Eur Biophys J (2004) 33: 233–237
DOI 10.1007/s00249-003-0348-1

ARTICLE

L. Whitmore · B. A. Wallace

Analysis of peptaibol sequence composition: implications for in vivo synthesis and channel formation

Received: 17 June 2003 / Accepted: 27 July 2003 / Published online: 8 October 2003

Abstract The sequence entries in the Peptaibol Database were analysed to provide information on compositional features of this unusual family of peptides. The non-standard amino acid α-aminoisobutyric acid represents almost 40% of the residues in all the known sequences. Glutamine is the only significant polar residue in peptaibols, and the position and number of these residues appear to be related to their functional properties as ion channels. Aromatic residues are clustered at the termini, which may contribute to stabilization of the peptide vertically within the bilayer. The peptide chain length is strongly weighted towards the longer members of the family (16–20 residues) and likely to be an important feature in their mode of action as transmembrane permeabilizers. The significant skewing towards even numbers of residues and the bias in pairwise distributions of amino acids have implications for the nature of the in vivo synthesis of these peptides via large non-ribosomal protein complexes.

Keywords Bioinformatics · Database · Ion channels · Non-ribosomal synthesis · Peptide structure

Introduction

Peptaibols constitute a family of peptides that are isolated from soil fungi and exhibit antibacterial and antifungal properties. Many naturally occurring peptaibols have been isolated and sequenced. Their sequences have been collected in the Peptaibol Database (http://www.cryst.bbk.ac.uk/peptaibol) (Whitmore et al. 2003). Alignment of the sequences and a subsequent grouping into subfamilies (Chugh and Wallace 2001) have shown that there is a high degree of residue conservation amongst members of the subfamilies, especially in the relative sequence positions of the functionally important Gln residue.

Peptaibols are non-ribosomally synthesized on very large protein complexes, which apparently include different modules specific for the addition of different types of amino acids (Wiest et al. 2002). The amino acid compositions of peptaibols differ from ribosomally synthesized proteins owing to the presence of non-standard, non-coded amino acids, including large quantities of α-aminoisobutyric acid (Aib), plus isovaleric acid (Iva) and the imino acid hydroxyproline (Hyp).

Even though they are relatively small molecules, the peptaibols apparently have regular defined structures (Toniolo and Benedetti 1991). Few peptaibol crystal and NMR structures have been solved to date (Fox and Richards 1982; Esposito et al. 1987; Karle et al. 1991; Franklin et al. 1994; Toniolo et al. 1994; Karle et al.1998; Snook et al.1998; Anders et al. 2000; Balashova et al. 2000; Chugh et al. 2002; Shenkarev et al. 2002; Galbraith et al. 2003), but those that have show a high preference for helical conformations, with perturbations of the helix around the positions of the imino acid residues present. Based on theoretical studies (Prasad and Balaram 1984), the Aib residue has been proposed to be a strong helix-promoter. The structures of the peptaibols appear to represent a good experimental verification of this, and it is thought that their high Aib contents drive their overall structures to be predominantly helical, regardless of sequence.

It has been proposed that the "long" peptaibols (≥16 residues) insert into membranes as "barrel staves" (Boheim 1974) to form ion channels, thereby giving rise to their antibiotic properties. Their specificities towards different target organisms may depend on their sequences and chain lengths. Because the "short" peptaibols are insufficient in length to span lipid bilayers, they have been proposed to form different types of aggregates (perhaps involving two peptide molecules which meet in

Presented at the Biophysical Society Meeting on Ion Channels – from structure to disease held in May 2003, Rennes, France

L. Whitmore · B. A. Wallace (✉)
Department of Crystallography, Birkbeck College, University of London, London, WC1E 7HX, UK
E-mail: ubcg25a@mail.cryst.bbk.ac.uk
Fax: +44-207-6316803

the centre of the membrane) in order to function in a membrane-active fashion (Toniolo et al. 1996; Chugh and Wallace 2001).

The distributions and spatial positionings of amino acids across the whole family of peptaibols have not previously been addressed, especially with respect to possible implications for in vivo synthesis via synthetase complexes, and for ion channel formation propensity. The collection of sequences amassed in the Peptaibol Database is now numerous enough to allow such analyses.

Methods

As of 1 May 2003, the Peptaibol database contained 307 naturally occurring peptaibol sequences of varying lengths. Many other peptaibol analogues have been created synthetically; these are not included in the Peptaibol Database but are included in the Synthetic Antibiotic Peptides Database (Wade and Englund 2002). The synthetic peptaibols were not, however, considered in the present study as they do not represent in vivo-produced molecules, and so would skew analyses that examine the relationship of the sequences to the biosynthesis by synthetase complexes.

The entire content of the Peptaibol Database was downloaded and analysed using a simple script written in the Perl programming language. Where the C-terminal residue was modified (usually to an alcohol), that residue was considered to be regular for the purposes of the analysis (i.e. there was no distinction made between valine and valinol).

Results

Chain length

The 307 sequences in the Peptaibol Database contain a total of 5117 residues. They range in chain length from 5 to 20 residues. The mean chain length of 16.7 is unlikely to be of great significance as there is an uneven distribution of the number of sequences of each chain length. As shown in Fig. 1, some chain lengths are favoured, with the maximum chain length of 20 having the largest number of examples, representing more than 30% of the total sequences. Seventy percent of the sequences have chain lengths greater than or equal to 16 residues, in a helical conformation they would be sufficiently long to span at least some lipid bilayers.

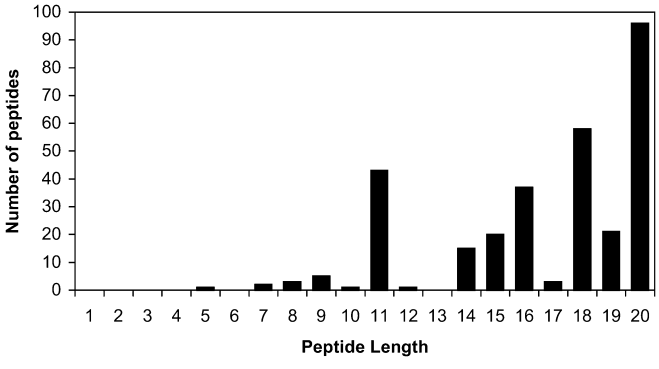

Fig. 1 The distribution of chain lengths amongst the peptaibols

The preference for longer chains lengths may be related to the membrane-spanning function of many of the peptaibols. The length distributions in the short peptaibols (with a peak at chain length 11) may be related to the alternative type of membrane insertion model proposed for them. In such cases, short peptaibols would stack pairwise in order to produce structures sufficiently long to span the bilayer, with their N-termini associated in the centre of the membrane, so lengths around 11 residues would be appropriate for this.

There is a strong bias in chain lengths for even numbers of residues over odd numbers, with 69% of all peptaibols having even numbers. When separated into long and short sequences, the skewing is even greater (89% of the long sequences have even chain lengths). This suggests that the synthetase complexes may have an internal pseudosymmetry. Studies on *Trichoderma virens* (Wiest et al. 2002) indicate that there is a single synthetase gene with 18 modules (one for each residue in the longest peptaibol made by this organism) which is responsible for the synthesis of all the peptaibol isoforms it produces. The sequence variants are largely due to multiple specificities of some of the modules, and, by implication, shorter peptaibols are made by the complexes using only some of the modules.

Residue types

Table 1 indicates the number and frequency of occurrences of each residue type. This survey shows that, approximately 40% of all peptaibol residues are Aib. Figure 2 shows the relationship between chain length and number of Aib residues. Each dot indicates that an example of that ratio exists and the larger dots show where significant populations of those examples occur. The Aib content ranges from 14% to 56%; the short peptaibols generally have a lower percentage of Aib residues than the long peptaibols, suggesting that more hydrophobic helix-promoting Aib residues are required to stabilize longer helices. Interestingly, Aib is never found in positions 18, 19 or 20 in the long peptaibols.

The second most frequent amino acid present is glutamine. Glutamine occurs in 86% of the entries and generally appears in specific places in the sequences, such as positions 6 or 7, positions $n-1$ or residue $n-6$, where n is the length of the chain. Residues at positions 6 and $n-6$ would lie at the same depth in the membrane in opposite leaflets of the bilayer. The Gln6 or Gln7 residue has been suggested to be the crucial residue in the control of ion flow through a transmembrane pore (Nagaoka et al. 1996). Considering the major groups in the subfamily (SF) classifications, peptaibols in SF1 have two or three Gln residues, whilst SF2 sequences have just one. As SF1 constitutes the bulk of peptaibol sequences, it is unsurprising that the overall percentage of glutamine residues is about 10%, or on average two residues in a 20-residue chain.

Table 1 Overall frequencies of amino acid occurrences. AHMO is 2-amino-6-hydroxy-4-methyl-8-oxodecanoic acid, Lxx and Vxx are not distinguished between the isomeric pairs leucine/isoleucine and valine/isovaline, respectively, and ETN is α-amino-α-ethyl-*n*-pentanoic acid

Residue	Occurrences	Percentage of all residues
Aib	1968	38.5
Gln	551	10.8
Leu	443	8.7
Ala	406	7.9
Pro	390	7.6
Val	281	5.5
Gly	226	4.4
Iva	174	3.4
Phe	164	3.2
Ile	109	2.1
Lxx	105	2.1
Hyp	101	2.0
Ser	57	1.1
Trp	40	0.8
Glu	38	0.7
Asn	30	0.6
Thr	17	0.3
AHMO	7	0.1
Vxx	7	0.1
EtN	2	0.0
Tyr	1	0.0
Arg	0	0.0
Asp	0	0.0
Cys	0	0.0
His	0	0.0
Lys	0	0.0
Met	0	0.0

Aromatic amino acids are found infrequently in peptaibols (they comprise only a total of 4% of the amino acids present) (Table 1). However, they are heavily biased in their positional distribution, being found almost exclusively at either the N- and C-termini, with only seven (out of a total of 204) examples of aromatic residues found at any other position in the sequences of any known peptaibol (Fig. 3). This places them at the bilayer interface, a position where aromatic amino acids are frequently found in large membrane proteins (Wallace and Janes 1999). It has been suggested (Schiffer et al. 1992) that aromatic amino acids may act to stabilize transmembrane segments by interaction with the polar head groups of the membrane lipids; this would be consistent with the present observation of their positional distributions.

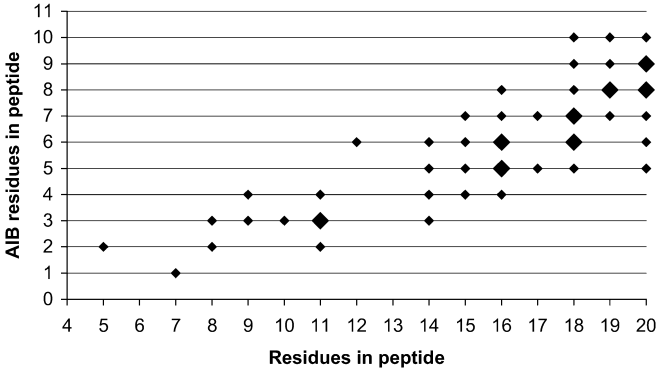

Fig. 2 The distribution of numbers of Aib residues as a function of peptaibol chain length. *Double sized points* are shown where there are more than 10 examples and where those examples constitute at least 25% of the total examples for that chain length

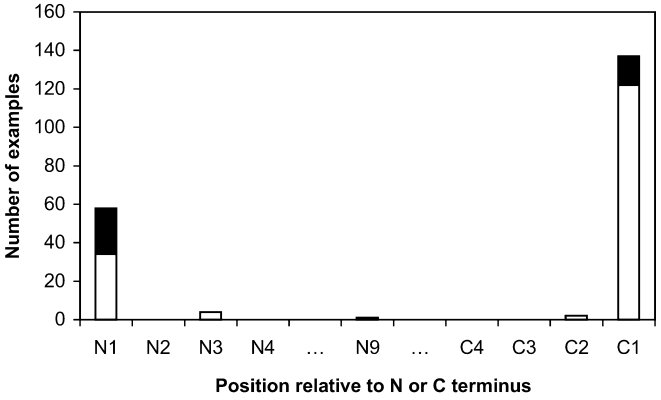

Fig. 3 The positional distributions of the aromatic amino acids from the N-terminus, N1, N2, etc., and from the C-terminus, C1, C2, etc. Because of the differences in chain lengths, each aromatic residue was only counted against the terminal it was closest to. The *open bars* represent phenylalanines and the *closed bars* represent tryptophans

The residues in the sequences that are not Aib or Gln are mainly non-polar or aliphatic; polar and charged amino acids other than glutamine are so underrepresented that altogether they only account for 4% of the total residues and several amino acids are not found at all (Table 1). A large aliphatic content would be expected for transmembrane peptides, the maximization of hydrophobic interactions providing a driving force for insertion into the membrane and association with other peptaibols to form channels.

The pairwise distributions of the amino acids are not random, with some pairs strongly favoured and some rarely or never found (Table 2). For example, prolines are almost exclusively (372/381) preceded by Aib, and hydroxyprolines (Hyp) are only preceded by one of the α,α-amino acids (Aib or Iva). Conversely, several of the more hydrophobic types, i.e. valine and leucine, are strongly biased against following Aib. Glycine frequently precedes leucine. These non-random pairwise distributions may suggest a mechanism by which biosynthetic incorporation of two-residue units are favoured, and perhaps also contribute to the bias towards even chain lengths.

A comparison that may be interesting in structural terms is the relative locations of the Gly and Pro residues, which can be responsible for producing either kinks in helices or β-turns. In all the peptaibol sequences examined, there were no examples of Gly and Pro residues being adjacent to each other, nor separated by one amino acid (i.e. Gly-Pro, Pro-Gly, Gly-X-Pro, Pro-X-Gly), nor Pro preceeding Gly by two residues (Pro-X-X-Gly), where X respresents any other amino acid. However, there are a large number (130) of examples of the Gly-X-X-Pro

Table 2 Frequencies of nearest-neighbour pairs of amino acids. The vertical column indicates the N-terminal residue of each pair

	Aib	Gln	Leu	Ala	Pro	Val	Gly	Iva	Phe	Ile	Lxx	Hyp	Ser	Trp	Glu	Asn	Thr	AHMO	Vxx	EtN	Tyr
Aib	447	245	87	244	372	72	184	72	10	38	8	75	46	0	21	30	0	0	3	0	0
Gln	186	87	67	9	0	31	7	48	69	17	5	0	0	12	7	0	0	0	2	0	0
Leu	319	7	8	6	0	2	0	10	0	4	0	0	0	0	0	0	0	0	0	0	0
Ala	232	57	26	49	8	5	6	10	3	3	0	0	3	0	0	0	0	0	2	0	0
Pro	53	1	95	9	0	117	0	13	41	5	39	0	6	0	0	0	0	7	0	0	1
Val	206	2	1	2	0	7	12	0	0	0	8	0	0	0	1	0	3	0	0	0	0
Gly	78	0	93	24	0	5	5	8	2	4	6	0	0	0	0	0	0	0	0	0	0
Iva	27	59	23	2	1	10	9	3	0	7	0	25	0	0	5	0	0	0	0	0	0
Phe	36	0	0	1	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0
Ile	32	19	13	4	0	2	0	0	0	7	0	0	0	0	3	0	13	0	0	0	0
Lxx	49	0	0	5	0	4	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0
Hyp	46	48	2	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Ser	7	0	12	30	0	2	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
Trp	0	0	0	8	0	4	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0
Glu	2	18	0	0	0	6	0	3	4	0	0	0	0	3	1	0	0	0	0	0	0
Asn	0	0	11	0	0	6	0	0	0	6	7	0	0	0	0	0	0	0	0	0	0
Thr	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AHMO	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vxx	0	3	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EtN	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tyr	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	1737	548	438	405	381	275	223	170	129	107	105	100	57	15	38	30	16	7	7	0	1

motif, which would be structurally compatible with a kink motif, but is very different than the classical turn (especially type II) pattern of X-Pro-Gly-X. This may be why the prolines in peptaibols tend to form distortions to the helical segments, but do not produce turns in any of the structures thus far seen.

Discussion

The large numbers of peptaibol sequences identified thus far are derived from a relatively small number of biological sources. They represent a rich source of "mutant" peptides which enable structure–function studies, especially of ion channels (Sansom 1993; Wallace 2000). Peptaibol function appears to stem from three key aspects of their sequences, namely that the positions of the important glutamine residues are conserved, that they contain enough Aib to force a helical configuration and that they are overwhelmingly aliphatic to promote membrane insertion. The wide variations in the peptaibol sequences indicate that the positionings of the aliphatic residues in the peptaibol sequences are relatively unimportant.

The abundance of the characteristic amino acid Aib present in the known sequences of peptaibols varies from 14% to 56%, with a mean of 38%. Long peptaibols generally have a higher percentage of Aib residues than shorter ones. Long peptaibols also are more frequently found in the database than short ones, with the maximum sequence length of 20 also being the most frequent length. The distribution of residues across all the sequences showed that glutamine was the only significant polar residue to be incorporated, a further indication that the peptaibol function is dictated by the positioning of glutamine residues. The sequences of peptaibols are mostly aliphatic, which gives impetus to membrane insertion and stability to transmembrane configurations. The stability may be augmented by the presence of aromatic amino acids at the termini, a frequent occurrence. The patterns of amino acid pairs observed and the high-frequency of even-numbered residue lengths suggest an internal pseudosymmetry present in the large synthetase complexes which produce the peptaibols in vivo.

Acknowledgements This work was supported, in part, by BBSRC grant B13586 to B.A.W.

References

Anders R, Ohlenschlager O, Soskic V, Wenschuh H, Heise B, Brown LR (2000) The NMR solution structure of the ion channel peptaibol chrysospermin C bound to dodecylphosphocholine micelles. Eur J Biochem 267:1784–1794

Balashova TA, Shenkarev, ZO, Tagaev AA, Ovchinnikova TV, Raap J, Arseniev AS (2000) NMR structure of the channel-former zervamicin IIB in isotropic solvents. FEBS Lett 466:333–336

Boheim G (1974) Statistical analysis of alamethicin channels in black lipid membranes. J Membr Biol 19:277–303

Chugh JK, Wallace BA (2001) Peptaibols: models for ion channels. Biochem Soc Trans 29:565–570

Chugh JK, Brückner H, Wallace BA. (2002) Model for a helical bundle channel based on the high resolution crystal structure of trichotoxin_A50E. Biochemistry 41:12934–12941

Esposito G, Carver JA, Boyd J, Campbell ID (1987) High resolution ^{1}H NMR study of the solution structure of alamethicin. Biochemistry 26:1043–1050

Fox RO Jr, Richards FM (1982) A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5 Å resolution. Nature 300:325–330

Franklin JC, Ellena JF, Jayasinghe S, Kelsh LP, Cafiso DS (1994) Structure of micelle-associated alamethicin from ^{1}H NMR: evidence for conformational heterogeneity in a voltage-gated peptide. Biochemistry 33:4036–4045

Galbraith TP, Harris R, Driscoll PC, Wallace BA (2003) Solution NMR studies of antiamoebin, a membrane channel-forming polypeptide. Biophys J 84:185–194

Karle IL, Flippen-Anderson JL, Agarwalla S, Balaram P (1991) Crystal structure of [leu1]-zervamicin, a membrane ion-channel peptide: implications for gating mechanisms. Proc Natl Acad Sci USA 88:5307–5311

Karle IL, Perozzo MA, Mishra VK, Balaram P (1998) Crystal structure of the channel-forming polypeptide antiamoebin in a membrane-mimetic environment. Proc Natl Acad Sci USA 95:5501–5504

Nagaoka Y, Iida A, Kambara T, Asami K, Fujita T (1996) Role of 61n (7) in the ion channel forming, properties of the peptaibol trichosporin-B-VIa. Chem Commun 9:1079–1080

Prasad BVV, Balaram P (1984) The stereochemistry of peptides containing alpha-aminoisobutyric acid. CRC Crit Rev Biochem 16:307–348

Sansom MSP (1993) Alamethicin and related peptaibols: model ion channels. Eur Biophys J 22:105–124

Schiffer M, Chang CH, Stevens FJ (1992) The functions of tryptophan residues in membrane proteins. Protein Eng 5:213–214

Shenkarev ZO, Balashova TA, Efremov RG, Yakimenko ZA, Ovchinnikova TV, Raap J, Arseniev AS (2002) Spatial structure of zervamicin IIB bound to DPC micelles: implications for voltage-gating. Biophys J 82:762–771

Snook CF, Woolley, GA, Oliva G, Pattabhi V, Wood SP, Blundell TL, Wallace BA (1998) The structure and function of antiamoebin I, a proline-rich membrane-active polypeptide. Structure 6:783–792

Toniolo C, Benedetti E (1991) The polypeptide 3_{10} helix. Trends Biochem Sci 16:350–353

Toniolo C, Peggion E, Crisma M, Formaggio F, Shui XQ, Eggleston DS (1994) Structure determination of racemic trichogin-A-IV using centrosymmetric crystals. Nat Struct Biol 1:908–914

Toniolo C, Crisma M, Formaggio F, Peggion C, Monaco V, Goulard C, Rebuffat S, Bodo B (1996) Effect of N-alpha-acyl chain length on the membrane-modifying properties of synthetic analogs of the lipopeptaibol trichogin_GA IV. J Am Chem Soc 118:4952–4958

Wade D, Englund J (2002) Synthetic antibiotic peptides database. Protein Pept Lett 9:53–57

Wallace BA (2000) Common features in gramicidin and other ion channels. Bioessays 22:227–234

Wallace BA, Janes RW (1999) Tryptophans in membrane proteins: X-ray crystallographic analyses. Adv Exp Med Biol 467:789–799

Whitmore L, Chugh JK, Snook CF, Wallace BA (2003) The peptaibol database: a sequence and structure resource. J Pept Sci (in press)

Wiest A, Grzegorski D, Xu BW, Goulard C, Rebuffat S, Ebbole DJ, Bodo B, Kenerley C (2002) Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase. J Biol Chem 277:20862–20868