Journal of Peptide Science
J. Peptide Sci. **9**: 663–665 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.533

JOURNAL OF PEPTIDE SCIENCE

The Peptaibol Database: A Sequence and Structure Resource‡

L. WHITMORE, J. K. CHUGH, C. F. SNOOK and B. A. WALLACE*

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

Abstract: The peptaibols are a large family of membrane-active peptides with considerable sequence homology, but with different biological properties and three-dimensional structures. They constitute a rich resource of naturally occurring 'mutants' which are potentially valuable for structure/function studies of ion channels.

A searchable on-line database of sequences and structures of the peptaibols has been created at http://www.cryst.bbk.ac.uk/peptaibol, as a resource for the biological and structural community. In this paper, the contents and organization of the website are discussed as well as procedures for submission of new entries to the database.

At present, more than 300 peptaibol sequences are stored in the database. Each sequence entry contains its full literature reference and information about its biological source. Tools are provided for searching for specific peptaibol sequences or groupings of sequences, and for locating peptaibols containing specified sequence motifs. In addition the website acts as a database for structural information. The coordinates of all currently available peptaibol x-ray and NMR structures are included and complemented, where appropriate, with molecular graphics illustrations. These include figures of model channel structures and comparisons between different peptaibol structures. The peptaibol database thus provides a tool for ready access to information and a means of investigating the sequences and structures of this class of polypeptides.

Keywords: bioinformatics; sequence; crystal structure; database; peptide; ion channel; alamethicin

INTRODUCTION/RESULTS/DISCUSSION

The peptaibols are a family of biologically active polypeptides ranging in length from 5 to 20 residues, that have been isolated from a wide range of fungal sources. Most are capable of acting as antibiotics against specific bacteria, parasites and/or viruses, and in general, their activities are derived from their membrane-active properties [1].

One of the prevailing characteristics of this family is the presence of an abundance of non-standard amino acid residues, including Aib (α-amino isobutyric acid or α-methyl alanine) and Iva (isovaleric acid or α-ethyl alanine). Due to these non-standard amino acids and their relatively short lengths, peptaibol sequences tend not to be lodged in traditional protein sequence databases such as SWISS-PROT or TREMBL [2], and because their intermediate size falls at the boundary between large and small molecules, their crystal structures may be deposited in either the Protein Data Bank [3] (PDB) or in the Cambridge Structural Database [4] (CSDB), or neither. Sometimes their NMR structures are deposited in the PDB, but not always. Thus, because mainstream protein informatics fails to adequately record and categorize members of the peptaibol family, the peptaibol database described in this paper was designed to provide a comprehensive resource for the naturally occurring peptaibol sequences and structures.

More than 300 sequences of naturally occurring peptaibols have been identified thus far, with the numbers increasing significantly on a yearly basis

* Correspondence to: Dr B. A. Wallace, Department of Crystallography, Birkbeck College, University of London, London WC1E 7HX, UK; e-mail: ubcg25a@mail.cryst.bbk.ac.uk
‡ Presented at Peptaibols, October 2002, Jena, Germany.
Contract/grant sponsor: BBSRC.

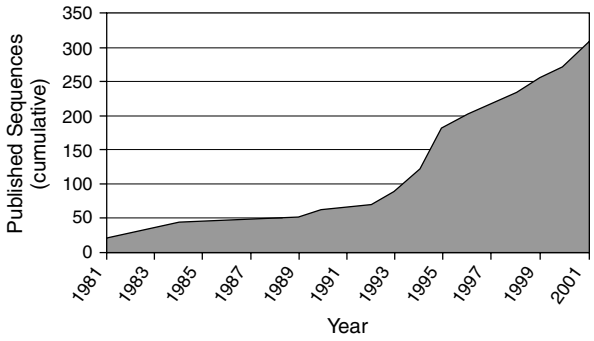

Figure 1 Plot showing cumulative number of naturally occurring peptaibol sequences published as a function of calendar year.

(Figure 1). Because the peptaibols have considerable sequence homology, but include a large number of sequence variants with different biological properties and three-dimensional structures, they constitute an excellent library of naturally occurring 'mutants' with which to investigate structure/function activity in ion channels [5]. Their potential utility has led us to produce a web-accessible database which includes sequence and structure information that is not easily found elsewhere for these unusual peptides.

Website Organization

The website, located at http://www.cryst.bbk.ac.uk/peptaibol, contains the following main sections: information, sequences and structures. The information sections are accessed via the 'Introduction', 'Information' and 'Nomenclature' hyperlinks. The sequence section is accessed either through the searching tools, or through links that return web pages containing the entire current database. The structure section is organized as a table, with entries for each available structure. Structural information, whether from crystallographic or NMR studies, is much rarer than sequence information for these molecules, meaning that a table of manageable size is a convenient means of providing all of the structural data.

Information

Given that the peptaibols inherently contain non-standard amino acid units, a clear and consistent naming system is described in the 'Nomenclature' section of the website. Currently, six non-standard or incompletely resolved residues are given the codes U, J, O, Z, Lx, Vx, respectively, for amino isobutyric acid (Aib), isovaline (Iva), hydroxyproline (Hyp), ethylnorvaline (EtNor), unresolved leucine/isoleucine and unresolved valine/isovaline.

The 'Information' section of the website features the aligned sequences of the peptaibols and the subfamily classification system [6].

Sequences

Sequence information can be obtained from the database in several ways: searching the database by peptaibol name, either in full or in part, will produce the names and sequences of all matching entries. Using the pull-down menu to search by peptaibol group name will result in the names and sequences for matches of that particular group only. Searching the database for sequences or fragments of sequences will generate a list of names and sequences of entries matching the search criteria. In addition to the searching tools, the whole of the current database may be displayed either as a list of names and sequences, names alone or names and literature references. Each peptaibol name is itself a link to the full record for that peptaibol, where the record comprises the name, the sequence, the biological source and the literature reference.

Structures

The structure section of the website includes not only coordinates, but also images of the structures. The coordinate data are in PDB-style format, readable by standard graphics packages such as Rasmol [7] and Chime, and include their ID codes from the original databases in which they were lodged (PDB or CSDB). Both x-ray and NMR structures are included. In the case of NMR structures, the coordinates of either each of the members of the ensemble of structures or only the average coordinates are included (whichever was provided in the original PDB entry).

The images presented in the structure section of the website give graphical meaning to the structural data and aim to highlight the important structural features of peptaibols. A number of the images are overlaid comparisons for the same molecule; as examples, the two crystal forms of antiamoebin, or the NMR and crystal structures of antiamoebin. Other images are of comparisons of different peptaibol structures (i.e. alamethicin, antiamoebin and zervamicin), highlighting structurally important features such as length and location of aromatic residues and imino acids [5]. Finally, multimeric

channel structures modelled based on the crystal structures are included where available, and in one case, a schematic diagram of membrane insertion based on the structures is presented.

Conclusions

The peptaibol database is a freely accessible online resource providing information on sequences, structures, subfamily classification, biological sources and references for more than 300 naturally occurring peptaibols.

Experimental

An online database has been compiled for the sequences and structures of naturally occurring peptaibols. Access is via a server located at: http://www.cryst.bbk.ac.uk/peptaibol. Access is freely available for non-commercial use; users should credit the site by referencing this publication, as well as, of course, the original sequence or structure papers.

Depositions. New sequences and structures may be deposited in the database by contacting the administrators of the database by e-mail (peptaibo@mail.cryst.bbk.ac.uk) or fax and providing copies of the relevant original publication or database entry. Only published sequences or structures of naturally occurring peptaibols are considered for inclusion.

Current holdings. The peptaibol database currently holds 307 sequence records, 9 coordinate sets, 11 molecular images, 5 comparative molecular images and a complete listing of the subfamily classification.

Acknowledgements

This work was supported, in part, by a grant from the BBSRC (to BAW). J.K.C. and C.F.S. were recipients of BBSRC studentships.

REFERENCES

1. Sansom MSP. Alamethicin and related peptaibols – model ion channels. *Eur. Biophys. J.* 1993; **22**: 105–124.
2. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 2000; **28**: 45–48.
3. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res.* 2000; **28**: 235–242.
4. Allen FH, Kennard O. 3D search and research using the Cambridge structural database. *Chem. Design Automat. News* 1993; **8**: 31–37.
5. Wallace BA. Common structural features in gramicidin and other ion channels. *Bioessays* 2000; **22**: 227–234.
6. Chugh JK, Wallace BA. Peptaibols: models for ion channels. *Biochem. Soc. Trans.* 2001; **29**: 565–570.
7. Sayle R, Milner-White EJ. RasMol: biomolecular graphics for all. *Trends Biochem. Sci.* 1995; **20**: 374–376.