcal mol⁻¹ Å² harmonic restraints on all heavy atoms, followed by constant pressure density equilibration of 1 ns, followed by cooling to 100 K at constant volume with harmonic restraints. The protein-ligand complexes were again subjected to 2,000 steps of energy minimization with restraints of 100 kcal mol⁻¹ Å² on protein heavy atoms, followed by 2,000 step minimization series with decreasing restraints of 50, 20, and 10 kcal mol⁻¹ Å². The systems were then reheated from 100 to 300 K under constant volume conditions over a period of 1 ns, with a harmonic restraint of 20 kcal mol⁻¹ Å² on protein and ligand heavy atoms, followed by 1 ns of constant pressure density equilibration with restraints, followed by 1 ns of equilibration with restraints on protein heavy atoms, followed by 1 ns of equilibration with 20 kcal mol⁻¹ Å² restraints on Cα atoms only. The systems were then equilibrated for 1 ns without any restraints and simulated for 100 ns under constant pressure (1 bar) and temperature (300 K) conditions, maintained with the Berendsen barostat and the Langevin thermostat. Collision frequencies for temperature coupling were set to 2 ps⁻¹; the pressure relaxation time was 2 ps. Both systems were simulated in four independent replicas with the
**ff14SB** force field. An 8.0 Å cutoff was used for van der Waals interactions. Long-range electrostatics were computed with the particle-mesh Ewald (PME) scheme. Bonds to hydrogen were constrained using the SHAKE algorithm, allowing for a 2 fs time step.

**System setup for thermodynamic integration simulations**

Trajectories were processed with *cpptraj V14.25* to perform roto-translational alignment and compute Cα and ligand heavy atom root-mean-square deviations (RMSDs). The instantaneous enthalpy of binding (ΔH) between protein and ligand was also monitored every picosecond; ΔH was computed with MMPBSA.py. Frames from the Bcl-xL – ABT-737 and Bcl-2 – ABT-263 trajectories were selected to serve as templates for modelling the 12 antiapoptotic protein-ligand complexes and the subsequent free energy calculations. The six compounds from Figure 3 were modeled bound to Bcl-xL and Bcl-2 using Schrodinger's Maestro suite. Our guiding principle in modelling was to bury hydrophobic moieties in hydrophobic pockets; polar fragments were anticipated to be solvent exposed. The placement of the azaindoloxy moiety of compound 8 was guided by the positioning of an analogous indoloxy moiety in the 4MAN crystal structure of Bcl-2 bound to a similar compound, i.e. nearly parallel to the nitroaryl fragment. Protonation states for the modeled ligands were assigned using ChemAxon's Calculator Plugins. Parameters were obtained from the general Amber force field (GAFF 1.7) with AM1-BCC charges using *antechamber*. The complexes were solvated in a cubic box with TIP4P water and a minimal wall distance of 13 Å; NaCl was added to a concentration of 0.15 M.

**Simulation protocol for thermodynamic integration simulations**

One-step thermodynamic integration with softcore potentials and linear scaling was performed on the ABT-737 and ABT-263 compounds bound to Bcl-xL and Bcl-2, transforming each of them into the six target compounds (Figure 3) bound to the two proteins, as well as isolated in solution. Henceforth, this is termed the “forward” transformation, as opposed to the reverse process of transforming the model ligand into the template molecule, which was also performed. Moreover, the absolute free energies of binding of the template ligands were computed by decoupling them from the computational box (for the reverse
transformation) and by inserting them into the computational box (for the forward transformation).

The solvated model complexes were subjected to 2,000 steps of energy minimization with a harmonic restraint of 20 kcal mol\(^{-1}\) Å\(^2\) on protein and ligand heavy atoms. The structures were heated from 100 to 300 K under constant volume conditions for 10 ps with 20 kcal mol\(^{-1}\) Å\(^2\) restraints on protein and ligand heavy atoms, followed by 10 ps of restrained density equilibration, and 2.5 ns of unrestrained constant pressure dynamics in each \(\lambda\)-window. The time step was set to 1 fs; bonds to hydrogen were not constrained during thermodynamic integration simulations. Transformations were carried out with 11 \(\lambda\)-windows for the forward (\(\lambda = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\)) and reverse (\(\lambda = 0.005, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.995\)) processes and default settings for the \(scalpha\) (0.5) and \(scbeta\) (12 Å\(^2\)) parameters, which control the softness of the potential. Energy minimization, heating, density equilibration, and production dynamics were all performed with potential energy functions, corresponding to the \(\lambda\)-value of every \(\lambda\)-window, thus avoiding any Hamiltonian lag.\(^{85}\) Forward, i.e. template-to-model, transformations were carried out with the \textit{pmemd} module of Amber, whereas reverse (model-to-template) transformations were performed with \textit{sander}. All transformations were carried out with the \textit{ff14SB} force field in three independent replicas under NPT conditions.

**Free energy analysis**

Autocorrelation times (\(\tau\)) for the free energy calculations were computed from the derivative of the potential with respect to \(\lambda\) (\(\delta V/\delta \lambda\)). The pressure-volume component of the free energy was considered marginal and ignored, as is typically the case in biomolecular simulations.\(^{86-89}\) Moreover, for the \textit{sander} transformations, per-residue energy decomposition was enabled, adding 1–4 energy terms to internal energy terms. Per-residue, i.e. residue-by-residue, decomposition provides a quantitative assessment of each individual residue’s energetic contribution to the binding process – favorable, unfavorable, or neutral.

Energy values for the protein residues from the last 1 ns of simulation in each \(\lambda\)-window were saved, whereas water and ions were disregarded in subsequent analysis. Per-residue energy values were compared for the model and template ligands bound to Bcl-xL and Bcl-2 for the three replicas by computing Pearson correlation coefficients. When comparing ligands bound to the same protein, Pearson correlations from all
protein residues were computed. When comparing ligands across proteins, values for residues 25 – 137 for Bcl-xL and 27 – 139 for Bcl-2 were used. Helix 1, which is not involved in binding, was discarded, i.e. comparisons were only made between the core fold residues among the two proteins, which have a 1:1 correspondence in sequence and structure. Correlation plots were generated with the R corrplot package.90,91 Tanimoto similarities between the eight ligands from Figure 3, topological fingerprint similarities, and MACCS key similarities were computed with RDKit. MACCS keys are constructed via a molecular perception algorithm,92 which uses atom typing and molecular context to translate ligand structures into bit vectors, where the presence of a particular property, e.g. a para-substituent, is indicated with a 1 in a specified position in the vector and its absence is indicated with a 0; vector similarity, then, corresponds to ligand similarity, which can thus be quantified using the difference between the bit vectors. Tanimoto coefficients and topological fingerprints are related measures of similarity, also based on bit vectors of molecular properties and connectivity.57 Tanimoto, topological, and MACCS key similarities were then compared to per-residue energy patterns, obtained from free energy analysis.

4D QSAR

In addition to the 2D QSAR analysis performed on a large set of compounds, we performed 4D QSAR on compounds 3, 4, 5, 6, 7, and 8 (see Figures 2 and 3), using the per-residue energy values, obtained from the thermodynamic integration calculations. Analogously to the 2D QSAR analysis, the descriptors, which in this case are the energy values, were scaled to unit variance, and used to perform partial least squares regression, fitting the data to the specificity. Training was performed on compounds 3, 5, and 7, testing was performed on compounds 4, 6, and 8 (Venetoclax, ABT-199), and vice versa. To verify the robustness of our results, for each case, 20,000 random combinations of energy values from replicas 1, 2, and 3 for each compound in the test and training set were used. Partial least squares regression was performed using energy values for residues 25 – 137 for Bcl-xL and residues 27 – 139 for Bcl-2.

Results

2D QSAR
The initial QSAR model, derived from the entire training set, performed reasonably well in the external validation on the 47-compound test set, with an overall $R^2$ between predicted and measured specificity of 0.48 (Figure 4 A). Most dual binders were predicted accurately, as were most Bcl-xL-selective compounds. Compounds A, B, C, and D, which belong to different chemical series, lie close to the $x = y$ identity line (see Figure 2 and 4 B), as do compounds 3, 4, N3C (ABT-737), and 1XJ (ABT-263), which belong to the arylsulphoneamidoaryl series, which dominates the set. Arylsulphoneamidoaryl compounds 5 and 6 have their specificity slightly shifted up towards the dual binders, likely because arylsulphoneamidoaryl dual binders dominate the set. Compounds E, F, and G – a set of tetrahydroisoquinoline derivatives – are well predicted, as similar compounds are present in the training set. Only the Bcl-2-selective arylsulphoneamidoaryls are mispredicted by more than 3 units of specificity, eroding overall $R^2$ and reducing the slope of the line of best fit for the predictions. Randomizing the training and test sets 20,000 times produced similar results, with $R^2$ values varying around 0.5. The standard deviations of predicted specificities for most compounds were around 0.5 with the notable exception of the Bcl-2-specific binders, where predicted specificities varied by up to 2 log units. Large decreases in predictive performance were mostly associated with ligand randomizations where training was performed on compounds from one chemical series and used to perform predictions on a very different series of compounds. Performing the analysis with reduced training and test sets (Figure 4 C) improves the prediction accuracy for compound B, which belongs to a fairly sparsely represented chemical series. Again, compounds E, F, and G are accurately predicted, as compounds from this series are present in the training set. There is a slight downshift in predicted specificities for arylsulphoneamidoaryls, as evidenced by compounds 3, 4, 5, 6, 7, 8, 1XJ, and N3C, likely because dual arylsulphoneamidoaryl binders do not dominate the “normalized” data set. Moreover, fluctuations in $R^2$ become more pronounced, as one or two severely mispredicted compounds influence $R^2$ much more dramatically with the reduced test set. Again, only large, Bcl-2-specific arylsulphoneamidoaryls were severely mispredicted. Indeed, removing the two Bcl-2-selective arylsulphoneamidoaryls from the test sets improved $R^2$ of the initial models to 0.70 and 0.90 for the 45 and 13 compound sets, respectively.

Template structure simulations
100-ns MD equilibration simulations were performed in quadruplicate on the template compounds bound to Bcl-xL and Bcl-2, in order to relieve any strain potentially present in the complexes and to assess the stability of the structures (Figure 5). The proteins were stable over the course of the 100-ns trajectories, with Cα RMSDs fluctuating around a steady value below 1.5 Å. The ABT-263 ligand remained highly stable throughout all four replicas, whereas ABT-737 appears to have a greater mobility in the hydrophobic groove of the Bcl-xL protein. We prioritized proximity to the crystal structure conformations of the ligands, followed by low Cα RMSD values for the proteins. MM-PBSA energies were also monitored to further assess the energetics of binding; these indicated stable binding throughout the trajectories. Suitable snapshots from the simulations (see Figure 5) were chosen to serve as templates for modeling the different protein-ligand complexes. These models were subsequently subjected to thermodynamic integration calculations.

**Alchemical transformations**

We next performed thermodynamic integration calculations on the ABT-737 and ABT-263 compounds, bound to Bcl-xL and Bcl-2, transforming each of them into the six compounds in Figure 3 bound to the two proteins, in triplicate. We also performed the reverse calculations, again in triplicate. Thus, relative binding affinities (ΔΔGs) were calculated. Moreover, we also decoupled/inserted the template ligands from/into the binding pocket and free in solution to compute their absolute binding affinities (ΔGs). Autocorrelation times varied from several picoseconds to several hundred picoseconds among the different λ-windows. Thus, performing long enough simulations (> 50τ) to obtain a sufficient number of truly uncorrelated samples in all λ-windows was unfeasible. We, therefore, present ΔGs and convergence plots from δV/δλ values taken every picosecond; these differ marginally from results sampled 10 ps apart. The means of the free energies for the transformations and their standard deviations are shown in Figure 6; the corresponding convergence and Cα RMSD plots are presented in Figures S1 – S8. The proteins were stable throughout the 2.5 ns of dynamics in each λ-window, with RMSDs typically remaining below 1 Å. For most of the ligands, the results from the forward and backward transformations are in good agreement, being of similar magnitude but opposite sign (Figure 6). This indicates that we have attained sufficient sampling, and confirms the equivalence of the sander and pmemd molecular dynamics engines in terms of output. In most
replicas, for the relative transformations, the $\Delta \Delta G$ converged to a value near the experimental $\Delta \Delta G$ over the course of the simulations, in certain cases within “chemical accuracy,” i.e. within 1 kcal mol$^{-1}$ (see Figures S1, S2, S5 and S6). Only for the transformations where the ligands were inserted into or decoupled from the computational box were computed $\Delta G$ values off from experimental ones by a greater margin, in certain cases by more than 10 kcal mol$^{-1}$. Moreover, with the exception of ligand 8 in isolation, these were the only simulations with large differences in computed $\Delta G$s among replicas (standard deviations $>$ 5 kcal mol$^{-1}$, see Figure 6). The standard deviations of computed $(\Delta)\Delta G$ values among replicas are several-fold larger than $(\Delta)\Delta G$ errors within replicas. The standard deviations, therefore, are the values we present (Figure 6) – most are around 1 kcal/mol.

**Energetic fingerprinting**

We next compared the protein per-residue energies from the *sander* simulations when the two proteins are bound to different compounds, as well as between Bcl-xL and Bcl-2 when bound to the same compound. To facilitate comparison, for each compound in each of the energetic similarity panels (Figure 7), we prepend the number of the compound (3, 4, 5, 6, 7, and 8) with N3 or 1X to designate which protein the examined compound is bound to (the N3 and 1X prefixes derive from the N3C and 1XJ template ligands bound to Bcl-xL and Bcl-2, respectively). Examining the correlation plots of the Bcl-xL per-residue energies when the protein is bound to different ligands (Figure 7 A) reveals that the strong dual binders – compounds 3 and 4 – are energetically highly similar to each other and to the Bcl-xL selective compounds – 5 and 6 – which are also potent Bcl-xL binders. The pattern becomes intuitive by noticing the large squares in the energetic similarity panels, ranging from compound 3 (labeled as N33 in the energy panels) to compound 6 (labeled as N36). Conversely, the dual and Bcl-xL-selective ligands are energetically distinct from the Bcl-2-specific compounds, which are highly similar to each other (small squares in the bottom left of the panels). Furthermore, compound 7, which is the weakest Bcl-xL binder in the entire set, has the least energetic similarity to any of the other remaining compounds in two of the three replicas. This appears to be the general pattern for the Bcl-2 complexes as well (Figure 7 B), with results being highly consistent among the three replicas. Moreover, if we compare energetic patterns in between proteins, i.e. correlate the energies of residues 25 – 137 of Bcl-xL with 27 – 139 of Bcl-2, it becomes evident that the dual binders elicit similar
energetic patterns in the two proteins (see the highlighted diagonals in Figure 8 A). Conversely, the Bcl-xL-specific binders appear to evoke less similar energetic responses, whereas the Bcl-2-selective compounds, like the dual binders, also elicit similar responses in Bcl-xL and Bcl-2. If we compare energetic similarities to Tanimoto, topological, or MACCS key similarities – typical metrics of compound likeness in chemoinformatics – an interesting pattern emerges – dual binders (compounds 3 and 4) are structurally and energetically similar, Bcl-xL-selective binders (5 and 6) are structurally similar, but energetically distinct, whereas Bcl-2-selective ligands (7 and 8) are structurally different, but energetically similar (compare the three squares lying on the highlighted diagonals in Figure 8 A). Finally, Bcl-xL-selective compounds are energetically least similar to Bcl-2-selective compounds than to any other compound in the dataset (see the off-diagonal squares in Figure 8 A).

4D QSAR

Training on per-residue energy values from MD simulations to predict ligand specificity demonstrates that the Bcl-2-selective ligands are predicted within 2 log units of experimental values; dual binders lie within or close to the square, delimited by the [-1,1] range; and Bcl-xL-specific binders are either well predicted or have slightly upshifted specificities, depending on the choice of test and training sets (Figure 9). Moreover, results are consistent over the two 20,000-interreplica randomizations, with mostly small standard deviations of predicted specificities in the range ~0.1 – 0.5.

Discussion

The antiapoptotic proteins remained stable throughout the 100-ns trajectories, in agreement with extensive previous simulations. The ABT-263 template compound remained stably bound to the Bcl-2 protein, whereas the highly similar ABT-737 template displayed greater mobility, particularly the fragment that fits into pocket 2 of the protein. This appears to be due to the topology of the Bcl-xL groove around the biphenyl fragment of ABT-737, which is shallower than that of Bcl-2, in part due to the A104D and S122R amino acid substitutions (nomenclature is Bcl-xL → Bcl-2; numbering corresponds to the canonical Bcl-xL sequence). Pocket 2 (labeled “p2” in Figure 1) accommodates the biphenyl fragment of ABT-737 and its
corresponding moieties in other ligands. Loss of this interaction, in turn, facilitates the loss of the intramolecular π-stacking interaction between the thiophenyl and nitrophenyl moieties in pocket 4 (labeled “p4” in Figure 1) – a characteristic feature of the thiophenyl-bearing sulphoneamidoaryl class of Bcl-inhibitors, observed in multiple X-ray and NMR structures. Given the relative shallowness of the Bcl-xL rim around p2 and the greater mobility of the template compound bound to Bcl-xL, observed in our simulations, we expect compounds of similar affinity for Bcl-xL and Bcl-2 to display greater $k_{\text{off}}$ rates for Bcl-xL than Bcl-2.

We have presented thermodynamic integration free energy calculations on large, complex drug and drug-like molecules. We find it reassuring that most of the relative transformations approach a value within ~2 kcal mol$^{-1}$ from the experimental $\Delta\Delta G$ (see Figure 6 B and the free energy convergence plots in Figures S1, S2, S5 and S6). As may be expected, only the absolute $\Delta G$ calculations are off by a greater margin, sometimes more than 10 kcal mol$^{-1}$. In part, this likely arises from force field inaccuracies. Added to this, insertion or deletion of the ligand is a much greater transformation than transforming the template compounds into the models or vice versa, and thus necessitates more sampling to alleviate any hysteresis. Finally, the alchemical transformations we have performed involve charged species. This introduces an error of a complex nature, involving periodicity-induced net-charge interactions, periodicity-induced net-charge under-solvation, discrete solvent effects, and residual integrated potential effects. This error is likely to be compounded by the circumstance that we are transforming the template ligands N3C and 1XJ, which have a net charge of +2, into the model ligands, which all have a net charge of +1, or vice versa. As all template and model ligands have a positive net charge, the error stemming from ligand charge is likely to largely cancel in the relative transformations, and judging by the respective free energy convergence plots (see Figures S1, S2, S5 and S6) it does. However, in the absolute free energy calculations, the transformed state has a charge of 0 (when the template ligand is decoupled from the simulation box) or +2 (when it is inserted into the simulation box), i.e. the change in charge in the template compound simulations (+2 to 0 or 0 to +2) is different from the change in charge in the relative transformations (+2 to +1 or +1 to +2), which diminishes error cancelation in the template simulations. Indeed, the $\Delta\Delta G$ curves converge to a value that is close to the experimental value (the purple horizontal lines in Figures S1, S2, S5 and S6) for compounds 3, 4, 5, 6, 7, and 8, but are off by a greater margin for the template compounds (N3C and 1XJ). Moreover, since all of the
template-to-model (and vice versa) transformations involve the same change in ligand charge (+2 to +1 and
+1 to +2, respectively), one can confidently compare these transformations among each other. In other
case, for any two of the model compounds, in a situation where experimental affinity measurements are not
available, for example, one would compare the calculated ∆∆Gs (as they are a measure of predicted affinity)
and subtract one ∆∆G from the other to discern which compound has the greater affinity for the target
protein. As all template-to-model and model-to-template transformations contain largely the same error
stemming from the change in ligand charge, it will mostly cancel during the subtraction, i.e. comparison.
However, the template compound simulations, where the ligand is decoupled or inserted from/into the
simulation box, involve a different charge transformation (+2 to 0 or 0 to +2), making it more difficult to
compare them to the rest of the simulations. This is a well established shortcoming of MD – small
perturbations can be assessed more reliably than large ones, relative energies can be computed much more
accurately than absolute ones. In this work, we have shown that large and complex drug molecules can be
reliably parameterized with a general force field, such as GAFF, and a rapid method for charge estimation,
such as AM1-BCC. Absolute binding energy calculations of charged species, however, necessitate further
theoretical and methodological developments.36,104-106

The foregoing error also influences the per-residue values obtained from energy decomposition. To
assess its influence, we also performed simulations on compounds 5 and 6 in the +2 protonation state (see
Figure 8 B). This state is less likely because the piperazinyl nitrogen is bound to electron withdrawing
groups. Moreover, protonation of the nitrogen atom stabilizes it in a tetrahedral geometry, which facilitated
inversions in the terminal moieties where the phenyl groups pointed outward into the solvent and the polar
atoms pointed inward into the hydrophobic groove. Such an arrangement seems unlikely. Conversely, the +1
protonation state stabilized the ligands in the groove. In the case of Bcl-xL, coupled with the relative
shallowness of the groove around pocket 2, it also increased the propensity of the terminal moieties of
ligands 5 and 6 to adopt an extended conformation along the surface of the groove. Such an arrangement has
been observed crystallographically in similar compounds bound to Bcl-xL.31 Nevertheless, simulating
compounds 5 and 6 in the +2 state reveals an interesting aspect of free energy calculations. It appears that the
charging/decharging error strongly influences the per-residue values. These are the only simulations where
the ligand net charge does not change during the alchemical transformation. The per-residue values are quite
different from the rest of the simulations, where one of the ligand end states has a charge of +2 or 0, the other
+1 or 0 (0 when the ligand is decoupled from the system ($\lambda = 1$) or inserted into the system ($\lambda = 0$)); this is
the only instance where anti-correlations appear (Figure 8 B). Since the error resulting from the change in
charge is present in all other transformations, we expect the energetic patterns reported herein to be bona
fide, particularly given the reproducibility across three independent replicas.

QSAR models derived from the arylsulfonamidearyl series of compounds performed well when
predicting properties of other arylsulfonamidearyls. Similarly, training on the tetrahydroisoquinolines
produced predictive models for other tetrahydroisoquinolines. However, training on one series of compounds
to predict the properties of another unsurprisingly leads to poor results. Moreover, Bcl-2-specific
arylsulfonamidearyls, of which there are only two in our entire dataset, are poorly predicted by the present
models. This is a well-known shortcoming of traditional QSAR models – they are much better at
interpolation than extrapolation$^{107,108}$ – and even the most sophisticated QSAR models may perform poorly
on challenging targets$^{109}$. Moreover, predictive performance is sensitive to the specific choice of test and
training compounds, very much so in the case of small datasets. Quite often, QSAR models suffer from
“activity cliffs”$^{110}$ where a small change in a compound, such as introduction or removal of a hydroxyl or a
methyl group, translates into a large change in activity. Most models, based on 2D descriptors, or 2D
measures of similarity, such as Tanimoto coefficients or MACCS keys, would experience little change upon
such an alteration in a large molecule. It appears that two-dimensional models and descriptors simply do not
contain in themselves all the information relevant to ligand activity. Similar findings have been reported
previously$^{111,112}$. In hindsight, this is not surprising – the descriptors used to construct low-dimensional
QSAR models are, by the very nature of the technique, invariant with respect to the macromolecule – a
major determinant of the binding process; they are also invariant with respect to time. Moreover, most 0-, 1-, and 2-dimensional descriptors do not discriminate between enantiomers. In stark contrast, free energy
calculations and energy decomposition should, at least in principle, provide a detailed understanding of the
binding process. Distinguishing enantiomers, for example, becomes trivial by the nature of the technique –
where one enantiomer makes favorable interactions, and the other does not, there will exist a difference in
potentials between the two systems, a difference that will manifest itself in a computationally detectable
$\Delta\Delta G$. Moreover, the difference in residue – ligand contacts should, again at least in principle, be evident on a
per-residue level; finer-grained decomposition is also possible.\textsuperscript{113}

Our work corroborates this line of reasoning. Despite the lower training:test set ratio in the 4D versus the 2D QSAR analysis (1:1 vs 3:1, respectively), the 4D models performed considerably better at estimating the activities for the challenging Bcl-2-selective binders, outperforming 2D models by around 1 log unit of specificity, with 2- to 10-fold lower standard deviations among the 20,000 combinations. Furthermore, while an analogous analysis on the same set of compounds using 2D descriptors exhibits a comparable performance in terms of absolute error (compare Figure 9 and S9), crucially, it again mispredicts the Bcl-2-selective compounds as dual binders. These results indicate that per-residue energies are a much more suitable, albeit more expensive, descriptor of activity. Indeed, 2D QSAR builds predictive models based on properties such as molecular weight, net charge, logP, the number of aliphatic or aromatic carbons in a compound, and molecular connectivity or branching\textsuperscript{57} – typically some linear or nonlinear combination thereof. However, these descriptors are too degenerate and too coarse to capture the fine difference in binding between paralogs. In stark contrast, molecular dynamics and free energy calculations provide a detailed, atomistic level of description of binding and energetics. It would then seem likely that the per-residue energies are a more fine-grained descriptor, which affords a higher-resolution account of the binding process. Indeed, atomistic simulation and decomposition analysis have already been successfully employed to unravel the complex binding patterns in the Bcl-2-family, explaining why certain protein – peptide pairs bind with high affinity, whereas others do not.\textsuperscript{17} Moreover, in contrast to low-level QSAR, properties computed from simulations are time-dependent. This is another key asset of MD, as monitoring RMSDs and free energies over time allows detecting the point of convergence in simulations of off-equilibrium models, generated in the absence of an X-ray or NMR structure.

We also note the remarkable accuracy of predicting specificities using the +2 protonation state for ligands 5 and 6 (Figure S10). Thus, despite the considerations we have discussed previously, we cannot completely rule out other factors perturbing the pK\textsubscript{a}'s of the ligands, stabilizing the +2 state. In such a scenario, the energetic pattern previously discussed still holds – the dual binders (3 and 4) are structurally and energetically similar; Bcl-xL-selective binders (5 and 6) are structurally similar, but energetically distinct; and, finally, Bcl-2-selective ligands (7 and 8) are structurally different, but energetically similar (see the correlation coefficients in Figure 8 B). In the +2 state, compounds 5 and 6 have a unique energetic
fingerprint, setting them apart from each other and the rest of the ligands, making specificity predictions using per-residue energies highly accurate – compounds 3, 4, 5, and 6 lie almost ideally on the identity line, with only 7 and 8 deviating more substantially from it – likely a consequence of the disparity in affinity for Bcl-2 between the training and test compound for the Bcl-2-selectives.

With Figure 9 and Figures S9 – S10, we provide a direct and fair comparison between the capabilities of 2D descriptors and models with 4D ones on the same set of compounds. While $R^2$ values from just a few data points are not informative in themselves, the key conclusion to be drawn from these plots and the plots in Figure 4 is that 2D models cannot correctly predict the selectivity of the Bcl-2-specific compounds, whereas 4D models are quite successful – they reproduce the correct trend in selectivity across the entire spectrum of compounds, with rather small quantitative errors for the Bcl-2-selectives. It is also instructive to consider the applicability domain of the models we present. One simple way of estimating the applicability domain is to use the standard deviation of predictions from multiple models (in our case - the 20,000 randomizations) - small variations indicate that the models are robust and being used within the boundaries of their applicability domain; large variations suggest otherwise. Using this metric, it becomes apparent that the applicability domain of the 2D QSAR models spans from the Bcl-xL-selective compounds up to and including the dual binders; it starts from the lower regions of the specificity axis and ends in the region around +1. Conversely, 4D models extend the applicability domain beyond +1, well into Bcl-2-specific territory. Indeed, comparison between the predicted and experimental values confirms this reasoning.

Other studies have also demonstrated that QSAR models based on ligand properties computed from MD trajectories can help to better predict ligand biological activity and resolve activity cliffs compared to 2D- and 3D-based models. What is particularly noteworthy about the study by Ash and Fourches on a set of kinase inhibitors is that for some of the descriptors, it is not the means of the descriptors that correlate with biological activity, but their standard deviations over the course of the simulations, with more active ligands displaying greater variance. The authors hypothesize that this is a result of the ligands adapting to the dynamic kinase pocket, and demonstrate how 4D QSAR models add an extra layer of information content over 3D models. Perhaps most important of all, they show that 4D models are capable of yielding information and features that no lower-level model is capable of providing. Furthermore, dynamics-based
pipelines such as *ensemble* Comparative Residue-Interaction Analysis (eCORIA\textsuperscript{115}) allow one to quantify the imperative properties and interactions in the binding process, which enhances the interpretability of modern QSAR models – an increasing requirement in legislation regulating the development of pharmaceuticals, agrochemicals, and cosmetics.

While characterizing a ligand's dynamic behavior in complex with a biological macromolecule is a prerequisite to fully understanding binding, understanding the macromolecule's behavior is also essential. In the case of Bcl-2-selective arylsulfonamidearyls, for instance, traditional QSAR models would perform poorly when predicting their specificity, as most presently available arylsulfonamidearyls are dual binders or Bcl-xL selective. MD simulations and free energy calculations, however, demonstrate that the Bcl-xL and Bcl-2 proteins differentiate between these ligands from dual and Bcl-xL selective binders, as evident from our current analysis, which can be viewed as an assessment of a protein's energetic response to a particular binding partner. Including the water molecules and ions in the per-residue analysis in such calculations would enable detecting bridging\textsuperscript{116} or trapped\textsuperscript{117} waters/ions, which are often important, but neglected, factors in the binding process. This may mitigate the need for GIST\textsuperscript{-118} or 3D-RISM-type calculations,\textsuperscript{119} and reduce the human workload, for example in a lead-optimization campaign, at no additional computational cost.

**Conclusions**

A means of characterizing protein – ligand/peptide/protein interactions is presented and used to demonstrate how macromolecules energetically distinguish between different binding partners. This work paves the way for developing and validating more predictive models and descriptors, based on per residue energies. Such energy-based four-dimensional descriptors\textsuperscript{120,121} would constitute a composite of multiple lower-level descriptors, encompassing in themselves aspects of the binding process such as steric and atom-type propensities for given ligands at given receptor locations, e.g. polar groups with positive partial charges in the vicinity of a key hot spot or specificity determinant.\textsuperscript{17,122} In the short-to-medium term, 4D QSAR and MD remain less accessible than 2D QSAR, as the former is several orders of magnitude more expensive. 2D QSAR requires seconds to minutes to perform on any personal computer, whereas simulations such as those reported here require several weeks of compute time on a large high-performance cluster. Furthermore, the
pre- and post-processing of thermodynamic integration simulations are labor intensive tasks that require
specialist knowledge. However, even at the present time, MD simulations and 4D QSAR have considerable
merits, which outweigh their greater expense. For example, they can steer hit-to-lead and lead-optimization
campaigns in an academic or industrial setting in a way that 2D QSAR cannot. We have shown that MD-
based QSAR outperforms 2D QSAR in predicting ligand selectivity, despite the lower training:test set ratio.
This holds not only in quantitative (absolute error and $R^2$) but, more importantly, in qualitative terms –
QSAR models based on low-level descriptors cannot predict the selectivity of compounds 7 and 8 for Bcl-2,
whereas energy-based models are quite successful at predicting specificity over the entire spectrum – from
Bcl-xL-selectives, through dual binders, to Bcl-2-specific molecules (compare Figures 4, 9, S9, and S10).
Indeed, MD- and ensemble-based pipelines extend the applicability domain of models well beyond what 2D
QSAR can provide. Remarkably, our 4D models correctly predict the selectivity of compound 8. This is
notable because compound 8 (Venetoclax, ABT-199) is a highly challenging target for traditional 2D QSAR
– it is the most Bcl-2-selective compound known to date. Given the constant ongoing improvements in force
fields,123 automation of free energy calculations and construction of optimal thermodynamic cycles,79,124 and
the increase in computing power at an ever decreasing cost, it is to be expected that higher-dimensional
descriptors and models will become more prevalent, at the expense of lower-level descriptors and models.

Supporting Information.
Free-energy convergence and RMSDs for alchemical calculations, and supplementary 2D/4D QSAR data
(PDF).

Funding sources
SMI is an A*STAR Research Attachment Programme (ARAP) studentship holder, which was awarded to
JW and PJB by the A*STAR Graduate Academy (A*GA) Singapore.

ORCID iDs:
The authors declare no conflict of interest.

We gratefully acknowledge James Baker and Nicholas Fowler for the many insightful discussions. This work made use of the facilities of the ARCHER UK National Supercomputing Service (www.archer.ac.uk); the N8 HPC Centre of Excellence, provided and funded by the N8 consortium and EPSRC (Grant No.EP/K000225/1), and coordinated by the Universities of Leeds and Manchester; and the Computational Shared Facility (CSF) at The University of Manchester. We thank the A*STAR Graduate Academy (A*GA) Singapore for funding.
References


21. Ivanov, S.; Dimitrov, I.; Doytchinova, I. Quantitative Prediction of Peptide Binding to HLA-DP1


43. Brady, R.M.; Vom, A.; Roy, M.J.; Toohey, N.; Smith, B.J.; Moss, R.M.; Hatzis, E.; Huang, D.C.; Parisot, J.P.; Yang, H.; Street, I.P.; Colman, P.M.; Czabotar, P.E.; Baell, J.B.; Lessene, G. De-Novo Designed Library of Benzoyleureas as Inhibitors of BCL-XL: Synthesis, Structural and Biochemical


66. Brandsdal, B.O.; Österberg, F.; Almlöf, M.; Freierberg, I.; Luzhkov, V.B.; Åqvist, J. Free Energy...


766 80. Schrödinger Release 2017-2: Maestro, Schrödinger, LLC, New York, NY, **2017**.


110. Cruz-Monteagudo, M.; Medina-Franco, J.L.; Perez-Castillo, Y.; Nicolotti, O.; Cordeiro, M.N.D.S.;


120. Pan, D.; Tseng, Y.; Hopfinger, A.J. Quantitative Structure-Based Design: Formalism and Application of Receptor-Dependent RD-4D-QSAR Analysis to a Set of Glucose Analogue Inhibitors of


Figure 1. Crystal structures of Bcl-xL and Bcl-2 bound to small molecule ligands. Bcl-xL bound to ABT-737 (left, PDB ID 2YXJ) and Bcl-2 bound to Navitoclax (ABT-263) (right, PDB ID 4LVT). The proteins are shown in surface representation, the ligands are in stick representation with carbon atoms in white, oxygen in red, nitrogen in blue, sulfur in yellow, chlorine in green, and fluorine in gray. Deep hydrophobic pockets on the surface of the proteins, occupied by ligand moieties, are labeled with p2 and p4.

Figure 2. Compounds included in QSAR analysis. All compounds discussed in this work plotted by experimentally measured specificity, molecular weight and computed logP (clogP). For clarity, Bcl-xL-specific ligands, dual binders, and Bcl-2-specific compounds are colored differently and the specificity region between -1 and 1 is highlighted. Compounds A, B, C, D, E, F,G, N3C (ABT-737), 1XJ (Navitoclax, ABT-263), and 3, 4, 5, 6, 7, and 8 (Venetoclax, ABT-199) are labeled, shown in bold, and discussed further in the text. N3C’s specificity was computed from IC$_{50}$ values reported by Sleebs et al.$^{35}$ As the affinities of compounds 5 and 8 for Bcl-xL and Bcl-2 were beyond the measuring sensitivity of the experimental setup ($K_i < 1$ nM,$^{31}$ $K_i < 0.01$ nM,$^8$ respectively), provisional specificities for those compounds were computed using values of 1 and 0.01 nM, respectively (also see Figure 3).

Figure 3. Structures of template and modeled compounds. Chemical formulae of the six model compounds (labeled 3, 4, 5, 6, 7, and 8) and the N3C (ABT-737) and 1XJ (Navitoclax, ABT-263) template compounds used in molecular dynamics simulations. Shown above every compound is its $K_i$ value (in nM) for Bcl-xL, followed by the $K_i$ value for Bcl-2 (in nM). Compounds 3, 4, N3C, and 1XJ are potent dual binders, compounds 5 and 6 are Bcl-xL selective, compounds 7 and 8 are Bcl-2 selective. The exact affinities of compounds 5 for Bcl-xL and N3C and 8 for Bcl-2 are above the measurement limit of the experimental procedures. Affinity data on compounds 3, 4, 5, and 6 is from Bruncko et al.$^{31}$ 7, 8 (Venetoclax, ABT-199), 1XJ, and N3C is from Souers et al.$^8$ IC$_{50}$ values for N3C (3/6.1 nM) are also reported in Sleebs et al.$^{35}$ Chiral carbon atoms marked with a black asterisk (*) were modeled in the R-configuration; such were the compounds used in the experimental affinity measurements. The chiral atom in compound 4, marked with a red asterisk (*), was modeled in the S-configuration; its configuration is not specified in the relevant publication. Thus, the S-configuration was chosen, as it shields the -CH$_3$ group from solvent to a greater
degree than the R-configuration. The nitrogen atoms in compounds 5 and 6, marked with a blue asterisk (*), were modeled in both the protonated and unprotonated state. Henceforth, results reported for compounds 5 and 6 correspond to the state where these nitrogens are unprotonated (ligand net charge +1), except where explicitly stated otherwise. The canonical groove of the antiapoptotic proteins lies below the piperazinyl and sulfonamidophenyl rings. Compounds 3, 4, 5, 6, 7, and 8 had their terminal phenyl groups oriented in pocket 2 of the antiapoptotic proteins (see Figure 1), and N and O atoms pointing outward into the solvent, with the exception of the spiro nitrogen of compound 4, which was also buried in the groove. The configurations of the spiro moieties of compounds 3 and 4 were modeled to be identical to the ones reported in the experimental study.31

**Figure 4. 2D QSAR results.** (A) R^2 between predicted and experimental specificity of the compounds in the external validation test sets plotted against the iteration number for the 20,000 iterations. N3C’s specificity was computed from IC_{50} values reported in Sleebs et al.,35 specificities for compounds 5 and 8 were computed using K_i values of 1.3^1 and 0.01^{18} nM for Bcl-xL and Bcl-2, respectively (see Figure 3). (B) Predicted specificity plotted against experimental values for the 47 compound test set. Values for compounds without marked standard deviations are from the first QSAR model, values for compounds 3, 4, 5, 6, 7, and 8, N3C, and 1XJ are averages over the 20,000 iterations; these do not differ significantly from values from the first model; also shown is the standard deviation of predicted specificities for these compounds over the 20,000 iterations. The x = y line is shown in purple, also shown is the line of best fit from the initial model; the corresponding line equation and R^2 are in the top left corner. The region between -1 and 1 predicted vs experimental specificity is highlighted with a square box. (C) The same plot as in (B) prepared for the 15 compound test set.

**Figure 5. Results from template equilibration simulations.** Cα RMSDs (top), ligand heavy atom RMSDs (middle), and computed MM-PBSA enthalpies of interaction (bottom) for the four 100-ns replicas of the Bcl-xL – ABT-737 (left) and Bcl-2 – ABT-263 (right) template structures. The vertical black lines indicate the time points chosen to be templates for the free energy calculations. Snapshots at t = 30 ns and t =
21 ns from replicas 2 and 1 were chosen for the Bcl-xL – ABT-737 and Bcl-2 – ABT-263 complexes, respectively.

Figure 6. Calculated free energies. (A) Computed ∆Gs from the 11 λ-window alchemical transformations, averaged over the three replicas; standard deviations are also shown. Ligand transformations in complex with Bcl-xL or Bcl-2 are labeled “CMP,” transformations where the ligand is free in solution are labeled “LIG.” ∆Gs were evaluated from 0 – 1500 ps and from 1500 to 2500 ps of production dynamics. For ease of comparison, pmemd results (labeled “Forward”) are multiplied by -1. Compounds 3, 4, 5, 6, 7, and 8 are prefixed with “N3” and “1X” to designate transformations from/to the N3C and 1XJ templates, respectively, i.e. N33 and 1X3 stand for transformations involving compound 3 using the N3C and 1XJ templates, analogously for N34 and 1X4, etc. (B) Computed ∆∆Gs from the 11 λ-window alchemical transformations, averaged over the three replicas; standard deviations are also shown. Differences between ligand transformations in complex with Bcl-xL or Bcl-2 (labeled “CMP” in panel A) and transformations where the ligand is free in solution (labeled “LIG” in panel A) are labeled as “CMP - LIG,” also shown is the experimentally measured free energy difference (last bar for every compound). ∆∆Gs were evaluated from 0 – 1500 ps and from 1500 to 2500 ps of production dynamics; computed values are directly comparable to experimental data. For ease of comparison, pmemd results (labeled “Forward”) are multiplied by -1. Compounds 3, 4, 5, 6, 7, and 8 are prefixed with “N3” and “1X” to designate transformations from/to the N3C and 1XJ templates, respectively, i.e. N33 and 1X3 stand for transformations involving compound 3 using the N3C and 1XJ templates, analogously for N34 and 1X4, etc.

Figure 7. Pearson correlations between per-residue energy values. (A) Pearson correlation coefficients between the per-residue energy values of Bcl-xL for the three replicas of the sander transformations (top row), and ligand Tanimoto, topological fingerprint, and MACCS key similarities (bottom row). Compounds 3, 4, 5, 6, 7, and 8 are prefixed with “N3” to designate transformations to the N3C template for the alchemical transformations, i.e. N33 stands for transformations involving compound 3 using the N3C template, analogously for N34, N35, etc.; correlations have been calculated from all protein residues using energy values from the latter 1 ns of production dynamics in every λ-window. (B) Pearson correlation
coefficients between the per-residue energy values of Bcl-2 for the three replicas of the *sander* transformations (top row), and ligand Tanimoto, topological fingerprint, and MACCS key similarities (bottom row). Compounds 3, 4, 5, 6, 7, and 8 are prefixed with “1X” to designate transformations to the 1XJ template for the alchemical transformations, i.e. 1X3 stands for transformations involving compound 3 using the 1XJ template, analogously for 1X4, 1X5, etc.; correlations have been calculated from all protein residues using energy values from the latter 1 ns of production dynamics in every $\lambda$-window.

**Figure 8. Pearson correlations between per-residue energy values.** (A) Pearson correlation coefficients between the per-residue energy values of Bcl-xL and Bcl-2 for the three replicas of the *sander* transformations. Compounds 3, 4, 5, 6, 7, and 8 are prefixed with “N3” or “1X” to designate transformations to the N3C and 1XJ templates, respectively, i.e. N33 and 1X3 stand for transformations involving compound 3 using the N3C and 1XJ templates, analogously for N34 and 1X4, etc.; correlations have been calculated from Bcl-xL residues 25 – 137 and Bcl-2 residues 27 – 139 using energy values from the latter 1 ns of production dynamics in every $\lambda$-window. The energetic responses of Bcl-xL and Bcl-2 when bound to the same ligand are compared (highlighted diagonals), as are the energetic responses between Bcl-xL and Bcl-2 bound to compounds 3 and 4, 5 and 6, 7 and 8 (squares lying on the diagonals), and compounds 5 and 6, and 7 and 8 (off-diagonal squares). (B) Pearson correlation coefficients between the per-residue energy values of Bcl-xL and Bcl-2 for the three replicas of the *sander* transformations. Compounds 3, 4, 5, 6, 7, and 8 are prefixed with “N3” or “1X” to designate transformations to the N3C and 1XJ templates, respectively, i.e. N33 and 1X3 stand for transformations involving compound 3 using the N3C and 1XJ templates, analogously for N34 and 1X4, etc.; correlations have been calculated from Bcl-xL residues 25 – 137 and Bcl-2 residues 27 – 139 using energy values from the latter 1 ns of production dynamics in every $\lambda$-window. Apart from the +1 protonation state, reported heretofore, compounds 5 and 6 are also examined in the +2 protonation state, where the piperazynyl nitrogen (see Figure 3) is protonated. The +1 and +2 states are designated with a corresponding suffix in this panel.

**Figure 9. 4D QSAR results.** (A) $R^2$ between predicted and experimental specificity of the compounds in the external validation test sets plotted against the iteration number for the 20,000 iterations. Experimental
specificities for compounds 5 and 8 were computed using $K_i$ values of $1.31 \times 10^{-3}$ nM for Bcl-xL and Bcl-2, respectively (see Figure 3). (B) Predicted specificity plotted against experimental values; training set consists of compounds 3, 5, and 7; test set includes ligands 4, 6, and 8; per-residue energy values for compounds 5 and 6 are from the +1 protonation state. Note that the training and test sets contain separate energy entries for the compounds bound to Bcl-xL and Bcl-2. Consequently, outputted predictions have two data points per compound; these are sufficiently close for meaningful analysis. This pertains to panels A and C as well. The values shown for the predicted specificities are averages over the 20,000 iterations; also shown are the standard deviations. The $x = y$ line is shown in purple, also shown is the line of best fit; the corresponding line equation and $R^2$ are in the top left corner. The region between -1 and 1 predicted vs experimental specificity is highlighted with a square box. (C) The same plot as in (B) prepared using ligands 4, 6, and 8 as the training set, and ligands 3, 5, and 7 as the test set; per-residue energy values for compounds 5 and 6 are from the +1 protonation state.