# Energetics and Dynamics Across the Bcl-2-Regulated Apoptotic Pathway Reveal Distinct Evolutionary Determinants of Specificity and Affinity

**Document Version**
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

**OPEN ACCESS**

**Energetics and Dynamics Across the Bcl-2-Family-Dependent Apoptosis Pathway Reveal Distinct Evolutionary Determinants of Specificity and Affinity**

Stefan M. Ivanov[1,2], Roland G. Huber[1], Jim Warwicker[2], Peter J. Bond[1,3*]

1 Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Matrix 07-01, 30 Biopolis Street, 138671 Singapore

2 Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

3 Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543

*corresponding author: peterjb@bii.a-star.edu.sg (Lead Contact)

## Summary

Critical regulatory pathways are replete with instances of intra- and interfamily protein-protein interactions due to the pervasiveness of gene duplication throughout evolution. Discerning the specificity determinants within these systems has proven a challenging task. Here, we present an energetic analysis of the specificity determinants within the Bcl-2 family of proteins – key regulators of the intrinsic apoptotic pathway – via a total of ~20 µs of simulation of 60 distinct protein-protein complexes. We demonstrate where affinity and specificity of protein-protein interactions arise across the family, and corroborate our conclusions with extensive experimental evidence. We identify energy and specificity hotspots, which may offer valuable guidance in the design of targeted therapeutics for manipulating the protein-protein interactions within the apoptosis-regulating pathway. Moreover, we propose a conceptual framework that allows us to quantify the relationship between sequence, structure and binding energetics. This approach may represent a general methodology for investigating other paralogous protein-protein interaction sites.

**Introduction**

Most proteins belong to families of evolutionarily and functionally related molecules, often arising from gene duplication (Friedman & Hughes, 2001). A classic example of such paralagous proteins are the human kinases, numbering over 500 (Manning et al., 2002). The specificity of biological pathways is thus striking, considering the thousands of potentially interacting macromolecules in a cell at any given time (Berggård et al., 2007). In general, protein interaction sites consist of tightly packed, structurally conserved regions or "hotspots" (Shoemaker & Panchenko, 2007; Ma et al., 2003). Hotspots tend to be enriched in tryptophan, tyrosine and arginine (Ma et al., 2003), and the most frequent residue pairs in the associated protein–protein complexes involve charged and aromatic residues (Gromiha et al., 2011; Gromiha et al., 2009). It has been suggested that polar residues at the interface cores confer rigidity, reducing the entropic loss upon binding, while the surrounding residues may form a "flexible cushion." A study of paralogous protein interfaces led to the proposal that binding affinity is determined mainly at the hub, whereas specificity is determined at the rim. Specificity between paralogs diverges at greatly differing rates, while interfaces evolve more slowly then the rest of the protein (Aiello & Caffrey, 2012). Explaining specificity within families of paralogs is particularly challenging, given that they usually share a common, conserved interface based on a conserved scaffold, for both interacting and non-interacting pairs (van Wijk et al., 2009; Kar et al., 2012).

In this work, we focus on the mechanisms by which a protein selects binding partners from a pool of closely related candidates. We begin with the assumption that the decisive factors determining binding versus non-binding in paralagous protein pairs are alterations in and around a common scaffold. We have opted to focus on the B-cell lymphoma-2 (Bcl-2) family of proteins, due to its physiological and clinical importance, as well as the abundance of structural and interaction data (Chen et al., 2005). The intrafamily interactions among Bcl-2-like proteins determine whether a cell undergoes apoptosis (Cory et al., 2003). The Bcl-2

family encompasses the antiapoptotic molecules Bcl-x$_L$, Bcl-2, Bcl-w, Mcl-1 and A1 (Cheng et al., 2001), and ~15 proapoptotic members. The antiapoptotic proteins have four Bcl-2 homology (BH) regions (BH1-4), as do the proapoptotic Bax and Bak (Kvansakul et al., 2008), which constitute a separate, Bax-like, subfamily. Most proapoptotic members (e.g., Noxa, Hrk, Bid, Puma, Bmf, Bik, and Bim) belong to the BH3-only subfamily (Happo et al., 2012).

Bax, Bak, and the antiapoptotic proteins consist of 7 or 8 amphipathic α-helices, clustered around a central hydrophobic α-helix (Suzuki et al., 2000) forming an exposed hydrophobic groove for binding the BH3 domain of proapoptotic proteins (Figure 1) (Petros et al., 2004). The core fold has 85-95 % structural overlap (Nguyen et al., 2011) across deposited structures in the PDB (Berman et al., 2000), and contains highly conserved regions including an invariant NWGR motif at the beginning of helix 5 (Day et al., 2008), and a conserved hydrophobic core which maintains the tryptophan in its position (Figure 1A).

In preapoptotic cells, the BH3 domains of proapoptotic Bak and Bax (Shamas-Din et al., 2011) are bound to the hydrophobic groove on the surface of the antiapoptotic proteins, rendering them inactive (Stewart et al., 2010). When an apoptosis signal reaches the cell, BH3-only proteins outcompete Bak and Bax for their antiapoptotic partners, freeing the formers' BH3 domains, which are then involved in homo- and possibly heterodimerization via a BH3 domain – hydrophobic groove interaction, leading to oligomeric pore formation in the outer mitochondrial membrane and subsequent apopotosis (Happo et al., 2012).

The binding mode between different pairs of proteins within the system is highly similar (Day et al., 2008; Smits et al., 2008; Czabotar et al., 2007): All BH3 peptides have four hydrophobic residues (positions 8, 12, 15, and 19, cf. Figure 1E) that fit into four hydrophobic pockets (labeled p1 – p4, see Figure 1B) on the surface of the groove, whilst an absolutely conserved aspartic acid (position 17) in the proapoptotic proteins forms a salt bridge with the arginine of the NWGR motif. Nevertheless, the affinities between the different BH3 peptides and the five antiapoptotic proteins span more than four orders of magnitude – from IC$_{50}$

values below 5 nM to >100 μM(Chen et al., 2005). As all antiapoptotic proteins in a cell must be neutralized for it to undergo apoptosis and not all BH3 peptides are omnibinders, their binding selectivity has implications for peptide-micking drugs that target this interaction (Czabotar et al., 2014).

In order to elucidate the origins of affinity and specificity across a paralogous set of interacting and non-interacting pairs, we now report a computational study of the Bcl-2 family. Guided by a dataset of experimentally measured binding affinities, we have modeled a total of 60 different complexes (see Table 1 and Figure 1) of BH3 peptides (or "ligands") onto templates of peptide-bound antiapoptotic proteins (or "receptors"). For each complex, and also for the constituent isolated ligands/receptors, triplicate MD simulations were carried out (amounting to 180 x 100-ns complex trajectories, 15 x 100-ns receptor trajectories, and 39 x 100-ns ligand trajectories), enabling accurate calculation of the enthalpies of each protein-protein interaction and decomposition on a per-residue basis. We demonstrate that in the antiapoptotic proteins, pockets provide affinity, but not specificity. Energetic recognition patterns are shown to be the most adaptable feature in a hierarchy of structure, sequence and energy conservation. We posit that the groove – BH3 helix case discussed here may be representative of a pattern on the relationship between structure, sequence, and binding energetics in protein families, and present a method to characterize energy and specificity hotspots that can be utilized in targeting paralogous protein–protein interactions.

## Results

### *Structural Stability of the Modeled Complexes*

Cα root-mean-square deviation (RMSD) values were measured for all simulation systems, and these indicated that the complexes were stable (Figure S1A), along with the core regions of receptor (Figure S1B) and ligand (Figures S1C, S1D). For the antiapoptotic protein components, structural variability was concentrated primarily in the loops connecting the

helices (Figure S1B). Ligands tended to display higher RMSD values, but the increased dynamics originated from the termini (Figure S1C). If RMSDs between positions 8 and 20 are considered, RMSD values generally tend to vary within a small window of around 0.3 Å with mean values between 0.2 – 0.7 Å (Figure S1D). Moreover, RMSD variations between replicas were small for the majority of complexes. When simulated in isolation, the receptors maintained their structure (Figure S1E), whereas the peptides unfolded in agreement with experiment (Chen et al., 2005). In order to optimize the signal/noise ratio for the energy calculations, we based our subsequent analyses on the latter 60 ns of each trajectory.

### *Energetic Basis for Protein-Protein Affinities*

We next utilized Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) calculations to calculate the enthalpy of binding (ΔH) for each replica, and decomposed the results on a per-residue basis. In order to discern the origins of affinity and specificity, we looked at the per-residue ΔH contributions across the whole set of simulations. We compared the means and variances of per-residue ΔH observed in the trajectory sets for the five receptors, each interacting with the same set of ligands. If a residue consistently contributed a high ΔH value with low variance, this indicates that it is an important site for generating affinity. Conversely, residues that show a high variance across the set of interactions with different ligands are likely to be involved in determining binding specificity. We subsequently mapped the mean per-residue ΔH values and their variances onto the surface of each complex, in order to discern the main contributors to affinity and specificity for the five receptor – ligand sets (Figures 2A and 2B, respectively). The underlying numerical values are given in Supplementary Tables 1 and 2.

It is evident that for the Bcls (Bcl-xL, Bcl-2, and Bcl-w) and Mcl-1, affinity originates predominantly from the region around the NWGR motif of the receptor, particularly the arginine, which forms a salt bridge with the aspartic acid in position 17 of the ligand.

Furthermore, in the Bcls, there exists a conserved glutamic acid (E129/E136/E85, respectively), which contacts ligand positions 6, 10, and 13, which are typically positively charged or polar (cf. Figure 1). Correspondingly, for the peptides bound to Bcl-x$_L$, Bcl-2, and Bcl-w, it is these three residues, along with D17, that are the greatest contributors to affinity. In Mcl-1 and A1, the glutamic acid has been substituted by H233 and K77, respectively. As evident from Table 2, the most conserved residues account for around 45 – 55% of total receptor contribution to binding, with the NWGR motif alone responsible for 25 – 35%.

### *Energetic Basis for Protein-Protein Specificities*

For the Bcl receptors specificity is greatest at the rim around pockets 3 and 4, and a patch surrounding the conserved glutamate in the receptor, E129/E136/E85 (Figure 2B; cf. Figure 1). For the ligands, specificity is highest at the N-terminal half of the peptides, at positions 6, 10, and 13, which contact this patch, and position 18, which contacts the aforementioned rim. In Mcl-1 and A1 the rim is much shallower, especially around pocket 4 (Czabotar et al., 2007), and is a lot less discriminating than in the other antiapoptotic proteins, whereas the NWGR motif and its adjacent residues appear to take on a greater role in determining specificity as they contact ligand residues 16, 19, and 20. Due to the increased ligand flexibility in the absence of a receptor, the results from MM-PBSA calculations on complex, receptor, and ligand trajectories (the "three-trajectory" approach) point to a greater number of residues being involved in determining specificity then the MM-PBSA data relying solely upon complex trajectories. A complete sampling of ligand conformations in isolation would require orders of magnitude longer dynamics than could typically be accessed computationally. Importantly, however, the results from the three-trajectory MM-PBSA calculations are consistent with the forgoing data on specificity and affinity (Figure S2).

As previously stated, ligand residues 6, 10, and 13 contact a conserved glutamic acid in Bcl-x$_L$, Bcl-2, and Bcl-w (E129/E136/E85, respectively). When two of those positions are

positively charged (cf. Figure 1B), this allowed the formation of a highly favorable salt-linked triad (Horovitz et al., 1990) between them and the glutamic acid. Moreover, when the remaining residue is also capable of hydrogen bonding to this glutamic acid, the latter hydrogen bond became coupled to the triad, further strengthening binding (Figure 4). Although position 14 remained oriented towards the solvent throughout most of our simulations, it is possible that it may also participate in binding through E129/E136/E85 or D133/D140/G89 (cf. Figure 1). Interestingly, the side chain of R13 in the ligand could simultaneously hydrogen bond to the backbone and side chain of the glutamic acid residue (Figure 4). Positively charged residues, especially KR and RR combinations for positions 13 and 14, are commonly found in these positions. In Mcl-1 and A1, the glutamic acid has been substituted by histidine and lysine, respectively, greatly reducing the hydrogen bonding potential between ligand and receptor. Consequently, in the Mcl-1 and A1 – ligand trajectories, R13 could only form hydrogen bonds with receptor backbone atoms, resulting in much less favorable interactions with the antiapoptotic protein. The importance of the 6–10–13 – receptor residue coupling is also reinforced by the fact that all weak binders (i.e., peptides in receptor-ligand complexes with $pIC_{50} < 6$) have one or more residues in positions 6, 10, or 13 which are incapable of participating in an interaction with this key receptor residue (cf. Figure 1 and Table 1).

### *Energetic Correlation Analysis*

Across the five trajectory sets, we correlated ΔH values for each ligand position with every other, i.e., each of the latter 26 rows in Tables S1 and S2 with each of the remaining 25. Correlating ΔH values for peptide positions 1 through 26 among each other reveals that in ligands bound to Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1, there seem to exist two regions of energetic correlation (Figure 3A & S3). The first one extends up to around position 15, which fits into pocket 3. Past that, there is a C-terminal region of somewhat weaker energetic

correlation. It is possible that this is due to the 6–10–13 and the 16–19–20 couplings (the latter of which is achieved through the NWGR motif and its adjacent residues coming into contact with the ligand residues), and the clamping effect exerted on the bound peptides by the protein rim. In A1, however, there appears to be an almost uninterrupted region of helix-like energetic correlation spanning most of the peptide length (Figure 3B). This is likely because the rim in A1 is much shallower, particularly around pocket 4, making ligand structure and properties more pronounced and important for binding A1 than the other antiapoptotic proteins. This implies that helix stability *per se* would offer greater gains in affinity to A1 than the other proteins. Given that Mcl-1's rim is shallower than those of the Bcls, but less so than A1, we anticipate helix stability to have an effect intermediate in magnitude between those in A1 and the Bcls. Indeed, in two Mcl-1 trajectories and five A1 trajectories, we observed disengagement of ~10 C-terminal peptide residues from the proteins. Although the three-trajectory MM-PBSA results are somewhat harder to interpret, they are consistent with these findings (Figure S3).

## Discussion

In this study, we systematically investigated where affinity and specificity originate within a family of proteins by a careful analysis of binding energetics across a diverse set of complexes. Moreover, we showed how the behavior of ligands differs according to which receptor they are complexed with. A caveat of our analysis is that it has been performed exclusively on homology models. However, they are in excellent agreement with multiple existing structures (RMSD ~0.4 – 1Å), with recently published ones (Robin et al., 20015; Kim et al., 2015; Rajan et al., 2015) only reinforcing confidence in our models. Other potential limitations are the limited sampling afforded by explicit solvent simulations, the fidelity of the force field parameters, and the reliability of MM-PBSA results, omitting entropic contributions (Hansen & van Gunsteren, 2014). Nevertheless, our results are in good agreement with

multiple experimental studies and provide, to our knowledge, the first quantitative assessment across the family of the contributions of different regions in each receptor and ligand to binding. For example, most of the receptor residues deemed critical to BH3 peptide binding in an alanine scan study (Campbell et al., 2015), all of which are highly conserved, are prominent contributors to binding in our energetic analysis. That study and others (Day et al., 2008; Ku et al., 2011; Fletcher et al., 2008) have shown that the D17 – R (from NWGR) interaction is critical in multiple peptide – protein pairs, in accord with our data, which suggests that typically it is the greatest single contributor to binding. The significance of the 6–10–13 coupling through the E129/E136/E85/H233/K77 residue is clearly demonstrated by the observation that mutating the glutamate in the Bcls is detrimental to BH3 binding, whereas mutating the corresponding histidine in Mcl-1 to alanine strengthens binding to peptides which carry positive charges in positions 6, 10, and/or 13 (Campbell et al., 2015).

Mutating Bim residues 6 and 10 to glutamate strengthens binding to Mcl-1, whereas the I6E mutation weakens binding to Bcl-$x_L$; Q10E has little effect on Bcl-$x_L$ binding. Mutating Bim positions 13 and 14 to glutamate weakens binding to Mcl-1. This is likely because they contact a highly conserved aspartate located four positions C-terminal to H233. This aspartate is highly conserved among all antiapoptotic proteins except Bcl-w (Figure 1D). However, the R13E and R14E substitutions practically abolish Bim binding to Bcl-$x_L$ (Boersma et al., 2008), suggesting another route to the design of Mcl-1 selective peptides and peptidomimetics (Smits et al., 2008; Lee et al., 2008).

Mutating Bim position 13 to an acidic residue weakens binding to A1, rather than enhancing it (DeBartolo et al., 2012). This is likely due to the aforementioned aspartate (D81 in mouse A1), as well as a unique feature of A1, residue E78, which is involved in forming pocket 2 and is buried in all human and murine A1 – BH3 X-ray structures (Herman et al., 2008). This residue is a leucine in the Bcls (L130 in Bcl-$x_L$, cf. Figure 1D) and a valine in Mcl-1. Indeed, it is the only pocket-forming residue with a high variance in ΔH values (Figure 2B and

Figure S2B). Our simulations demonstrate that positions 10 and 13 are in greater proximity to D81 and this glutamate, rather than the preceding lysine, and that position 6 appears in a more favorable position to interact with K77. Thus, we anticipate that an acidic residue in position 6 would either strengthen binding to mouse A1 or at least offer greater selectivity for A1 than Bcl-xL, Bcl-2, and Bcl-w. Moreover, we expect that acidic residues in positions 10 and 13 would cause a greater decrease in affinity to the Bcls than A1, opening up an avenue for the design of A1-selective molecules. Finally, we believe that our suggested mutations should have an effect on binding affinity towards Mcl-1, which is intermediate in magnitude between A1 and the Bcls.

Bad is the only BH3 sequence that does not bind Mcl-1. Moreover, its affinity to A1 seems to be only slightly above the detection limit of the affinity measurements (cf. Table 1; Chen et al., 2005). This is likely because of the excess positive charge in 6-10-13 (greatest among all the ligands), which is paired with H233/K77 in Mcl-1/A1, and the peculiarity of Bad residues 16 and 20, which are unique. In particular, all peptides have a glycine or an alanine in position 16, except for Bad, which has a serine. Its side chain is in proximity to that of T247/T91 (in Mcl-1 and A1, respectively) and NWGR and several adjacent residues, which helps explain why serine seems to be disfavored at this position whereas glycine and alanine are favored. T247/T91, located three positions C-terminal to the NWGR motif, seem to be more restrictive of binding than the corresponding alanines in Bcl-xL, Bcl-2, and Bcl-w (A142/A149/A98, Figure 1D), as those proteins better tolerate mutations to serine in peptide position 16. Indeed, the packing in this region is very dense, which is likely the reason mutating position 16 to any other residue weakens binding (DeBartolo et al., 2012) and mutating the glycine from NWGR even to alanine abolishes antiapoptotic activity (Yin et al., 1994; Sedlak et al., 1995). Moreover, Bad has a valine in position 20, unlike any of the other BH3 sequences under study, which have polar or charged residues in this position (D, N, or H). In our simulations, G245 of Mcl-1 is involved in an intermolecular N-capping interaction with

the ligand residue in position 20, helping maintain the ligand tethered to the receptor. Other authors have described this N-capping interaction as well (Day et al., 2008). In the receptor – Bad trajectories, where a valine stands at position 20, however, no such interaction is possible and in two of the Mcl-1 and A1 simulations the C-terminus disengages from the receptor. This led to the breaking of the key D – R salt bridge, which is the reason position 17 and the arginine from NWGR in Mcl-1 and A1 seem so variable in terms of energetics. Experimental evidence also demonstrates that the antiapoptotic proteins have a high preference for polar and charged residues in ligand position 20, with Mcl-1 (data not available for A1) being particularly selective for D, E, H, and N (DeBartolo et al., 2012). We expect A1 to display an identical preference and believe that this heightened selectivity in Mcl-1 is due to the shallowness of the rim, which makes the NWGR motif and its adjacent residues critical in terms of providing affinity and, as a consequence, specificity.

Our data suggest that in Bcl-x$_L$, Bcl-2, and Bcl-w, the rim around pockets 3 and 4 provides more specificity than affinity (cf. Figure 2 and Figure S2). This is corroborated by experiments which demonstrate that mutating Noxa residue 18, which contacts the foregoing rim, from a lysine to a glutamate transforms Noxa from a non-binder to a weak binder to Bcl-x$_L$ and Bcl-w. It seems that this mutation alone is not enough to achieve detectable binding to Bcl-2 (cf. Table 1). Typically, position 18 is an acidic residue, which contacts R100/R107/R56 from the rim in Bcl-x$_L$, Bcl-2, and Bcl-w. Only in Noxa is position 18 positively charged (cf. Figure 1E). Notably, Noxa is the only ligand that does not bind to these three proteins (cf. Table 1). In Mcl-1 and A1, the arginine has been mutated to N204 or E47, respectively.

For the Bcls in isolation, the calculated RMSD values seemed to be slightly higher than the complexed molecules, hinting at the stabilizing effect the peptides exert when bound (Figure S1). This has been observed previously for Bcl-x$_L$ (Guo et al., 2015). Compared to the Bcls, mouse Mcl-1 and A1 seem to be more stable in isolation, which agrees with the observation that they experience very little backbone conformational changes when binding

different BH3 peptides (Day et al., 2008; Smits et al., 2008; Day et al., 2005), contrasting with Bcl-x$_L$'s notable structural plasticity (Lee et al., 2009; Moldoveanu et al., 2014)., The ligands unfolded when not bound, in agreement with circular dichroism data (Chen et al., 2005).

It has previously been observed that helix stability is a factor contributing to affinity (Modi et al., 2012). Based on our simulations and energy correlation analysis, we may add that C-terminal helicity contributes to binding by stabilizing the D17 – R (from the NWGR motif) and position 19 – pocket 4 interactions. Correspondingly, lower helix stability would facilitate the loss of these intermolecular interactions and would decrease binding affinity. Similarly, N-terminal stability of the peptide helix would help maintain peptide – receptor interactions in this region and the key hydrophobic residue – pocket 1 interaction. From our analysis of the crucial interactions, we predict that the Bad mutations S16G and V20N should enhance binding to Mcl-1 and A1, as would mutating residues H233 (Mcl-1) and K77 (A1) to acidic amino acids. Further, we expect that mutations in the key acidic residues in the three Bcls (E129/E136/E85) should weaken or completely abolish binding to most of the BH3 domains reviewed here. We also anticipate that mutating R100/R107/R56 in the Bcls to acidic amino acids would weaken binding to the peptides with an acidic residue in position 18 and strengthen binding to Noxa, which has a lysine in this position. Lastly, the E47K or E47R mutations in A1 should decrease affinity for Noxa and enhance binding to most of the remaining peptides.

We have presented a detailed analysis of the specificity determinants and energetic contributions for the groove – BH3 peptide interaction. We have opted to discuss energies in relative, rather than absolute, terms, so as to make our conclusions insensitive to the choice of MM-PBSA parameters. An important conclusion to be drawn from our work is that the highly conserved pockets provide affinity, but not specificity. Aiello et al. previously investigated the balance between functionally conserved (i.e., binding the same ligand) and divergent interfaces in structural terms. (Aiello & Cafferey, 2012) Their analysis found that optimized

hydrogen bonding networks in the rim regions of the binding pocket are important in specific interfaces, whereas functionally conserved interfaces tend to draw a larger portion of their total affinity from the central hub region. Their conclusion is consistent with the energetic analysis we performed.

The wealth and fine-grained nature of the energy data presented in this study allows us to explore the connection between conservation of sequence and of binding energetics. All investigated complexes have a similar fold and binding mode. Hence, observed correlations directly relate sequence to energy. In order to quantify these relations, we set out to construct an "energetic fingerprint" for each complex (see SI for further details). We then correlated these energetic fingerprints among our simulations, grouped either by common ligand (Figure S4A) or by common receptor (Figure S4B). These similarity maps of energies were then compared to maps of sequence identity (Figure S4, green). Hence we now have a (semi)-quantitative approach that reveals to what degree similarity in sequence results in similarity in binding energetics (Figure S4). Careful inspection of the plots reveals that there are cases with a strong link between sequence and energy similarity (e.g., ligands Bak, Bim, Bad, Puma, Bmf and Noxa, receptors Bcl-xL, Bcl-w). However, in several cases, such a direct link is less apparent (e.g., ligands Bax, Bik, Bid and Hrk, receptor Mcl-1). The absence of strong correlation in some cases allows us to rationalize the efficiency of gene duplication as a means by which specific pathways emerge. Although greater divergence in sequence is usually accompanied by greater divergence in the interaction energy patterns, in some cases even slight changes in sequence can lead to large changes in interaction patterns. From an evolutionary perspective, this discontinuity could rapidly alter the specificity or promiscuity of an interface. This would indicate that energetic recognition patterns are the most adaptable feature in a hierarchy of structure, sequence and energy conservation. We posit that the groove – BH3 peptide example presented here is a manifestation of a more general pattern on the relationship between structure, sequence, and binding energetics. Indeed, instances

14

where a pool of structurally similar small molecules/peptides/proteins bind a well defined region on a set of structurally similar protein partners are found in all domains of life and physiological pathways (Friedman & Hughes, 2001). Our hypothesis may provide an attractive framework to investigate in a similar manner physiologically and therapeutically relevant systems, e.g., the bZIP transcription factors (Nair & Burley, 2003) and EGF receptors (Arkhipov et al., 2014), which have been implicated in malignant cellular proliferation; histidine kinase – response regulator protein interactions, central to signal transduction in bacterial cells (Casino et al., 2009); Toll-like receptors (Berglund et al., 2015) and MHC proteins (Patronov et al., 2012; Ivanov et al., 2012), both of which regulate immunity; and the E2 – E3 enzyme interaction, part of the ubiquitination pathway (Kar et al., 2012).

**Experimental Procedures**

We modeled human Bim, Bad, Bid, Puma, Bik, Hrk, Noxa, and three Noxa mutants, as well as mouse Bmf bound to human Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1 and A1 (See Table 1 and Figure 1). Additionally, we modeled human Bax and Bak with human Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1. The following template structures were used: 2XA0 (Ku et al., 2011), 4CIM (Lee et al., 2014), 3PL7 (Czabotar et al., 2011), 2ROC (Day et al., 2008), and 2VOF (Smits et al., 2008). Terminal BH3 residues were modeled using MODELLER 9.14 (Webb & Sali, 2014). Any mutations in the template antiapoptotic proteins were reverted back to wild type; BH3 sequences were modeled onto the BH3 template using in-house code. Briefly, the positions of backbone atoms were kept fixed, as were side chains in residues identical between model and template. Side chains for non-identical residues were re-packed (Bougouffa & Warwicker, 2008) using an adaptation (Cole & Warwicker, 2002) of a self-consistent mean-field method for rotamer selection from a rotamer library (Koehl & Delarue, 1994).

The resulting complexes were solvated with TIP3P water (Jorgensen et al., 1983) using the *tleap* module of Amber14 (Case et al., 2005) with a minimum wall distance of 12 Å. NaCl

was added to neutralize system charge, to a concentration of 0.15 M. After 1,000 steps of minimization, the systems were gradually heated from 0 to 300 K over a period of 150 ps, applying weak restraints to the protein. A 150 ps density-equilibration with restraints was followed by 2 ns of unrestrained constant pressure equilibration at 300 K. The protonation state of the solute and ionic strength, and temperature were set to match the conditions under which the $pIC_{50}$ values were obtained. 100 ns of production dynamics were then carried out in triplicate at a pressure of 1 bar and temperature of 300 K, maintained respectively with the Berendsen barostat and Langevin thermostat. An identical protocol was utilized to simulate the individual components of the complexes. Bonds to hydrogen were constrained using the SHAKE algorithm (Ciccotti & Ryckaert, 1986), thus allowing for a 2 fs time step. An 8.0 Å cutoff was used for Lennard-Jones interactions, and long-range electrostatics were computed with the Particle mesh Ewald scheme (Darden et al., 1993). All simulations were carried out using the ff14SB force field (Maier et al., 2015); trajectories were processed with *cpptraj V14.25* (Roe & Cheatham, 2013).

For each complex simulation, the enthalpy of interaction between the antiapoptotic protein and the bound BH3 helix was computed with the Amber14 MMPBSA.py script (Miller et al., 2012) using both the "one-trajectory" and "three-trajectory" approach. MM-PBSA calculations were performed using Bondi radii (Bondi, 1964) and default settings for the nonpolar decomposition scheme, surface tension, cavity offset, and external and internal dielectric constants. We adjusted the setting for the ionic strength to the one used during $IC_{50}$ measurements in our reference dataset. Per-residue energy decompositions were also performed, adding 1-4 energy terms to internal energy terms. For each 100 ns MD run, free energy calculations were performed on the latter 60 ns of dynamics. Snapshots for PBSA calculations were taken every 6 frames (60 ps apart), producing 1,000 frames per trajectory.

As we were primarily interested in relative rather than absolute binding energies (Homeyer & Gohlke, 2012; Huber et al., 2013), we chose to omit entropy calculations from our

analysis. This decision is reinforced by published calorimetric data, which demonstrates that BH3 helix binding is an enthalpically driven process (Day et al., 2008). Finally, the means and variance of the per-residue $\Delta H$ values were computed for the 39-trajectory sets for Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1, and the 24-trajectory set for mouse A1.

The absolute values of the computed energy terms are sensitive to the choice of atomic radii and nonpolar decomposition scheme in the MM-PBSA approach, whereas their relative values have been shown to be insensitive to these parameters (Kumari et al., 2014). Our results support this conclusion and demonstrate that the difference in computed $\Delta H$ values for a trajectory using *bondi* and *mbondi2* radii (Onufriev et al., 2004) is around 4 to 5 kcal/mol. The chosen scheme for computing $\Delta G_{nonpolar}$ yielded $\Delta H$ values which are of similar magnitude to calorimetric data (Day et al., 2008) ($\sim$ -10 to -25 kcal/mol), whereas the alternative scheme, where $\Delta G_{nonpolar}$ is linearly dependent upon solvent accessible surface area, significantly overestimated $\Delta H$ ($\sim$ -80 to -100 kcal/mol). Our work, therefore, corroborates the benefit of decomposing $\Delta G_{nonpolar}$ into a dispersive (attractive) and cavitation (repulsive) term (Tan et al., 2007).

For each complex, we used the per-residue interaction energies derived from our MM-PBSA calculations and represented them as a $\sim$150-dimensional vector. Analogously to ideas used to compare specificity patterns of proteases (Fuchs et al., 2013), we calculated the inner product of the respective vectors to quantify the similarity between different energy patterns. This measure is 1 if the patterns are identical, 0 if the patterns are orthogonal (i.e., no energy contributions are in common between paired patterns), and -1 if the patterns are inverted. Sequence identities were calculated omitting insertions and deletions. All energies were compared and subsequently plotted in groups of common ligands (Figure S4A) or common receptor (Figure S4B).

**Author Contributions**

SMI and RGH performed the experiments. All authors analyzed the data, wrote the paper, and designed the research.

## Acknowledgements

## References

Aiello, D., & Caffrey, D. R. (2012). Evolution of specific protein-protein interaction sites following gene duplication. *Journal of Molecular Biology*, *423*(2), 257–72.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Yeh, L.-S. L. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research*, *32,* D115–119.

Arkhipov, A., Shan, Y., Kim, E. T., & Shaw, D. E. (2014). Membrane interaction of bound ligands contributes to the negative binding cooperativity of the EGF receptor. *PLoS Computational Biology*, *10*(7), e1003742.

Berggård, T., Linse, S., & James, P. (2007). Methods for the detection and analysis of protein-protein interactions. *Proteomics*, *7*(16), 2833–42.

Berglund, N. A., Kargas, V., Ortiz-Suarez, M. L., & Bond, P. J. (2015). The role of protein-protein interactions in Toll-like receptor function. *Progress in Biophysics and Molecular Biology*, *119*(1), 72–83.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H.,  Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–42.

Boersma, M. D., Sadowsky, J. D., Tomita, Y. A., & Gellman, S. H. (2008). Hydrophile scanning as a complement to alanine scanning for exploring and manipulating protein-protein recognition: application to the Bim BH3 domain. *Protein Sci.*, *17*, 1232–1240.

Bondi, A. (1964). van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, *68*, 441–451.

Bougouffa, S., & Warwicker, J. (2008). Volume-based solvation models out-perform area-based models in combined studies of wild-type and mutated protein-protein interfaces. *BMC Bioinformatics*, *9*, 448.

Campbell, S. T., Carlson, K. J., Buchholz, C. J., Helmers, M. R., & Ghosh, I. (2015). Mapping the

BH3 Binding Interface of Bcl-xL, Bcl-2, and Mcl-1 Using Split-Luciferase Reassembly. *Biochemistry*, *54*, 2632–2643.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, *26*, 1668–1688.

Casino, P., Rubio, V., & Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell*, *139*(2), 325–36.

Chen, L., Willis, S. N., Wei, A., Smith, B. J., Fletcher, J. I., Hinds, M. G., Huang, D. C. S. (2005). Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. *Molecular Cell*, *17*, 393–403.

Cheng, E. H. Y., Wei, M. C., Weiler, S., Flavell, R. a., Mak, T. W., Lindsten, T., & Korsmeyer, S. J. (2001). BCL-2, BCL-X L sequester BH3 domain-only molecules preventing BAX- and BAK-mediated mitochondrial apoptosis. *Molecular Cell*, *8*, 705–711.

Ciccotti, G., & Ryckaert, J. P. (1986). Molecular dynamics simulation of rigid molecules. *Computer Physics Reports*, *4*, 346–392.

Cole, C., & Warwicker, J. I. M. (2002). Side-chain conformational entropy at protein – protein interfaces. *Protein Science*, *11*, 2860–2870.

Cory, S., Huang, D. C. S., & Adams, J. M. (2003). The Bcl-2 family: roles in cell survival and oncogenesis. *Oncogene*, *22*, 8590–8607.

Czabotar, P. E., Lee, E. F., Thompson, G. V., Wardak, A. Z., Fairlie, W. D., & Colman, P. M. (2011). Mutation to bax beyond the BH3 domain disrupts interactions with pro-survival proteins and promotes apoptosis. *Journal of Biological Chemistry*, *286*(9), 7123–7131.

Czabotar, P. E., Lee, E. F., van Delft, M. F., Day, C. L., Smith, B. J., Huang, D. C. S., Colman, P. M. (2007). Structural insights into the degradation of Mcl-1 induced by BH3 domains.

20

*Proceedings of the National Academy of Sciences of the United States of America*, *104*(15), 6217–6222.

Czabotar, P. E., Lessene, G., Strasser, A., & Adams, J. M. (2014). Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nature Reviews. Molecular Cell Biology*, *15*(1), 49–63.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(1993), 10089.

Day, C. L., Chen, L., Richardson, S. J., Harrison, P. J., Huang, D. C. S., & Hinds, M. G. (2005). Solution Structure of Prosurvival Mcl-1 and Characterization of Its Binding by Proapoptotic BH3-only Ligands. *Journal of Biological Chemistry*, *280*(6), 4738–4744.

Day, C. L., Smits, C., Fan, F. C., Lee, E. F., Fairlie, W. D., & Hinds, M. G. (2008). Structure of the BH3 Domains from the p53-Inducible BH3-Only Proteins Noxa and Puma in Complex with Mcl-1. *Journal of Molecular Biology*, *380*(5), 958–971.

DeBartolo, J., Dutta, S., Reich, L., & Keating, A. E. (2012). Predictive Bcl-2 family binding models rooted in experiment or structure. *Journal of Molecular Biology*, *422*(1), 124–144.

Fletcher, J. I., Meusburger, S., Hawkins, C. J., Ri, D. T., Lee, E. F., Fairlie, W. D., Adams, J. M. (2008). Apoptosis is triggered when prosurvival Bcl-2 proteins cannot restrain Bax. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 18081–18087.

Friedman, R., & Hughes, A. L. (2001). Gene duplication and the structure of eukaryotic genomes. Genome Research, 11, 373–381.

Fuchs, J. E., von Grafenstein, S., Huber, R. G., Kramer, C., Liedl, K. R. (2013) Substrate-Driven Mapping of the Degradome by Comparison of Sequence Logos. PLoS Computational Biology, 9, e1003353.

Gromiha, M. M., Saranya, N., Selvaraj, S., Jayaram, B., & Fukui, K. (2011). Sequence and

    structural features of binding site residues in protein-protein complexes: comparison

    with protein-nucleic acid complexes. *Proteome Science*, *9*

Gromiha, M. M., Yokota, K., & Fukui, K. (2009). Energy based approach for understanding the

    recognition mechanism in protein-protein complexes. *Molecular Biosystems*, *5*, 1779–

    1786.

Guo, Z., Thorarensen, A., Che, J., & Xing, L. (2015). Target the more druggable protein states in

    a highly dynamic protein-protein interaction system. *Journal of Chemical Information and*

    *Modeling*. 56 (1), pp 35–45

Hansen, N., & van Gunsteren, W. F. (2014). Practical aspects of free-energy calculations: A

    review. *Journal of Chemical Theory and Computation*, *10*(7), 2632–2647.

Happo, L., Strasser, A., & Cory, S. (2012). BH3-only proteins in apoptosis at a glance. *Journal of*

    *Cell Science*, *125*, 1081–1087.

Herman, M. D., Nyman, T., Welin, M., Lehtiö, L., Flodin, S., Trésaugues, L., Nordlund, P. (2008).

    Completing the family portrait of the anti-apoptotic Bcl-2 proteins: Crystal structure of

    human Bfl-1 in complex with Bim. *FEBS Letters*, *582*, 3590–3594.

Homeyer, N., & Gohlke, H. (2012). Free Energy Calculations by the Molecular Mechanics

    Poisson–Boltzmann Surface Area Method. *Molecular Informatics*, *31*, 114–122.

Horovitz, A., Serrano, L., Avron, B., Bycroft, M., & Fersht, A. R. (1990). Strength and co-

    operativity of contributions to surface salt bridges to protein stability. *Journal of*

    *Molecular Biology*, *216*, 1031–1044.

Huber, R. G., Fuchs, J. E., von Grafenstein, S., Laner, M., Wallnoefer, H. G., Abdelkader, N., Liedl,

    K. R. (2013). Entropy from state probabilities: hydration entropy of cations. *The Journal*

    *of Physical Chemistry. B*, *117*(1), 6466–72.

Ivanov, S., Dimitrov, I., & Doytchinova, I. (2013). Quantitative Prediction of Peptide Binding to HLA-DP1 Protein, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *10*(3), 811–815.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(1983), 926.

Kar, G., Keskin, O., Nussinov, R., & Gursoy, A. (2012). Human proteome-scale structural modeling of E2-E3 interactions exploiting interface motifs. *Journal of Proteome Research*, *11*(2), 1196–207.

Kim J. S., Ku B., Woo T. G., et al. Conversion of cell-survival activity of Akt into apoptotic death of cancer cells by two mutations on the BIM BH3 domain. *Cell Death Dis*. 2015;6:e1804.

Koehl, P., & Delarue, M. (1994). Application of a Self-consistent Mean Field Theory to Predict Protein Side-chains Conformation and Estimate Their Conformational Entropy. *Journal of Molecular Biology*, *239*(2), 249–275.

Ku, B., Liang, C., Jung, J. U., & Oh, B.-H. (2011). Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. *Cell Research*, *21*(4), 627–41.

Kumari, R., Kumar, R., & Lynn, A. (2014). g-mmpbsa - a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*, 1951–1962.

Kvansakul, M., Yang, H., Fairlie, W. D., Czabotar, P. E., Fischer, S. F., Perugini, M. A, Colman, P. M. (2008). Vaccinia virus anti-apoptotic F1L is a novel Bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands. *Cell Death and Differentiation*, *15*, 1564–1571.

Lee, E. F., Czabotar, P. E., van Delft, M. F., Michalak, E. M., Boyle, M. J., Willis, S. N., Fairlie, W. D.

(2008). A novel BH3 ligand that selectively targets Mcl-1 reveals that apoptosis can proceed without Mcl-1 degradation. *Journal of Cell Biology*, *180*(2), 341–355.

Lee, E. F., Czabotar, P. E., Yang, H., Sleebs, B. E., Lessene, G., Colman, P. M., Fairlie, W. D. (2009). Conformational changes in Bcl-2 pro-survival proteins determine their capacity to bind ligands. *Journal of Biological Chemistry*, *284*(44), 30508–30517.

Lee, E. F., Dewson, G., Evangelista, M., Pettikiriarachchi, A., Gold, G. J., Zhu, H., Fairlie, W. D. (2014). The functional differences between pro-survival and pro-apoptotic B cell lymphoma 2 (Bcl-2) proteins depend on structural differences in their Bcl-2 homology 3 (BH3) domains. *The Journal of Biological Chemistry*, *289*(52), 3601–17.

Ma, B., Elkayam, T., Wolfson, H., & Nussinov, R. (2003). Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10), 5772–7.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11 (8), pp 3696–3713.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. *Science (New York, N.Y.)*, *298*(5600), 1912–34.

Miller, B. R., Mcgee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.*, *8*(9), pp 3314–3321.

Modi, V., Lama, D., & Sankararamakrishnan, R. (2012). Relationship between helix stability and binding affinities: molecular dynamics simulations of Bfl-1 / A1- binding pro-

apoptotic BH3 peptide helices in explicit solvent. *Journal of Biomolecular Structure and Dynamics*, (March 2013), 37–41.

Moldoveanu, T., Follis, A. V., Kriwacki, R. W., & Green, D. R. (2014). Many players in BCL-2 family affairs. *Trends in Biochemical Sciences*, *39*(3), 101–111.

Nair, S. K., & Burley, S. K. (2003). X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors. *Cell*, *112*(2), 193–205.

Nguyen, M. N., Tan, K. P., & Madhusudhan, M. S. (2011). CLICK--topology-independent comparison of biomolecular 3D structures. *Nucleic Acids Research*, *39*(May), W24–W28.

Onufriev, A., Bashford, D., & Case, D. A. (2004). Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Structure, Function and Genetics*, *55*, 383–394.

Patronov, A., Dimitrov, I., Flower, D. R., & Doytchinova, I. (2012). Peptide binding to HLA-DP proteins at pH 5.0 and pH 7.0: a quantitative molecular docking study. *BMC Structural Biology*, *12*(1), 20.

Petros, A. M., Olejniczak, E. T., & Fesik, S. W. (2004). Structural biology of the Bcl-2 family of proteins. *Biochimica et Biophysica Acta*, *1644*, 83–94.

Rajan S., Choi M., Baek K., Yoon H. S. BH3 induced conformational changes in Bcl-Xl revealed by crystal structure and comparative analysis. *Proteins*. 2015;83(7):1262-1272.

Robin A. Y., Krishna Kumar K., Westphal D., et al. Crystal structure of Bax bound to the BH3 peptide of Bim identifies important contacts for interaction. *Cell Death Dis.* 2015;6(7):e1809.

Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and*

*Computation*, *9*(7), 3084–3095.

Sedlak, T. W., Oltvai, Z. N., Yang, E., Wang, K., Boise, L. H., Thompson, C. B., & Korsmeyer, S. J. (1995). Multiple Bcl-2 family members demonstrate selective dimerizations with Bax. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(17), 7834–8.

Shamas-Din, A., Brahmbhatt, H., Leber, B., & Andrews, D. W. (2011). BH3-only proteins: Orchestrators of apoptosis. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1813*(4), 508–520.

Shoemaker, B. A., & Panchenko, A. R. (2007). Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Computational Biology*, *3*(3), e42.

Smits, C., Czabotar, P. E., Hinds, M. G., & Day, C. L. (2008). Structural Plasticity Underpins Promiscuous Binding of the Prosurvival Protein A1. *Structure*, *16*, 818–829.

Stewart, M. L., Fire, E., Keating, A. E., & Walensky, L. D. (2010). The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nature Chemical Biology*, *6*(8), 595–601.

Suzuki, M., Youle, R. J., & Tjandra, N. (2000). Structure of Bax: coregulation of dimer formation and intracellular localization. *Cell*, *103*(4), 645–54.

Tan, C., Tan, Y. H., & Luo, R. (2007). Implicit nonpolar solvent models. *Journal of Physical Chemistry B*, *111*(2), 12263–12274.

van Wijk, S. J. L., de Vries, S. J., Kemmeren, P., Huang, A., Boelens, R., Bonvin, A. M. J. J., & Timmers, H. T. M. (2009). A comprehensive framework of E2-RING E3 interactions of the human ubiquitin-proteasome system. *Molecular Systems Biology*, *5*(295), 295.

Warwicker, J., & Watson, H. C. (1982). Calculation of the electric potential in the active site cleft due to α-helix dipoles. *Journal of Molecular Biology*, *157*, 671–679.

Webb, B., & Sali, A. (2014). *Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics* (Vol. Chapter 5, Unit 5.6).

Willis, S. N., Chen, L., Dewson, G., Wei, A., Naik, E., Fletcher, J. I., Huang, D. C. S. (2005). Proapoptotic Bak is sequestered by Mcl-1 and Bcl-xL, but not Bcl-2, until displaced by BH3-only proteins. *Genes and Development*, *19*, 1294–1305.

Yin, X. M., Oltvai, Z. N., & Korsmeyer, S. J. (1994). BH1 and BH2 domains of Bcl-2 are required for inhibition of apoptosis and heterodimerization with Bax. *Nature*, *369*, 321–323.
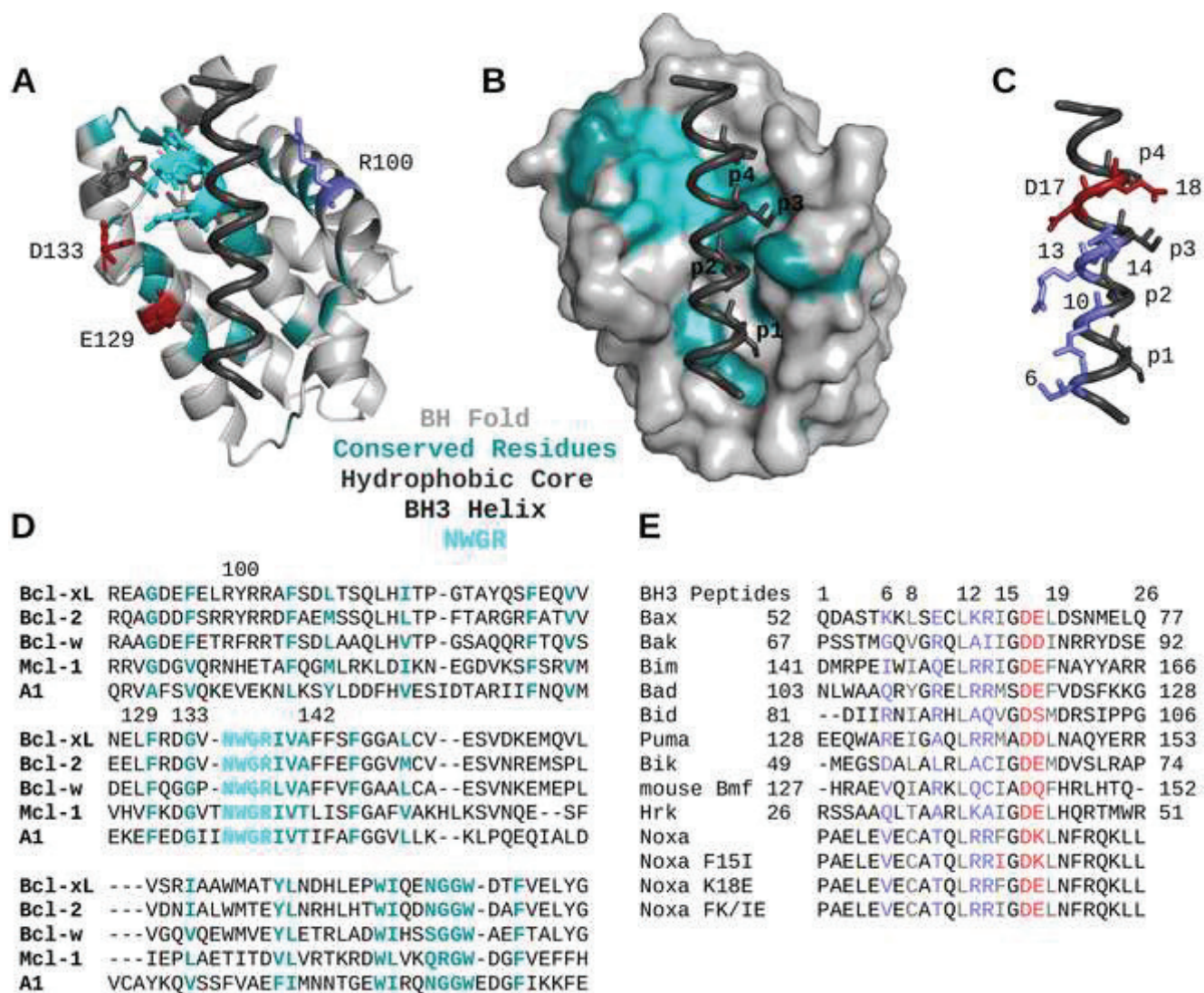
**Figure Legends**

**Figure 1. Summary of conserved structure and sequence properties across the Bcl-2-family-dependent apoptosis pathway. (A)** Bcl-$x_L$ – Bim complex in cartoon representation with key residues in stick representation – NWGR motif (cyan), hydrophobic core around the tryptophan (dark gray), E129 and D133 (red), and R100 (blue). **(B)** Same complex in surface representation for Bcl-$x_L$; also labeled are the 4 hydrophobic pockets and the peptide residues (in stick representation) that fit into them. **(C)** Bim BH3 peptide with key residues labeled and in stick representation. **(D)** Sequence alignment of the fold-forming portions of the 5 antiapoptotic proteins with the most conserved regions highlighted. Receptor residues are referred to by their canonical Uniprot (Apweiler et al., 2004) numbering throughout this report; numbering in the figure corresponds to Bcl-$x_L$. **(E)** Sequence alignment of the BH3 peptides used in this study and their location in the full-length proteins. Pocket residues (positions 8, 12, 15, and 19) are highlighted in gray, positions 6, 10, and 13 are highlighted in blue, and positions 17 and 18 are in red. All sequences are human, except Bmf, which is from mouse. All sequences are identical to the canonical sequences, deposited in Uniprot, except for a single mutation in Hrk (L15I). The sequences we have modeled are identical to the ones used during $pIC_{50}$ measurements, except for Bax. In the Bax affinity measurements, the authors used 34-mer peptides (Fletcher et al., 2008), whereas we have simulated the 26-residue-long Bax BH3 peptide. See also Figure S1.

**Figure 2. Sources of affinity and specificity assessed via energetics analysis, based on protein-protein complex trajectories. (A)** Antiapoptotic protein – BH3 peptide complexes colored by average per-residue ΔH values. **(B)** Antiapoptotic protein – BH3 peptide complexes colored by the variance of per-residue ΔH values. Averages and variance were calculated across 39-trajectory sets for Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1 (13 ligands x 3 replicas), and across 24 trajectories for mouse A1 (8 ligands x 3 replicas). Ligand N-termini are at the bottom of the figures, C-termini are at the top. ΔH was computed from complex trajectories only. See also Figure S2.

**Figure 3. Energy correlation analysis, based on protein-protein complex trajectories. (A)** Energy correlation analysis performed among the 26 ligand residues across the four 39-trajectory sets (13 ligands x 3 replicas). BH3 ligands seem to display two regions of energetic correlation – an N-terminal one, spanning up to around position 15 (colored orange in the structure to the right), and a C-terminal one (colored gray). **(B)** Energy correlation analysis performed among the 26 ligand residues across the 24-trajectory set for A1 (8 ligands x 3 replicas). BH3 ligands seem to display an almost uninterrupted region of helix-like energetic correlation, spanning most of the peptide length (colored orange in the structure to the right). ΔH was computed from complex trajectories only. See also Figures S3 and S4.

**Figure 4. Key interactions highlighted in a snapshot from a Bcl-$x_L$ – Bad trajectory.** The complex is in cartoon representation with Bcl-$x_L$ colored gray, Bad colored dark gray, and key residues in stick representation. Q6, R10, and R13 of Bad are in blue, D17 is in red, E129 of Bcl-$x_L$ is in green, and R139 (from the NWGR motif) is in cyan, with nitrogen atoms in blue and oxygen atoms in red. Also labeled are the peptide termini. Bcl-$x_L$ residue E129 simultaneously forms three salt-linked triads with 6Q, 10R, and R13 of Bad. Additionally, R13 simultaneously hydrogen bonds to the side chain and backbone of E129. The key D – R salt bridge is also present.
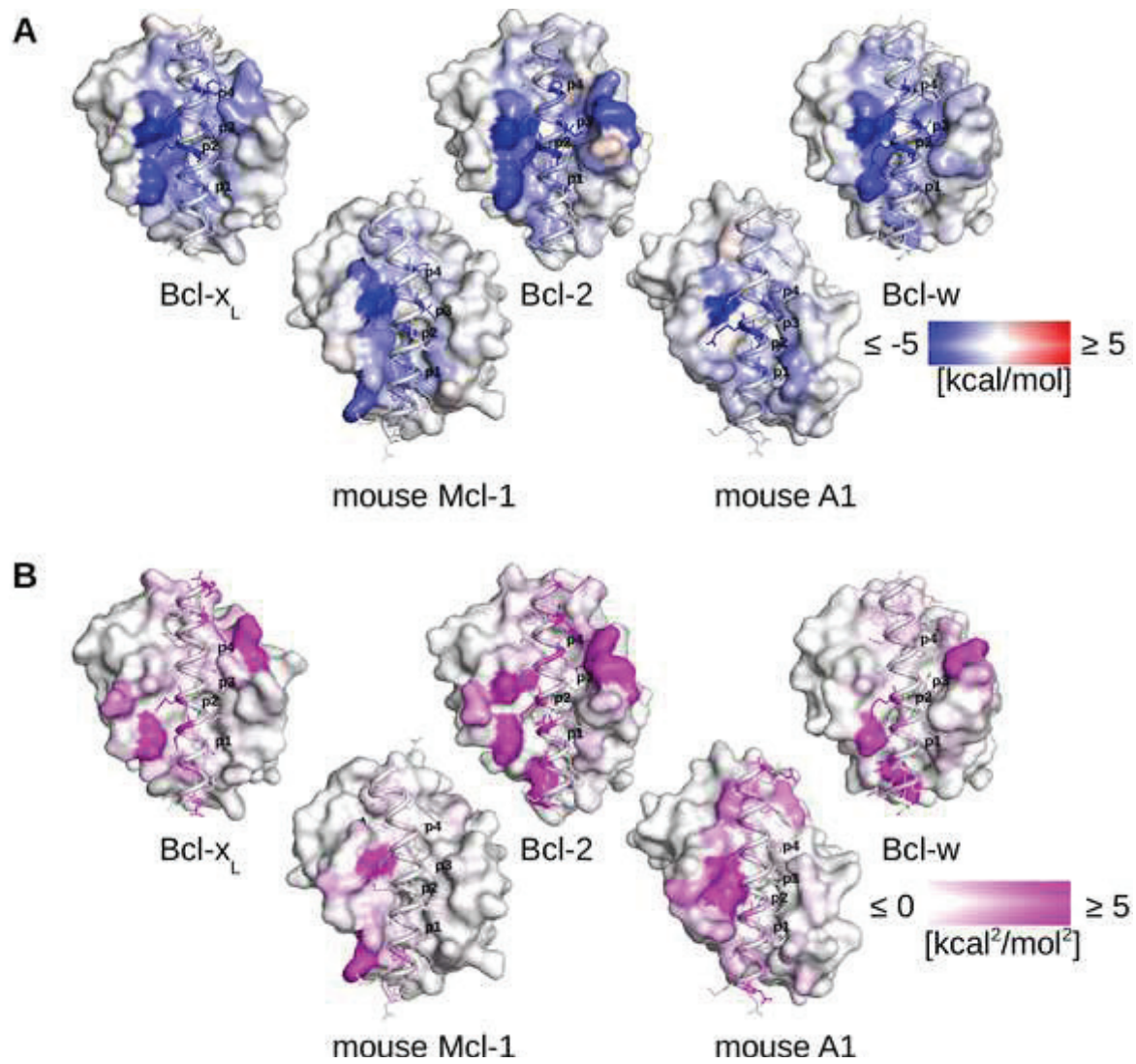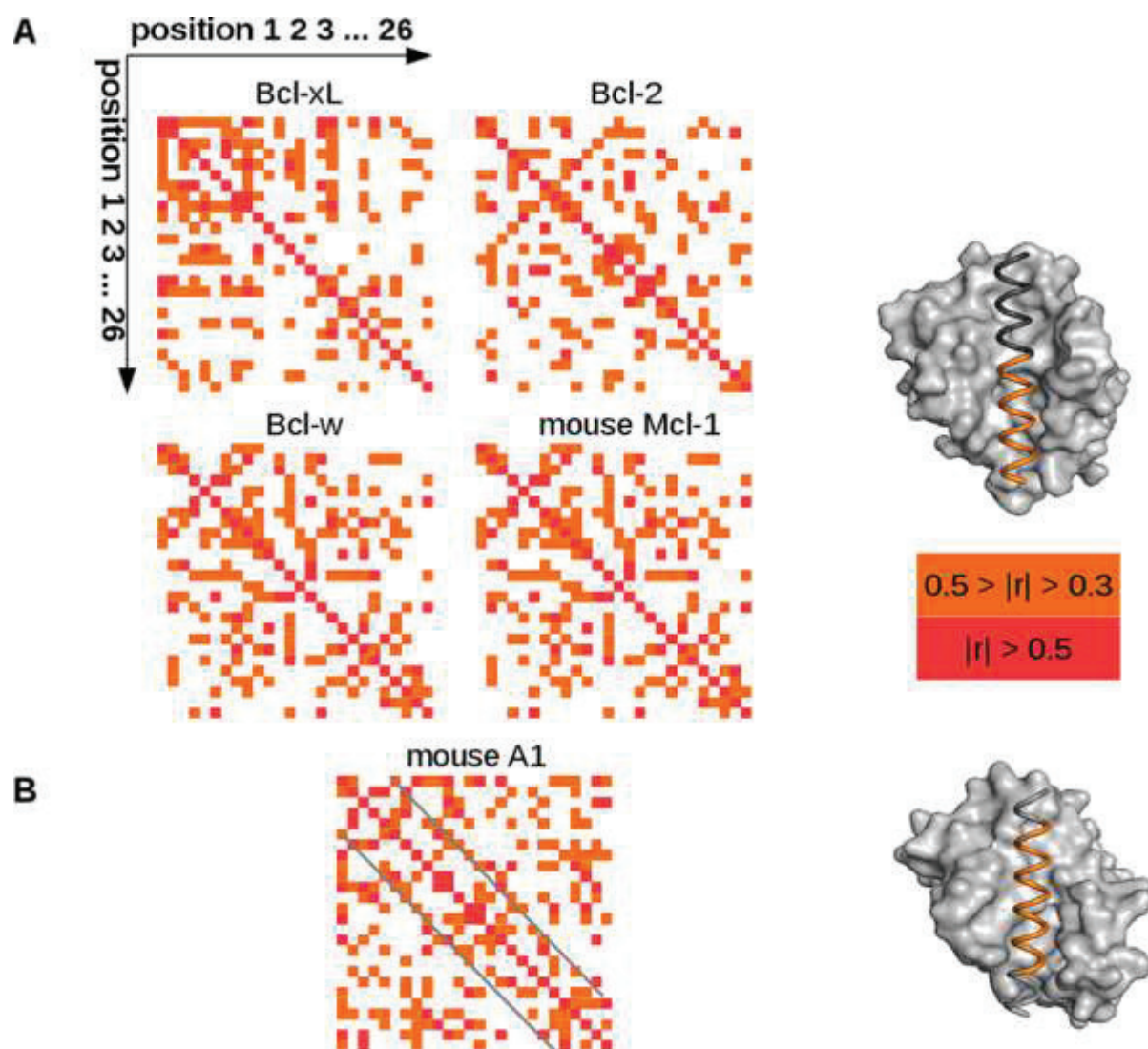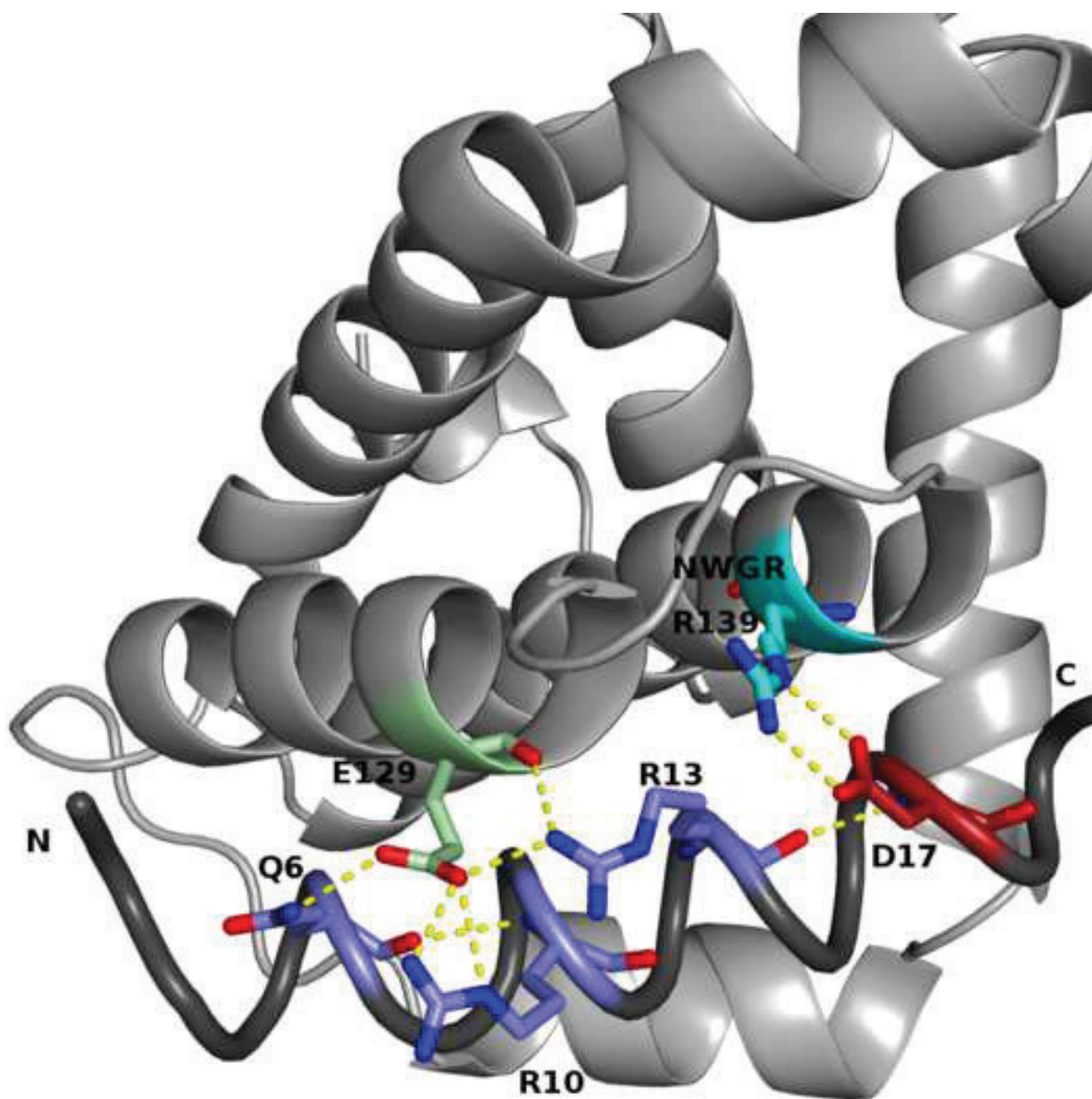
Figure 1

**A**

R100
D133
E129

BH Fold
Conserved Residues
Hydrophobic Core
BH3 Helix
NWGR

**B**

p4
p3
p2
p1

**C**

p4
D17    18
13    14    p3
10    p2
6    p1

**D**

```
                    100
Bcl-xL  REAGDEFELRYRRAFSDLTSQLHITP-GTAYQSFEQVV
Bcl-2   RQAGDDFSRRYRRDFAEMSSQLHLTP-FTARGRFATVV
Bcl-w   RAAGDEFETRFRRTFSDLAAQLHVTP-GSAQQRFTQVS
Mcl-1   RRVGDGVQRNHETAFQGMLRKLDIKN-EGDVKSFSRVM
A1      QRVAFSVQKEVEKNLKSYLDDFHVESIDTARIIFNQVM

          129 133      142
Bcl-xL  NELFRDGV-NWGRIVAFFSFGGALCV--ESVDKEMQVL
Bcl-2   EELFRDGV-NWGRIVAFFEFGGVMCV--ESVNREMSPL
Bcl-w   DELFQGGP-NWGRLVAFFVFGAALCA--ESVNKEMEPL
Mcl-1   VHVFKDGVTNWGRIVTLISFGAFVAKHLKSVNQE--SF
A1      EKEFEDGIINWGRIVTIFAFGGVLLK--KLPQEQIALD

Bcl-xL  ---VSRIAAWMATYLNDHLEPWIQENGGW-DTFVELYG
Bcl-2   ---VDNIALWMTEYLNRHLHTWIQDNGGW-DAFVELYG
Bcl-w   ---VGQVQEWMVEYLETRLADWIHSSGGW-AEFTALYG
Mcl-1   ---IEPLAETITDVLVRTKRDWLVKQRGW-DGFVEFFH
A1      VCAYKQVSSFVAEFIMNNTGEWIRQNGGWEDGFIKKFE
```

**E**

```
BH3 Peptides  1    6 8  12 15  19    26
Bax        52  QDASTKKLSECLKRIGDELDSNMELQ 77
Bak        67  PSSTMGQVGRQLAIIGDDINRRYDSE 92
Bim       141  DMRPEIWIAQELRRIGDEFNAYYARR 166
Bad       103  NLWAAQRYGRELRRMSDEFVDSFKKG 128
Bid        81  --DIIRNIARHLAQVGDSMDRSIPPG 106
Puma      128  EEQWAREIGAQLRRMADDLNAQYERR 153
Bik        49  -MEGSDALALRLACIGDEMDVSLRAP 74
mouse Bmf 127  -HRAEVQIARKLQCIADQFHRLHTQ- 152
Hrk        26  RSSAAQLTAARLKAIGDELHQRTMWR 51
Noxa           PAELEVECATQLRRFGDKLNFRQKLL
Noxa F15I      PAELEVECATQLRRIGDKLNFRQKLL
Noxa K18E      PAELEVECATQLRRFGDELNFRQKLL
Noxa FK/IE     PAELEVECATQLRRIGDELNFRQKLL
```

Figure 2

**A**

Bcl-x$_L$

mouse Mcl-1

Bcl-2

mouse A1

Bcl-w

$\leq$ -5 [color scale] $\geq$ 5

[kcal/mol]

**B**

Bcl-x$_L$

mouse Mcl-1

Bcl-2

mouse A1

Bcl-w

$\leq$ 0 [color scale] $\geq$ 5

[kcal$^2$/mol$^2$]

Figure 3

Figure 4

Table 1

| pIC$_{50}$ [M] | Bcl-xL | Bcl-2 | Bcl-w | Mouse Mcl-1 | Mouse A1 |
|---|---|---|---|---|---|
| **Bax** | 6.89 | 7.00 | 7.23 | 7.92 | N/A |
| **Bak** | 7.30 | < 6.00 | 6.30 | 8.00 | N/A |
| **Bim** | > 8.30 | > 8.30 | > 8.30 | > 8.30 | > 8.30 |
| **Bad** | 8.28 | 7.80 | 7.52 | < 4.00 | 4.82 |
| **Bid** | 7.09 | 5.17 | 7.40 | 5.68 | 8.03 |
| **Puma** | 8.20 | > 8.30 | 8.29 | > 8.30 | 8.24 |
| **Bik** | 7.37 | 6.08 | 7.92 | 5.77 | 7.24 |
| **Mouse Bmf** | 8.01 | > 8.30 | 8.01 | 5.96 | 5.74 |
| **Hrk** | > 8.30 | 6.49 | 7.31 | 6.43 | 7.34 |
| **Noxa** | < 4.00 | < 4.00 | < 4.00 | 7.22 | 6.74 |
| **Noxa K18E** | 5.30 | < 4.00 | 4.05 | 7.46 | N/A |
| **Noxa F15I** | 6.00 | < 4.00 | 5.00 | 7.60 | N/A |
| **Noxa FK/IE** | 6.96 | 4.96 | 6.30 | 7.62 | N/A |

**Table 1. pIC$_{50}$ values for different BH3 peptide – antiapoptotic protein interactions.** All sequences are human, except where explicitly stated otherwise. Bax data is from Fletcher et al. (2008); Bak data is from Willis et al. (2005); the remaining data is from Chen et al. (2005).

Table 2

| | Bcl-x$_L$ | Bcl-2 | Bcl-w | Mouse Mcl-1 | Mouse A1 |
|---|---|---|---|---|---|
| Conserved Residues | 53% | 44% | 48% | 55% | 56% |
| NWGR Motif | 31% | 25% | 28% | 35% | 29% |

**Table 2 (See also Tables S1-S2). Energetic contributions to binding (as a percentage of total receptor contribution) for the conserved residues (highlighted in Figure 1) and the NWGR motif.** Data shown are averages over 39 trajectories for Bcl-x$_L$, Bcl-2, Bcl-w, and Mcl-1 (13 ligands x 3 replicas) and 24 for A1 (8 ligands x 3 replicas).

**Energetics and Dynamics Across the Bcl-2-Family-Dependent Apoptosis Pathway Reveal Distinct Evolutionary Determinants of Specificity and Affinity – Supplemental Information**

Stefan M. Ivanov[1,2], Roland G. Huber[1], Jim Warwicker[2], Peter J. Bond[1,3*]

1 Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Matrix 07-01, 30 Biopolis Street, 138671 Singapore

2 Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

3 Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543

*corresponding author: peterjb@bii.a-star.edu.sg (Lead Contact)
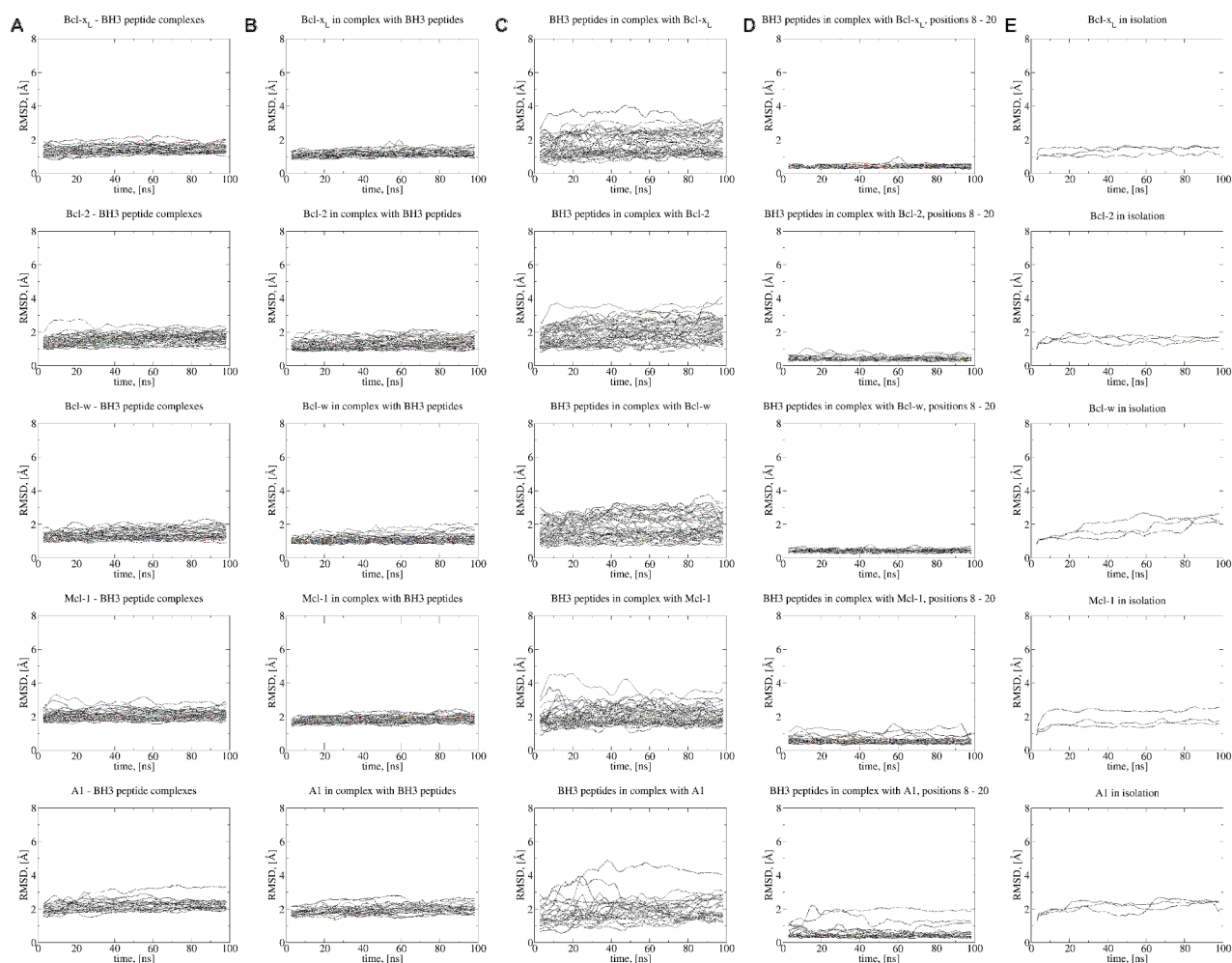
**Figure S1 (Related to Figure 1). Root-mean-squared deviation of different systems. (A)** Cα RMSD values for the entire antiapoptotic protein – BH3 peptide complexes. Data shown are 5-ns running averages of the complex RMSD values for each trajectory of every complex, i.e., 39 trajectories for Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1 – BH3 peptide complexes, and 24 trajectories for the mouse A1 – BH3 peptide complexes. **(B)** Cα RMSD values for the antiapoptotic proteins. Data shown are 5-ns running averages of the receptor RMSD values for each trajectory of every complex, i.e., 39 trajectories for Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1 – BH3 peptide complexes, and 24 trajectories for the mouse A1 – BH3 peptide complexes. **(C)** Cα RMSD values for the BH3 peptides. Data shown are 5-ns running averages of the ligand RMSD values for each trajectory of every complex, i.e., 39 trajectories for Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1 – BH3 peptide complexes, and 24 trajectories for the mouse A1 – BH3 peptide complexes. **(D)** Cα RMSD values for the core residues of the BH3 peptides (positions 8 - 20). Data shown are 5-ns running averages of the ligand core RMSD values for each trajectory of every complex, i.e., 39 trajectories for Bcl-$x_L$, Bcl-2, Bcl-w, and mouse Mcl-1 – BH3 peptide complexes, and 24 trajectories for the mouse A1 – BH3 peptide complexes. **(E)** Cα RMSD values for the antiapoptotic proteins in isolation. Each protein was simulated in triplicate; data shown are 5-ns running averages.
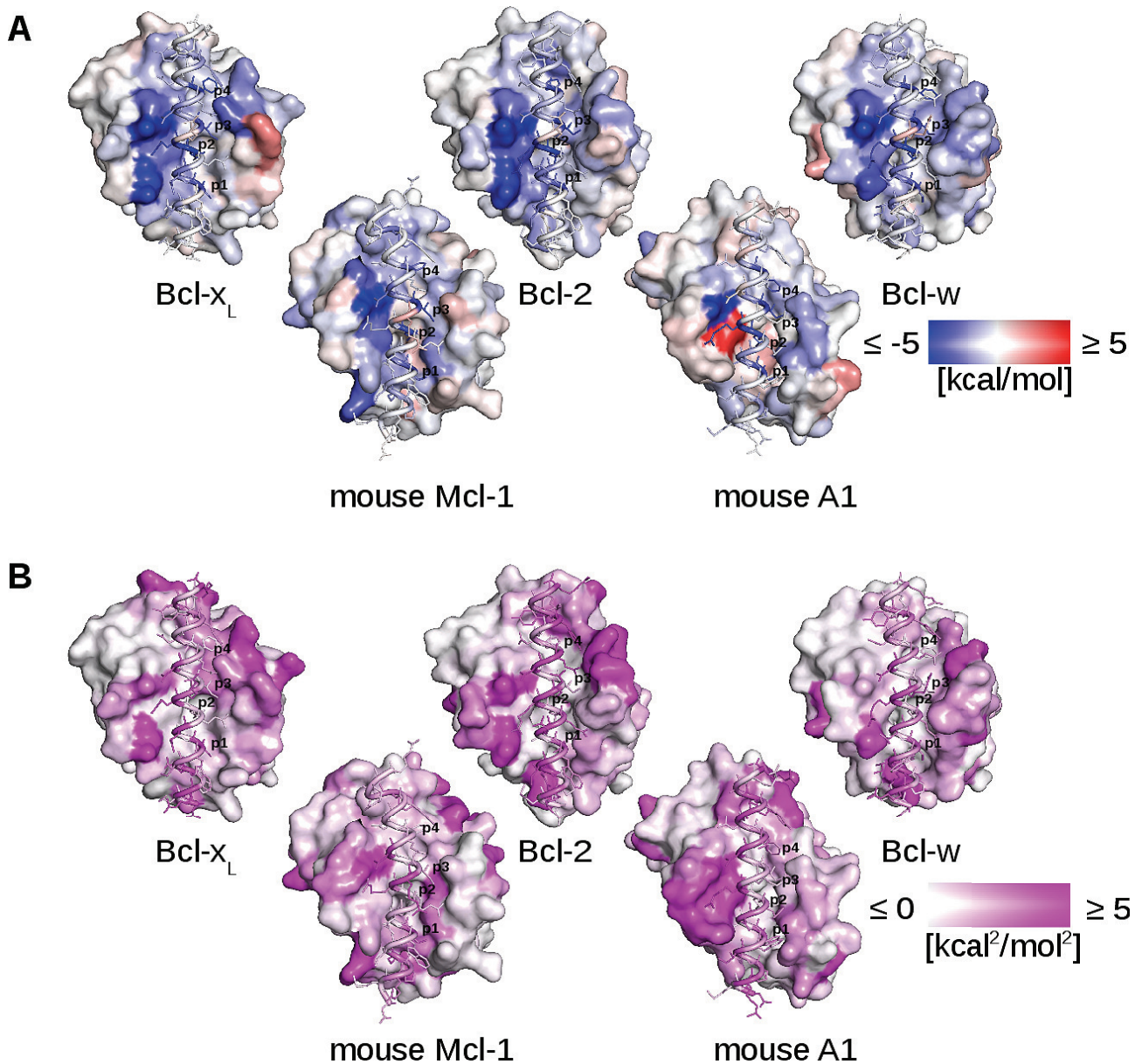
2

**Figure S2 (Related to Figure 2). Sources of affinity and specificity assessed via energetics analysis, based on trajectories of complex, receptor, and ligand. (A)** Antiapoptotic protein – BH3 peptide complexes colored by average per-residue ΔH values. **(B)** Antiapoptotic protein – BH3 peptide complexes colored by the variance of per-residue ΔH values. Averages and variance were computed across 39-trajectory sets for Bcl-x$_L$, Bcl-2, Bcl-w, and mouse Mcl-1, and across 24 trajectories for mouse A1. Ligand N-termini are at the bottom of the figures, C-termini are at the top. ΔH was calculated from complex, receptor, and ligand trajectories.
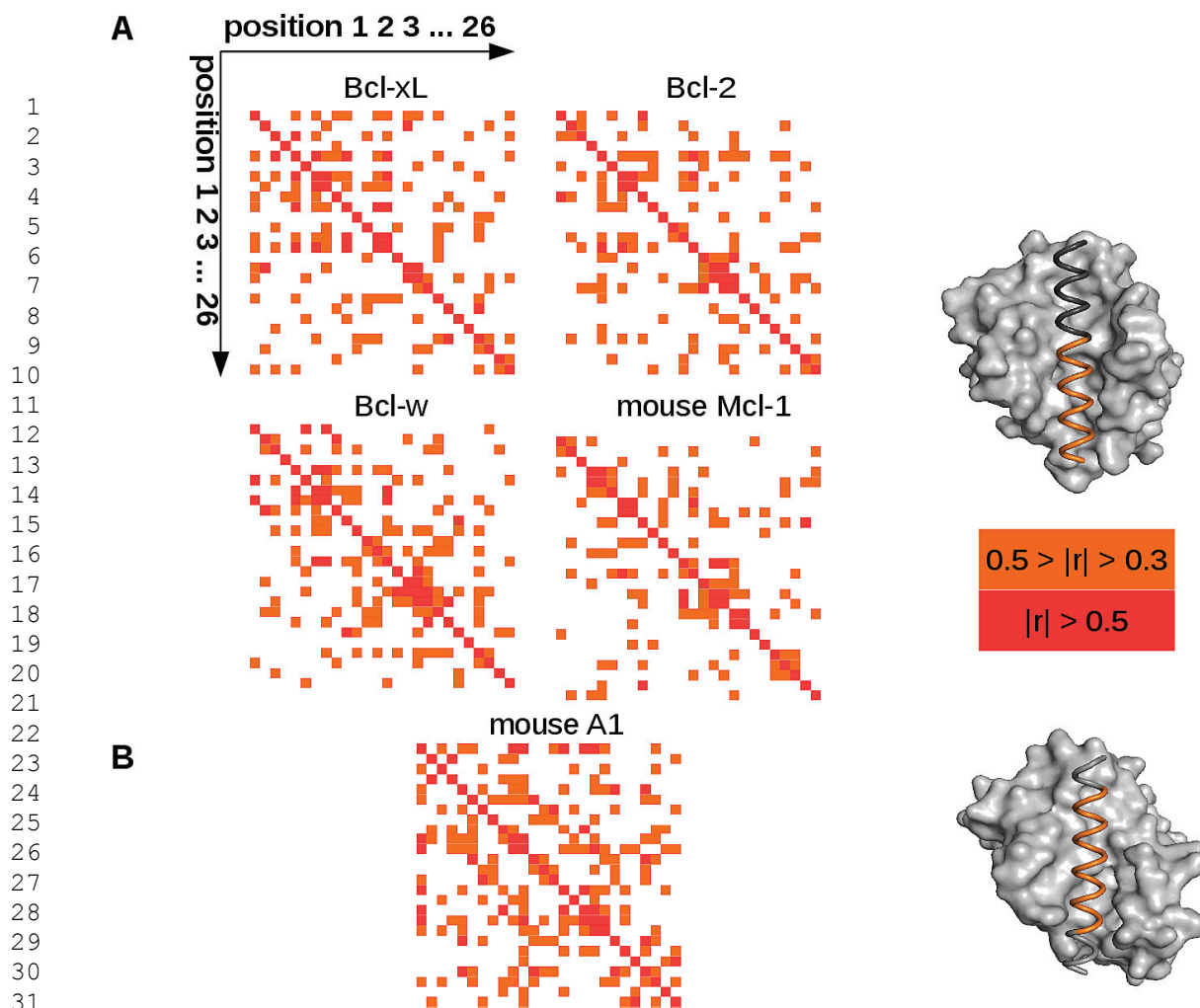
**Figure S3 (Related to Figure 3). Energy correlation analysis, based on trajectories of complex, receptor, and ligand. (A)** Energy correlation analysis performed for the 26 ligand residues across 39-trajectory sets. BH3 ligands seem to display two regions of energetic correlation – an N-terminal one, spanning up to around position 15 (colored orange in the structure to the right), and a C-terminal one (colored gray). **(B)** Energy correlation analysis performed for the 26 ligand residues across the 24-trajectory set for A1. BH3 ligands seem to display an almost uninterrupted region of helix-like energetic correlation, spanning most of the peptide length (colored orange in the structure to the right). ΔH was computed from complex, receptor, and ligand trajectories.
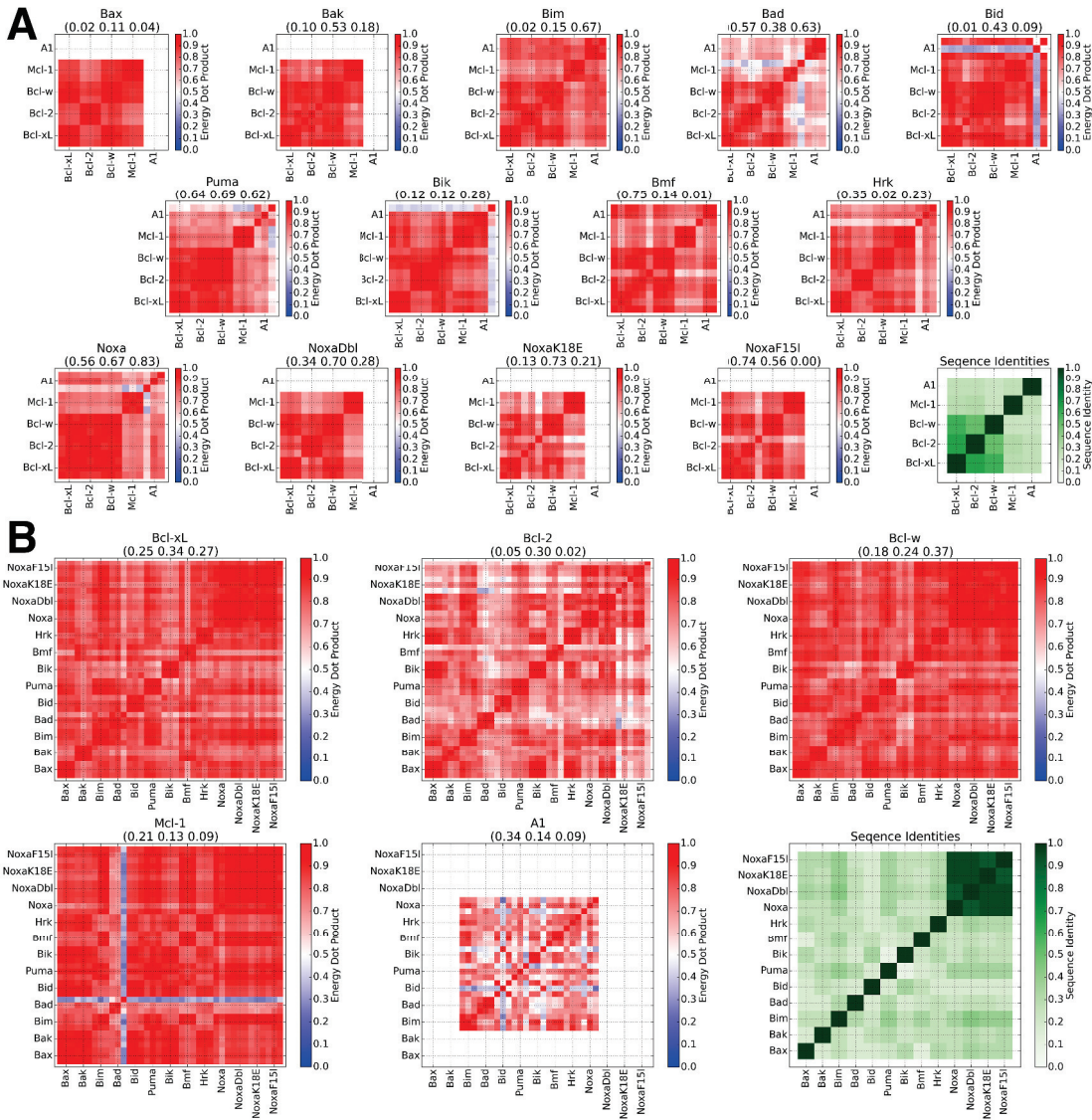
**Figure S4 (Related to Figure 3). Comparison of energy pattern similarity with sequence identity. (A)** Similarity of complex energies grouped by ligand simulations as an inner product of energy pattern vectors. Sequence identities of the respective receptors are shown in the bottom right corner in green. Correlation coefficients for the patterns (separately for replicas 1, 2 and 3) are shown in the plot titles. In **(B)** the same data is shown but grouped by receptor. The sequence identity of the ligands is shown in green.

5

**Table S1 (Related to Table 2). Per-residue ΔH values (kcal/mol) based on protein-protein complex trajectories.** The receptor NWGR and ligand pocket residues are highlighted in cyan and gray, respectively. The means and variance are computed by row.

**Table S2 (Related to Table 2). Per-residue ΔH values (kcal/mol) based on trajectories of complex, receptor, and ligand.** The receptor NWGR and ligand pocket residues are highlighted in cyan and gray, respectively. The means and variance are computed by row.

Supplemental Table 1

Click here to access/download
**Supplemental Movies & Spreadsheets**
table S1.xlsx

Click here to access/download
**Supplemental Movies & Spreadsheets
table S2.xlsx**