

The Rho GTPase inactivation domain in *Vibrio cholerae* MARTX toxin has a circularly permuted papain-like thiol protease fold

Jimin Pei^{1*} and Nick V. Grishin^{1,2}

¹Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9050

²Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9050

ABSTRACT

A Rho GTPase inactivation domain (RID) has been discovered in the multifunctional, autoprocessing RTX toxin RtxA from *Vibrio cholerae*. The RID domain causes actin depolymerization and rounding of host cells through inactivation of the small Rho GTPases Rho, Rac, and Cdc42. With only a few toxin proteins containing RID domains in the current sequence database, the structure and molecular mechanisms of this domain are unknown. Using comparative sequence and structural analyses, we report homology inference, fold recognition, and active site prediction for RID domains. Remote homologs of RID domains were identified in two other experimentally characterized bacterial virulence factors: IcsB of *Shigella flexneri* and BopA of *Burkholderia pseudomallei*, as well as in a group of uncharacterized bacterial membrane proteins. IcsB plays an important role in helping *Shigella* to evade the host autophagy defense system. RID domain homologs share a conserved diad of cysteine and histidine residues, and are predicted to adopt a circularly permuted papain-like thiol protease fold. RID domains from MARTX toxins and virulence factors IcsB and BopA thus could function as proteases or acyltransferases acting on host molecules. Our results provide structural and mechanistic insights into several important proteins functioning in bacterial pathogenesis.

Proteins 2009; 77:413–419.
© 2009 Wiley-Liss, Inc.

Key words: Rho GTPase inactivation; cysteine protease domain; papain-like fold; multifunctional, autoprocessing RTX toxins; *Shigella* virulence factor IcsB; structure prediction; homology inference.

INTRODUCTION

Certain gram-negative bacterial pathogens use Type I secretion systems to export large exoproteins called RTX (repeats-in-toxin) toxins, including *Escherichia coli* α -hemolysin, *Pasteurella haeflolytica* leukotoxin, and *Bordetella pertussis* adenylate cyclase toxin.¹ These toxins are characterized by repeats of a glycine and aspartate-rich, calcium-binding sequence motif. A family of multifunctional, autoprocessing RTX toxins (MARTX) such as RtxA from *Vibrio cholerae* (VcRtxA)² has been recently identified.³ MARTX toxins have several distinct types of repeats present at the N- and C-termini, and a middle mosaic region containing multiple domains with virulence activity.³ They also possess a cysteine protease domain with a caspase-like fold that is essential for autoprocessing of these toxins to release the virulence activity domains to the cytosol of host cells.^{4,5} Two distinct virulence activity domains in VcRtxA have been characterized that each can cause the host cell rounding phenotype: an actin cross-linking domain (ACD)^{6,7} and a Rho GTPase inactivation domain (RID).⁸ While the ACD domain modifies the cytoskeleton by directly acting on the actin molecules, the RID domain induces actin depolymerization and rounding of host cells through inactivation of the small Rho GTPases Rho, Rac and Cdc42.⁸ The structural and molecular mechanisms of RID domains are unknown. In this study, we report distant homology inference, fold recognition, and active site prediction for RID domains through comparative sequence and structural analyses. We found that RID domains are remotely related to two experimentally characterized virulence factors: IcsB of *Shigella flexneri*^{9,10} and BopA of *Burkholderia pseudomallei*,^{11,12} as well as a group of uncharacterized bacterial membrane proteins. RID domain homologs are predicted to adopt a circularly permuted papain-like thiol protease fold with a conserved cysteine and histidine catalytic diad. Our prediction results suggest that RID domains in MARTX toxins and virulence factors IcsB and BopA could function as proteolytic enzymes or acyltransferases acting on host molecules.

Grant sponsor: NIH; Grant number: GM67165; Grant sponsor: Welch foundation; Grant number: 11505 to NVG.

*Correspondence to: Jimin Pei, Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9050. E-mail: jpei@chop.swmed.edu

Received 30 January 2009; Revised 12 March 2009; Accepted 14 March 2009

Published online 31 March 2009 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.22447

MATERIALS AND METHODS

The PSI-BLAST program¹³ was used to search for homologs of the experimentally determined RID domain-containing region of *V. cholerae* RtxA (residues 2552-3099, gil153817921) against the NCBI nonredundant database (November 26 2008; 7,365,651 sequences; 2,547,387,767 letters), with an inclusion e-value cutoff of 0.001. Found homologs were clustered and one representative sequence from each group was used to initiate new PSI-BLAST iterations to ensure maximum coverage. Manual inspections of PSI-BLAST hits above the default e-value cutoff were conducted to search for remote homologs of RID domains, and the same PSI-BLAST search strategy was used for these remote homologs. RID homologs were submitted to the web server of HHpred,¹⁴ a sequence similarity search method based on profile-profile comparison,¹⁵ to search for distant relationships against a nonredundant structure database (pdb70) and Pfam¹⁶ database (version 23.0). Domain architecture analyses were made by submitting sequences to protein domain database servers such as CDD,¹⁷ Pfam and SMART.¹⁸ Multiple sequence alignments were constructed by using the PROMALS3D program.¹⁹ Manual adjustment of the alignments was made with guidance from available three-dimensional structures and secondary structure predictions by PSIPRED.²⁰

RESULTS AND DISCUSSIONS

Defining the RID domain region

The RID domain-containing region was originally mapped to residues 2552-3099 in VcRtxA (NCBI gene identification (gi) number 4455065).⁸ This region alone can cause Rho GTPase inactivation and cell rounding.⁸ The PSI-BLAST¹³ program was used to search for homologs of residues 2552-3099 in VcRtxA. The first 100 residues (2552-2651) showed significant sequence similarity to several domains in known structures. These domains are also from bacterial toxins including the *Pasteurella multocida* toxin PMT [Protein Data Bank (PDB) id: 2ebf]²¹ and *Clostridium difficile* toxin B (PDB id: 2bvl).²² They adopt a fold of a four-helix bundle and have been shown to associate with anionic lipids and contribute to membrane localization of these toxins. This helical domain itself in VcRtxA is unlikely to convey the activity of Rho GTPase inactivation as such an activity has not been reported in its homologs in other toxins. The C-terminal regions of this helical domain in both PMT and toxin B contain putative virulence activity domains. For example, a glycosyltransferase domain responsible for inactivation of RhoA is present right after the helical domain in toxin B.²² Thus the C-terminal portion (residues 2653-3099) after this helical domain in the experimentally determined RID region in VcRtxA is

likely to harbor the virulence activity of Rho GTPase inactivation. We refer to this region (residues 2653-3099) as the RID domain in this study, as opposed to the originally proposed region of residues 2552-3099.⁸

Sequence similarity searches for RID domains

PSI-BLAST searches were conducted using the redefined region of VcRtxA RID domain. With an e-value inclusion threshold of 0.001, these searches converged to 18 domains from 17 proteins in the NCBI nonredundant sequence database. Domain content analysis revealed that these proteins are MARTX toxins from various strains of *Vibrio cholerae* and a few other bacterial pathogens such as *Vibrio vulnificus*, *Listonella anguillarum*, and *Proteus mirabilis*. The limited number of RID domains in these toxins could have prevented generation of diverse sequence profiles for detection of remote homologs above the significance cutoff.

To investigate if remote homologs of RID domains exist, we inspected PSI-BLAST hits above the default e-value cutoff (0.001). A protein named BopA from *Burkholderia thailandensis* was found to have a marginal e-value of 0.05. This protein shares a sequence motif present in all RID domains that contains a conserved cysteine residue. PSI-BLAST searches using BopA as the query found closely related sequences in various *Burkholderia* species as well as an experimentally characterized protein IcsB from *Shigella flexaneri*.¹⁰ RID domains were also found as significant hits (e-value <0.001) using BopA as the query, suggesting that they are evolutionarily related.

Another marginally significant PSI-BLAST hit of the VcRtxA RID domain is a hypothetical protein from *Methylobacterium radiotolerans* (gil170748086, e-value: 0.062). The region of this PSI-BLAST hit (residues 330-442) contains the same sequence motif with a conserved cysteine. PSI-BLAST iterations from this protein detected a group of bacterial proteins as well as the RID domains as significant hits (e-value <0.001). Most of these bacterial proteins are annotated as hypothetical proteins, and a few of them are annotated as “membrane-bound protease” (e.g., gil154248066) or “Zn-dependent proteases” (e.g., gil163857530).

The RID domains and the homologous regions in the detected remote homologs do not contain known domains in public protein domain databases, such as CDD,¹⁷ Pfam¹⁶ and SMART.¹⁸ None of their structures has been solved in the current PDB database. We used the HHpred server¹⁴ of profile-profile comparison to search for their remote homologs against a nonredundant structure database (pdb70) and Pfam database. Weak sequence similarities were detected against a protein named PPPDE1 (permuted papain fold peptidases of dsRNA viruses and eukaryotes) with a known structure (PDB id: 3ebq, solved by Structural Genomics

Consortium, unpublished results). For example, the hypothetical protein from *Methylobacterium radiotolerans* (gil170748086) found 3ebq with a probability score of 47.2. Originally described in a sequence analysis work,²³ PPPDE family proteins have a circularly permuted papain-like fold with a conserved cysteine and histidine diad for protease or acyltransferase activity. In the HHpred alignment, the conserved cysteine residue in RID domain homologs was aligned to the catalytic cysteine in the PPPDE1 protein. Multiple sequence alignments¹⁹ of RID domain homologs also revealed an invariant histidine residue N-terminal to the conserved cysteine residue (see Fig. 1), which is also reminiscent of the active site arrangement of a circularly permuted papain-like fold.

Sequence and structure characterization of RID domain homologs

In the MEROPS²⁴ peptidase database, papain-like thiol proteases belong to clan CA. With 24 distinct families, this clan represents the most divergent group of evolutionarily related cysteine peptidases. The circularly permuted papain-like proteases are grouped in clan CE, currently comprising six peptidase families. Several bacterial virulence factors, such as *Yersinia* YopJ (peptidase family C55), *Yersinia* YopT and *Pseudomonas* AvrPphB (peptidase family C58), have been shown to be (circularly permuted) papain-like proteases.^{25,26} In the Structural Classification of Proteins (SCOP) database,²⁷ papain-like proteases and circularly permuted papain-like proteases are grouped in the same superfamily under the “cysteine proteinases” fold, suggesting that they are evolutionarily related. Aside from structural similarities, significant sequence similarities have been observed for papain-like proteases and their circularly permuted versions.^{23,28} Catalytically active members of these proteins possess a conserved cysteine residue (in rare cases substituted by a serine residue) that uses the sidechain thiol group as the nucleophile to attack the peptide bond of the substrate. A conserved histidine residue functions as a general base/acid for proton transfer. These domains basically act as acylhydrolases such as peptidases or acyltransferases. For example, the SCOP fold of “cysteine proteinases” also includes transglutaminases that transfer the γ -carboxyl group of glutamine to the ϵ -amino group of lysine or other primary amines.^{29,30}

The structural core of the papain-like fold consists of a mainly antiparallel β -sheet forming a barrel-like subdomain, and one or several surrounding α -helices. The catalytic histidine and cysteine are located at the beginning of one β -strand and at the beginning of one α -helix, respectively. In papain-like proteases, the catalytic cysteine resides N-terminal to the catalytic histidine [Fig. 2(b)], whereas in circularly-permuted papain-like proteases the order is reversed [Fig. 2(a)]. Regions outside the struc-

tural core exhibit great diversity in different structures.²⁸ For example, two circularly permuted papain-like structures PPPDE1 (PDB id: 3ebq) and Senp2 (PDB id:1th0)³¹ only share three β -strands (Fig. 1, b1, b2, b3) and one α -helix (Fig. 1, A2) as superimposable core secondary structural elements.

Multiple sequence alignments of RID domain homologs and several known structures suggest the presence of core secondary structural elements of a circularly permuted papain-like fold in RID domains and their remote homologs (see Fig. 1). The sequences of RID domain homologs are most similar to structures 3ebq (a PPPDE family member) and 2if6 (a bacterial hypothetical protein, solved by New York and Structural Genomix Research Consortium, unpublished results). Among the core secondary structural elements, strands b1, b2, and b3 form a β -meander motif (see Fig. 2). Strand b4 is antiparallel to b1 and completes the β -barrel-like structure. Like PPPDE proteases, two α -helices (A1 and A2) follow the β -sheet in RID domain homologs (see Fig. 1). Catalytic histidine and cysteine are located at the beginning of strand b2 and helix A2, respectively. An asparagine residue before the catalytic cysteine serves as the N-cap for helix A2 in 3ebq. This residue is also conserved in RID domain homologs (see Fig. 1). For papain-like thiol proteases, usually a conserved polar residue (Asx/Glx) in strand b3 forms an interaction with the catalytic histidine. However, the corresponding regions in the PPPDE family (3ebq) and RID domain homologs do not have such a conserved Asx/Glx residue (Fig. 1, a conserved serine in strand b3 is present in RID domains instead), suggesting a catalytic diad (Cys-His) instead of a catalytic triad (Cys-His-Asx/Glx) functioning in these proteins. Another highly conserved residue in both RID homologs and the PPPDE family is a tyrosine (see Fig. 1) located several residues before the catalytic cysteine. Papain family proteases have a conserved glutamine residue in such a position, which has been shown to be part of the oxyanion hole that stabilizes the main-chain carbonyl group of the P1 residue of the substrate.^{21,32}

A distinct feature of RID domain homologs is the long insertion between strands b3 and b4 (see Fig. 1). In particular, these insertion regions in the RID domains of MARTX toxins are more than 150 residues, and mainly consist of predicted α -helices. They could form a (sub)-domain that contributes to substrate binding or interactions with other proteins. Corresponding regions in virulence factors IcsB and BopA are also relatively long (>90 residues, Fig. 1).

Domain architecture and cellular functions of RID domain homologs

Three groups of RID domain homologs were detected: RID domains in MARTX toxins, bacterial virulence

factors IcsB and BopA, and a group of closely related bacterial proteins.

RID domains

RID domains were discovered in several MARTX toxins. MARTX toxins share a similar overall domain organization³ with several types of repeats present in the N- and C-terminal regions. A cysteine protease domain with a caspase-like fold⁴ is present before the C-terminal repeats and is responsible for autoprocessing of these proteins to release domains with virulence activity inside host cell cytosol.³ The middle region has a mosaic domain structure and contains a variety of putative virulence activity domains in different MARTX toxins.³ The RID domain is one of the experimentally demonstrated virulence activity domains.⁸ Sequence database searches revealed a limited phyletic distribution of RID domains from closely related gram-negative pathogenic species such as *V. cholerae*, *V. vulnificus*, *L. anguillarum*, and *P. mirabilis*. RID domain-containing MARTX toxins were also found in the unreleased *Xenorhabdus bovienii* and *Xenorhabdus nematophila* genomes.³ The MARTX from *P. mirabilis* has two copies of RID domains.

Rho GTPases are targeted by various bacterial virulence factors to modify the plasticity of actin cytoskeleton.³³ As molecular switches, Rho GTPases cycle between GDP-bound inactive state and GTP-bound active state.³⁴ GAPs (GTPase-activating proteins) facilitate the hydrolysis of GTP to GDP for Rho GTPases. On the other hand, GEFs (GDP-GTP exchange factors) activate Rho GTPases by releasing GDP and allowing binding of GTP. Inactivation of Rho GTPases by the VcRtxA RID domain leads to actin depolymerization and rounding of host cells.⁸ Four mechanisms of Rho GTPase inactivation used by bacterial virulence factors have been discovered, including covalent modification of Rho-like GTPases, proteolytic cleavage of Rho-like GTPases, mimicry of host cell GAPs, and dephosphorylation of proteins regulating upstream signaling pathways.^{8,33} RID domain-induced cell rounding and actin depolymerization were prevented and rapidly reversed by CNF1-induced constitutive activation of Rho GTPases, suggesting that the RID domain does not directly act on Rho GTPases by proteolytic cleavage or glycosylation.⁸ Experiments also showed that the RID domain lacks GAP activity and phosphatase activity.⁸ Thus, the RID domain might adopt a different mechanism of Rho GTPase inactivation than the four known inactivation schemes. Several possible mechanisms of the RID domain have been proposed, including regulation of the activation state of Rho GTPases, activation of GAPs, inactivation of GEFs, and interference with signaling pathways upstream of GAPs and GEFs.⁸ The prediction of a circularly permuted papain-like fold and the presence of conserved catalytic residues suggest that the RID domain carries out its function by using a peptidase or

acyltransferase reaction. For example, one possible mechanism of the RID domain could be inactivation of GEFs by proteolytic cleavage. The substrate(s) of the RID domain remain to be experimentally discovered. The presence of a membrane-binding helical domain just N-terminal to the RID domain in MARTX toxins suggests that the RID substrate(s) have a membrane localization. Interestingly, structural studies revealed that a highly divergent papain-like protease domain is also present C-terminal to such a helical domain in the *P. multocida* toxin PMT.²¹

IcsB and BopA

The *Shigella* IcsB protein has been characterized as a virulence factor delivered to host cells by the Type III secretion system.⁹ A subsequent functional study suggests that IcsB is important for *Shigella* to escape autophagy,¹⁰ a host defense system that engulfs and sequesters bacteria in membrane-bound organelles such as phagosomes. When IcsB is mutated, the VirG protein (also called IcsA) of *Shigella* induces autophagy by binding to the host autophagy protein Atg5.¹⁰ In nonmutant *Shigella*, IcsB inhibits such a binding event and thus prevents autophagy.¹⁰ Interestingly, the function of VirG is to induce actin polymerization at one end of the bacteria to facilitate their motility inside the host cells. Therefore IcsB also has a function related to actin cytoskeleton. Contrary to the actin depolymerization phenotype caused by the RID domain, IcsB indirectly facilitates actin polymerization by protecting VirG from being recognized by the host autophagy machinery. The molecular mechanism of IcsB has not been revealed by experimental studies. Our prediction results suggest that IcsB could help protect VirG by proteolysis of Atg5, or via an acyltransferase activity that abolishes the interaction between VirG and Atg5.

The BopA protein from *Burkholderia pseudomallei*, a relatively close homolog of IcsB, is also a putative Type III secreted virulence factor.¹¹ Recent studies showed that BopA has a similar function to IcsB in mediating bacterial evasion of autophagy.¹² Close homologs of IcsB and BopA were only identified in *Shigella* and *Burkholderia* species. IcsB and BopA are about 500 amino acid residues in length. The regions homologous to RID domains are located at the C-termini of these proteins. The N-terminal regions of about 120 residues in these proteins harbor a SicP binding domain (Pfam entry PF09119). SicP is a chaperone that maintains the stability of certain bacterial proteins and assists their secretion.³⁵ The *Shigella flexneri* IpgA protein, encoded by a gene immediately downstream of the *icsB* gene, serves as a chaperone required for stabilization and secretion of IcsB.⁹ Similarly, the BicP protein of *B. pseudomallei* is a putative chaperone for BopA.¹² Thus the SicP binding domains in IcsB and BopA proteins may facilitate

chaperone-binding and play a role in their stabilization and secretion.

Uncharacterized bacterial membrane proteins

A third group of RID domain homologs are from bacterial proteins mostly annotated as hypothetical proteins. Although a couple of these proteins are annotated as “membrane bound protease” or “Zn-dependent proteases”, we found no experimental studies on them, nor did protein domain database searches reveal any known protease domains in them. These proteins also have a limited phyletic distribution. They are mainly from the proteobacteria species, and a couple of them are from the cyanobacteria species. Most of these proteins have a signal peptide and several transmembrane segments at the N-termini, suggesting that they are secreted membrane proteins. The regions homologous to RID domains are located at the C-termini of these proteins. As an exception, a relatively divergent member in this group from *Trichodesmium erythraeum* does not have transmembrane regions (gil113475387, Fig. 1). Its open reading frame starts right from the predicted β -strand b1, which indirectly confirms the domain boundaries of other RID domains. The cellular functions of these putative peptidases are yet to be revealed by experimental studies.

CONCLUSIONS

We report homology inference, fold recognition, and active site prediction for RID domains present in certain multifunctional, autoprocessing RTX toxins from bacterial pathogens. Remote homologs of RID domains were found in bacterial virulence factors IcsB of *Shigella flexneri* and BopA of *Burkholderia pseudomallei*, as well as in a group of uncharacterized bacterial membrane proteins. RID domain homologs are predicted to adopt a circularly permuted papain-like thiol protease fold with a conserved cysteine and histidine catalytic diad. RID domains of MARTX toxins and IcsB/BopA could function as proteolytic enzymes or acyltransferases acting on host molecules. Our computational analyses offer insights into the structural mechanisms of these bacterial virulence factors, generate relevant hypotheses, and facilitate experimental design for them.

ACKNOWLEDGMENTS

The authors thank Lisa Kinch and Dorothee Staber for critical reading of the manuscript and helpful suggestions.

REFERENCES

- Czuprynski CJ, Welch RA. Biological effects of RTX toxins: the possible role of lipopolysaccharide. *Trends Microbiol* 1995;3:480–483.
- Lin W, Fullner KJ, Clayton R, Sexton JA, Rogers MB, Calia KE, Calderwood SB, Fraser C, Mekalanos JJ. Identification of a *vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc Natl Acad Sci USA* 1999;96:1071–1076.
- Satchell KJ. MARTX, multifunctional autoprocessing repeats-in-toxin toxins. *Infect Immun* 2007;75:5079–5084.
- Lupardus PJ, Shen A, Bogyo M, Garcia KC. Small molecule-induced allosteric activation of the *Vibrio cholerae* RTX cysteine protease domain. *Science* 2008;322:265–268.
- Sheahan KL, Cordero CL, Satchell KJ. Autoprocessing of the *Vibrio cholerae* RTX toxin by the cysteine protease domain. *EMBO J* 2007;26:2552–2561.
- Cordero CL, Kudryashov DS, Reisler E, Satchell KJ. The Actin cross-linking domain of the *Vibrio cholerae* RTX toxin directly catalyzes the covalent cross-linking of actin. *J Biol Chem* 2006;281:32366–32374.
- Fullner KJ, Mekalanos JJ. In vivo covalent cross-linking of cellular actin by the *Vibrio cholerae* RTX toxin. *EMBO J* 2000;19:5315–5323.
- Sheahan KL, Satchell KJ. Inactivation of small Rho GTPases by the multifunctional RTX toxin from *Vibrio cholerae*. *Cell Microbiol* 2007;9:1324–1335.
- Ogawa M, Suzuki T, Tatsuno I, Abe H, Sasakawa C. IcsB, secreted via the type III secretion system, is chaperoned by IpgA and required at the post-invasion stage of *Shigella* pathogenicity. *Mol Microbiol* 2003;48:913–931.
- Ogawa M, Yoshimori T, Suzuki T, Sagara H, Mizushima N, Sasakawa C. Escape of intracellular *Shigella* from autophagy. *Science* 2005;307:727–731.
- Stevens MP, Haque A, Atkins T, Hill J, Wood MW, Easton A, Nelson M, Underwood-Fowler C, Titball RW, Bancroft GJ, Galyov EE. Attenuated virulence and protective efficacy of a *Burkholderia pseudomallei* bsa type III secretion mutant in murine models of melioidosis. *Microbiology* 2004;150 (Part 8):2669–2676.
- Cullinane M, Gong L, Li X, Lazar-Adler N, Tra T, Wolvetang E, Prescott M, Boyce JD, Devenish RJ, Adler B. Stimulation of autophagy suppresses the intracellular survival of *Burkholderia pseudomallei* in mammalian cell lines. *Autophagy* 2008;4:744–753.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
- Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005;33 (Web Server issue):W244–W248.
- Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 2005;21:951–960.
- Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. The Pfam protein families database. *Nucleic Acids Res* 2008;36 (Database issue):D281–D288.
- Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 2009;7 (Database issue):D205–D210.
- Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. *Nucleic Acids Res* 2009;37 (Database issue):D229–D232.
- Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 2008;36:2295–2300.
- Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;292:195–202.
- Kitadokoro K, Kamitani S, Miyazawa M, Hanajima-Ozawa M, Fukui A, Miyake M, Horiguchi Y. Crystal structures reveal a

- thiol protease-like catalytic triad in the C-terminal region of *Pasteurella multocida* toxin. Proc Natl Acad Sci USA 2007;104:5139–5144.
22. Ziegler MO, Jank T, Aktories K, Schulz GE. Conformational changes and reaction of clostridial glycosylating toxins. J Mol Biol 2008;377:1346–1356.
 23. Iyer LM, Koonin EV, Aravind L. Novel predicted peptidases with a potential role in the ubiquitin signaling pathway. Cell Cycle 2004;3:1440–1450.
 24. Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ. MEROPS: the peptidase database. Nucleic Acids Res 2008;36 (Database issue): D320–D325.
 25. Orth K, Xu Z, Mudgett MB, Bao ZQ, Palmer LE, Bliska JB, Mangel WF, Staskawicz B, Dixon JE. Disruption of signaling by *Yersinia* effector YopJ, a ubiquitin-like protein protease. Science 2000;290:1594–1597.
 26. Shao F, Merritt PM, Bao Z, Innes RW, Dixon JE. A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. Cell 2002;109:575–588.
 27. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
 28. Anantharaman V, Aravind L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol 2003;4:R11.
 29. Lorand L, Conrad SM. Transglutaminases. Mol Cell Biochem 1984; 58:9–35.
 30. Makarova KS, Aravind L, Koonin EV. A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. Protein Sci 1999;8:1714–1719.
 31. Reverter D, Lima CD. A basis for SUMO protease specificity provided by analysis of human Senp2 and a Senp2-SUMO complex. Structure 2004;12:1519–1531.
 32. Menard R, Carriere J, Laflamme P, Plouffe C, Khouri HE, Vernet T, Tessier DC, Thomas DY, Storer AC. Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain. Biochemistry 1991;30:8924–8928.
 33. Boquet P, Lemichez E. Bacterial virulence factors targeting Rho GTPases: parasitism or symbiosis? Trends Cell Biol 2003;13:238–246.
 34. Jaffe AB, Hall A. Rho GTPases: biochemistry and biology. Annu Rev Cell Dev Biol 2005;21:247–269.
 35. Stebbins CE, Galan JE. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial Type III secretion. Nature 2001; 414:77–81.