Structural and Functional Similarity between the Bacterial Type III Secretion System Needle Protein PrgI and the Eukaryotic Apoptosis Bcl-2 Proteins

Matthew D. Shortridge, Robert Powers*
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska, United States of America

Abstract

Background: Functional similarity is challenging to identify when global sequence and structure similarity is low. Active-sites or functionally relevant regions are evolutionarily more stable relative to the remainder of a protein structure and provide an alternative means to identify functional similarity between proteins. We recently developed the FAST-NMR methodology to discover biochemical functions or functional hypotheses of proteins of unknown function by experimentally identifying ligand binding sites. FAST-NMR utilizes our CPASS software and database to assign a function based on a similarity in the structure and sequence of ligand binding sites between proteins of known and unknown function.

Methodology/Principal Findings: The PrgI protein from Salmonella typhimurium forms the needle complex in the type III secretion system (T3SS). A FAST-NMR screen identified a similarity between the ligand binding sites of PrgI and the Bcl-2 apoptosis protein Bcl-xL. These ligand binding sites correlate with known protein-protein binding interfaces required for oligomerization. Both proteins form membrane pores through this oligomerization to release effector proteins to stimulate cell death. Structural analysis indicates an overlap between the PrgI structure and the pore forming motif of Bcl-xL. A sequence alignment indicates conservation between the PrgI and Bcl-xL ligand binding sites and pore formation regions. This active-site similarity was then used to verify that chelerythrine, a known Bcl-xL inhibitor, also binds PrgI.

Conclusions/Significance: A structural and functional relationship between the bacterial T3SS and eukaryotic apoptosis was identified using our FAST-NMR ligand affinity screen in combination with a bioinformatic analysis based on our CPASS program. A similarity between PrgI and Bcl-xL is not readily apparent using traditional global sequence and structure analysis, but was only identified because of conservation in ligand binding sites. These results demonstrate the unique opportunity that ligand-binding sites provide for the identification of functional relationships when global sequence and structural information is limited.

Introduction

Functional regions of a protein are evolutionarily more stable relative to the remainder of the protein [1,2]. The evolutionary correlation between ligand binding sites, ligand structure and protein function has also been demonstrated by a network of ligand-binding-site similarity described by Park & Kim [3]. A variety of computational methods have attempted to exploit the evolutionary stability of functional regions by identifying ligand binding sites as a method to predict function [4,5]. Unfortunately, the combined requirements of predicting the ligand, the binding site, and a similarity to an annotated protein lead to a high level of ambiguity. Recently, we developed the Functional Annotation Screening Technology by Nuclear Magnetic Resonance (FAST-NMR) assay to experimentally identify ligand binding sites to annotate proteins of unknown function [5–7]. FAST-NMR utilizes the Comparison of Protein Active Site Structures (CPASS) software and database to identify similar sequence and structure characteristics between experimentally identified ligand binding sites for proteins of known and unknown function [8]. Applying the FAST-NMR method on previously annotated systems also enables experimental ligand binding site data to identify functional relationships that otherwise would not be recognized based solely on global sequence and structure similarity.

The type three secretion system (T3SS) is composed of 20–25 different proteins, which are assembled in a highly choreographed mechanism similar to the assembly of flagella [9–11]. In Salmonella typhimurium, the needle complex is responsible for puncturing a host’s cell membrane to allow effector proteins (SipB, SipC, SipD) from S. typhimurium to enter the host [12]. Many of these effectors can activate bacterial induced apoptosis of a host’s cell by interacting with caspase-1 [13] in a mechanism similar to
apoptosis in eukaryotic cells [14]. The needle complex is a large homomultimer composed of ~120 repeated copies of the monomeric protein Prgl, a small helical protein of 83 amino acids [15]. The monomeric form of Prgl is a helix-turn-helix motif with two symmetrically charged surfaces and a conserved loop region, PxxP domain, which are important for needle assembly [15–17]. The charged surfaces of Prgl responsible for needle assembly also provide a potential binding site for small molecule ligands. This makes Prgl an attractive drug target to disrupt the formation of the needle complex and prevent infection by S. typhimurium. However, to date there have been no reported ligands that bind to either region of this protein-protein interaction site.

The Prgl needle complex protein from S. typhimurium T3SS was screened in our FAST-NMR assay, which resulted in the identification of a functional similarity between the ligand binding sites of Prgl and the anti-apoptosis protein Bcl-xL. Additionally, Dali [18] and T-Coffee [19] analysis found regions of structure and sequence similarity between the two proteins consistent with the FAST-NMR results. The predicted active-site similarity between Prgl and Bcl-xL was also used to experimentally verify that chelerythrine [20], a ligand known to inhibit Bcl-xL and induce apoptosis, also binds Prgl. These results provide experimental evidence that suggest a functional relationship between the bacterial type III secretion systems and apoptosis. This is consistent with a general conservation in function between Prgl and the Bcl-2 family of proteins that includes Bcl-xL; both form membrane pores through oligomerization using a conserved helix-turn-helix motif to release effectors to stimulate cell death.

Results

Results from the FAST-NMR screen

The needle complex protein, Prgl, from S. typhimurium is an attractive antibacterial target because the protein is exposed to the cell surface and blocking this target could prevent injection of virulence factors into the host [21]. The interaction of Prgl with the host membrane stimulates the delivery of effectors from the bacteria into the host cytosol to induce cell death. Recently an NMR structure was determined for a monomeric form of Prgl [15], which enabled the screening of Prgl using the FAST-NMR assay [6]. FAST-NMR combines NMR ligand affinity screening [22] using a fragment-based functional library [23] with structural biology and bioinformatics [8] to rapidly determine protein-ligand complexes [24] and infer functional relationship between proteins based on similarities in functional epitopes. Also, the resulting protein-ligand co-structure provides a valuable starting point for structure-based drug design.

FAST-NMR applies a tiered approach to screening [22] to minimize resources and increase throughput (Figure 1). First, Prgl was screened with the functional chemical library using 1D 1H NMR line-broadening experiments. Five compounds (L-carnitine inner salt, didecyldimethylammonium bromide, 1-methylimidazole, methiothepin mesylate salt, sucrose) were found to bind Prgl by showing a significant decrease in 1H peak intensity upon addition of 25 μM of Prgl. This was determined by comparing normalized 1H ligand peak intensities between the free and bound NMR spectra (Figure 1A). However, the secondary 2D 15N-1H HSQC experiments identified the lipid derivative didecyldimethylammonium bromide (DDAB) as the only specific Prgl binder (Figure 1B) based on the observation of a significant number of chemical shift changes in the spectrum. The remaining four compounds elicited no change in chemical shifts in the Prgl 2D 15N-1H HSQC spectrum, which suggest the compounds bound non-specifically to Prgl. Prgl was found to bind DDAB with a Kd of 553 μM as calculated by a 1D 1H line broadening method [25]. Finding a lipid derivative that specifically binds to Prgl is consistent with the protein’s function; sensing new host cells and signaling secretion through an interaction with the host membrane [26].

Chemical shift perturbations (CSPs) in the 2D 15N-1H HSQC experiments between free Prgl and the complex identified the Prgl residues that bind DDAB. Mapping these CSPs onto the Prgl surface identified the DDAB binding site as corresponding to residues at the bifurcation point of the two helices (Figure 1C). Specifically, residues S6, L9, S13, K15, and D17 of helix 1 and N59, V63, K66, V67, F68, K69, D70, D72, A73 and L76 of helix 2 showed significant CSPs in the presence of DDAB as calculated by equation 1. This ligand binding site has been shown to be

![Figure 1. Identification of Prgl Binding Ligands.](image)
important for the formation of the T3SS needle complex in which PrgI forms a repeating coiled-coils structure [10]. According to recent alanine scanning and structural studies, the surface residues in the region between the bifurcation point of the two helices and the conserved loop region, PxxP domain, are important for needle assembly [15–17]. These residues bind to the backside of the bifurcation point of the two helices in a stacked N-terminus to C-terminus manner [15–17].

The PrgI residues exhibiting significant CSPs upon binding DDAB were used to guide and filter a molecular docking simulation based on our method to rapidly determine protein-ligand co-structures [24]. AutoDock 4.0 [27] was used to calculate 100 docked structures within a 3D grid defined by the CSPs. Our AutoDock Filter program (ADF) selected the best conformation based on consistency with the magnitude of chemical shift changes [24]. The ligand is expected to be closest to the protein residues that incurred the largest CSPs. The best PrgI-DDAB docked structure is shown in Figure 1C, where DDAB adopts an extended conformation that straddles both helices of PrgI.

Analysis of CPASS and Structure Similarity Results

Comparison of Protein Active Site Structures (CPASS) analysis of the PrgI-DDAB complex identified a human Bcl-2 protein family member (the anti-apoptosis regulating protein Bcl-xL (PDB-ID:1YSN) complexed to an acyl-sulfonamide-based inhibitor (ABT-737)) [28] as the top hit based on a ligand binding-site CPASS similarity score of 37.7%. The CPASS alignment is shown in Figure 2A and is based on maximizing the spatial orientation of similar residue types between the two ligand-binding sites. All other proteins with a CPASS similarity >30% were also evaluated, but Bcl-xL was the only protein that gave a reliable CPASS score and showed some level of structure or sequence similarity to PrgI. It is important to note that the CPASS identified similarity between PrgI and Bcl-xL was fundamentally dependent on the existence of a Bcl-xL-ligand complex in the PDB. Ligand complexes for other members of the Bcl-2 protein family (Bax, Bid) currently do not exist.

While DDAB and ABT-737 are distinctly different ligands, the compounds share strong similarities in their mode of protein interactions. ABT-737 binds Bcl-xL edge-on in an elongated conformation where a minimal number of atoms contact the hydrophobic binding cleft of Bcl-xL. In this manner, DDAB mimics this edge contact interaction of ABT-737 with the similar hydrophobic binding cleft in PrgI. Also, ABT-737 binds in a protein-protein binding interface similar to DDAB, where inhibiting protein interactions is the drug’s mechanism of action.

Figure 2. Active Site Similarity between PrgI and Bcl-xL. (A) CPASS alignment of the S. typhimurium PrgI active-site complexed to DDAB with the active-site of human Bcl-2 protein (Bcl-xL) complexed with acyl-sulfonamide-based inhibitor. The residues aligned by CPASS are labeled and colored blue in the structures. The active site sequence alignment is also shown below the structures. The ligands are colored yellow. (B) Overlay of the human Bcl-2 protein (red) with S. typhimurium PrgI (turquoise) based on a DaliLite [29] alignment. (C) Multiple-sequence alignment of the three known T3SS structures of S. typhimurium PrgI, B. pseudomallei BsaL, and S. flexneri MxiH with the human Bcl-2 protein (Bcl-xL). The reliability of the each amino acid alignment is color-coded from blue (poor) to red (good) using the CORE index [35]. The consensus alignment received a score of 69, where a perfect alignment receives a score of 100. doi:10.1371/journal.pone.0007442.g002
in cancer cells [28]. Thus, the PrgI and Bcl-xL ligand binding-sites are functionally similar.

A pairwise structure alignment using DaliLite [29] yielded a non-significant Z-score of 1.4 and only 6% sequence identity between PrgI (PDB ID:2JOW) and Bcl-xL (PDB ID:1YSN). Nevertheless, the helix-turn-helix structure of PrgI (residues S13-V65) overlaps the buried helix-turn-helix motif (N136-I182) in Bcl-xL that corresponds to helices α5 (residues W137-D156) and α6 (residues L162-D176) (Figure 2B). A focused pairwise comparison between the full PrgI protein and the α5 and α6 helices of Bcl-xL gave a low but significant Z-score of 3.3 with a root-mean-square-difference (r.m.s.d) of 3.1 Å. The sequence identity also increases from 6% to 9% between the full and focused pairwise alignments, respectively.

While there is an overlap between the DaliLite alignment of PrgI with Bcl-xL and the protein ligand binding sites identified by CPASS, these sites are not identical. This arises because the CPASS similarity is not confined by the primary sequence of the two proteins, but simply captures the spatial orientation of conserved residues around a ligand binding site. This is illustrated by the non-sequential sequence alignment of the PrgI and Bcl-xL ligand binding sites in Figure 2. The exclusion of the sequence connectivity as a constraint to determine an alignment illustrates the advantage of CPASS in identifying a functional relationship and the limitation of global sequence and structure alignments such as Dali and BLAST [29,30].

Sequence Similarity Results

A BLAST [30] homology search using the PrgI and Bcl-xL sequences did not yield any significant information relating PrgI to Bcl-xL. The Bcl-xL sequence only identified homology to other Bcl-2 proteins. Similarly, the PrgI sequence was only aligned to other T3SS needle proteins. This is consistent with a ClustalW2 [31] sequence alignment between PrgI and Bcl-xL that resulted in a low 14.3% sequence similarity, which falls below the twilight zone of sequence similarity [32]. Also, focused BLAST searches did not provide any new information. Searching microbial genomes using the Bcl-2 sequences or searching the human genome with T3SS sequences did not identify any sequence alignments with significant E-values. Thus, global sequence alignments did not readily result in identifying any relationship between T3SS and apoptosis proteins. This highlights the power of active site similarity searches to identify potentially new functional similarities in proteins.

Hidden Markov model (HMM) methods [33] provide an alternative and more robust approach to identify homology between distantly related proteins with low sequence similarity relative to traditional BLAST searches. The T-Coffee web server (http://www.tcoffee.org/) provides a consensus sequence alignment (M-Coffee) using multiple HMM protocols [19]. A reliable alignment of conserved residues (Figure 2C) was obtained between the known T3SS structures of PrgI (PDB ID:2JOW), BsaL (PDB ID:2GOU) from Burkholderia pseudomallei, and MxiH (PDB ID:2CA5) from Shigella flexneri with the human Bcl-xL (PDB ID:1YSN) protein. The multiple-sequence alignment was obtained using EXPRESSO (3DCoffee) [34] that combines structural information with a HMM sequence alignment method. The reliability of the per residue alignment is color-coded using the CORE index (consistency of overall residue evaluation) [33], where the majority of residues were in the average to good range. The alignment of Bcl-xL with the three T3SS structures received a CORE index score of 53, where a score ≥50 indicates a 90% probability of being correctly aligned [19]. For comparison, the alignment of the three known human T3SS proteins resulted in a range of scores from 72 to 76. Thus, PrgI aligns preferentially to the other T3SS proteins, but its alignment to the pore forming helices in Bcl-xL is significant and reliable. Importantly, the sequence alignment of PrgI with Bcl-xL encompasses the same residues involved in the ligand binding sites identified by CPASS and the structural similarity identified by DaliLite.

Identification of a Second PrgI Ligand Binding Site

The identification of a compound that binds similarly to both PrgI and Bcl-xL would further establish a functional relationship between these two proteins. The BindingDB [36] was used to identify potential inhibitors of PrgI based on the CPASS predicted active site similarity with Bcl-xL. A total of 71 ligands were reported to bind Bcl-xL. A majority of the compounds were piperazin derivates and were not readily available. Two compounds, chelerythrine and sanquinarine were identified as having affinity to Bcl-xL and were both available from commercial suppliers. Chelerythrine was selected over sanquinarine based on previous NMR screening and docking studies that suggested chelerythrine binds between α4, α5 and α6 of Bcl-xL [37]. This region of Bcl-xL was predicted to overlap with PrgI based on the pairwise Dali alignment (Figure 2B). Conversely, sanquinarine bound the BH3 binding cleft of Bcl-xL and thus was not selected for this secondary binding analysis [37].

A comparison between the free and chelerythrine bound PrgI 2D 15N-1H HSQC spectra (Figure 3A) identified a chelerythrine binding site on PrgI (Figure 3B). The PrgI residues that exhibited chemical shift changes upon binding chelerythrine include residues A14, K15 in helix 1 and residues Y57, N59, A60, V65, K66, V67, F68, and D72 in helix 2. The AutoDock/ADF docked structure of PrgI with chelerythrine suggests PrgI residues K15 and Y57 are the most important residues for chelerythrine binding based on a close contact with the ligand (Figure 3B). Many of the residues that show significant CSPs for PrgI bound to chelerythrine overlap with the DDAB residues, however, the chelerythrine binding site is on the opposite face of PrgI (Figure 4). This indicates that there are two ligand binding sites on PrgI that is consistent with the two known protein-protein interaction sites for PrgI self-oligomerization. The chelerythrine AutoDock docking energy decreased significantly compared to DDAB, −0.43 to −5.29 kcal/mol, respectively.

The binding site of chelerythrine on PrgI is nearly identical to the binding site of chelerythrine to Bcl-xL (Figure 3C and 3D). In Bcl-xL, the chelerythrine binding site is described as being located in the BH groove of helix α4, α5 and α6, which is composed of residues F131, R132, V135, Y173 and H177 (Figure 3C) [37]. Pairwise structure analysis between PrgI and Bcl-xL shows that Y173 of Bcl-xL and Y57 on PrgI are overlapping residues and K15 from PrgI is proximal to V135 from Bcl-xL (Figure 3E). The primary difference between the two proteins is the lack of α6-helix 4 in PrgI, where helix 4 of Bcl-xL appears to act as a ‘cap’ encasing the ligand and effecting its relative binding orientation. Chelerythrine binds flat in the PrgI binding site, while the compound points into the corresponding Bcl-xL binding site partially overlaying helix α4. Again, both of these structures are docked models based on NMR CSPs and require a high-resolution x-ray or NMR structure to confirm the conformation of the chelerythrine binding site. It is paramount to note that this similarity in chelerythrine binding between the two proteins would have not been discovered if it was not for the identification of the initial conserved ligand binding site between PrgI and Bcl-xL using the FAST-NMR method in combination with the CPASS database.
Ligand Binding Similarity of the Bcl-2 Family of Proteins with PrgI

A structural and functional similarity between PrgI, a type three secretion system protein, and Bcl-xL, a member of the Bcl-2 family of proteins involved in eukaryotic apoptosis, was identified from a FAST-NMR ligand affinity screen in combination with a bioinformatic analysis. This association is fundamentally based on the similarity in ligand binding sites depicted in Figure 2A, where the conserved helix-turn-helix motif simply provides secondary support of a PrgI and Bcl-2 functional link. While similar active sites provide a measure of functional similarity, inferring homology based solely on the observation of a similar helix-turn-helix motif is questionable. The helix-turn-helix is a common motif and without a global sequence similarity, an evolutionary lineage based solely on active site similarity cannot be readily established. However, identifying similar ligand-binding sites between the two proteins does provide support that the proteins share a common function and are expected to bind similar ligands.

The initial identification of the conserved DDAB ligand binding site between Bcl-xL and PrgI was used to predict, test and confirm that chelerythrine binds PrgI in a similar manner to Bcl-xL. This further supports the structural and functional similarity between PrgI and Bcl-xL, but also demonstrates the utility of active site similarity as a predictive tool for ligand binding. Chelerythrine was

Figure 3. Verification that the Bcl-xL inhibitor chelerythrine also binds PrgI. (A). Expanded overlay of the 2D $^{15}$N-$^1$H HSQC spectra for free PrgI (black) and PrgI bound to chelerythrine (blue). CSPs greater than one standard deviation are boxed. (B) An AutoDock/ADF docked structure of PrgI complexed with chelerythrine based on the observed CSPs from (A). (C) The Bcl-xL region shown to bind chelerythrine is highlighted while the remaining protein structure is transparent. Chelerythrine is colored yellow and is drawn with licorice bonds. Side-chains for Y173 and V135 are shown as licorice bonds and colored grey. (D) A ribbon diagram of the AutoDock/ADF docked PrgI-chelerythrine co-structure. The PrgI-chelerythrine binding region that overlaps with Bcl-xL is highlighted. Chelerythrine is colored yellow and is drawn with licorice bonds. Side-chains for Y57 and K15 are shown as licorice bonds and colored grey. (E) An expanded view of the overlay of Bcl-xL (red) with PrgI (blue) illustrating the structural similarity of the chelerythrine binding sites.

doi:10.1371/journal.pone.0007442.g003

Discussion

Ligand Binding Similarity of the Bcl-2 Family of Proteins with PrgI

A structural and functional similarity between PrgI, a type three secretion system protein, and Bcl-xL, a member of the Bcl-2 family of proteins involved in eukaryotic apoptosis, was identified from a FAST-NMR ligand affinity screen in combination with a bioinformatic analysis. This association is fundamentally based on the similarity in ligand binding sites depicted in Figure 2A, where the conserved helix-turn-helix motif simply provides secondary support of a PrgI and Bcl-2 functional link. While similar active sites provide a measure of functional similarity, inferring homology based solely on the observation of a similar helix-turn-helix motif is questionable. The helix-turn-helix is a common motif and without a global sequence similarity, an evolutionary lineage based solely on active site similarity cannot be readily established. However, identifying similar ligand-binding sites between the two proteins does provide support that the proteins share a common function and are expected to bind similar ligands.

The initial identification of the conserved DDAB ligand binding site between Bcl-xL and PrgI was used to predict, test and confirm that chelerythrine binds PrgI in a similar manner to Bcl-xL. This further supports the structural and functional similarity between PrgI and Bcl-xL, but also demonstrates the utility of active site similarity as a predictive tool for ligand binding. Chelerythrine was

Figure 3. Verification that the Bcl-xL inhibitor chelerythrine also binds PrgI. (A). Expanded overlay of the 2D $^{15}$N-$^1$H HSQC spectra for free PrgI (black) and PrgI bound to chelerythrine (blue). CSPs greater than one standard deviation are boxed. (B) An AutoDock/ADF docked structure of PrgI complexed with chelerythrine based on the observed CSPs from (A). (C) The Bcl-xL region shown to bind chelerythrine is highlighted while the remaining protein structure is transparent. Chelerythrine is colored yellow and is drawn with licorice bonds. Side-chains for Y173 and V135 are shown as licorice bonds and colored grey. (D) A ribbon diagram of the AutoDock/ADF docked PrgI-chelerythrine co-structure. The PrgI-chelerythrine binding region that overlaps with Bcl-xL is highlighted. Chelerythrine is colored yellow and is drawn with licorice bonds. Side-chains for Y57 and K15 are shown as licorice bonds and colored grey. (E) An expanded view of the overlay of Bcl-xL (red) with PrgI (blue) illustrating the structural similarity of the chelerythrine binding sites.

doi:10.1371/journal.pone.0007442.g003

Discussion

Ligand Binding Similarity of the Bcl-2 Family of Proteins with PrgI

A structural and functional similarity between PrgI, a type three secretion system protein, and Bcl-xL, a member of the Bcl-2 family of proteins involved in eukaryotic apoptosis, was identified from a FAST-NMR ligand affinity screen in combination with a bioinformatic analysis. This association is fundamentally based on the similarity in ligand binding sites depicted in Figure 2A, where the conserved helix-turn-helix motif simply provides secondary support of a PrgI and Bcl-2 functional link. While similar active sites provide a measure of functional similarity, inferring homology based solely on the observation of a similar helix-turn-helix motif is questionable. The helix-turn-helix is a common motif and without a global sequence similarity, an evolutionary lineage based solely on active site similarity cannot be readily established. However, identifying similar ligand-binding sites between the two proteins does provide support that the proteins share a common function and are expected to bind similar ligands.

The initial identification of the conserved DDAB ligand binding site between Bcl-xL and PrgI was used to predict, test and confirm that chelerythrine binds PrgI in a similar manner to Bcl-xL. This further supports the structural and functional similarity between PrgI and Bcl-xL, but also demonstrates the utility of active site similarity as a predictive tool for ligand binding. Chelerythrine was
only tested for Prgl binding because of the proposed active site similarity with Bcl-xL. Thus, these studies have identified the first known ligands to bind Prgl (DDAB and chelerythrine). Both ligand binding sites are associated with the functionally important Prgl self-oligomerization sites. Therefore, compounds based on either the DDAB or chelerythrine scaffold may disrupt Prgl oligomerization. These compounds may serve as valuable chemical leads to develop novel antibiotics. Additionally, since the ligands bind in separate locations on the Prgl surface (Figure 4), the compounds present two distinct approaches for developing drugs targeting Prgl. Unfortunately, because chelerythrine also binds Bcl-xL it is reasonable to expect that an antibiotic designed using chelerythrine as a scaffold may produce undesirable off-target side effects. This issue may be minimized or eliminated by simply improving the Prgl binding affinity for chelerythrine derivatives. This illustrates another important feature of our FAST-NMR protocol; active site similarity is a useful tool to predict potential side effects due to off target inhibition, in addition to predicting potential drug leads. While computational methods for predicting potential drug toxicity [38] are useful because of their speed, validation requires experimental methods such as the FAST-NMR approach.

Functional Similarity of the Bcl-2 Family of Proteins with Prgl

The Bcl-2 family of proteins are essential for eukaryotic apoptosis; where Bcl-xL is responsible for repressing cell death activity [14]. The in vivo binding partners of Bcl-xL include the pro-apoptosis proteins Bax, Bak and Bid. It has been shown that expression levels of repressor (Bcl-xL) and pro-apoptosis proteins (Bax, Bak and Bid) are reciprocal in nature suggesting precise regulation of eukaryotic apoptosis [39]. A combination of mutational and structural based work has shown that the BH3 binding domain of Bcl-xL is critical for binding interactions with other Bcl-2 proteins and apoptosis regulation [39].

The structure of Bcl-xL very closely resembles the structures of Bax, Bid, Bcl-2, and other members of the Bcl-2 family of proteins, which all resemble pore-forming domains of bacterial toxins [40–42]. Bcl-2, Bcl-xL, Bax and the truncated active form of Bid (tBid) or inhibit (Bcl-xL) Bcl-xL oligomerization. This pore structure requires Bid to induce pore formation [46]. Thus, a protein interaction with other members of the Bcl-2 family is required to either promote (Bax) or inhibit (Bcl-xL) Bcl-xL oligomerization. It is also interesting that Prgl was found to bind to a lipid analog and lipids have been found to play a role in Bax oligomerization [46].

Importantly, the experimentally observed ligand-binding sites for both Prgl and Bcl-xL are functionally equivalent and within the conserved helix-turn-helix motif. Both sites correspond to functionally critical protein-protein interaction sites required for oligomerization and pore formation. The DDAB binding site on Prgl overlaps with key residues involved in Prgl oligomerization and needle assembly. Similarly, ABT-737 is an inhibitor of apoptosis and functions by inhibiting Bcl-xL protein interactions [51]. Thus, the similarity in the ligand-binding sites helps establish a functional link between the two proteins.

Structural Similarity of the Bcl-2 Family of Proteins with Prgl

The Bax pore-forming domain is conserved in Bcl-xL, Bcl-2 and Bid [40,52] and corresponds to the helix-turn-helix motif (helices α5 and α6) that was identified by CPASS to be similar to Prgl.
(Figure 2A). Also, a comparison of the Bcl-XL and PrgI structure by DaliLite resulted in the alignment of the PrgI structure with this conserved Bcl-2 helix-turn-helix motif (Figure 2B). Additionally, a multiple sequence alignment indicated a reliable similarity between T3SS needle-forming proteins and the Bcl-2 pore-forming region (Figure 2C). Thus, the PrgI structure can be viewed as a minimalistic version of the Bcl-2 structure, and it corresponds to the functionally essential and conserved core pore-forming domain.

Gene duplication along with insertion and/or deletions of substructures into variable genetic regions is known methods for the evolution of protein function [53,54]. These processes may explain the evolution of the Bcl-2 family of proteins from a smaller PrgI-like ancestor. Since the PrgI structure overlaps with residues N136 to I182, this may suggest N- and C-terminal insertions generated a Bcl-2 protein from a PrgI-like ancestor. This is consistent with the hypothesis proposed by Aouacheria et al. [55], where the ancestral toxic pore forming domain (helices β5 and β6) required developing a means to prevent inappropriate apoptosis and to regulate cell death.

Presumably, a main function of the N- and C-terminal inserts into a PrgI-like ancestor would be to stabilize the monomer form of Bax until an apoptotic signal occurs. In effect, the insertions would provide a stronger control over the pore formation process. This is consistent with what has been experimentally observed, both the N-terminus and C-terminus residues of Bax are essential to maintain the monomer form of Bax in the cytosol [42,56,57]. Deletion of the first 20 amino acids from the N-terminus results in Bax being localized to the mitochondria [56,57]. Similarly, the Bax structure indicates the C-terminal hydrophobic helix β9 is bent in a hydrophobic groove, but contains some critical solvent exposed polar residues that are necessary to maintain solubility [42]. In fact, a model for the translocation of Bax from the cytosol to the mitochondria requires a conformational change in Bax that opens up helix β9 and exposes the pore forming region composed of helices β5 and β6 [42,58]. Deletions of 21 residues from the C-terminus, which includes part of helix β6, prevents oligomerization [59].

While Bax oligomerizes to form a circular pore structure containing ~22 copies, this oligomerization process does not extend to form layers like the PrgI needle structure. The conformational change in Bax results in the globular domain remaining in the cytosol and sterically prevents oligomerization perpendicular to the membrane [60]. Thus, the structural insert that maintains a monomer Bax in the cytosol also prevents an unnecessary linear extension of the Bax oligomer out of the mitochondria membrane. Conversely, regulating PrgI oligomerization is not necessary since the assembly of the T3SS system is not detrimental to the cell. Therefore, a minimal pore-forming structure is all that is necessary for the T3SS system. The length of the PrgI needle is controlled by the proper assembly of the inner rod (PrgJ) that requires the InvJ protein [61]. The deletion of InvJ results in long non-functional needles.

An Evolutionary Relationship between T3SS and Eukaryotic Apoptosis?

Based on the observed similarity in the structure and function between PrgI and the Bcl-2 protein family it is tempting to hypothesize that the proteins share a common ancestor. The structural comparison of PrgI with the Bcl-2 family of proteins discussed above suggests a possible evolutionary path. A common ancestral protein has been suggested for the Bcl-2 protein family, where pore formation using helices β5 and β6 is the ancestral protein’s predicted primary function [55]. Similarly, T3SS are also predicted to evolve from a single gene [62] that is a simple but versatile export system [63]. Again, the helix-turn-helix is a common and ancient motif [64] demonstrating both its diverse utility and evolutionary stability. Thus, it is plausible that a simple and ancient PrgI-like protein could be an evolutionary precursor to both the Bcl-2 protein family and PrgI. It also appears unlikely that PrgI and the Bcl-2 protein family would evolve through a convergent process since the helix-turn-helix is such a simple and ancient motif [64] and essential to the function of both proteins.

Evolving a readily available helix-turn-helix protein into either PrgI or the Bcl-2 protein family seems like a simpler path than the conversion of a uniquely distinct fold to incorporate a core helix-turn-helix motif. Also, the evolution of proteins from simple structural components has been previously proposed [65] and is consistent with other general evolutionary trends where complex systems evolve from simpler systems [66].

By analogy, the sharing of a common ancestor by PrgI and the Bcl-2 family of proteins would imply an evolutionary relationship between the T3SS and eukaryotic apoptosis systems. T3SS is a prime example of a vestigial system and an important illustration of the stepwise evolution of the flagella machinery [67,68]. Therefore, it is reasonable to expect that other systems will be identified that share an evolutionary relationship with T3SS. T3SS is also an ancient system and clearly predates the origin of the mitochondria from prokaryote endosymbiosis [69,70]. α-proteobacteria [69], which are close relatives of the mitochondria, are known to contain T3SS [63,71,72]. Could an obsolete T3SS system contribute valuable components to the eukaryotic apoptosis system after endosymbiosis? An evolutionary link has already been observed between a mitochondrial and T3SS protein [73,74]. Furthermore, a detailed analysis of the origin of apoptotic proteins suggests a pivotal role for bacterial proteins in the evolution of eukaryotic apoptosis [75].

**Materials and Methods**

**FAST-NMR Screen of PrgI**

The *Salmonella typhimurium* type three secretion protein (T3SS) PrgI was screened with a functional library [23] using our FAST-NMR assay [5,6]. Unlabeled and 15N labeled monomeric PrgI was graciously provided by Dr. Roberto DeGuzman (University of Kansas) along with the assigned 15N-1H HSQC spectrum. Sample preparation and experimental parameters for the NMR screening were executed in the same manner as described previously [6]. A total of 113 1D 1H NMR line-broadening spectra were collected to identify 5 binders from the functional chemical library of 437 compounds. Measurement of binding dissociation constants were completed as described previously [25]. Secondary 2D 15N-1H HSQC NMR experiments were collected only for the 5 compounds identified as binders in the line-broadening experiments. Chemical shift perturbations (CSPs) (eq 1) from the 2D 15N-1H HSQC NMR experiments were used to identify the PrgI ligand binding site, where only residues with a CSP greater than one standard deviation from the mean were used:

\[
CSP = \left( \frac{\Delta NH^2 + \left( \frac{\Delta 15N}{5} \right)^2}{2} \right)^{1/2}
\]

where \(\Delta NH\) is the difference between free and bound 1H amide chemical shifts (ppm) and \(\Delta 15N\) is the difference between free and bound 15N chemical shifts (ppm).
Our rapid approach to determine a ligand binding orientation was employed to determine a PrgI co-structure in the same manner as described previously [24]. The CSPs minimize the search space by using a significantly reduced AutoDock 3D grid. AutoDock 4.0 was used to generate 100 docked PrgI-ligand co-structures using the Lamarckian search algorithm with a population size of 300 and 500,000 energy evaluations [27]. Our AutoDockFilter (ADF) program then uses an NMR energy function based on the magnitude of CSPs to select the best ligand conformation.

\[
E_{\text{NMR}} = k \sum \left( \frac{d_{\text{obs}} - d_{\text{exp}}}{d_{\text{exp}} < d_{i}} \right)^2 A_{\text{Dist}} = \left\{ \begin{array}{ll}
0 & \text{if } d_{i} \leq d_{\text{exp}} \\
\left( d_{\text{exp}} - d_{i} \right)^2 & \text{otherwise}
\end{array} \right.
\]

where ADF calculates a pseudo-distance \( d_{\text{exp}} \) based on the magnitude of the NH CSP, which is then compared to the shortest distance \( d_{i} \) between any atom in the residue that incurred an NH CSP and any atom in each docked ligand conformer. Comparison of these CSP-directed and selected ligand-docked structures with experimental x-ray and NMR structures has yielded an overall average rmsd of 1.72 ± 0.74 Å A [24].

A co-structure of the lipid derivative DDAB bound to PrgI was uploaded to the CPASS database (http://bionmr-cl.unl.edu/CPASS/cpass.htm) to identify proteins with similar ligand binding sites by maximizing an rmsd weighted BLOSUM62 [76,77] scoring function \( S_{\text{rmsd}} \).

\[
S_{\text{rmsd}} = \sum_{i,j} \left( \frac{d_{\text{rmsd}}} {d_{\text{rmsd}} < 1A} \right)^2 p_{ij} \cdot \text{Armsd}_{ij} = \left\{ \begin{array}{ll}
0 & \text{if } \text{rmsd}_{ij} \leq 1A \\
\left( \text{rmsd}_{ij} - 1 \right)^2 & \text{otherwise}
\end{array} \right.
\]

where active site \( a \) contains \( n \) residues and is compared to active site \( b \) from the CPASS database which contains \( m \) residues, \( p_{ij} \) is the BLOSUM62 probability for amino-acid replacement for residue \( i \) from active site \( a \) with residue \( j \) from active site \( b \), \( d_{\text{rmsd}} \) is the corrected root-mean-square difference in the Cα coordinate positions between residues \( i \) and \( j \), and \( d_{\text{min}} / d_{i} \) is the ratio of the shortest distance to the ligand among all amino-acids in the active site compared to the current amino-acid's shortest distance to the ligand. \( S_{\text{rmsd}} \) is only summed over the optimal alignment for residue \( i \) from active site \( a \) with residue \( j \) from active site \( b \). It is not summed over all possible combinations of \( i \) and \( j \). If the number of residues are not identical between active sites \( a \) and \( b \) (\( n \neq m \)), then the additional residues will not have a corresponding match. Each residue can only be used once in the alignment. If active site \( a \) contains unmatched residues, then no contribution is made to \( S_{\text{rmsd}} \) which effectively reduces the maximal possible score that can be achieved for active site \( a \). Currently, there are ~35,000 protein-ligand structures in the CPASS database. CPASS runs on a 16 node Beowulf Linux cluster, requires approximately 40 sec for each pair-wise comparison and takes ~24 hrs to complete a full search against the entire database.

Structure Similarity Searching

Native protein structures for PrgI (PDB ID: 2JOW) [15] and Bcl-XL (PDB ID: 1YSN) [28] were uploaded to the DaliLite [29] server (http://www.ebi.ac.uk/DaliLite/) to identify regions of structure homology between the two proteins. To identify structure similarity and possible homologies between other proteins within the PDB, the structures were also uploaded to the full Dali [10] web server (http://www.ebi.ac.uk/dali/). A truncated version of the Bcl-XL structure was generated by identifying the amino acids within regions of structure similarity and removing these residues from the native PDB file. The truncated PDB file was searched for regions of similarity using the DaliLite web server (http://www.ebi.ac.uk/Tools/dalilite/index.html).

Sequence Similarity Searching using BLAST and T-Coffee

Sequences from the T3SS and apoptosis regulation were downloaded from the NCBI server (http://www.ncbi.nlm.nih.gov/) and included PrgI (gi|16766179), InvG (gi|16766190) and InvG (gi|474941) from S. typhimurium, and Bcl-xL (gi|510901), Bak1 (gi|82571450), Bid (gi|4577361) and Bag (gi|231632) from Homo sapiens respectively. A full BLAST search was completed using these sequences associated with both systems as queries [30]. All BLAST sequence searches used default web settings. In addition, the sequences and structures for Bcl-xL (PDB-ID: 1YSN), S. typhimurium PrgI (PDB-ID: 2JOW), B. pseudomallei BsaL (PDB-ID: 2GOU) and S. flexneri MxiH (PDB-ID: 2CA5) obtained from the PDB were uploaded to the T-Coffee [19] web server (http://www.tcoffee.org/) to obtain a multiple sequence alignment using the EXPRESSO(3DCoffee) software [34]. Only the sequence region of the Bcl-xL structure that contained the pore-forming domain and yielded the highest alignment score was used for the multiple sequence alignment.

Secondary Binding Site Similarity between Bcl-xL and PrgI

To further support a structural and functional similarity between Bcl-xL and PrgI, the BindingDB [36] (http://www.bindingdb.org/) was searched for commercially available compounds to test for binding to PrgI. The free 15N-H HSQC spectrum was collected using 100 µM 15N labeled PrgI in 20 mM bis-Tris buffer with 150 µM sodium chloride at pH 7.5. A second PrgI sample was prepared in the same manner as above with the addition of 500 µM chelerythrine to generate the bound 15N-H HSQC spectrum. Chemical shift perturbations and a PrgI-chelerythrine docked co-structure were determined as described previously [24] and was compared to the Bcl-xL-chelerythrine model [37].

Acknowledgments

We would like to thank Dr. Roberto De Guzman from the University of Kansas for supplying the PrgI samples and the assigned 2D 15N-1H HSQC spectrum. We would also like to thank Dr. Henry Mok from the National University of Singapore for providing us with the coordinates for the Bcl-xL-chelerythrine model.

Author Contributions

Conceived and designed the experiments: RP. Performed the experiments: MDS. Analyzed the data: MDS RP. Wrote the paper: MDS RP.

References

1. Gerlt JA, Babitt PC (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. Annu Rev Biochem 70: 209–246.
2. Munro LA, Shakhnovich EI (1999) Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. J Mol Biol 291: 177–190.
3. Park K, Kim D (2008) Binding similarity network of ligand. Proteins 71: 960–971.
4. Campbell SJ, Gold ND, Jackson RM, Westhead DR (2003) Ligand binding: functional site location, similarity and docking.Curr Opin Struct Biol 13: 389–395.
5. Powers R, Mercier KA, Copeland JC (2006) The application of FAST-NMR for the identification of novel drug discovery targets. Drug Discov Today 13: 172–179.
6. Mercier KA, Baran M, Ramanathan V, Revez P, Xiao R, et al. (2006) FAST-NMR: functional annotation screening technology using NMR spectroscopy. J Am Chem Soc 128: 15292–15299.
7. Powers R (2007) Functional genomics and NMR spectroscopy. Comb Chem High Throughput Screen 10: 676–697.
8. Powers R, Copeland JC, Germer K, Mercier KA, Ramanathan V, et al. (2006) Comparison of protein active site structures for functional annotation of proteins against drug design. Trends Biotechnol 24: 124–135.
9. Cornelis GR, Van Gijsegem F (2000) Assembly and function of type III secretion systems. Annu Rev Microbiol 54: 735–774.
10. Kimbrough TG, Miller SI (2002) Assembly of the type III secretion needle complex of Salmonella typhimurium. Microbes Infect 3(10): 635–643.
11. Macnab RM (2003) How bacteria assemble flagella. Annu Rev Microbiol 57: 77–100.
12. Galan JE (2001) Salmonella interactions with host cells: type III secretion at work. Annu Rev Cell Dev Biol 17: 53–86.
13. Wang Y, Ouallet AN, Egan CW, Rathinavelan T, Ivan W, et al. (2007) Differences in the electrostatic surfaces of the type III secretion proteins PygB, BssK, and MobX. J Mol Biol 371: 1304–1314.
14. Duan JE, Rivero P, Cordes FS, Johnson S, Kenjale R, et al. (2006) Molecular model of a type III secretion system needle: Implications for host-cell sensing. Proc Natl Acad Sci U S A 103: 12529–12533.
15. Torrellas J, Jackson MW, Przewoski JW, Plano GV (2005) The Yersinia pestis type III secretion needle plays a role in the regulation of Yop secretion. Mol Microbiol 57: 1719–1734.
16. Dietmann S, Park J, Notredame C, Heger A, Lappe M, et al. (2001) A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3. Nucleic Acids Res 29: 55–57.
17. Poirier O, O'Toole E, Notredame C (2003) Tcoffee: a web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res 31: 5503–5506.
18. Han SL, Lee LC, Tan KO, Yang LK, Lee AS, et al. (2001) Identification of chelerythrine as an inhibitor of Bcl-XL function. J Biol Chem 276: 20453–20456.
19. Muller S, Feldman MF, Cornelis GR (2001) The type III secretion system of Gram-negative bacteria: a potential therapeutic target? Expert Opin Ther Targets 5: 527–539.
20. Mercier KA, Shortridge MD, Powers R (2009) A multi-step NMR screen for the identification and evaluation of chemical leads for drug discovery. Comb Chem High Throughput Screen 12: 203–209.
21. Mercier KA, Germer K, Powers R (2006) Design and characterization of a functional library for NMR screening against novel protein targets. Comb Chem High Throughput Screen 9: 513–534.
22. Stark J, Powers R (2000) Rapid protein-ligand construct uses chemical shift perturbations. J Am Chem Soc 122: 5334–5345.
23. Shortridge MD, Hage DS, Harbison GS, Powers R (2000) Estimating protein-ligand binding affinity using high-throughput screening by NMR. J Comb Chem 10: 948–958.
24. Kubori T, Sokhan A, Aizawa SI, Galan JE (2000) Molecular characterization and assembly of the needle complex of the Salmonella typhimurium type III protein secretion system. Proc Natl Acad Sci U S A 97: 10225–10230.
25. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. (1998) Automated docking: a protein modeling method. J Comput Chem 19: 1639–1662.
26. Ottersdorff T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, et al. (2002) The origin and early evolution of the Bcl-2 family. J Biol Chem 277: 5334–5345.
27. Holm L, Park J (2007) DALI Lite workbench for protein structure comparison. Nucleic Acids Res 35: 3389–3402.
28. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Chustal W and Chustal X version 2.0. Bioinformatics 23: 2947–2948.
29. Root B (1999) Twilight zone of protein sequence alignments. Protein Eng 12: 85–94.
30. Bystroff C, Krogh A (2008) Hidden Markov Models for prediction of protein features. Methods Mol Biol 413: 173–198.
31. Armougom F, Moretti S, Poirier O, Audic S, Dumaz P, et al. (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Clustal. 34: W604–W608.
32. Rost B (1999) Twilight zone of protein sequence alignments. Protein Eng 12: 85–94.
33. Zhang YH, Bhuiya A, Wan KF, Lee MC, Chan SL, et al. (2006) Chelerythrine and sanguinarine dock at distinct sites on Bcl-XL that are not the classic BH3 binding cleft. J Biol Chem 281: 536–549.
34. Xie L, Li J, Bourne PE, (2009) Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CELT inhibitors. PLoS Comput Biol 5: e1000387.
73. Pallen Mark J, Bailey Christopher M, Beaton Scott A (2006) Evolutionary links between Flh/YscL-like proteins from bacterial type III secretion systems and second-stalk components of the FoF1 and vacuolar ATPases. Protein Sci 15: 935–941.
74. Mulkidjanian AY, Makarova KS, Galperin MY, Koonin EV (2007) Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. Nat Rev Microbiol 5: 892–899.
75. Koonin EV, Aravind L (2002) Origins and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ 9: 394–404.
76. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89: 10915–10919.
77. Henikoff S, Henikoff JG (1993) Performance evaluation of amino acid substitution matrices. Proteins 17: 49–61.
78. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33–58, 27–38.
79. Nicholls A, Honig B (1991) A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J Comput Chem 12: 435–445.
80. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25: 1605–1612.