The Peptaibol Database: a database for sequences and structures of naturally occurring peptaibols

Lee Whitmore and B. A. Wallace*

Department of Crystallography, Birkbeck College, University of London, London WC1E 7HX, UK

ABSTRACT

The Peptaibol Database is a sequence and structure resource for the unusual class of peptides known as peptaibols. These peptides exhibit antibiotic and membrane channel-forming activities. The database includes sequence, biological source and bibliographical data for the naturally occurring peptaibols. Information is also collated for the growing number of peptaibol 3D structures determined by either crystallography or NMR spectroscopy. The database can be obtained as a whole or can be queried by name, group, sequence motif, biological origin and/or literature reference. The Peptaibol Database can be freely accessed at http://www.cryst.bbk.ac.uk/peptaibol.

INTRODUCTION

Peptaibols are a family of peptides characterized by short chain lengths (≤20 residues), C-terminal alcohol residues and high levels of non-standard amino acids, principally α-aminoisobutyric acid (Aib), isovaleric acid (Iva) and the imino acid hydroxyproline (Hyp). Naturally occurring peptaibols are isolated from soil fungi and often exhibit antibacterial and antifungal properties. These peptides are synthesized non-ribosomally by large multidomain enzymes (1).

The unusual amino acids and short lengths of the peptaibols (as well as the absence of genetic coding of the sequences) has meant that such molecules have not generally been included in protein sequence databases such as Swiss-Prot and TrEMBL (2). When the Peptaibol Database was first created in 1997, it was the only database that included these peptide sequences. Advances in sequence database design and software since that time has meant that the unusual sequence characteristics are now less of a problem for inclusion in standard protein sequence databases. However, whilst most peptaibol sequences can now be found using the Sequence Retrieval System (3) by searching across multiple databases, the results are scattered and occur in different formats with different approaches employed to deal with the non-standard residues.

The antibiotic functions of peptaibols arise from their membrane insertion and pore-forming abilities. The intermediate size and hydrophobic nature of these peptides has meant that very few of their structures have been solved, either by crystallographic or NMR techniques. Some of their coordinates have been deposited in the macromolecular Protein Data Bank (PDB) (4) and some in the small molecule Cambridge Structural Database (5). All of these structures are available in the Peptaibol Database.

DATABASE CONTENT AND ACCESS

The Peptaibol Database was created in order to store all sequence and structure information for peptaibols in a single place with a consistent formatting scheme. Synthetic peptaibols and peptaibol analogues are not included in the database, but may be found in the Synthetic Antibiotic Peptides Database (6).

The database is arranged into two sections, 'Sequences' and 'Structures'; a navigation bar appears on every page that links to these sections and to information pages that include background material on peptaibols, nomenclature definitions of non-standard residues, and subfamily groups and sequence alignments (7).

The 'Sequences' section of the database allows searches of the database by peptaibol name, group, sequence motif and biological source. Each searching option is furnished with instructions and there is a quick-search box for frequent users. Results of searches provide further links to the full record of each individual peptaibol matched.

The 'Structures' section of the database is currently arranged as a table due to the relatively small number of available peptaibol structures. Each entry in the table has a link to downloadable PDB-style (4) coordinates marked 'crystal' or 'nmr' as appropriate. Molecular graphics visualizations of the structures are also provided.

There are currently 307 sequence entries and nine structure entries (of seven different peptaibols) in the database. Statistical analyses on peptaibol sequence features have been derived from use of this database (8).

AVAILABILITY

The Peptaibol Database is freely available at http://www.cryst.bbk.ac.uk/peptaibol. All comments, queries, requests and corrections should be sent by email to peptaibo@mail.cryst.bbk.ac.uk. Depositions of new sequence or structure entries of naturally occurring peptaibols should be sent to this address via email and should be accompanied by a reference to a published, peer-reviewed article.

*To whom correspondence should be addressed. Fax: +44 207 6316803; Email: ubcg25a@mail.cryst.bbk.ac.uk

ACKNOWLEDGEMENTS

REFERENCES

1. Wiest,A., Grzegorski,D., Xu,B.W., Goulard,C., Rebuffat,S., Ebbole,D.J., Bodo,B. and Kenerley,C. (2002) Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase. *J. Biol. Chem.*, **277**, 20862–20868.
2. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.
3. Zdobnov,E.M., Lopez,R., Apweiler,R. and Etzold,T. (2002) The EBI SRS server—new features. *Bioinformatics*, **18**, 1149–1150.
4. Westbrook,J., Feng,Z., Chen,L., Yang,H. and Berman,H.M. (2003) The Protein Data Bank and structural genomics. *Nucleic Acids Res.*, **31**, 489–491.
5. Allen,F.H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B*, **58**, 380–388.
6. Wade,D. and Englund,J. (2002) Synthetic antibiotic peptides database. *Prot. Pept. Lett.*, **9**, 53–57.
7. Chugh,J.K. and Wallace,B.A. (2001) Peptaibols: models for ion channels. *Biochem. Soc. Trans.*, **29**, 565–570.
8. Whitmore,L. and Wallace,B.A. (2003) Analysis of peptaibol sequence composition: implications for *in vivo* synthesis and channel formation *Eur. Biophys. J.*, DOI: 10.1007/s00249-003-0348-1.