

# DAMPD: a manually curated antimicrobial peptide database

Vijayaraghava Seshadri Sundararajan<sup>1</sup>, Musa Nur Gabere<sup>1</sup>, Ashley Pretorius<sup>1</sup>, Saleem Adam<sup>1</sup>, Alan Christoffels<sup>1</sup>, Minna Lehvälaiho<sup>2</sup>, John A. C. Archer<sup>2</sup> and Vladimir B. Bajic<sup>2,\*</sup>

<sup>1</sup>South African National Bioinformatics Institute, The University of the Western Cape, 7535 Bellville, South Africa and <sup>2</sup>Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

Received September 30, 2011; Revised and Accepted October 26, 2011

## ABSTRACT

**The demand for antimicrobial peptides (AMPs) is rising because of the increased occurrence of pathogens that are tolerant or resistant to conventional antibiotics. Since naturally occurring AMPs could serve as templates for the development of new anti-infectious agents to which pathogens are not resistant, a resource that contains relevant information on AMP is of great interest. To that extent, we developed the Dragon Antimicrobial Peptide Database (DAMPD, <http://apps.sanbi.ac.za/dampd>) that contains 1232 manually curated AMPs. DAMPD is an update and a replacement of the ANTIMIC database. In DAMPD an integrated interface allows in a simple fashion querying based on taxonomy, species, AMP family, citation, keywords and a combination of search terms and fields (Advanced Search). A number of tools such as Blast, ClustalW, HMMER, Hydrocalculator, SignalP, AMP predictor, as well as a number of other resources that provide additional information about the results are also provided and integrated into DAMPD to augment biological analysis of AMPs.**

## INTRODUCTION

Antimicrobial peptides (AMPs) are recognized for their significant role in the innate immune response and are found in bacteria, fungi, animals and plants (1–7). AMPs are short [6–100 amino acid residues (8,9)] ribosomally-produced peptides that are post-translationally activated

by proteolytic cleavage. With few exceptions, AMPs are cationic and possess a significant proportion ( $\geq 30\%$ ) of hydrophobic residues (10,11). Their secondary structure generally adopts one of four structural motifs: (i) an  $\alpha$ -helical structure, (ii)  $\beta$ -stranded structure due to the presence of two or more disulfide bonds, (iii)  $\beta$ -hairpin structure or loop due to the presence of a single disulfide bond and/or cyclization of the peptide chain and (iv) an extended structure (12). Mature AMPs form amphipathic structures that associate via electrostatic interactions between positively charged AMP regions and negatively charged phospholipids of the cell membrane (4,14) which is thought to be necessary for antimicrobial activity. However, AMP modes of action can be divided into membrane disruptive or non-disruptive categories (8,13–17) indicating that multiple modes of action following membrane association exist. In addition, mammalian AMPs exhibit chemokine-like and immunomodulatory activities (18,19) that can integrate innate and adaptive immune responses to microbial infection. Measurements of non-synonymous and synonymous mutation rates in mammalian AMP exons and comparative genomic studies indicate that mammalian AMP genes are under positive selection and are among the most rapidly evolving group of mammalian genes known (20). The combination of a broad spectrum antimicrobial activities targeted at non-protein cellular components with localized, high-level expression at the site of infection, makes AMPs highly effective antimicrobial agents with significant potential as a source of new antimicrobial drugs (21) such as new more effective antitubercular agents active against multidrug resistant (MDR) and extensively drug resistant (XDR) *Mycobacterium tuberculosis* complex pathogens (22).

\*To whom correspondence should be addressed. Tel: +966 5447 00088 ; Email: vladimir.bajic@kaust.edu.sa

Present addresses:

Vijayaraghava Seshadri Sundararajan, School of Computer Engineering, Nanyang Technological University, 639 798 Singapore.

Ashley Pretorius, Department of Biotechnology, The University of the Western Cape, 7535 Bellville, South Africa.

Although several AMP related databases exist such as APD (23), AMSdb (14), BACTIBASE (24), Defensin knowledgebase (25), PenBase (26), Peptaibol Database (27), SAPD (28), AMPer (29), CyBase (30), BAGEL (31), Minicope (The Innate immunity defense peptides MiniCOPE Dictionary, <http://www.copewithcytokines.de/cope.cgi?key=Innate%20immunity%20defense%20peptides%20MiniCOPE%20Dictionary>), CAMP (32), PhytAMP (33) and RAPD (34), each has certain shortcomings, such as covering only specific AMP families or containing a limited collection of AMP families; many are lacking manually curated AMPs or do not have tools for exploration of relevant AMP characteristics (detailed in [Supplementary Table S1](#), in Section 5).

These observations combined with a frequent update of peptide information in major (<http://apps.sanbi.ac.za/dampd/Link.php>) databases, motivated us to retrieve peptides from UniProt (36) and GenBank (37), select those that are AMPs based on manual curation and develop a new database, Dragon Antimicrobial Peptide Database (DAMPD) as an update and extension to our earlier published ANTIMIC (35) database. DAMPD contains information on UniProt reviewed and UniProt non-reviewed (i.e. putative) natural AMPs. The utility of DAMPD is enriched by integration of several tools such as BLAST (38,39), ClustalW (40), HMMER (42), Hydrocalculator (43), SignalP (44,45), AMP predictor, as well as links to several resources that can provide additional information on results generated by DAMPD and used to explore characteristics of AMPs and enhance search for novel AMPs thus supporting biological analysis of AMPs.

## POPULATION OF DATABASE

AMPs were retrieved from UniProt (search: Antimicrobial [KW-0929]) AND (existence: evidence at protein level OR existence: evidence at transcript level). On 9 September 2011 UniProt we retrieved 1483 (PE1) and 682 (PE2) peptides. We manually curated these entries selecting only peptides that are experimentally validated. This finally resulted in 1232 (out of 2165) manually curated peptides that have experimentally proven/confirmed antimicrobial activity to be included into DAMPD database. We used the latest information on each of these 1232 peptides from UniProt and re-build our latest database as a version 'DAMPD DB 09\_Sep\_2011 (1232)'. To populate DAMPD, we searched for AMPs from different databases, as well as from journals. If a database entry has a keyword indicating the peptide has antimicrobial qualities, this may be an assumption derived from sequence similarity. Antimicrobial activity can be sensitive to even slight changes in the peptide, and just one amino acid difference can mean that the peptide is inactive. Since we only wanted true AMPs, we checked the research articles and made sure each peptide did in fact have an experimentally proven antimicrobial activity. We also added in peptides that UniProt has not yet annotated to be

antimicrobial. All annotations were verified from the original articles.

## DATABASE SYSTEM

DAMPD is the collection of manually curated AMPs. An integrated system driven through MySQL (5.0), PHP (5.2.4), and Perl (Ver. 5.8.8) was developed to handle the storage of information on these peptides. Each peptide in DAMPD database has a unique accession number (e.g. DAMPD:0001). AMPs in DAMPD include peptides containing precursor (477 AMPs), as well as mature peptide parts (755 AMPs). Peptide entries were cross-referenced to external resources and linked to graphical views. This section details on sub-databases, multiple catalogs, tools and graphics views.

### Sub-databases

AMPs are retrieved from UniProt where the entries are categorized as reviewed and non-reviewed. Our 1232 AMPs are thus split into 1113 reviewed and 119 non-reviewed entries we named Swiss-Prot\_AMP and TrEMBL\_AMP, respectively. They form two sub-databases in DAMPD one with reviewed and the other with non-reviewed UniProt AMPs. The total collection of these AMPs is termed as 'UniProt\_AMP'. We could not get any new peptide (not already included in the above) from GenBank. Consequently, DAMPD contains all AMPs from UniProt and GenBank.

### Multiple search capabilities

DAMPD has six search capabilities. These six search tools can operate as independent search engines to interrogate the database or executed as part of a more complex query. Five of these search utilities are based on catalogs that are created as vocabularies of terms from taxonomies, AMP families, species, keywords, and citations of 1232 peptide entries to ease browsing of the database.

*Taxonomy catalog.* Organisms are classified in a hierarchical tree structure. Taxonomy database contains every node (taxon) of the tree.

*Species catalog.* UniProt brings out annotation on species on each peptide that we compiled into a catalog for specific search by species.

*Keyword catalog.* UniProt entries are tagged with keywords that can be used to retrieve particular subsets of entries.

*Family catalog.* UniProt general annotation provides the family details on each for known AMPs. We collected family, super family, sub-family information and build family catalog.

*Citation catalog.* UniProt keeps publications with title (RT, example: protein interaction); author name (RA, examples: Ashburner, Sanger F., Pierson L.S. III), journal (RL, example: J. Exp. Biol.), year of publication

(YR, example: 1951) are cataloged for search on any of them.

*Advanced search.* This search category allows a combination of search terms, search fields and search values. Users can query the database using field names which are not listed in the other catalogs.

### Tools and resources

Several analytical tools and relevant resources are integrated into DAMPD to support exploration and biological analysis of AMPs (<http://apps.sanbi.ac.za/dampd/BioTools.php>). For the same purpose we provided links to a number of additional resources.

*Standalone version of integrated tools.* The DAMPD database tools can also operate on a standalone basis. For example, one can perform an alignment of antimicrobial sequences or any other protein/DNA sequence using ClustalW (40). NJplot (41) is used to draw phylogenetic tree of the aligned sequence generated by ClustalW (40). HMMER (42) allows the user to tentatively classify unknown sequences into a particular antimicrobial family using two ways: (i) the user can either use 27 predefined antimicrobial libraries of profiles or (ii) use their own generated profiles. The physicochemical properties of the peptides such as hydrophobicity, net charge, percentage of hydrophobic residues, mean hydrophobicity and mean hydrophobic moment can be calculated using the Hydrocalculator (43). Hydrophobicity of amino acid residues influences protein folding, protein subunits interactions binding to receptors and interactions of proteins with cell membranes (43). Hydrophobic moment of a sequence gives an indication as to how the hydrophobicity's of its constituent residues of a particular segment of the sequences happen to be folded into a particular conformation, i.e.  $\alpha$ -helix and  $\beta$ -helix. SignalP (44,45) can be used to predict the signal cleavage site of the peptide from different organisms. This is useful to determine the mature part of the peptide that has the activity.

*Catalog-integrated tools.* Each catalog page contains integrated tools such as BLAST, ClustalW, HMMER and Hydrocalculator and SignalP. When the user performs a search, the result page shows the summary of peptide information. The user can choose to process the entire result set or select individual sequences from the result set. The integrated tools are implemented in this framework.

*Other resources.* The retrieved peptides in DAMPD searches can be linked to other databases to provide additional information on these peptides. These resources are described here: *ProtParam* (46) computes the physico-chemical properties of a sequence. *Compute PI/MW* (47,48) requires the user to choose or enter a Swiss-Prot/TrEMBL accession number. *ProtScale* (46) generates a profile of each type of amino acid on a protein. *PeptideMass* (46,49) uses a Swiss-Prot/TrEMBL accession number assigned to a protein to generate peptide information. *PeptideCutter* (46) requires the end user to

enter an accession number used by Swiss-Prot/TrEMBL to uniquely identify proteins. *ModBase* (50) provides predicted protein structure models. *SMART* (51) a Simple Modular Architecture Research Tool maps a protein sequence to its catalog of target domains. *InterProt* (52) uses a host of member databases to generate protein signatures, which are used as a basis to identify distant relationships between potentially novel sequences. *Pfam* (53) is a database of protein family classification, protein domain data and multiple sequence alignments generated using Hidden Markov models. *Prosite* (54) is a database, which contains descriptions and documentation relating to amino acid profiles, protein domains, families and functional sites. *ProtoNet* (55) is a database of computationally derived protein structures, which have been clustered and then hierarchically structured using data, derived from Swiss-Prot/TrEMBL.

### DAMPD VERSUS ANTIMIC

We transformed and upgraded the ANTIMIC database which contained 1799 peptides from UniProt and Gen-Bank, to DAMPD resource. DAMPD is aimed to be one stop web portal user system for AMPs. The capabilities of DAMPD are enriched with several tools that can enhance AMP studies. One of these is the AMP predictor (based on support vector machines, SVM) that can classify very accurately a peptide into a family of AMPs (out of 27 AMP families), a feature that currently no other tool and database has. Users can search the DB either for 'Reviewed' or 'Non-Reviewed', or both classes of peptides. Catalogs help user to search database differently through various aspects like 'Keywords, Taxonomy, Citations, Family and Species'. The system is capable of updates with the latest information on 1232 peptides whenever these UniProt entries are updated. 'Help' pages are provided to give explanations on the use and access to the database. 'Links to other antimicrobial databases' provide direct access to information from other relevant AMP resources. We further enabled users to view the results of their queries either on their computer screen or to receive them by email. We regularly download all peptides with 'keyword: antimicrobial' from UniProt and GenBank, manually verify them as explained earlier and add only those peptide entries which are experimentally validated.

### DAMPD VERSUS OTHER AMP DATABASES

*Supplementary Table 1* provides a short comparison of DAMPD and currently available AMP databases and resources. Significant improvements available in DAMPD include the combination of BLAST, ClustalW, Hydrocalculator, SignalP, AMP prediction using HMMER and SVM-based predictor of AMPs operating on a database of experimentally validated peptides. These features are combined with multilevel catalog searching. The current DAMPD version has 145 entry keyword catalog entries, 943 taxonomy catalog entries. The AMP family catalog has main and sub families of 128 entries, and the species catalog possess 406 entries.

## CONCLUSION

Most of the other ‘antimicrobial’ databases to date are becoming outdated and not regularly maintained. We integrated retrieving, creation of catalogs, database version as semi-automatic process which helps us in updating DAMPD within a day. This process automatically retrieved the new peptides of ‘antimicrobial’ category, but the final inclusion into DAMPD requires checking by a domain expert. DAMPD will be updated regularly on a bi-monthly basis. In the near future we intend to add text-mining capabilities to identify potential AMPs from texts that possibly are not yet annotated as AMPs. We believe that our DAMPD will be a useful resource for researchers in this domain.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR online: Supplementary table S1.

## FUNDING

Funding to A.C., V.S.S., M.N.G., A.P. and S.A. provided through grants from South African Research Chairs Initiative of the department of Science and Technology, as well as from National Research Foundation of South Africa. Funding for open access charge: Base Research Funds of V.B.B. at King Abdullah University of Science and Technology.

*Conflict of interest statement.* None declared.

## REFERENCES

- Hoffmann, J.A., Reichhart, J.M. and Hetru, C. (1996) Innate immunity in higher insects. *Curr. Opin. Immunol.*, **8**, 8–13.
- Garcia-Olmedo, F., Molina, A., Alamillo, J.M. and Rodriguez-Palenzuela, P. (1998) Plant DEFENSE PEPTIDES. *Biopolymers*, **47**, 479–491.
- Vizioli, J. and Salzet, M. (2002) Antimicrobial peptides from animals: focus on invertebrates. *Trends Pharmacol. Sci.*, **23**, 494–496.
- Zasloff, M. (2002) Antimicrobial peptides of multicellular organisms. *Nature*, **415**, 389–395.
- Brogden, K.A., Ackermann, M., McCray, P.B. and Tack, B.F. (2003) Antimicrobial peptides in animals and their role in host defences. *Int. J. Antimicrob. Agents*, **22**, 465–478.
- Ganz, T. (2003) Defensins: antimicrobial peptides of innate immunity. *Nature Rev. Immunol.*, **3**, 710–720.
- Lehrer, R.I. (2004) Primate defensins. *Nat. Rev. Microbiol.*, **2**, 727–738.
- Brogden, K.A. (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.*, **3**, 238–250.
- Giuliani, A., Pirri, G. and Nicoletto, S.F. (2007) Antimicrobial peptides: an overview of a promising class of therapeutics. *Cent. Eur. J. Biol.*, **2**, 1–33.
- Hancock, R.E.W. and Lehrer, R. (1998) Cationic peptides: a new source of antibiotics. *Trends Biotechnol.*, **16**, 82–88.
- Zasloff, M. (2002) Antimicrobial peptides of multicellular organisms. *Nature*, **415**, 389–395.
- Dhople, V., Krukemeyer, A. and Ramamoorthy, A. (2006) The human beta-defensin-3, an antibacterial peptide with multiple biological functions. *Biochim. Biophys. Acta Biomemb.*, **1758**, 1499–1512.
- Yeaman, M.R. and Yount, N.Y. (2003) Mechanisms of antimicrobial peptide action and resistance. *Pharmacol. Rev.*, **55**, 27–55.
- Tossi, A. and Sandri, L. (2002) Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. *Curr. Pharm. Des.*, **8**, 743–761.
- Lohner, K. and Blondelle, S.E. (2005) Molecular mechanisms of membrane perturbation by antimicrobial peptides and the use of biophysical studies in the design of novel peptide antibiotics. *Comb Chem High Throughput Screen*, **8**, 241–256.
- Yount, N.Y., Bayer, A.S., Xiong, Y.Q. and Yeaman, M.R. (2006) Advances in antimicrobial peptide immunobiology. *Biopolymers*, **84**, 435–458.
- Sahl, H., Pag, U., Bonness, S., Wagner, S., Antcheva, N. and Tossi, A. (2005) Mammalian defensins: structures and mechanism of antibiotic activity. *J. Leukocyte Biol.*, **77**, 466–475.
- Yang, D., Biragyn, A., Kwak, L.W. and Oppenheim, J.J. (2002) Mammalian defensins in immunity: more than just microbicidal. *Trends Immunol.*, **23**, 291–296.
- Dürr, M. and Peschel, A. (2002) Chemokines meet defensins -the merging concepts of chemoattractants and antimicrobial peptides in host defense. *Infect. Immun.*, **70**, 6515–6517.
- Peschel, A. and Sahl, H.G. (2006) The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat. Rev. Microbiol.*, **4**, 529–536.
- Pag, U. and Sahl, H.G. (2002) Multiple activities in lantibiotics - models for the design of novel antibiotics? *Curr. Pharm. Des.*, **8**, 815–833.
- Yew, W.W., Lange, C. and Leung, C.C. (2011) Treatment of tuberculosis: update 2010. *Eur. Respir. J.*, **37**, 441–32.
- Wang, G., Li, X. and Wang, Z. (2009) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.*, **37**, D933–D937.
- Hammami, R., Zouhir, A., Hamida, J.B. and Fliss, I. (2007) BACTIBASE: a new web-accessible database for bacteriocin characterization. *BMC Microbiol.*, **7**, 89.
- Seebah, S., Suresh, A., Zhuo, S., Choong, Y.H., Chua, H., Chuon, D., Beuerman, R. and Verma, C. (2007) Defensins knowledgebase: a manually curated database and information source focused on the defensins family of antimicrobial peptides. *Nucleic Acids Res.*, **35**, D265–D268.
- Gueguen, Y., Garnier, J., Robert, L., Lefranc, M.P., Mougnot, I., Lorgeril, J., Janech, M., Gross, P.S., Warr, G.W., Cuthbertson, B. et al. (2006) PenBase, the shrimp antimicrobial peptide pnaeicid database: sequence-based classification and recommended nomenclature. *Dev. Compar. Immunol.*, **30**, 283–288.
- Whitmore, L. and Wallace, B.A. (2004) The peptaibol database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Res.*, **32**, D593–D594.
- Wade, D. and Englund, J. (2002) Synthetic antibiotic peptides database. *Prot. Pept. Lett.*, **9**, 53–57.
- Fjell, C.D., Hancock, R.E.W. and Cherkasov, A. (2007) AMPer: a database and an automated discovery tool for antimicrobial peptides. *Bioinformatics*, **23**, 1148–1155.
- Wang, C.K., Kaas, Q., Chiche, L. and Craik, D.J. (2008) CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.*, **36**, D206–D210.
- de Jong, A., van Heel, A.J., Kok, J. and Kuipers, O.P. (2010) BAGEL2: mining for bacteriocins in genomic data. *Nucleic Acids Res.*, **38**, W647–W651.
- Thomas, S., Karnik, S., Barai, R.S., Jayaraman, V.K. and Idicula-Thomas, S. (2010) CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Res.*, **38**, D774–D780.
- Hammami, R., Ben Hamida, J., Vergoten, G. and Fliss, I. (2009) PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic Acids Res.*, **37**, D963–D968.
- Li, Y. and Chen, Z. (2008) RAPD: a database of recombinantly-produced antimicrobial peptides. *FEMS Microbiol. Lett.*, **289**, 126–129.
- Brahmachary, M., Krishnan, S.P., Koh, J.L., Khan, A.M., Seah, S.H., Tan, T.W., Brusica, V. and Bajic, V.B. (2004) ANTIMIC: a database of antimicrobial sequences. *Nucleic Acids Res.*, **32**, D586–D589.

36. Apweiler,R., Martin,M.J., O'Donovan,C., Magrane,M., Alam-Faruque,Y., Antunes,R., Barrell,D., Bely,B., Bingley,M., Binns,D. *et al.* (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.*, **39**, D214–D219.
37. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2011) GenBank. *Nucleic Acids Res.*, **39**, D32–D37.
38. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J Mol Biol.*, **215**, 403–410.
39. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
40. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
41. Perrière,G. and Gouy,M. (1996) WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie.*, **78**, 364–369.
42. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics.*, **14**, 755–763.
43. Tossi,A., Sandri,L. and Giangaspero,A. (2002) New consensus hydrophobicity scale extended to non-proteinogenic amino acids. In *Peptides 2002: Proceedings of the twenty-seventh European peptide symposium*. Edizioni Ziino, Napoli, Italy, pp. 416–417.
44. Bendtsen,J.D., Nielsen,H., von Heijne,G. and Brunak,S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, **340**, 783–795.
45. Petersen,T.N., Brunak,S., von Heijne,G. and Nielsen,H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.
46. Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S., Wilkins,M.R., Appel,R.D. and Bairoch,A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker,J.M. (ed.), *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, USA, pp. 571–607.
47. Bjellqvist,B., Hughes,G.J., Pasquali,Ch., Paquet,N., Ravier,F., Sanchez,J.-Ch., Frutiger,S. and Hochstrasser,D.F. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**, 1023–1031.
48. Bjellqvist,B., Basse,B., Olsen,E. and Celis,J.E. (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis*, **15**, 529–539.
49. Wilkins,M.R., Lindskog,I., Gasteiger,E., Bairoch,A., Sanchez,J.-C., Hochstrasser,D.F. and Appel,R.D. (1997) Detailed peptide characterization using PEPTIDEMASS - a World-Wide Web accessible tool. *Electrophoresis*, **18**, 403–408.
50. Pieper,U., Webb,B.M., Barkan,D.T., Schneidman-Duhovny,D., Schlessinger,A., Braberg,H., Yang,Z., Meng,E.C., Pettersen,E.F., Huang,C.C. *et al.* (2011) ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.*, **39**, D465–D474.
51. Letunic,I., Doerks,T. and Bork,P. (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res.*, **37**, D229–D232.
52. Hunter,S., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Das,U., Daugherty,L., Duquenne,L. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **37**, D211–D215.
53. Finn,R.D., Mistry,J., Tate,J., Coggill,P., Heger,A., Pollington,J.E., Gavin,O.L., Gunasekaran,P., Ceric,G., Forslund,K. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.
54. Sigrist,C.J., Cerutti,L., de Castro,E., Langendijk-Genevaux,P.S., Bulliard,V., Bairoch,A. and Hulo,N. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.*, **38**, D161–D166.
55. Kaplan,N., Sasson,O., Inbar,U., Friedlich,M., Fromer,M., Fleischer,H., Portugaly,E., Linial,N. and Linial,M. (2005) ProtoNet 4.0: a hierarchical classification of one million protein sequences. *Nucleic Acids Res.*, **33**, D216–D218.