The BID domain of type IV secretion substrates forms a conserved four-helix bundle topped with a hook

Frédéric V. Stanger[1,2,5,6], Tjaart A.P. de Beer[1,6], David M. Dranow[3,6], Tilman Schirmer[2,*], Isabelle Phan[4] and Christoph Dehio[1,*]

[1]Focal Area Infection Biology and [2]Focal Area Structural Biology and Biophysics, Biozentrum, University of Basel, Basel, Switzerland

[3]Seattle Structural Genomics Center for Infectious Disease, Seattle, Washington, USA, and Beryllium Discovery Corp., Bainbridge Island, Washington, USA

[4]Seattle Structural Genomics Center for Infectious Disease, Seattle, Washington, USA, and The Center for Infectious Disease Research, Seattle, Washington, USA

[5]Present address: Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[6]Equal contribution

*Corresponding authors:

Prof. Christoph Dehio
Biozentrum, University of Basel
Klingelbergstrasse 70
CH-4056 Basel, Switzerland
Tel.: +41 61 267 2140
Fax.: +41 61 267 2118
e-mail: christoph.dehio@unibas.ch

Prof. Tilman Schirmer
Biozentrum, University of Basel
Klingelbergstrasse 70
CH-4056 Basel, Switzerland
Tel.: +41 61 267 20 89
Fax.: +41 61 267 2109
e-mail: tilman.schirmer@unibas.ch

Summary

The BID domain functions as secretion signal in a subfamily of protein substrates of bacterial type IV secretion (T4S) systems. It mediates transfer of (i) relaxases and the attached DNA during bacterial conjugation, and (ii) numerous *Bartonella* effector proteins (Beps) during protein transfer into host cells infected by pathogenic *Bartonella* species. Furthermore, BID domains of Beps have often evolved secondary effector functions within host cells. Here, we provide crystal structures for three representative BID domains and describe a novel conserved fold characterized by a compact, antiparallel four-helix bundle topped with a hook. The conserved hydrophobic core provides a rigid scaffold to a surface that, despite a few conserved exposed residues and similarities in charge distribution, displays significant variability. We propose that the genuine function of BID domains as T4S signal may primarily depend on their rigid structure, while the plasticity of their surface may facilitate adaptation to secondary effector functions.

Highlights

- **BID domains share a novel fold with a compact four-helix bundle and a hook**
- **The rigid fold is determined by a conserved core of hydrophobic amino acids**
- **Charged surface areas and few exposed residues are preserved**
- **Low surface conservation facilitates the evolution of secondary functions**

eTOC Blurb

Stanger, de Beer, Dranow *et al.* describe the novel BID domain fold, revealing a compact four-helix bundle. Their analyses suggest that the conserved shape of BID domains is critical to function as secretion signal, while low surface conservation facilitates the evolution of secondary functions.

Introduction

Bacterial type IV secretion (T4S) systems are supramolecular protein assemblies that mediate contact-dependent (i) inter-bacterial transfer of relaxases and the covalently attached DNA into recipient cells during bacterial conjugation and (ii) inter-kingdom transfer of host cell-targeted effector proteins of pathogenic bacteria such as *Helicobacter pylori*, *Legionella pneumophila*, *Brucella* spp. and *Bartonella* spp. (Christie et al., 2014). T4S substrates harbor a C-terminal non-cleavable T4S signal that is considered to interact with the T4S coupling protein (T4CP), a T4S system-associated ATPase (Christie et al., 2014). T4S signals are typically only a few tens of amino acids long and consist of clusters of positively charged or hydrophobic residues (Christie et al., 2014). However, a subfamily of T4S systems prominently found in the α-proteobacteria display T4S signals with a more complex bipartite structure composed of the approximately 140 amino acid long BID (Bep Intracellular Delivery) domain and a short positively charged C-terminal segment that is similar to the genuine T4S signal of other T4S systems (Schulein et al., 2005). T4CPs that are associated with BID domain-containing effectors form a monophyletic cluster within the phylogenetic tree of T4CPs (Schulein et al., 2005), indicating that this sublineage evolved and maintained specific adaptations to facilitate interaction with the BID domain.

BID domains are found in relaxases and in the Beps (*Bartonella* effector proteins) representing numerous host cell-targeted effectors of pathogenic *Bartonella* spp. Beps are translocated by the VirB T4S system and its associated T4CP VirD4 (Saenz et al., 2007). Beps have evolved by gene duplication, diversification and reshuffling from a single ancestor containing an N-terminal FIC (filamentation induced by cAMP) domain, a C-terminal BID-domain and a connecting central OB (oligonucleotide/oligosaccharide binding) fold, resulting in diverse derived modular architectures (Engel et al., 2011; Saenz et al., 2007). Providing a striking example of parallel evolution, this process occurred independently in two distinct lineages of *Bartonella*, giving rise to Bep1 to Bep10 in lineage

3 (L3) and BepA to BepJ in lineage 4 (L4) (Engel et al., 2011). In the N-terminal part, Beps harbor either a FIC-OB fold, tandem-repeat tyrosine-phosphorylation motifs or additional BID domains (Engel et al., 2011), which mediate diverse effector functions within host cells (Harms et al., 2016; Selbach et al., 2009; Siamer and Dehio, 2015). The original function of the C-terminal BID domain present in each Bep is to facilitate protein transfer via the VirB/VirD4 T4S system. However, several studies in the model pathogen *Bartonella henselae* and other L4 species showed that individual BID domains, including those in multi-BID domain architectures, have secondarily evolved discrete effector functions within host cells (Siamer and Dehio, 2015) that are considered to be mediated by specific protein-protein interactions with host proteins. The single BID domain of BepA binds to host adenylyl cyclase and potentiates GαS-dependent cyclic-AMP production, ultimately resulting in inhibition of apoptosis (Pulliainen et al., 2012). The BID domains of BepE are required for normal migration of host cells during infection *in vitro* and *in vivo* (Okujava et al., 2014), and those of BepF or BepG trigger actin-dependent uptake of bacterial aggregates into a unique cellular structure known as invasome (Rhomberg et al., 2009; Truttmann et al., 2011).

To pave the way for addressing structure/function-related questions concerning the BID domain, we determined crystal structures of BID domains from three different *Bartonella* effector proteins, describe the new fold, and analyze site-specific determinants for structure and potential function by sequence comparison.

Results

Structure determination

Crystals of the BID domains from *Bartonella rochalimae* Bep6 (*Bro*Bep6_tBID1), *Bartonella clarridgeiae* Bep9 (*Bcl*Bep9_tBID1) and *B. henselae* BepE (*Bhe*BepE_BID1) were obtained and complete datasets collected to a resolution of 2.2 Å or higher (Table 1).

Since there were no homologs in the PDB with significant sequence identity (>25%), the crystal structure of *Bro*Bep6_tBID1 was determined by SeMet-SAD phasing. The resulting model was then used for the structure determination of *Bcl*Bep9_tBID1 and *Bhe*BepE_BID1 by molecular replacement. Refinement yielded BID domain models with R_{work}/R_{free} (%) values of 17.7/20.3, 17.1/21.4 and 18.3/22.5, respectively. Data collection and final refinement statistics are given in Table 1.

BID domain structures and sequence conservation

All three BID domains are folded to an anti-parallel four-helix bundle and adopt an elongated shape with a length of 70 Å and a diameter of 25 Å (Figure 1B-D). Superposition of residues 319-413 of *Bro*Bep6_tBID1 with the corresponding residues of *Bcl*Bep9_tBID1 and *Bhe*BepE_BID1 yielded a root mean square deviation of 1.15 and 1.76 Å for 95 Cα-atoms (sequence identity of 36% and 21%, respectively). The three structures are virtually identical in their core but display significant conformational variability at the extremities of the polypeptide chain (Figure 1B-D and S1). Noteworthy, the first and last helices (α1 and α4, respectively) can adopt either a straight or a kinked conformation. In *Bhe*BepE_BID1 the kinks in helices α1 and α4 coincide with proline residues P154 and P253, respectively (Figure S1C). Except stated otherwise, in the following we take the BID structure of *Bro*Bep6 as reference and its corresponding residue numbering.

Strikingly, 3D structure comparison using the DaliLite server (Holm and Park, 2000) and several other servers (see Supplemental experimental procedures) revealed no significant structural homology to any other known structures. Some structural similarity between the BID domain structures and various other α-helix bundles or coiled-coils structures were found (best Z-scores between 6.6 and 6.8), but with superposition of only some of the α-helices at best. For example, the best hit for our reference structure is a four-helix bundle

of a histidine kinase sensor domain (PDB: 3I9Y) (Moore and Hendrickson, 2009), of which only one helix superposed with helix α4 of the BID domain. Thus, we can safely describe the BID domain as a novel fold.

The N- and C-termini of the BID domain are located at the same pole with the α3 - α4 linker located at the opposite pole (Figure 1, S1). This helix linker (residues 368-384 of *Bro*Bep6_tBID1) adopts a well-defined, apparently conserved structure (Figure 1E) and comprises a short 3_{10} helix (η1) and a β-hairpin (β1-β2). Because of its shape, we named this structure the hook. A detailed view of the hook of our reference structure, *Bro*Bep6_tBID1, is shown in Figure 1F. The η1 helix is initiated by a conserved proline (P368) and ends with a conserved serine (S370). S370 interacts with the main-chain carbonyl of residue 367 (immediately preceding the 3_{10} helix), which explains the conservation of this serine. The η1 helix is followed by a short loop that leads to the β1-β2 β-hairpin located at the top of the structure that further connects to helix α4. The hydrophobic side chain of residue L374 stabilizes the loop by burying itself in the BID core. Strikingly, residue G376 at the C-terminal end of this short loop is strictly conserved in BID domains, which allows the subsequent β-hairpin to closely pack with the N-terminal end of helix α4. The β1-β2 hairpin itself contains a regular reverse turn (β-turn) stabilized by an H-bond between carbonyl 379 (i) and amide 382 (i+3) as shown in Figure 1F. The β-turn is of type II' in *Bro*Bep6_tBID1 and of type I' in *Bcl*Bep9_tBID1 and *Bhe*BepE_BID1, with a glycine at position i+1 and i+2, respectively (Sibanda and Thornton, 1985). In addition to the main chain - main chain interactions between the β1 and β2 strands, the β-hairpin forms an H-bond between the side chain of S384 (i+5) with the main chain amide of residue 377 (i-2) of our reference structure (Figure 1F). Interestingly, a serine (S384 of *Bro*Bep6_tBID1), asparagine (N150 of *Bcl*Bep9_tBID1) or threonine (T218 of *Bhe*BepE_BID1) can occupy this position and allows the formation of the aforementioned interaction. A few residues at the surface of the hook appear partially conserved,

potentially for functional reasons. This includes two positively charged residues (K383 and K388), a negatively charged residue (E391) and two hydrophobic residues (I379 and I382). Noteworthy, the two aforementioned isoleucines are variable and any medium to large hydrophobic side-chain seems to be accommodated at these positions. We anticipate that the hook may constitute (part of) the interface for the contact with the T4CP. The sequence alignment of the three BID domains reveals only a few more conserved residues (Figure 1G). These are mostly hydrophobic, located at the center of the α-helical bundle and probably crucial for the integrity of the four-helix bundle (Figure 1G, H).

The BID fold: conserved but specialized

To gain insight into the conservation of the BID domain we performed a BLAST search and retrieved 351 sequences (with less than 90% redundancy), all Beps or relaxases. The neighbor-joining distance based tree of the BID domain sequences is clearly divided into two classes, representing relaxases and Beps (Figure 2B; a high resolution image with individual species names and sequence references is shown in Figure S3). Based on the Bep and relaxase multi-domain architectures (Figure 2A) as well as the clusters seen in the neighbor-joining tree, we devised a systematic nomenclature to classify the BID domains (Figure 2C). The BIDs are divided into terminal (tBIDx) and non-terminal BIDx, with “x” indicating the order of the BID domain from the N-terminus. For the Beps, the tBIDx class is subdivided into tBIDx domains found in either the “ancestral” FIC-OB-tBIDx architecture or the diverse “derived” domain arrangements. In relaxases, which have either one or two BID domains, the tBIDx class is subdivided into tBID1 and tBID2, while the BIDx class consists exclusively of BID1 domains. In Beps, the ancestral tBID1 subclass forms a distinct cluster, which, in comparison to other subclasses, is well conserved at both the N- and C-termini as revealed by the respective sequence logos in Figure 3A. It is tempting to speculate that these conserved residues may interact with the adjacent OB

fold as deduced by the short length (five residues) of the connecting segment. The "derived" tBIDx subclass, covering all remaining C-terminal BID domains of Beps, forms a separate cluster. Compared to ancestral tBID1, derived tBIDx have less conserved domain borders, which, due to the lack of an adjacent OB-fold, may reflect the lack of conserved inter-domain interactions. Given that only the most C-terminal BID domain constitutes the T4S signal (Schulein et al., 2005), the additional BID domains present in multi-BID domain Beps (BIDx class) are likely released from selection pressure and may thus more easily have adapted to novel functions. Consistent with this notion, the BIDx class does not form a uniform cluster (Figure 2B). The BID domains of relaxases cluster into three discrete subclasses tBID1, BID1 and tBID2 (Figure 2B). A few tBID1 domains derived from homologues of the VbhT toxin (Engel et al., 2012; Harms et al., 2015). Due to the low number of sequences, this group was not investigated further in the current study.

The overall conservation as well as the conservation within the six defined subclasses was mapped to the BID domain sequence and structure (Figure 3). Upon mapping the ConSurf conservation scores to the structure, it appears that the overall conservation at the surface (Figure 3B) is rather low compared to the conservation of the buried residues (median ConSurf score for surface residues is 5.0 *vs.* 8.0 for buried residues in *Bro*Bep6_tBID1*,* Figure 3C). However, a single hotspot appears highly conserved in all BID domains: the P_{368}xxxxxL_{374}[A/R/K]G_{376} motif located directly at the N-terminus of the short β-hairpin (β1-β2) at the tip of the BID domain (Figure 1F, 3B). This region, as already mentioned in the previous paragraph, may interact with the T4CP. Additionally, several prolines are very well conserved and appear critical for the proper folding of BID domains, *e.g.* P354 in *Bro*Bep6_tBID1 (structurally homologous to P120 in *Bcl*Bep9_tBID1 and P188 in *Bhe*BepE_BID1 in Beps) and P368 at the start of η1 in both Beps and relaxases.

When looking at the six BID domain subclasses individually, some specific conservation patterns appear. In the Beps, only the ancestral tBID1 subclass has a highly conserved

L_{298}IPxE_{302} motif right at the N-terminus that could potentially interact with the preceding OB fold. There is an additional conserved R_{425}xxxx[V/I]xxP_{433} motif located ~140 residues further downstream. In Beps, an additional R_{387}[K/R]xAE_{391} motif occurs right after the hook. In contrast, the relaxases are more conserved across the subclasses and share a very prominent [V/I]$_{429}$P[A/G]LS_{433} motif at the C-terminus. The relaxase BID1 class has a similar motif, L_{296}[I/L]PP_{299}, to the ancestral tBID1 domain whereas the relaxase tBID1 and tBID2 domains have a M_{296}[V/L]A[G/A]$_{299}$ motif. As in the Beps, there is a conserved motif after the hook, although in relaxases this motif is R_{387}xxA_{390}. For relaxases, residue numbers correspond to the numbering in the alignment of Figure 3A.

The sequence alignment combined with the three BID domain structures provide a solid basis to refine the boundaries of the BID domain described initially by Schulein *et al.* (Schulein et al., 2005). As a general domain definition for the BID fold (based on the structural superimposition), we propose to use the highly conserved proline located at the N-terminus of helix $\eta 1$ as an “anchor” (P368 in *Bro*Bep6_tBID1, P134 in *Bcl*Bep9_tBID1 and P202 in *Bhe*BepE_BID1) and define the domain boundaries as ~50 amino acids in the N-terminal direction and ~50 amino acids in the C-terminal direction, resulting in a BID domain of ~100 amino acids (Figure 1G, highlighted in beige). This domain definition is based on the best-conserved superimposable part of the BID domain and thus excludes the variable N-terminal part of helix $\alpha 1$ and the variable C-terminal part of $\alpha 4$. Noteworthy, the neighbor-joining distance based tree of BID domain trimmed to the new boundaries (Figure S4) closely resembles the tree shown in Figure 2B (compare Figure S3 and S4).

The sequence and structure analyses have shown low sequence conservation over the entire BID domain (on average ~14% between Beps and relaxases). To assess whether at least the surface charge distribution is conserved, electrostatic surface calculations were performed using APBS and PDB2PQR. Figure 4A-C shows that in all three BID domain structures, the electrostatic potential on the surface appears to be rather consistent. It

consists of two highly positively charged areas that are separated by a small patch of negative charges, mostly generated by E310, E317 and E344. The hook region in all three structures is highly positively charged, suggesting that it may interact with a negatively charged partner. This analysis was expanded to homology models of BID domains of other Beps, i.e. *Bhe*BepA_tBID1 (ancestral), *Bhe*BepE_tBID2 (derived) and *Bro*Bep9_BID1, revealing similar surface properties (Figure 4D-F), suggesting that this feature was acquired early during evolution of the BID domain. Due to low sequence similarity between the BID domains of Beps and relaxases (on average ~14%) and uncertain placement of indels (Figure 3A), no accurate relaxase homology models could be built for surface charge conservation comparisons.

Discussion

The C-terminal BID domain and adjacent positively charged tail sequence function as an evolutionary conserved bipartite signal for T4S in both Beps and a subset of relaxases (Schulein et al., 2005), likely by mediating protein-protein interaction with the T4CP as initial step of the T4S process (Schroder et al., 2011). The determination of three BID domain structures reveals a well-conserved novel fold formed by a four-helix bundle (Figure 1) lacking significant structural homology to known protein structures. The core of the domain is formed by highly conserved apolar residues that likely provide rigidity to the domain (Figure 3). Despite the constraint to maintain a functional T4S signal, residue conservation at the surface of the protein appears to be generally low, even though charge distribution is preserved (Figures 3 and 4). In particular, a conserved structural feature of the BID domain, that due to its shape we named the hook, is invariantly positively charged. Further to the positively charged tail sequence adjacent to the BID domain, we thus envisage that the hook may serve as anchoring point for conserved interactions of the BID domain with the T4CP. Future structure/function-related studies should characterize the interaction interface of the hook and possibly other surfaces of the BID domain with the T4CP and address possible cooperative binding of the adjacent tail sequence.

The discrete clustering of defined sub-classes of BID domains based on their origin (from relaxases or Beps) and position in multi-domain architectures (terminal or non-terminal) probably reflects more their evolutionary history than their function (Figure 2). The exception may be the less conserved non-terminal BID domains (BIDx) of Beps derived from domain duplication that – likely as they are relieved from selection pressure to maintain interaction with the T4CP – diversified faster facilitating the evolution of novel effector functions within host cells (Siamer and Dehio, 2015). However, also some of the more conserved terminal BID domains (tBIDx) have evolved secondary effector functions. As an example, the interaction of *Bh*BepA_tBIDx with the C2 subunit of human adenylyl

cyclase has been demonstrated (Pulliainen et al., 2012), but the interaction surface remains unknown.

The sole other study revealing the 3D structure of a secretion signal is of the TSA domain of the conjugative relaxase TraI of the R1 plasmid (Redzej et al., 2013), revealing a globular structure that is in contrast with the elongated BID domain. TSA forms, similarly to the BID domain, a domain with both termini in proximity. The BID domain (Figure S5A) shares some structural features with the unrelated oligomeric proteins IpaD (Figure S5B) and prefoldin (Figure S5C, D). IpaD consists of an α-helical bundle with a small β-sheet at its top (Fig. S5B) that is located at the tip of the type III secretion system's needle of *Shigella flexneri*. Upon IpaD oligomerization, IpaB and IpaC are recruited and translocated in a contact-dependent manner into host cell membranes to form a pore (Cheung et al., 2015). Interestingly, the N-terminal domain of IpaD acts as an intramolecular chaperone that prevents premature oligomerization (Johnson et al., 2007). Could the BID domain represent also an intramolecular chaperon? Secreted substrates can potentially cross the inner membrane through the VirB4/TrwK hexamer (Low et al., 2014). The inner diameter of the TrwK hexamer has been measured to 42 Å (Peña et al., 2012). Therefore, the dimensions of the BID domains, measured to an apparent width of ~25 Å, appear compatible with translocation via a T4S system in a fully folded state, while globular protein domains should require at least partial unfolding during translocation (Christie et al., 2014). BID domains may thus act as folding seeds to refold Beps or relaxases from their C-terminus, thereby acting as intramolecular chaperones upon transfer into target cells. Prefoldin found in archaea or eukaryotes is formed by two long coiled-coils topped by one or two β-sheets that are reminiscent of the hook of BID domains (Siegert et al., 2000) (Figure S5C). Prefoldin oligomerizes to form hexameric rings containing a large cavity that captures unfolded protein or folding intermediates (Siegert et al., 2000). Thus, it is tempting to speculate that multi-BID domain effectors, *e.g.* BepF or BepG (Figure 2A),

may form BID oligomers in a similar manner as prefoldin forms its cavity (Figure S5D).

Our structural work sets the stage for follow-up structure-function studies that will, amongst others, aim at investigating the interactions of the BID domain with (i) T4CP and other T4S system components during protein translocation, and upon translocation with (ii) other domains encoded in the same polypeptide chain, *e.g.* FIC domain, OB fold, tandem-repeat tyrosine-phosphorylation motifs and additional BID domains, as well as (iii) in cases of the evolution of secondary effector functions in individual BID domains, with their specific interaction partners.

Experimental Procedures

Protein expression and purification

The BID domains from *Bartonella rochalimae* Bep6 (*Bro*Bep6_tBID1), *Bartonella clarridgeiae* Bep9 (*Bcl*Bep9_tBID1) and *Bartonella henselae* BepE (*Bhe*BepE_BID1) were cloned, expressed and purified as described in detail in the Supplemental Experimental Procedures.

Protein crystallization, x-ray data collection and structure determination

*Bro*Bep6_tBID1, *Bcl*Bep9_tBID1 and *Bhe*BepE_BID1 were crystallized, x-ray data collected and structures determined as described in detail in the Supplemental Experimental Procedures. All data reduction and refinement statistics are reported in Table 1.

Sequence analysis

The sequence dataset used in this analysis was generated by BLAST against the UniProtKB database. All the sequences of our working dataset were then aligned using ClustalX 2.0. Neighbor-joining distance based trees were constructed and visualized with iTOL Sequence logos were generated with Weblogos and alignments visualized with Aline. Conservation scores were generated using ConSurf. Electrostatic potentials were calculated with the APBS-Tools and PDB2PQR plugins for PyMOL using the default settings. Details on the matrices and cut-off used and references are given in the Supplemental Experimental Procedures.

Accession Numbers

The coordinates and structure factors of *Bro*Bep6_tBID1, *Bcl*Bep9_tBID1 and *Bhe*BepE_BID1 have been deposited in the Protein Data Bank with accession numbers

4YK1, 4YK2 and 4YK3, respectively. Sequence abbreviations are given in the Supplemental Information.

Supplemental Information

Supplemental Information includes four figures, supplemental experimental procedures and supplemental references.

Author contributions

All authors contributed to the design of experiments. Proteins were cloned, expressed, purified and x-ray data were collected by the Seattle Structural Genomics Center for Infectious Diseases. D.M.D. crystallized the proteins, harvested the crystals, processed the x-ray data, determined and refined crystal structures. F.V.S. and T.A.P.d.B performed the phylogenetic and structural conservation analyses and the related bioinformatics analysis. C.D. provided input into the biological implications of the work, and the initial constructs were suggested by his laboratory. F.V.S., T.A.P.d.B., T.S. and C.D. wrote the manuscript with contributions from D.M.D. and I.P..

Acknowledgements

We acknowledge Peter J. Myler and the SSGCID, Seattle, US for support of the project. This work was supported by the NIH/NIAID contracts no. HHSN272200700057C and HHSN272201200025C, the ERC AdG FICModFun 340330 (to CD), and SNF grants 310030B-149886 (to CD), and 31003A_138414 (to TS). We thank the staff of beam-lines 08ID-1 of the Canadian Light Source (Saskatchewan, Canada) and 21-ID-G of the Advanced Photon Source (Argonne National Laboratory, USA) for their excellent support.

References

Cheung, M., Shen, D.K., Makino, F., Kato, T., Roehrich, A.D., Martinez-Argudo, I., Walker, M.L., Murillo, I., Liu, X., Pain, M.*, et al.* (2015). Three-dimensional electron microscopy reconstruction and cysteine-mediated crosslinking provide a model of the type III secretion system needle tip complex. Molecular microbiology *95*, 31-50.

Christie, P.J., Whitaker, N., and Gonzalez-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. Biochimica et biophysica acta *1843*, 1578-1591.

Engel, P., Goepfert, A., Stanger, F.V., Harms, A., Schmidt, A., Schirmer, T., and Dehio, C. (2012). Adenylylation control by intra- or intermolecular active-site obstruction in Fic proteins. Nature *482*, 107-110.

Engel, P., Salzburger, W., Liesch, M., Chang, C.C., Maruyama, S., Lanz, C., Calteau, A., Lajus, A., Medigue, C., Schuster, S.C.*, et al.* (2011). Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen Bartonella. PLoS genetics *7*, e1001296.

Harms, A., Stanger, F.V., and Dehio, C. (2016). Biological Diversity and Molecular Plasticity of FIC Domain Proteins. Annual review of microbiology *70*.

Harms, A., Stanger, F.V., Scheu, P.D., de Jong, I.G., Goepfert, A., Glatter, T., Gerdes, K., Schirmer, T., and Dehio, C. (2015). Adenylylation of Gyrase and Topo IV by FicT Toxins Disrupts Bacterial DNA Topology. Cell reports *12*, 1497-1507.

Holm, L., and Park, J. (2000). DaliLite workbench for protein structure comparison. Bioinformatics *16*, 566-567.

Johnson, S., Roversi, P., Espina, M., Olive, A., Deane, J.E., Birket, S., Field, T., Picking, W.D., Blocker, A.J., Galyov, E.E.*, et al.* (2007). Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. The Journal of biological chemistry *282*, 4035-4044.

Low, H.H., Gubellini, F., Rivera-Calzada, A., Braun, N., Connery, S., Dujeancourt, A., Lu, F., Redzej, A., Fronzes, R., Orlova, E.V., *et al.* (2014). Structure of a type IV secretion system. Nature *508*, 550-553.

Moore, J.O., and Hendrickson, W.A. (2009). Structural analysis of sensor domains from the TMAO-responsive histidine kinase receptor TorS. Structure *17*, 1195-1204.

Okujava, R., Guye, P., Lu, Y.Y., Mistl, C., Polus, F., Vayssier-Taussat, M., Halin, C., Rolink, A.G., and Dehio, C. (2014). A translocated effector required for Bartonella dissemination from derma to blood safeguards migratory host cells from damage by co-translocated effectors. PLoS pathogens *10*, e1004187.

Peña, A., Matilla, I., Martín-Benito, J., Valpuesta, J.M., Carrascosa, J.L., la Cruz, De, F., Cabezon, E., Arechaga, I. (2012). The hexameric structure of a conjugative VirB4 protein ATPase provides new insights for a functional and phylogenetic relationship with DNA translocases. The Journal of biological chemistry 287, 39925–39932.

Pulliainen, A.T., Pieles, K., Brand, C.S., Hauert, B., Bohm, A., Quebatte, M., Wepf, A., Gstaiger, M., Aebersold, R., Dessauer, C.W., *et al.* (2012). Bacterial effector binds host cell adenylyl cyclase to potentiate Galphas-dependent cAMP production. Proc. Natl. Acad. Sci. USA *109*, 9581-9586.

Redzej, A., Ilangovan, A., Lang, S., Gruber, C.J., Topf, M., Zangger, K., Zechner, E.L., Waksman, G. (2013). Structure of a translocation signal domain mediating conjugative transfer by type IV secretion systems. Molecular microbiology 89, 324–333.

Rhomberg, T.A., Truttmann, M.C., Guye, P., Ellner, Y., and Dehio, C. (2009). A translocated protein of Bartonella henselae interferes with endocytic uptake of individual bacteria and triggers uptake of large bacterial aggregates via the invasome. Cellular microbiology *11*, 927-945.

Saenz, H.L., Engel, P., Stoeckli, M.C., Lanz, C., Raddatz, G., Vayssier-Taussat, M., Birtles, R., Schuster, S.C., and Dehio, C. (2007). Genomic analysis of Bartonella identifies type IV secretion systems as host adaptability factors. Nature genetics *39*, 1469-1476.

Schroder, G., Schuelein, R., Quebatte, M., and Dehio, C. (2011). Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen Bartonella henselae. Proc. Natl. Acad. Sci. USA *108*, 14643-14648.

Schulein, R., Guye, P., Rhomberg, T.A., Schmid, M.C., Schroder, G., Vergunst, A.C., Carena, I., and Dehio, C. (2005). A bipartite signal mediates the transfer of type IV secretion substrates of Bartonella henselae into human cells. Proc. Natl. Acad. Sci. USA *102*, 856-861.

Selbach, M., Paul, F.E., Brandt, S., Guye, P., Daumke, O., Backert, S., Dehio, C., and Mann, M. (2009). Host cell interactome of tyrosine-phosphorylated bacterial proteins. Cell host & microbe *5*, 397-403.

Siamer, S., and Dehio, C. (2015). New insights into the role of Bartonella effector proteins in pathogenesis. Current opinion in microbiology *23*, 80-85.

Sibanda, B.L., and Thornton, J.M. (1985). Beta-hairpin families in globular proteins. Nature *316*, 170-174.

Siegert, R., Leroux, M.R., Scheufler, C., Hartl, F.U., and Moarefi, I. (2000). Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. Cell *103*, 621-632.

Truttmann, M.C., Guye, P., and Dehio, C. (2011). BID-F1 and BID-F2 domains of Bartonella henselae effector protein BepF trigger together with BepC the formation of invasome structures. PloS one *6*, e25106.

Figure Legends

Figure 1. Crystal structures of the BID domain of *Bartonella* effector proteins reveal a novel fold. (A) Domain architectures of Bep6 from *B. rochalimae* (*Bro*Bep6_tBID1), Bep9 from *B. clarridgeiae* (*Bcl*Bep9_tBID1) and BepE from *B. henselae* (*Bhe*BepE_BID1) (from left to right). Vertical bars represent tandem-repeat tyrosine-phosphorylation motifs. Side views of the crystal structures of (B) *Bro*Bep6_tBID1, (C) *Bcl*Bep9_tBID1 and (D) *Bhe*BepE_BID1. (E) Superposition of the three structures shown in (B), (C) and (D), highlighting the similarity and compact nature of the BID fold. A stereoview is provided in Figure S1D. (F) Detailed structure of the α3-α4 linker, referred to as the hook, of *Bro*Bep6_tBID1 as representative for the three BID domain structures. The Cα-trace is shown as cartoon and H-bonds are shown as black dashed lines. The position of glycine residues is highlighted by spheres at their respective Cα positions. Residues discussed in the text are shown in stick representation. Secondary structure elements are labeled in light pink. (G) Sequence logo of three structures shown in panels B-D with secondary structure elements of the reference structure indicated. Residues of structural importance are marked with black triangles and residues of potential functional relevance are marked with red triangles. Additionally, green arches indicate kinks in helices α1 and α4. The core of the BID domain is highlighted in beige. (H) Structure of *Bro*Bep6_tBID1 with residues conserved amongst the three structures shown as sticks. See also Figure S1 and S2.

Figure 2. BID domain arrangements and subclasses in Beps and relaxases.
(A) The BID domain architecture as seen in two relaxases (TraA from *Agrobacterium fabrum* and Riorf112 from *Agrobacterium rhizogenes*), a VbhT homologue (VbhT from *B. schoenbuchensis*) and representative *B. henselae* Beps, with vertical black lines indicating Tyr phosphorylation motifs and +++ the positively charged C-terminus. (MobA:

mobilisation proteins of the MobA/MobL family, AAA_30: ATPases Associated with diverse cellular Activities (AAA) domain, FIC: Filamentation induced by cAMP domain, OB: Oligonucleotide/Oligosaccharide binding fold). (B) Simplified neighbor-joining distance based tree representation of the multiple sequence alignment of the BID domains from Beps and relaxases. The branches corresponding to Bep_tBIDx (ancestral) are colored in pink, Bep_tBIDx (derived) in blue and Bep_BIDx in green (Engel et al., 2011). The relaxase tBID1 domains are colored in yellow, tBID2 in brown and the BID1 in red. See Figure S3 for the full high-resolution tree shown in panel (B) with species names and UniProt IDs. (C) Classification scheme of BID domains developed in this study. See also Figure S3 and S4.

Figure 3. Conservation analysis of BID domains. (A) Comparison of residue conservation in BID subclasses defined in Figure 2C. For each subclass, the sequence logo is shown and the background is colored using the ConSurf conservation color scale ranging from cyan to magenta for low to high conservation. An overall sequence logo is shown above the subclasses that indicates only a low amount of globally conserved residues. (B-C) Overall conservation score mapped to the *Bro*Bep6_tBID1 structure. (B) Surface representation colored by overall conservation. The L_{374}[A/R/K]G_{376} motif appears at the start of the hook. (C) Ribbon representation with conserved residues (ConSurf score ≥ 8) shown as sticks. (D) Surface representation of *Bro*Bep6_tBID1 colored by conservation of the BID domain in an ancestral tBID1 arrangement (last row of the alignment in panel A). The N- and C-terminus appear well conserved when compared to the overall conservation shown in (B).

Figure 4. Electrostatic potential of experimentally determined and modeled BID domain structures. The potentials have been mapped onto the respective protein

surfaces with color code as defined in the inset. Protein backbones are shown in cartoon representation. (A) *Bro*Bep6_tBID1 (ancestral) (PDB: 4YK1), (B) *Bcl*Bep9_tBID1 (derived) (PDB: 4YK2), (C) *Bhe*BepE_BID1 (PDB: 4YK3), and homology models of (D) *Bhe*BepA_tBID1 (ancestral), (E) *Bhe*BepE_tBID2 (derived) and (F) *Bro*Bep9_BID1.

Figure 1

Click here to download Figure Figure1.tif

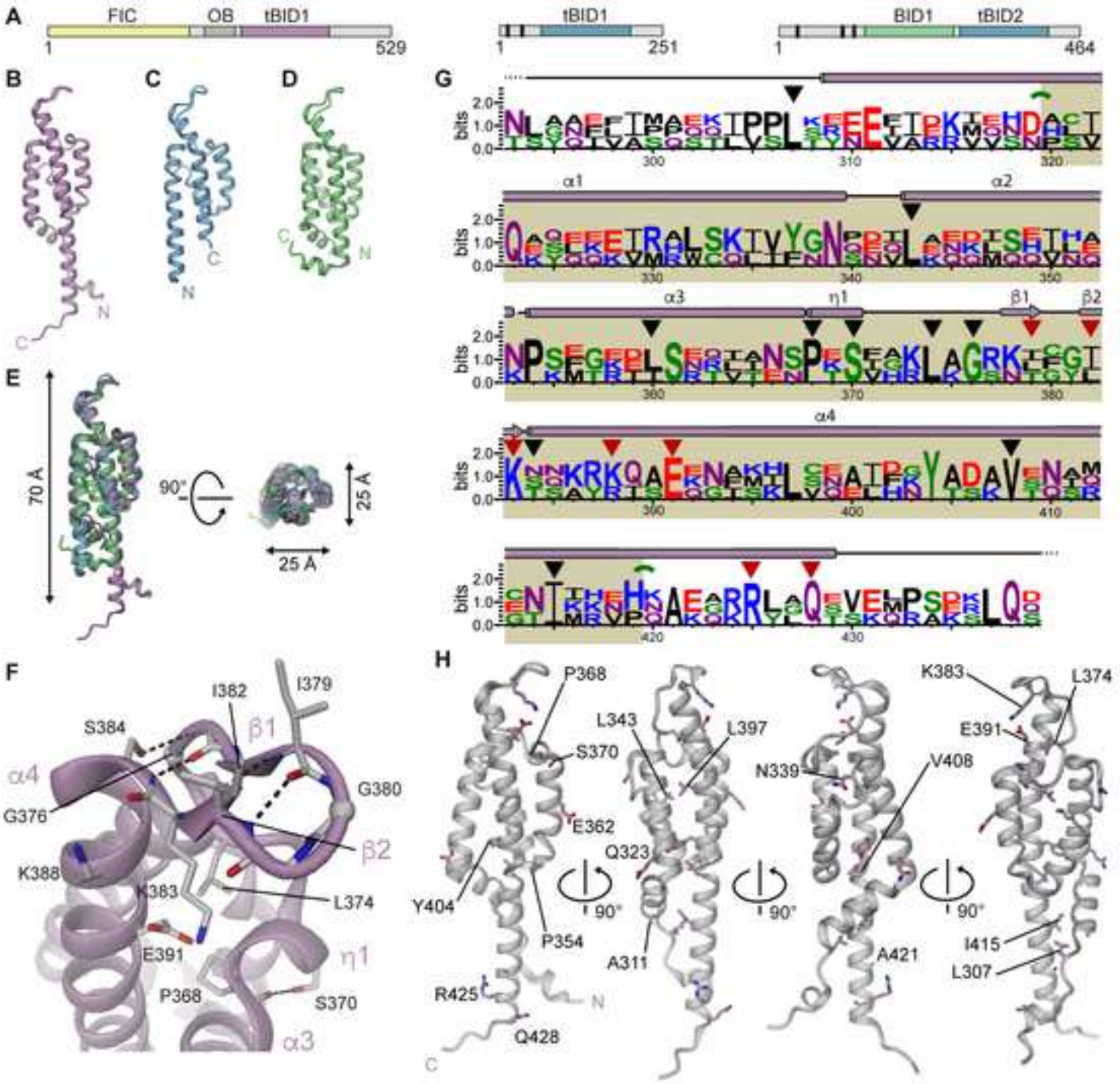

Figure 2

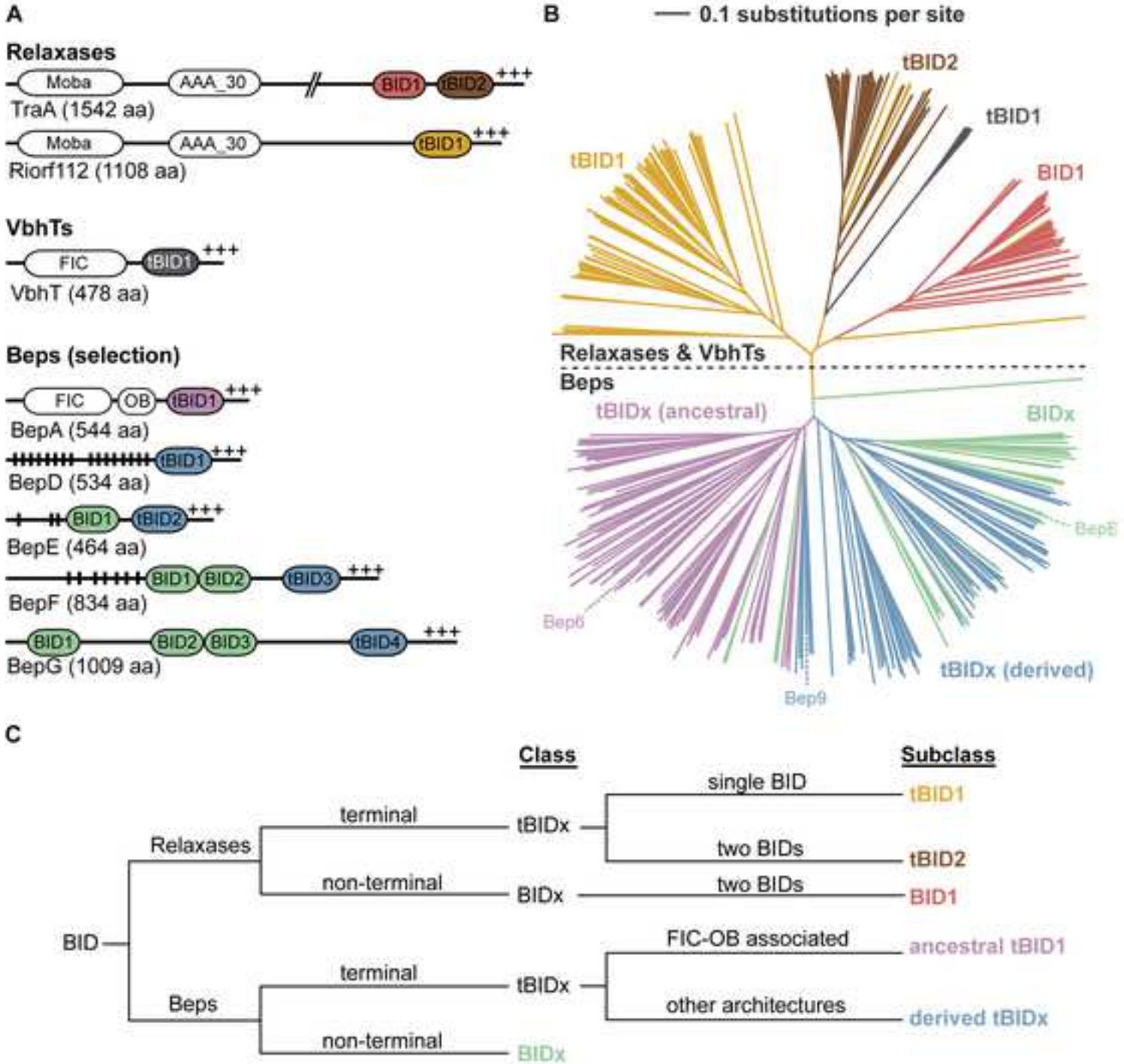

Figure 3

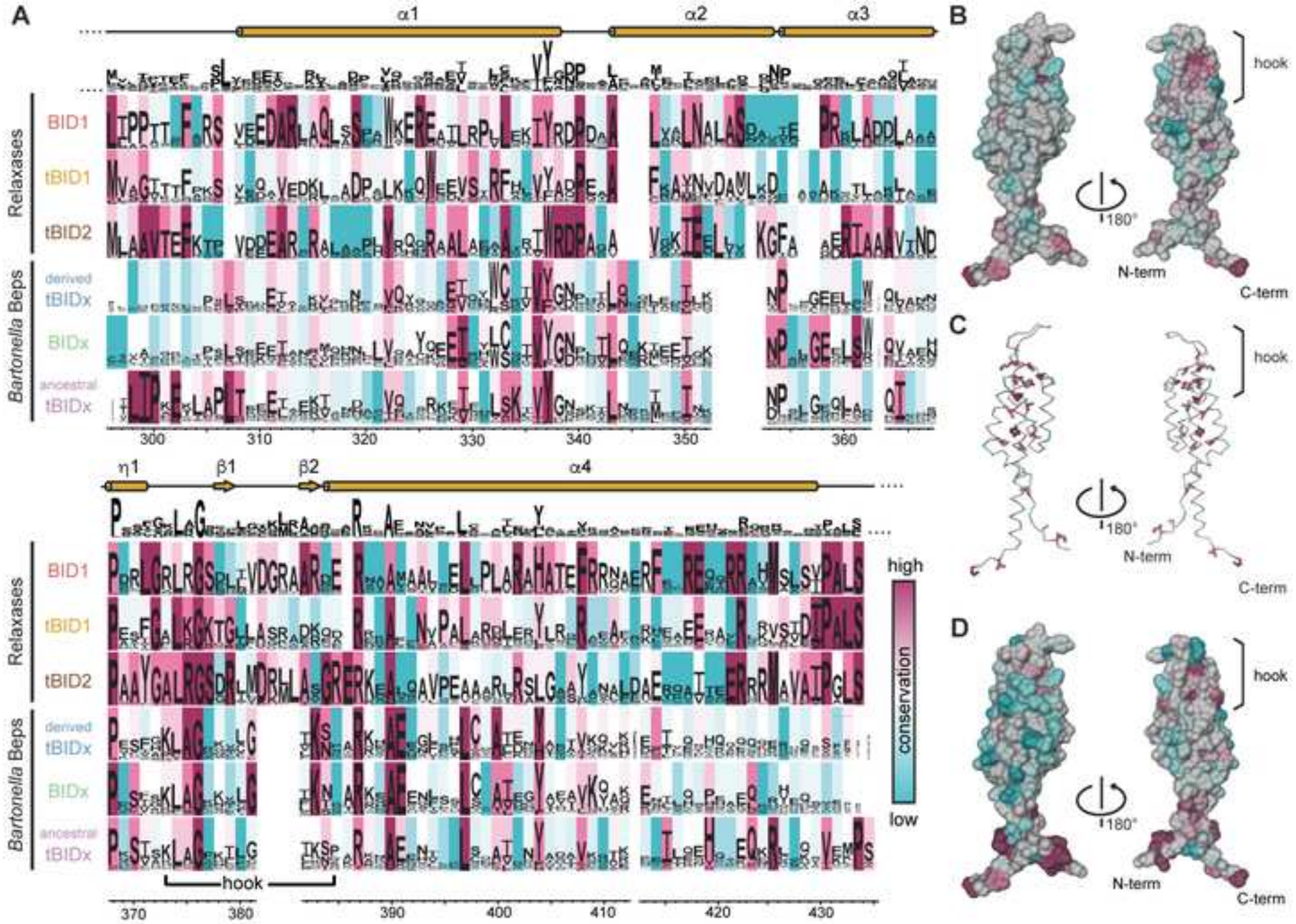

Figure 4

Click here to download Figure Fig4_final_rev2-01.tif

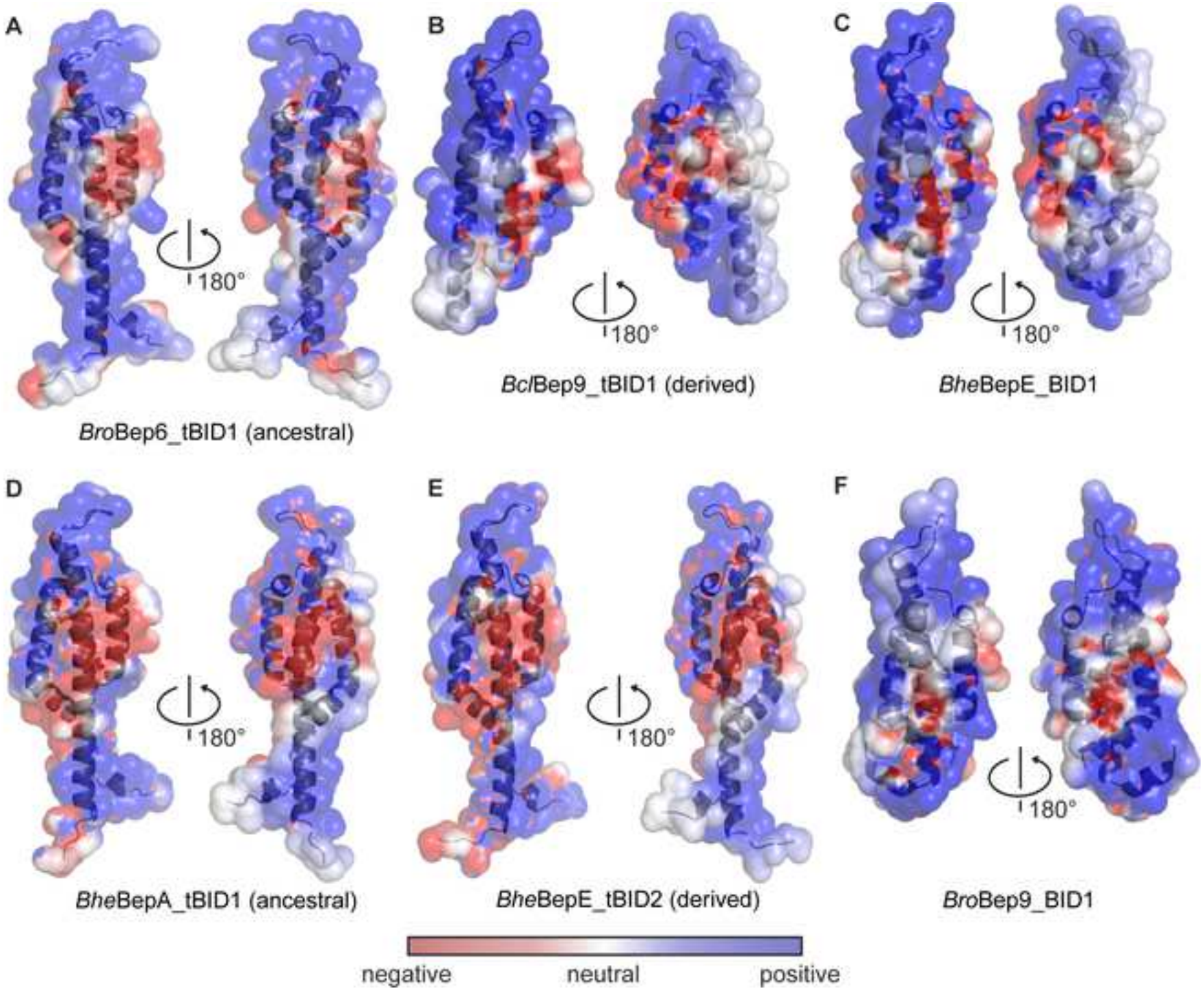

Table 1. Crystallographic table.

	*Bro*Bep6_tBID1	*Bcl*Bep9_tBID1	*Bhe*BepE_BID1
PDB ID code	4YK1	4YK2	4YK3
Data collection statistics			
X-ray source	CLS 08ID-1	APS 21-ID-G	APS 21-ID-G
X-ray detector	Mar300 CCD	Mar300 CCD	Mar300 CCD
Wavelength (Å)	0.9796	0.9786	0.9786
Space group	P 4_1 2 2	P 2_1 2_1 2	P 1 2_1 1
Cell dimensions a,b,c (Å), β (°)	79.25, 79.25, 85.97, 90.00	76.68, 62.31, 71.04, 90.00	58.07, 77.42, 67.51, 93.66
Resolution limits (Å)	50.0 - 2.10 (2.15 - 2.10)	50.0 - 2.05 (2.10 - 2.05)	50.0 - 2.20 (2.26 - 2.20)
R_{merge} † (%)	7.0 (58.2)	6.1 (54.8)	4.4 (52.9)
R_{meas} ‡ (%)	7.2 (60.2)	6.8 (61.1)	5.0 (59.6)
CC ½ (%)	100 (94.3)	99.9 (89.0)	99.9 (84.2)
⟨I / σ(I)⟩	29.99 (5.32)	19.33 (3.18)	23.49 (2.91)
Wilson B-factor	25.9	19.7	33.7
Total reflections	473’920 (34’650)	106’806 (7’902)	140’015 (10’368)
Unique reflections	30’572 (2’234)	21’863 (1’597)	30’337 (2’230)
Multiplicity	15.5 (13.7)	4.9 (4.9)	4.6 (4.6)
Completeness (%)	100.0 (100.0)	99.2 (99.3)	99.8 (99.9)
Mosaicity (°)	0.190	0.161	0.273
Refinement statistics			
R_{work} * (%)	17.7 (20.0)	17.1 (21.2)	18.3 (23.2)
R_{free} ** (%)	20.3 (22.4)	21.4 (24.0)	22.5 (25.5)
No. of non-H atoms	1’174	1’943	3’340
Macromolecules	1’075	1’733	3’204
Ligands	0	39	12
Solvent	99	171	124
Protein residues	137	213	434
rmsd from ideal			
Bond lengths (Å)	0.008	0.007	0.008
Bond angles (°)	0.93	0.98	0.97
Ramachandran favored *** (%)	99	100	99
Ramachandran allowed *** (%)	0.72	0.45	1.2
Ramachandran outliers *** (%)	0	0	0
Clashscore ***	2.77	3.11	2.91
Average B values ($Å^2$)	42.51	37.04	54.01
Macromolecules	42.17	34.95	54.19
Ligands		110.70	61.34
Solvent	46.23	41.36	48.63
TLS groups	4	9	18

Numbers in parentheses refer to the outmost shell.

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection and $\langle I(hkl)\rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

‡ $R_{meas} = \sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection, $\langle I(hkl)\rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections and N is the number of observations of intensity I(hkl).

* $R_{work} = \sum_{hkl}|| F_{obs}| - |F_{calc}|| / \sum_{hkl}|F_{obs}|$

** R_{free} is the R value calculated for 5% of the data set that was not included in the refinement.

*** Molprobity

Inventory of Supplemental Information

Figure S1, related to Figure 1.

Figure S2, related to Figure 1.

Figure S3, related to Figure 2.

Figure S4, related to Figure 2.

Figure S5, related to Discussion.

Supplemental experimental procedures.

Supplemental references.

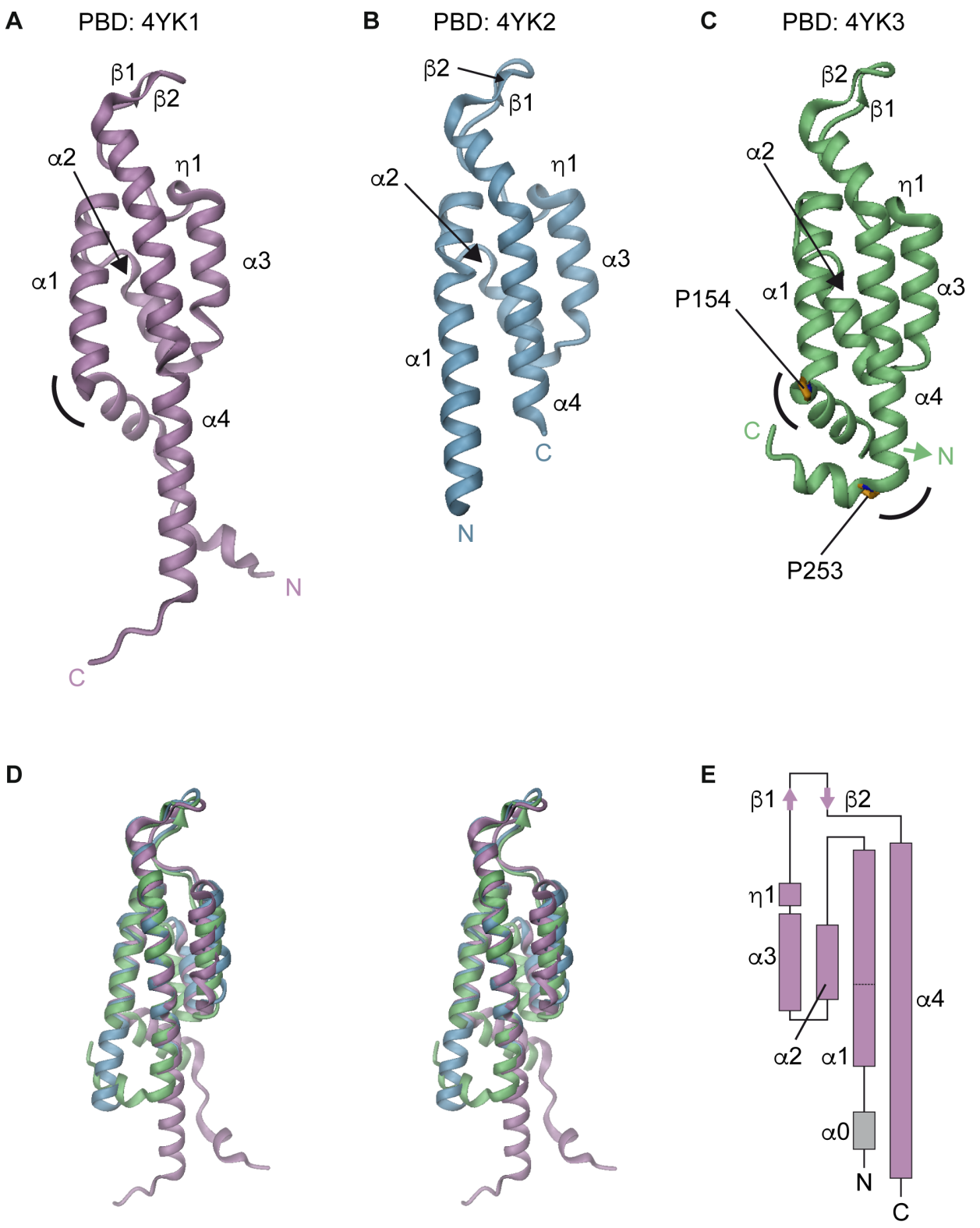

Figure S1, related to Figure 1. Secondary structure element annotations for the three BID structures. The BID domain structures are colored according to their BID domain subclass (see Figure 1A and 3B) and the numbering of the secondary structure

elements is shown. (A) tBID1 domain from *Bro*Bep6, (B) tBID1 domain from *Bcl*Bep9 and (C) BID1 domain from *Bhe*BepE. Prolines P154 and P253 located respectively at the kink of helices α1 and α4 are shown as sticks with their carbon atoms colored in orange. A black arch indicate the position of the kinks in helix α1 and α4 in panels A and C. (D) Stereoview of the three superposed BID domains shown in (A-C). (E) Topology diagram of the structure of *Bro*Bep6_tBID1 (PDB: 4YK1) shown in panel A.

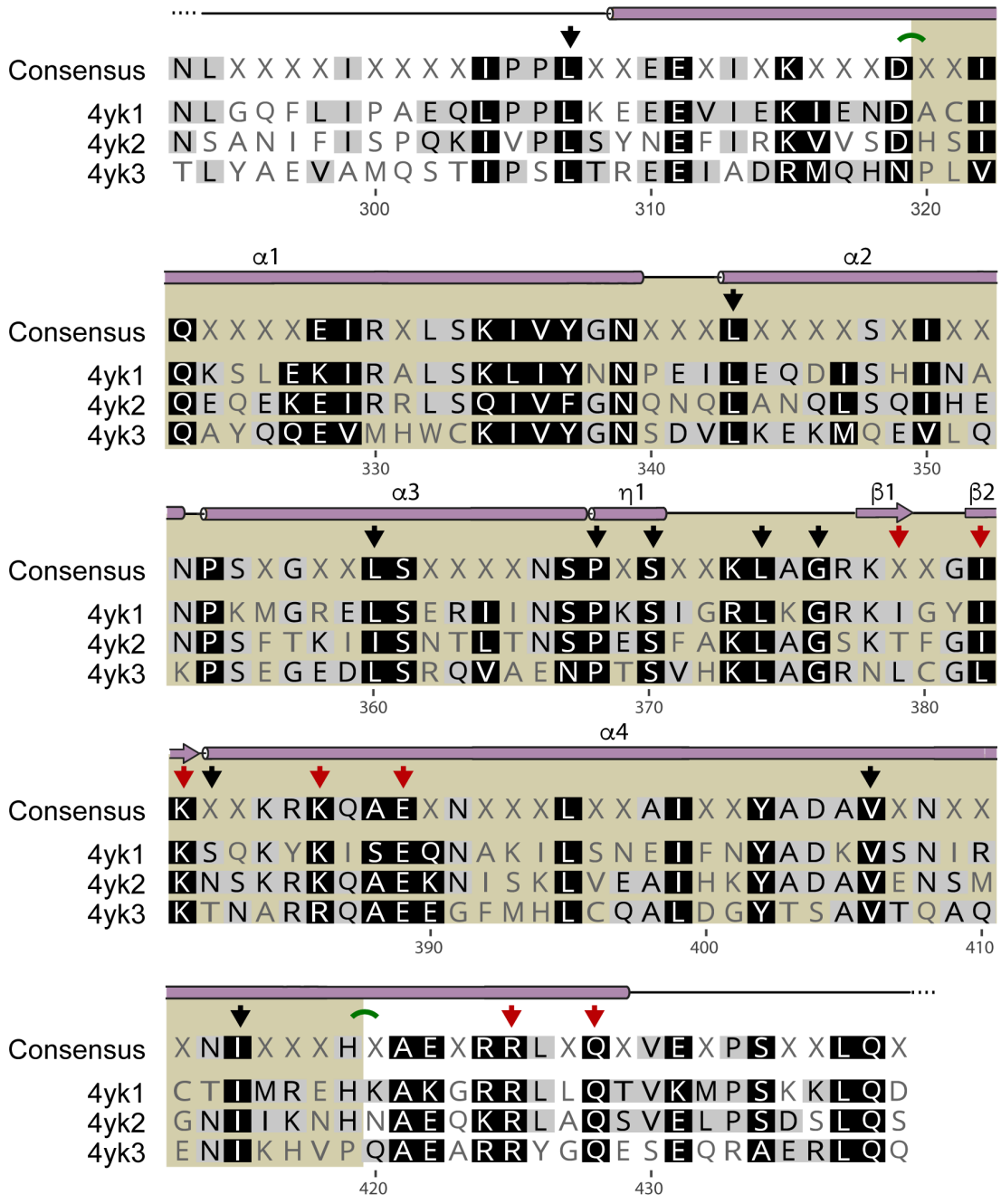

Figure S2, related to Figure 1. Sequence alignment of the three structures determined in this study. The consensus and secondary structure elements of the reference structure (PDB: 4YK1) are indicated on top of the alignment. Residues of structural importance are marked with black triangles and residues of potential functional relevance are marked with red triangles. Additionally, green arches indicate kinks in helices α1 and α4. The core of the BID domain is highlighted in beige.

Figure S3, related to Figure 2. High-resolution image of the neighbor-joining distance based tree shown in Figure 2. This tree includes the individual species names and sequence references.

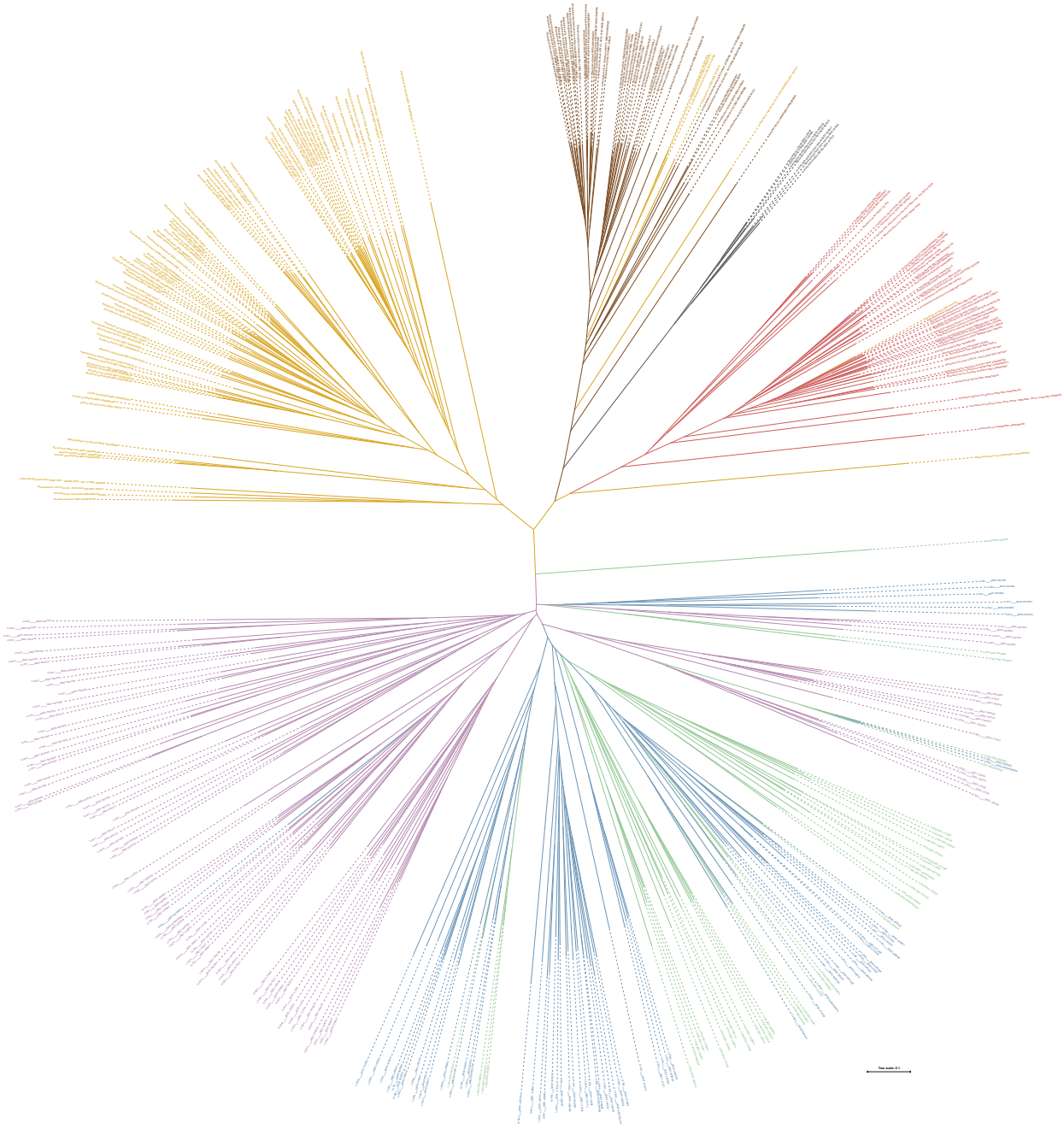

Figure S4, related to Figure 2. High-resolution image of the neighbor-joining distance based tree of the core BID domain. This tree includes the individual species names and sequence references**.**

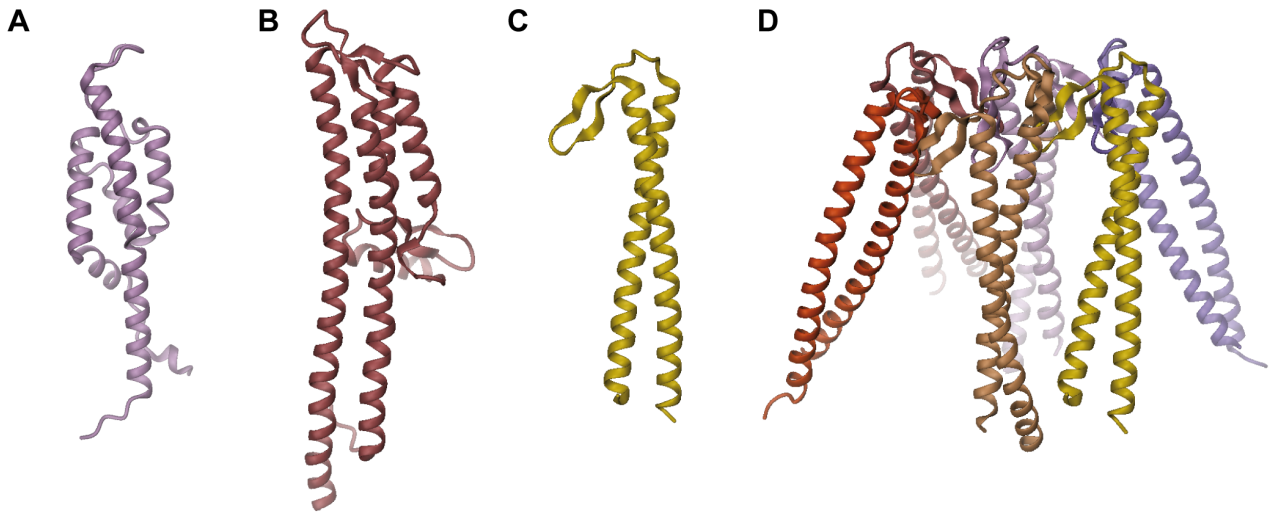

Figure S5, related to Discussion. Structural comparison of the BID domain with resembling folds. (A) Structure of the BID domain of *Bro*Bep6 as shown in Figure 1A as reference. (B) Monomer of IpaD as observed in the tetrameric structure of IpaD from *Shigella flexneri* (PDB:4D3E) in a similar orientation as the BID domain shown in A. For clarity, only the monomer is shown. (C) Structure of one monomer of archaeal prefoldin (*Methanothermobacter thermautotrophicus*) as observed in the hexametric structure (PDB: 1FXK) displayed in (D).

Supplemental experimental procedures

Protein Expression and Purification

The BID domains from three different *Bartonella* effector proteins, *Bartonella rochalim*ae Bep6 (UniProt: E6YLF3 residues 298-434), *Bartonella clarridgeiae* Bep9 (UniProt: E6YIM5 residues 64-201), and *Bartonella henselae* BepE (UniProt: Q5QT01 residues 131-268) were introduced by ligation independent cloning (Aslanidis and de Jong, 1990) into the *E. coli* expression vector BG1861, which results in the fusion of a non-cleavable His_6-tag to the N-terminus of the BID domains (Myler et al., 2009) yielding *Bro*Bep6_tBID1, *Bcl*Bep9_tBID1 and *Bhe*BepE_BID1, respectively. *Bro*Bep6_tBID1 was overexpressed using BL21(DE3)-pLysS *E. coli* cells in M9 media supplemented with SeMet and induced with 1 mM IPTG overnight at 16˚C, shaking at 220 rpm. Cells were harvested by centrifugation and frozen at -80˚C. The ~10 g bacterial pellet was resuspended in 50 ml of buffer containing 25 mM Tris pH 8.0, 200 mM NaCl, 50 mM arginine, 10 mM imidazole, 0.25% glycerol, 1 mM TCEP (VWR), 1% CHAPS (JT Baker), 1/2 tablet of EDTA-free protease inhibitor (Roche), 75 U benzonase (Novagen), 75 mg lysozyme (Sigma) and sonicated at 4˚C for 45 minutes. The resulting slurry was clarified by centrifugation at 4˚C for 30 minutes and the supernatant was the loaded onto a HiTrap Ni Chelating Column (GE Healthcare) attached to an AEKTA FPLC and washed with buffer A (25 mM Tris pH 8.0, 200 mM NaCl, 50 mM arginine, 0.25% glycerol, 1 mM TCEP) at 4˚C. The protein was eluted with a gradient of eluting buffer (25 mM Tris pH 8.0, 200 mM NaCl, 500 mM imidazole, 1 mM TCEP). The eluted protein was pooled, concentrated to 22.36 mg/ml via centrifugation using a 3 kDa molecular weight cutoff membrane (Amicon). The protein was then loaded onto a Sephacryl S-100 (GE Healthcare) size exclusion chromatography column pre-equilibrated with a buffer containing 25 mM Tris pH 8.0, 200 mM NaCl, 1%

glycerol, 1 mM TCEP. The protein was concentrated to a final concentration of 19.68 mg/mL.

After transformation into chemically competent *E. coli* BL21(DE3) Rosetta cells, starter cultures for each *Bcl*Bep9_tBID1 or *Bhe*BepE_BID1 construct were grown for 18 hours at 37°C. The protein was expressed in a LEX bioreactor in the presence of ampicillin (50 $\mu g.ml^{-1}$) (Studier, 2005). The cells were grown for 24 hours at 25°C and the temperature was reduced to 15°C for another 60 hours. The pellet was flash frozen in liquid nitrogen and stored at -80°C. Cells were resuspended in lysis buffer (20 mM HEPES pH 7.4, 300 mM NaCl, 5% glycerol, 30 mM imidazole, 0.5% CHAPS, 10 mM $MgCl_2$, 3 mM β-mercaptoethanol, 1.3 µg/ml protease-inhibitor cocktail, 0.05 mg/ml lysozyme) at 4°C. The cells were sonicated and incubated with Benzonase (20 µL of 25 unit/µL) at 37°C for 40 minutes. The soluble fraction was loaded onto a 5 mL Ni-NTA His-Trap FF column (GE Biosciences, Piscataway, New Jersey, USA). The column was washed with binding buffer (20 mM HEPES pH 7.0, 300 mM NaCl, 5% glycerol, 30 mM imidazole, 1 mM TCEP) and eluted with 500 mM imidazole in the same buffer. The collected protein was concentrated and further resolved by size-exclusion chromatography (SEC) using a Hiload 26/60 Superdex 75 prep grade column (GE Biosciences) pre-equilibrated with a buffer containing 25 mM HEPES pH 7.0, 500 mM NaCl, 5% glycerol, 0.25% azide and 2 mM DTT for *Bcl*Bep9_tBID1 and 20 mM HEPES pH 7.0, 300 mM NaCl, 5% glycerol and 1 mM TCEP for *Bhe*BepE_BID1. Peak fractions were collected and pooled based on purity-profile assessment by SDS-PAGE. *Bcl*Bep9_tBID1 was concentrated to 24.7 mg/mL and *Bhe*BepE_BID1 was concentrated to 28.7 mg/mL. All concentrated pure proteins were flash-frozen in liquid nitrogen and stored at -80°C. The three purified BID domains eluted as monomers from gel filtration columns.

Protein Crystallization

All proteins were thawed and crystallized using the sitting drop vapor diffusion method at 289 K with 0.4 µl protein and 0.4 µl precipitant equilibrated against 80 µl of reservoir solution. *Bro*Bep6_tBID1 crystals grew within days (in 0.1 M sodium cacodylate-HCl pH 6.5 and 1 M sodium citrate tribasic (MCSG3 (Anatrace) A1)) and were then soaked in a reservoir solution supplemented with 20% (v/v) ethylene glycol and subsequently flash frozen in liquid nitrogen. *Bcl*Bep9_tBID1 crystals grew within days in 200 mM ammonium sulfate, 100 mM sodium citrate-HCl pH 5.6, 25% (w/v) PEG 4000 (MCSG1 (Anatrace) C8) and were immediately harvested into liquid nitrogen for flash freezing. *Bhe*BepE_BID1crystals grew in 200 mM $MgCl_2$, 100 mM HEPES-NaOH, pH 7.5, 25% (w/v) PEG 3350 (MCSG1 (Anatrace) A9). Crystals grew within days and were harvested and soaked in a solution containing the prior crystallization solution supplemented with 15% (v/v) ethylene glycol before flash freezing

X-ray data collection and structure determination

Data for *Bro*Bep6_tBID1 was collected at the Canadian Macromolecular Crystallization Facility beamline 08ID-1 with a Marmosaic 300 CCD detector. Data for *Bcl*Bep9_tBID1 and *Bhe*Bep6_BID1 were collected at the Advanced Photon Source on beamline 21-ID-G on a Marmosaic 300 CCD detector. All data were reduced using XDS/XSCALE (Kabsch, 2010). For *Bro*Bep6_tBID1 Friedel pairs were not merged, and the unmerged data provided an anomalous signal that was used to phase the data of *Bro*Bep6_tBID1 with Phaser (McCoy et al., 2007) from the CCP4 program suite (Winn et al., 2011). Density modification was performed with Parrot (Zhang et al., 1997) on the resulting electron density and the initial model was built into this modified map with ArpWarp (Morris et al., 2003). This structure was then used as a model for molecular replacement to determine the structure of *Bcl*Bep9_tBID1 using MR-Rosetta (Terwilliger et al., 2012). *Bcl*Bep9_tBID1 was then used as a molecular replacement model to determine the structure of

*Bhe*BepE_BID1 using MR-Rosetta. All structures were completed using iterative rounds of refinement in Phenix (Adams et al., 2010) followed by manual structure rebuilding with COOT (Emsley et al., 2010). All models were quality checked by Molprobity (Chen et al., 2010). All data reduction and refinement statistics are reported in Table 1.

Sequence and structure analysis

To generate our BID sequence working dataset, the sequences of the three newly determined BID domain structures were searched in Uniprot (UniProt Consortium, 2015) against the UniprotKB database using BLAST (Altschul et al., 1990) with a maximal e-value threshold of $1e^{-3}$. This resulted in 203, 196 and 197 homologous sequences for *Bro*Bep6_tBID1, *Bcl*Bep9_tBID1 and *Bhe*BepE_BID1, respectively. To remove redundant sequences (90% level of redundancy), we merged the three datasets and obtained 211 unique sequences. The FIC, OB and BID domains were then annotated using Geneious v7.1.7; Biomatters. Similarly, we BLASTed the BID domain of the relaxase of At-pRi1724 and the first and second BID domains of pATC58 (Schulein et al., 2005). After merging the three relaxase datasets, it resulted in 140 unique sequences. Combining the *Bartonella* and the relaxases a total of 351 sequences were retrieved.

Relaxase sequences were annotated with their species name, followed by the domain classification and ending with the UniProt sequence reference. For the *Bartonella* species, we annotated them with the lineage (L3/L4) followed by a three letter abbreviation for the species (see accession numbers section) followed by the domain classification and then the UniProt sequence reference.

All the sequences of our working dataset were then aligned using ClustalX 2.0 (Larkin et al., 2007) with a gap opening penalty of 10 and a gap extension penalty of 0.2 using BLOSUM matrices (S. Henikoff and J. G. Henikoff, 1992). Neighbor-joining distance based trees were constructed and visualized with iTOL (Letunic and Bork, 2011). Sequence

logos were generated with Weblogos (Crooks et al., 2004) and alignments visualized with Aline (Bond and Schüttelkopf, 2009). Conservation scores were generated using ConSurf (Ashkenazy et al., 2010). Electrostatics were calculated with the APBS-Tools and PDB2PQR (Dolinsky et al., 2007; 2004) plugins for PyMOL using the default settings.

The following servers were used to compare the structure of the newly determined BID domain structures and revealed no structural homology to any known structure currently available: ProFunc (Laskowski et al., 2005), InterProScan (Jones et al., 2014), PDBeFold (Krissinel and Henrick, 2004), MarkUs (Fischer et al., 2011) and ProBIS (Konc and Janezic, 2010).

Abbreviations

The abbreviations for the *Bartonella* sequences are: *Bartonella clarridgeiae* – L3_*Bcl*, *Bartonella rochalimae* - L3_*Bro*, *Bartonella sp.* AR 15-3 - L3_B15, *Bartonella sp.* 1-1C - L3_B11, *Bartonella alsatica* - L4_*Bal*, *Bartonella birtlesii* - L4_*Bbi*, *Bartonella doshiae* - L4_*Bdo*, *Bartonella elizabethae* - L4_*Bel*, *Bartonella henselae* - L4_*Bhe*, *Bartonella grahamii* - L4_*Bgr*, *Bartonella koehlerae* - L4_*Bko*, *Bartonella quintana* - L4_*Bqu*, *Bartonella rattimassiliensis* - L4_*Bra*, *Bartonella taylorii* - L4_*Bta*, *Bartonella tribocorum* - L4_*Btr*, *Bartonella washoensis* 085-0475 - L4_*Bwa*085, *Bartonella washoensis* Sb944nv - L4_*Bwa*Sb, *Bartonella washoensis* - L4_*Bwa*, *Bartonella sp. DB5-6* - L4_*Bdb*, *Bartonella vinsonii* subsp. Arupensis - L4_*Bva*, *Bartonella vinsonii* subsp. Berkhoffii - L4_*Bvb*.

Accession numbers

A0A024J204, I3QKD8, P55418, J0Q2C9, J0QJW8, J0R9F3, K0PZL2, E6YHI3, E6YHI2, H0HGV1, J1J4K1, W3TX69, J1JMC5, J1JQX9, A0A087LYR2, A0A060I368, E6YW78, M5JSQ2, F8C170, E6YS53, E6YS54, A0A031LX45, J0K3V5, A0A0Q7Y262, J0ZU19, W8IF22, K0Q5J8, J1K5A2, B9JPA6, J0QS37, H0GBP1, J1J5R5, J0R175, J0WFK0,

I3XGX9, E6YWF3, J0YIF8, J1JJY7, W1L862, Q5QT03, Q5QT02, Q5QT01, Q2KC95, Q6G2A9, A0A0T6XQ69, A0A011UFZ1, Q5QT04, A0A0Q7XCR7, B9JE50, I3QKE3, A0A0T7G559, I5BT15, I3QKE4, A7IQA6, J1JMD0, F7UHD1, J1IRW7, A0A0N1KXS5, A8W094, A0A0B4XBP6, E6YV21, J0YV99, A0A0Q8GC00, J0ZTN1, E6YLF6, A0A0L6JZ23, A0A0H1A1D3, N6VH47, N6UQE3, E6YLF3, B3Q3U2, N6VL01, W0WS63, J0PVI6, J0ZAV1, J0Q230, E6YMR5, J0YNJ0, E6YMR3, J0PUH0, A0A0T2Z4E6, E6YIM7, E6YIM4, E6YIM5, J1IY74, A0A0B4X7E2, A0A0Q6EN91, E6YLF2, F5JH52, A0A0Q6M3Q3, E6YHH2, E6YIM8, A9IWP4, A6WWM4, A9IWP7, J0YVD3, J1JNI6, E6YLF1, J1IT50, J1IX26, J2VIX2, E6YHH5, J0QJW4, J0QUJ6, E6YQI1, J1IY69, U4V598, J1IY65, A0A0Q8MLR0, W6S1J9, E6YGF5, J0YVG6, E6YQP8, J0ZFI4, C6B2F3, C6AAI2, A0A0K8J1V5, J0PVQ4, S5RTV3, J0R9E8, E6YQP7, K0VCU5, A0A0N1DQV8, E6YU24, W3TV45, N6VKZ4, A0A0K8J190, J1T3N3, J0PTX3, N6UCU4, J0ZD53, J0R910, A9IWQ2, J1JPU6, A9IWQ5, A9IWQ4, F7XZ18, A0A0D0KVF8, E6YHH3, E6YQQ0, J5PM42, B9K3V6, J0PP03, H4F2V9, A0A0A8G8B4, E6YQQ1, N6VN16, A0A0Q9BL84, A0A0Q9BG22, A0A0N0LBI4, S5SQB1, A0A0Q9DY10, J0PTT5, J5M525, A9IWQ0, A0A0Q6SVL8, A0A098RKN6, J0PVK0, J1IZ97, A0A0T2Z5T2, K2Q8T9, J1IT55, S3H423, A0A081D1E6, V5RE37, A0A0Q6NLD2, F7UHX4, E6YNY8, N6U8Q7, F4Y9H6, A0A060ICQ2, F4Y9H5, G6XZP2, J0R1K0, C6AF22, J0R905, E6YGD8, A0A0Q5ZXD3, A0A0E4G0H7, A0A011TSF9, A0A0Q6FQR7, J0CZE6, C3KR00, J1JNJ1, J1K3Z1, A0A0T6Y901, A0A0Q5KGE3, N6UWX7, J1IZB4, A0A0Q5YGP8, A0A037XVE4, E6YTB2, J1JPU2, A0A0N1A5D8, S5S8T3, W8FIM2, A0A021WX37, J0CGM2, A0A0Q9E4S5, J0GWB6, J0QUK2, A0A0Q5YY14, A0A0Q6QSS1, A0A0F4G1B7, A0A0F5LV30, A0A0Q9EQJ3, E6YWF1, E6YLB1, A6UKI1, W3TUS8, I3X6L2, A0A0Q8Z9E6, A0A0Q5YEQ1, A0A021XCC5, A0A067WBA3, A0A067W5H3, W3TWZ4, A0A0Q7YDX6, J0ZTR0, A0A0A8GKW7, J1J5M3, J0QPN2, J0UVQ2, J0BV06, J0QMC4, J1J5P5, A0A0T6ZRF3, E6YKB8, J0Q110, A0A0F5FF40, A0A0F5LQS9, A0A0D0L722, J0YUB9,

A0A0N1MD99, A0A0T6XR82, G9AI24, A0A072C238, E6YV77, E6YV76, U1XGP9, A0A0L8BI81, A0A067W308, E6YV78, J0PP81, M3K8X5, A0A072BS58, F7XGT1, C6AET1, C6AET0, E6YIF3, J0QQF2, C6AES7, A0A072CI45, C6AES4, F0LG45, A9IWP9, Q84HT7, Q5QT00, L0LPV7, Q7D3W2, E6YPK5, A0A0F5PRS7, C6AET2, B9K371, A0A067UDA2, A0A037XPJ5, J0ZFH9, M5JU27, Q5QSZ9, E6YHH4, Q1QF79, A0A0A8GNY4, E6YHH8, J2L8H2, M3KFC2, J0ZFH4, Q11MR5, N6UWX1, A0A0Q6NR71, Q1ML87, A0A098RMF6, A0A0Q7RJG4, A0A0Q8B230, A0A081D006, E6YFW2, J0ZAV5, A0A0Q8G7Q3, E6YV82, I9WXX0, A0A072C5B2, J0Q0Q2, J1JLW6, B9QRK8, W3TUX5, A0A067WDS6, B9K4J7, N6UQF1, J0Q8S5, J1K3V6, A0A0T6YLC2, B3Q2J2, R4IL57, C6AES8, J0QHJ3, E6YGE9, A0A0T6YKC1, E6YMI0, C6AES9.

Supplemental references

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403–410.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38*, W529-533.

Aslanidis, C., de Jong, P.J., 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18, 6069–6074.

Bond, C.S., Schüttelkopf, A.W., 2009. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr D Biol Crystallogr 65, 510–

512.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12–21.

Crooks, G.E., Hon, G., Chandonia, J.-M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A., 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res *32*, W665–7.

Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., Baker, N.A., 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35, W522–5.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501.

Fischer, M., Zhang, Q.C., Dey, F., Chen, B.Y., Honig, B., Petrey, D., 2011. MarkUs: a server to navigate sequence-structure-function space. Nucleic Acids Res 39, W357–61.

Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences 89, 10915–10919.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Kabsch, W., 2010. XDS. Acta Crystallogr D Biol Crystallogr 66, 125–132.

Konc, J., Janezic, D., 2010. ProBiS: a web server for detection of structurally similar

protein binding sites. Nucleic Acids Res 38, W436–W440.

Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60, 2256–2268.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Laskowski, R.A., Watson, J.D., Thornton, J.M., 2005. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res 33, W89–93.

Letunic, I., Bork, P., 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39, W475–8.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J Appl Crystallogr 40, 658–674.

Morris, R.J., Perrakis, A., Lamzin, V.S., 2003. ARP/wARP and automatic interpretation of protein electron density maps. Meth. Enzymol. 374, 229–244.

Myler, P.J., Stacy, R., Stewart, L., Staker, B.L., Van Voorhis, W.C., Varani, G., Buchko, G.W., 2009. The Seattle Structural Genomics Center for Infectious Disease (SSGCID). Infect Disord Drug Targets 9, 493–506.

Studier, F., 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif.

Terwilliger, T.C., Dimaio, F., Read, R.J., Baker, D., Bunkóczi, G., Adams, P.D., Grosse-Kunstleve, R.W., Afonine, P.V., Echols, N., 2012. phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. J Struct Funct Genomics 13, 81–90.

UniProt Consortium, 2015. UniProt: a hub for protein information. Nucleic Acids Res 43, D204–12.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67, 235–242.

Zhang, K.Y., Cowtan, K., Main, P., 1997. Combining constraints for electron-density modification. Meth. Enzymol. 277, 53–64.