Wade Research Foundation Reports (2007) 4 (4)

The PRINCETON Peptide

David Wade, Wade Research Foundation, P.O. Box 257, Princeton, New Jersey 08542

Abstract

This article proposes the chemical synthesis and study of a peptide that has an amino acid (AA) sequence with one letter abbreviations that correspond to the name, PRINCETON: Proline (P)-Arginine (R)-Isoleucine (I)-Asparagine (N)-Cysteine (C)-Glutamic acid (E)-Threonine (T)-Ornithine (O)-Asparagine (N). In this sequence, eight of the nine AAs are represented by official International Union of Pure and Applied Chemistry (IUPAC) one letter abbreviations for AAs, and the letter, O, represents Ornithine. A possible structure for the peptide is described, the existence of similar AA sequences in nature are discussed, and it is concluded that the peptide would probably exhibit biological activities.

Introduction

Princeton is a town in central New Jersey, USA, that is located on what was originally a path used by the Lenni Lenape Indians to travel between the Raritan and Delaware Rivers [1]. The first European settler to the area arrived around 1683, and by 1724 the name, Princeton, was in use. Princeton became the home of the College of New Jersey in 1756, and the college changed its name to Princeton University in 1896 [1, 2]. The university became internationally famous, and, as a result, the name, Princeton, is highly recognizable. Consequently, the name, Princeton, is an ideal candidate for use in testing the name-to-peptide approach for generating novel peptides [3-5].

Among the most important biomolecules of life are proteins, polymers of AAs that are held together by chemical bonds, called peptide bonds [6]. They have been compared to "beads on a string", where the beads are AAs, and the beads plus string is the protein. Proteins come in a variety of sizes, ranging from polymers containing only 2 AAs to polymers containing hundreds of AAs or more. Proteins that contain less than 100 AAs are usually referred to as peptides (Figure 1). There are numerous proteins and peptides in the human body, where they perform functions vital for life. For example, the hormone, insulin, is a peptide containing 51 AAs that is involved in the regulation of carbohydrate and lipid metabolism, and associated with diabetes.

There are about 20 different AAs that occur naturally in proteins, and when describing the AA composition of proteins, chemists commonly use one letter abbreviations that correspond to letters of the English alphabet. These abbreviations have been officially defined by the IUPAC-International Union of Biochemistry and Molecular Biology, Joint Commission on Biochemical Nomenclature. They are widely used in biomedical research, and can be found in any textbook of biochemistry and on the internet [6, 7].

Figure 1. The relationship between AAs, peptides, and proteins.

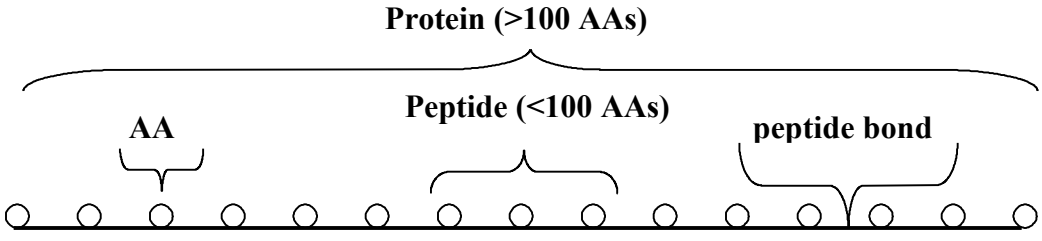
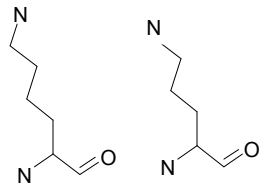


Figure 2. The chemical structures of Lysine (left) and Ornithine (right). Carbon and hydrogen atoms are not shown.



The letter, O, has not been officially assigned to any AA by the IUPAC. However, it is commonly used in the biochemical literature as an abbreviation for the AA, Ornithine [3-5]. The reason that the IUPAC has not assigned O to Ornithine is that this AA is not genetically encoded, and does not occur in proteins that are biosynthesized on ribosomes. Ornithine does occur in proteins that are biosynthesized by other methods (e.g., the bacterial antibiotic, ramoplanin [8]), and it is important to human life. The body makes Ornithine from another AA, Arginine, and then uses the Ornithine to detoxify ammonia that also is made by the body [6]. Ornithine is structurally and chemically almost identical to the AA, Lysine (Figure 2), which is found in natural proteins. The side chain of Ornithine contains one less methylene group ($-CH_2-$) than does the side chain of Lysine, and the side chain amino groups of Ornithine and Lysine have nearly identical pK values ($-NH_3^+ \leftrightarrow -NH_2$; pK = 10.6-10.7) [5]. If the letter, O, is used as an abbreviation for Ornithine, it then becomes possible to use the letter sequence of the name, Princeton, to create a peptide (Figure 3). Based on the IUPAC definitions, and the assignment of O to Ornithine, the single letters of the name, Princeton, would correspond to the amino acid sequence: Proline (P)-Arginine (R)-Isoleucine (I)-Asparagine (N)-Cysteine (C)-Glutamic acid (E)-Threonine (T)- Ornithine (O)-Asparagine (N).

In addition to the presence of Ornithine, the PRINCETON peptide also contains the AA, Cysteine (Figure 3). The sulfhydryl (-SH) group of Cysteine can oxidize to form a disulfide bond (-S-S-). Consequently, two PRINCETON peptides could undergo oxidation of their Cysteine sulfhydryl groups to form an intermolecular disulfide bond and a dimeric molecule, $(PRINCETON)_2$ (Figure 3).

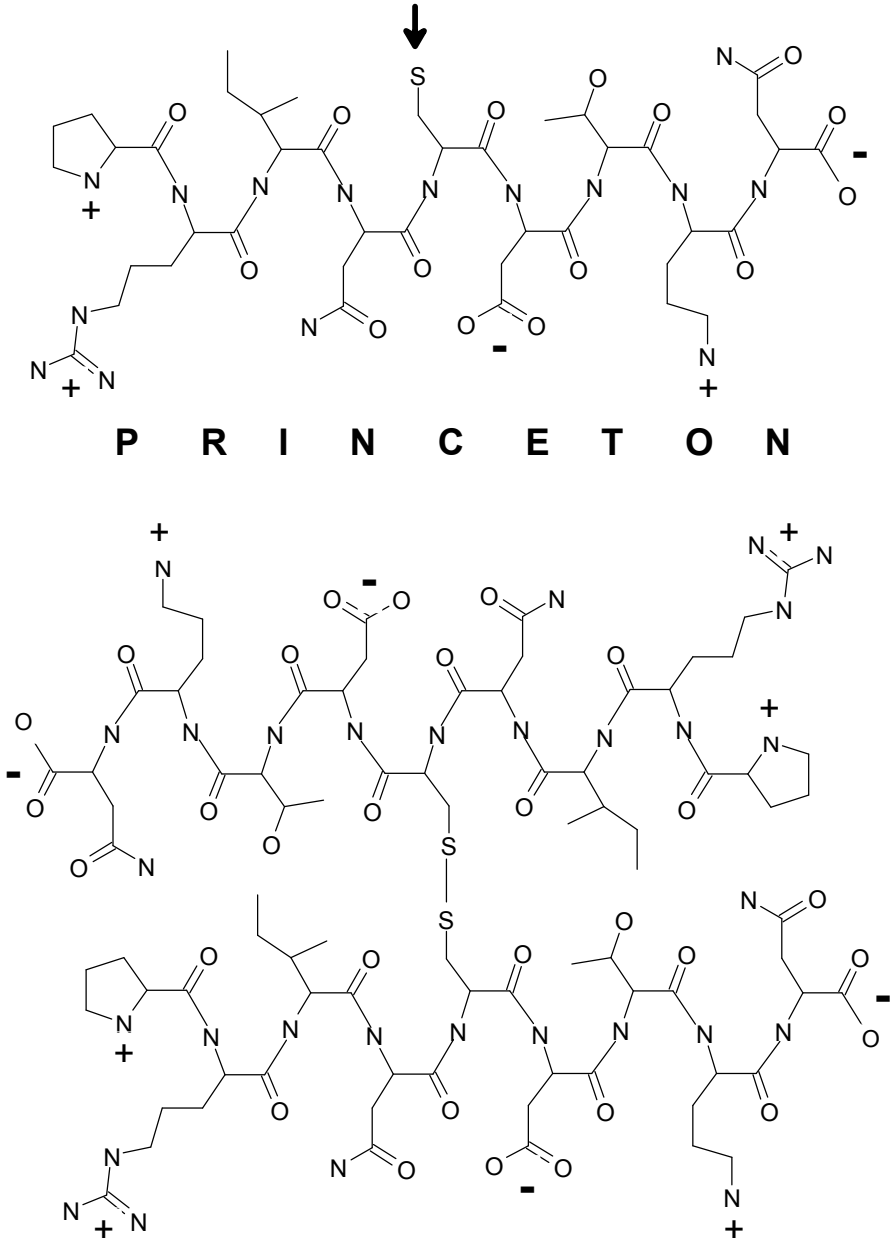
Due to technological advances developed by R.B. Merrifield (1984 Nobel Prize, Chemistry), it is possible to rapidly synthesize almost any peptide. This technology enables the synthesis of naturally occurring peptides, and also the creation of peptides that do not occur in nature [5, 9]. For example, a peptide corresponding to the name, Princeton, could be synthesized in substantial quantities in less than a day.

Methods

Two dimensional (2D) models of peptides were made with the ISIS™/Draw 2.4 program (MDL Information Systems, Inc.). Three dimensional models (3D) of the PRINCETON peptide were made by first modeling the peptide, PRINCETKN, where K, the IUPAC abbreviation for the AA, Lysine, replaces the O of Ornithine. Modeling was done with the Deep View/Swiss-PdbViewer v. 3.7 program (http://www.expasy.org/spdbv/), and the peptide was modeled as a cylindrical α-helix. The 3D coordinate, or Protein Data Bank (PDB), file for PRINCETKN was then modified to convert the atomic coordinates for Lysine into coordinates for Ornithine. This was done by relabeling all LYS atoms as ORN, relabeling the epsilon carbon in Lysine's side

Wade Research Foundation Reports (2007) 4 (4)

Figure 3. 2D representations of the chemical structures of the peptide, PRINCETON (top), with an arrow showing the location of the sulfhydryl group of Cysteine, and the dimeric form of the peptide, $(PRINCETON)_2$ (bottom), containing an intermolecular disulfide bond. The single letter abbreviations for each AA of PRINCETON are shown directly under the alpha carbons of each AA in the peptide. The dimeric peptide is shown with the monomers arranged in an antiparallel orientation. The average molecular mass of the monomeric peptide would be 1060, and that of the dimeric peptide would be 2118. Charges on AA side chains at pH 7 are shown, and the net charges on the peptides would be +1 for the monomer and +2 for the dimer.



Methods (Con't.)

chain as a nitrogen, and finally removing the Lysine side chain nitrogen with its hydrogens. The resulting 3D coordinate file was then used as a PDF file with Deep View/Swiss-PdbViewer to create a 3D model of the PRINCETON peptide, as an α-helix. Some torsional adjustments were made in the default positions of certain AA side chains of the peptide (i.e., Arginine and Glutamic acid), and the peptide was subjected to several steps of energy minimization to produce a more realistic 3D model. The resulting PDF file for PRINCETON was used to create stick figure models of the peptide with the RasWin Molecular Graphics, Windows version 2.6-ucb program (http://mc2.cchem.berkeley.edu/Rasmol/v2.6/) [10], and the Microsoft Paint version 5.1 program (Microsoft Corp.). Electrostatic potential diagrams were made with the Deep View/Swiss-PdbViewer v. 3.7 program, and the Microsoft Paint version 5.1 program.

Average molecular masses for the PRINCETON and $(PRINCETON)_2$ peptides were calculated by use of the ProtParam tool on the ExPASy Proteomics Server website (http://expasy.org/tools/protparam.html).

Protein database searches were done using the Basic Local Alignment Search Tool (BLAST) program for short, nearly exact matches, of the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/BLAST/) [11].

Results and Discussion

Figure 3 shows 2D representations of the PRINCETON peptide, and a dimeric peptide, $(PRINCETON)_2$, that would be formed by oxidation of the sulfhydryl groups in the Cysteine side chains of PRINCETON. The monomeric peptide would have an average molecular mass of 1060 and a net charge of +1 at pH 7, whereas the dimeric peptide would have an average molecular mass of 2118 and a net charge of +2.

Figure 4 shows two, 2D views of 3D stick figure models of the PRINCETON peptide as a cylindrically shaped, α-helix, the type of 3D structure that peptides are known to adopt when in contact with other molecules or structures, such as biological membranes.

Figure 5 shows a 2D view of an electrostatic potential model of the PRINCETON peptide in the same α-helical and longitudinal orientation shown in Figure 4. The information shown in this model can be helpful in predicting how the peptide might interact with other molecules.

When a new AA sequence is obtained, it is often of interest to know if the sequence occurs in natural proteins. Such information may be helpful in determining whether or not the new sequence might exhibit biological activities. A BLAST search of the AA sequence in the NCBI protein databases will provide such information. However, due to the fact that the PRINCETON sequence contain the letter, O, which does not occur among the AA sequences in protein databases, the BLAST search algorithm will interpret this sequence as PRINCET_N. As indicated above, Lysine is structurally and chemically nearly identical to Ornithine, and the one letter abbreviation for Lysine, K, does occur in protein databases. Consequently, replacement of the letter, O, in the search sequence, PRINCETON, with the letter, K, will form the new sequence analog, PRINCETKN, that represents a peptide that is very similar in structure to the peptide of interest, PRINCETON (i.e., only one more $–CH_2-$ group, and only a 1.3% increase in molecular mass). When the PRINCETKN sequence analog was used for a BLAST search of short, nearly exact matches among the nearly 5 million AA sequences of the NCBI protein databases, numerous partial matches were found (Table 1). These matches occurred in all types of living organisms, most had 67-78% sequence identity with the search sequence, PRINCETKN, and they covered all

Wade Research Foundation Reports (2007) 4 (4)

Figure 4. Stick figure models of the PRINCETON peptide as an α-helix, a 3D structure with a cylindrical shape that the peptide might form upon interaction with other biomolecules, such as those of biological membranes. The figures at the top (left and right) are the peptide model, and the figures at the bottom show the orientation of the helix cylinders in the peptide model directly above (bottom left, longitudinal; bottom right, cross sectional). The two figures differ only by a 90° rotation about the vertical axis. In the longitudinal view (top left), the amino (N-) terminal end of the peptide is on the left and the carboxyl (C-) terminal end of the peptide is on the right. In the cross sectional view (top right), the viewer is looking from the N-terminal end of the peptide toward the C-terminal end, along the helix axis. The color scheme is gray for carbon, blue for nitrogen, red for oxygen, yellow for sulfur (also highlighted with arrows), and white for hydrogen. Due to deficiencies in the molecular modeling program, the hydrogens on the side chain amino group ($-NH_3^+$) of Ornithine are missing. The single letter abbreviations for each AA are placed next to the AA side chains.

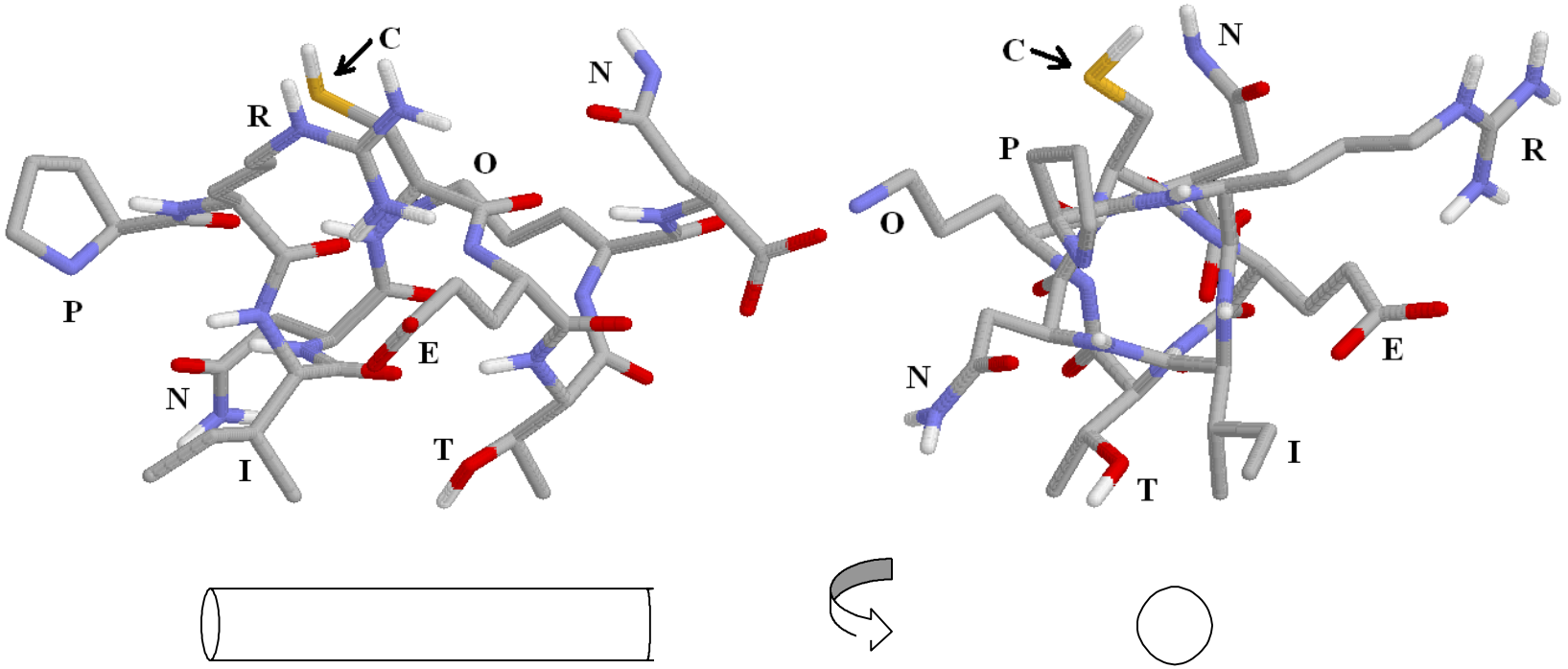
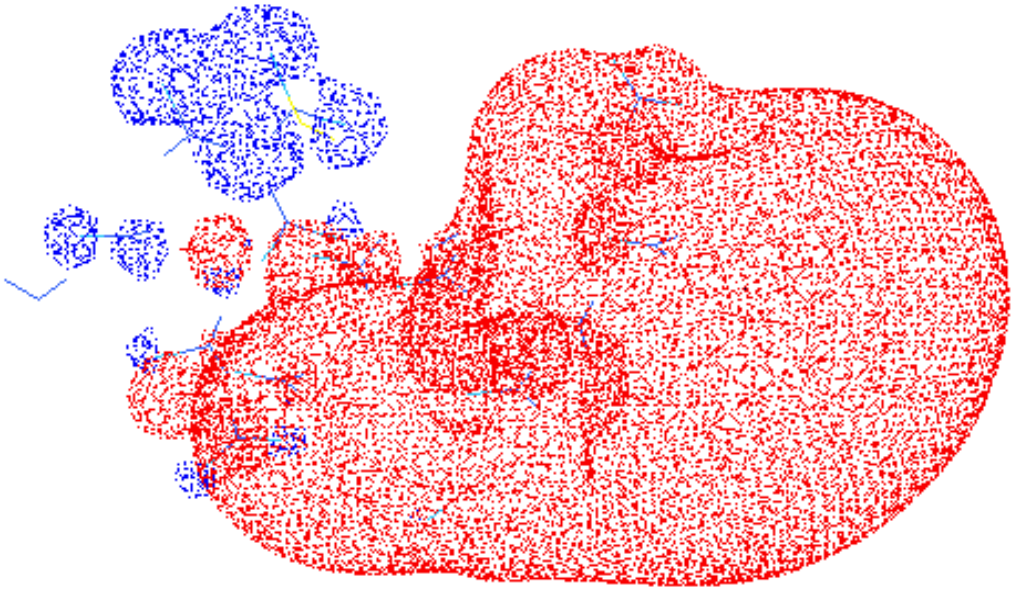


Figure 5. Electrostatic potential model of the PRINCETON peptide. The orientation of the peptide is similar to that of the longitudinal view of Figure 4, with the N-terminal end to the left and the C-terminal end to the right. Much of the carbon skeleton of the peptide is not visible since it is colored white, and, therefore, hidden by the background color. However, some nitrogens within the skeleton are visible as turquoise color, and sulfur is yellow. Blue spheres represent regions of positive electrostatic potential, and red spheres and areas represents regions of negative electrostatic potential. The electrostatic potential of a molecule is relevant to the manner in which it interacts with other molecules.



Results and Discussion (Con't.)

parts of the search sequence. Although the complete search sequence was not found in any protein of the database, all portions of the search sequence apparently do commonly occur in proteins of known biological functions. Consequently, a synthetic peptide with the analogous AA sequence, PRINCETON, would probably also exhibit biological activities.

There are no synthetic barriers to the creation of the PRINCETON peptide, and nearly identical AA sequences occur within many proteins. A small, Ornithine containing peptide of similar size, COLINPOWELL, was synthesized and found to exhibit biological activity in 50% of the tests to which it was subjected [5]. Therefore, creation of the PRINCETON peptide is feasible, and it would have a high probability of exhibiting biological activities.

References

1. A brief history of Princeton. Website of Princeton Township, Mercer County, New Jersey. (http://www.princetontwp.org/histofpt.html)
2. Princeton's history. Parents Handbook, 2005-06, Princeton University, Princeton, N.J. (http://www.princeton.edu/pr/pub/ph/05/03.htm)
3. Wade D. The name game: use of words composed of letters of the English alphabet as a source of novel bioactive peptides. Chemistry Preprint Archive (2003) 1: 159-170. (http://www.sciencedirect.com/preprintarchive)

Wade Research Foundation Reports (2007) 4 (4)

Table 1. Examples of the occurrence of **PRINCETKN,** a structural analog of **PRINCETON**, within proteins of the NCBI protein database.

Sequence found:	Protein, Source, and Database Accession Number:	% Identity:
P _ INCET _ N	Non-ribosomal peptide synthase [*Burkholderia pseudomallei* 305]; AAs 2599-2607 of 4248 AAs; ZP_01770412.1	78%
PRI _ _ ETKN	Glutathione S-transferase [*Parvibaculum lavamentivorans* DS-1]; AAs 127-135 of 208 AAs; ZP_01657967.1	78%
RIN _ ETKN	Arginyl-tRNA synthetase [*Streptococcus mutans* UA159]; AAs 195-202 of 563 AAs; NP_722386.1	78%
PRINCE	Casein kinase II regulatory subunit family protein [*Trichomonas vaginalis* G3]; AAs 127-132 of 218 AAs; EAY20064.1	67%
PRI _ CE _ K	Outer arm dynein beta heavy chain [*Paramecium tetraurelia*]; AAs 2120-2127 of 4588 AAs; AAA61680.1	67%
PRI _ C _ KN	ATPase, Na+/K+ transporting, beta 3 polypeptide, isoform CRA_e [*Homo sapiens*]; AAs 173-181 of 231 AAs; EAW78988.1	67%
P _ INCET	CD8 alpha2-2 [*Carassius auratus langsdorfii*]; AAs 155-161 of 214 AAs; BAD89374.1	67%
RINCET	Deoxyhypusine synthase [*Encephalitozoon cuniculi* GB-M1]; AAs 261-266 of 362 AAs ; XP_955645.1	67%
RIN _ E _ KN	E1-E2 ATPase family protein [*Tetrahymena thermophila* SB210]; AAs 113-120 of 1112 AAs; XP_001009340.1	67%
RI _ CETK	Rps12 [*Ostreococcus tauri*]; AAs 12-18 of 124 AAs; YP_717226.1	67%
INCETK	Heptosyltransferase [*Flavobacteria bacterium* BAL38]; AAs 123-128 of 344 AAs; ZP_01733097.1	67%
INCE _ KN	Cysteine desulfurase [*Theileria annulata* strain Ankara]; AAs 98-104 of 795 AAs; XP_954680.1	67%
I _ CETKN	Peptidyl-prolyl cis-trans isomerase, cyclophilin-binding [*Cytophaga hutchinsonii* ATCC 33406]; AAs 69-75 of 143 AAs; YP_678223.1	67%
INC _ TKN	Os03g0341200 [*Oryza sativa* (japonica cultivar-group)]; AAs 1279-1285 of 1571 AAs; NP_001050070.1	67%

References (Con't.)

4. Wade, D. The name game: use of words composed of letters of the English alphabet as a source of novel bioactive peptides. In Peptide Revolution: Genomics, Proteomics & Therapeutics, M. Chorev and T. K. Sawyer, eds., American Peptide Society, San Diego, CA, USA, 2004, pp. 580-581. (Copy available upon request to author.)
5. Wade, D., Yang, D., and Lea, M.A. Biological and structural properties of COLINPOWELL, a synthetic peptide amide. Wade Res. Found. Rep. (2004) 1: 2-35. (http://www.wade-research.com/images/COLINPOWELL_10-25-04_.pdf)
6. Rodwell, V.W., and Kennelly, P.J., Structures & Functions of Proteins & Enzymes: Amino Acids & Peptides, and Rodwell, V.W., Catabolism of Proteins & of Amino Acid Nitrogen, and Granner, D.K., The Diversity of the Endocrine System, in Harper's Illustrated Biochemistry, 26th Ed., R.K. Murray, D.K. Granner, P.A. Mayes, V.W. Rodwell, eds., McGraw-Hill, 2003, pp. 14-20, 245-247 449-450
7. Nomenclature and symbolism for amino acids and peptides. In Biochemical Nomenclature and Related Documents, 2nd ed., C. Liébecq, Ed., Portland Press, London, UK, 1992, pp. 39-69. (http://www.chem.qmul.ac.uk/iupac/AminoAcid/)
8. Kurz, M., and Guba, W. 3D structure of ramoplanin: a potent inhibitor of bacterial cell wall synthesis. Biochemistry (1996) 35: 12570-12575. (PDB i.d. code, 1DSR.)
9. Wade, D., Boman, A., Wåhlin, B, Drain, C.M., Andreu, D., Boman, H.G., and Merrifield, R.B. All D-amino acid containing channel-forming antibiotic peptides. Proc. Natl. Acad. Sci. USA (1990) 87: 4761-4765. (http://www.pnas.org/cgi/reprint/87/12/4761)
10. Sayle, R.A. and Milner, E.J., RasMol: Biomolecular graphics for all. Trends Biochem. Sci. (1995) 20: 374-376.
11. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. (1997) 25: 3389-3402.