Wade Research Foundation Reports (2007) 4 (5)

The KAROLINSKA 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, KAROLINSKA: Lysine (K)-Alanine (A)-Arginine (R)-Ornithine (O)-Leucine (L)-Isoleucine (I)-Asparagine (N)-Serine (S)-Lysine (K)-Alanine (A). In this sequence, nine of the ten 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

The Karolinska Institute is Sweden's largest institution for medical education and research, and one of Europe's largest medical universities [1]. Karolinska is also internationally famous for its connection to the Nobel Prizes. Its Nobel Assembly, composed of 50 senior faculty from Karolinska, have been selecting the yearly winners of the Nobel Prize in Physiology or Medicine since 1901 [2]. Consequently, the name, Karolinska, 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].

The letter, O, has not been officially assigned to any AA by the IUPAC. However, it is

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

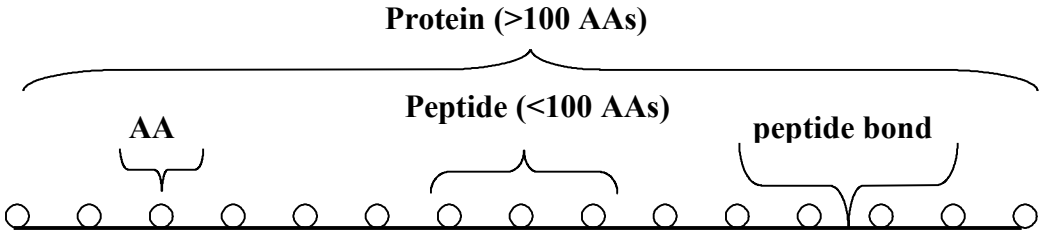
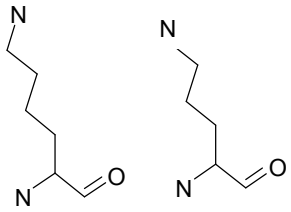


Figure 2. The chemical structures of Lysine (left) and Ornithine (right). Carbon and hydrogen atoms are not shown.



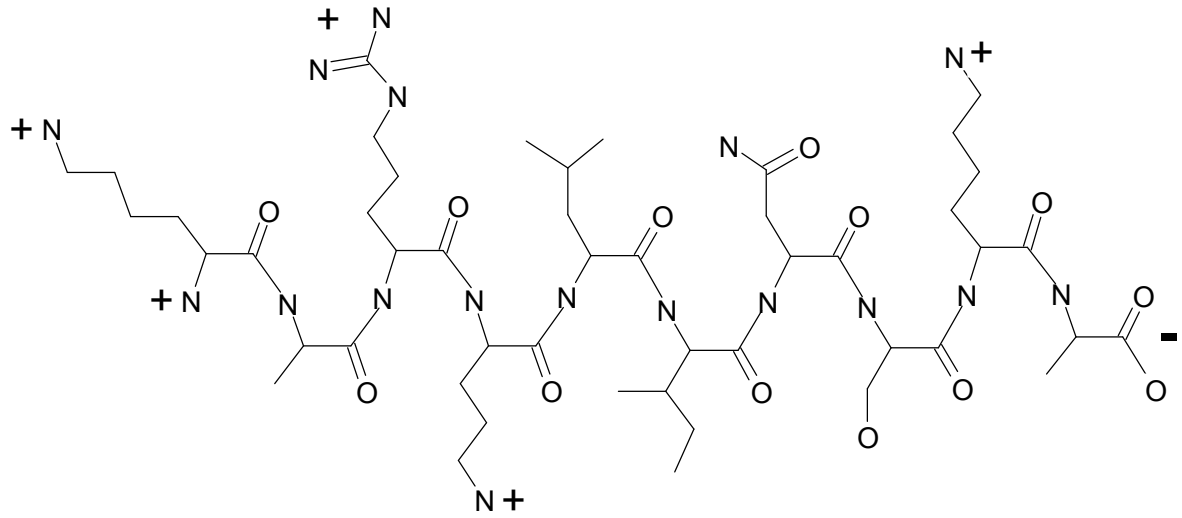
commonly used in the biochemical literature as an abbreviation for the AA, Ornithine [3-5]. The reason that the IUPAC has not assigned O to Ornithine is that this AA is not genetically encoded, and does not occur in proteins that are biosynthesized on ribosomes. Ornithine does occur in proteins that 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, Karolinska, to create a peptide (Figure 3). Based on the IUPAC definitions, and the assignment of O to Ornithine, the single letters of the name, Karolinska, would correspond to the amino acid sequence: Lysine (K)-Alanine (A)-Arginine (R)-Ornithine (O)-Leucine (L)-Isoleucine (I)-Asparagine (N)-Serine (S)-Lysine (K)-Alanine (A).

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, Karolinska, 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 (3D) models of the PRINCETON peptide were made by first modeling the peptide, KARKLINSKA, 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 type of 3D shape that peptides frequently adopt when in contact with other molecules, such as those of biological membranes. The 3D coordinate, or Protein Data Bank (PDB), file for KARKLINSKA was then modified to convert the atomic coordinates for Lysine, at position 4 in the peptide, into coordinates for Ornithine. This was done by relabeling all LYS-4 atoms as ORN, relabeling the epsilon carbon in Lysine-4's side chain as a nitrogen, and finally removing the Lysine-4 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 KAROLINSKA peptide, as an α-helix. The peptide was then subjected to several steps of

Figure 3. 2D representation of the chemical structure of the peptide, KAROLINSKA. The single letter abbreviations for each AA of KAROLINSKA are shown directly under the alpha carbons of each AA in the peptide. The average molecular mass of the peptide is 1114. Charges on AA side chains at pH 7 are shown, and the net charge on the peptide is +4.



K A R O L I N S K A

Methods (Con't.)

energy minimization to produce a more realistic 3D model. The resulting PDF file for KAROLINSKA 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.

The average molecular mass for KAROLINSKA was 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 KAROLINSKA peptide. The peptide would has an average molecular mass of 1114, and a net charge of +4 at pH 7.

Figure 4 shows two, 2D views of 3D stick figure models of the KAROLINSKA 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 KAROLINSKA 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

Wade Research Foundation Reports (2007) 4 (5)

Figure 4. Stick figure models of the KAROLINSKA 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, and red for oxygen. Hydrogens are not shown. The single letter abbreviations for each AA are placed next to the AA side chains.

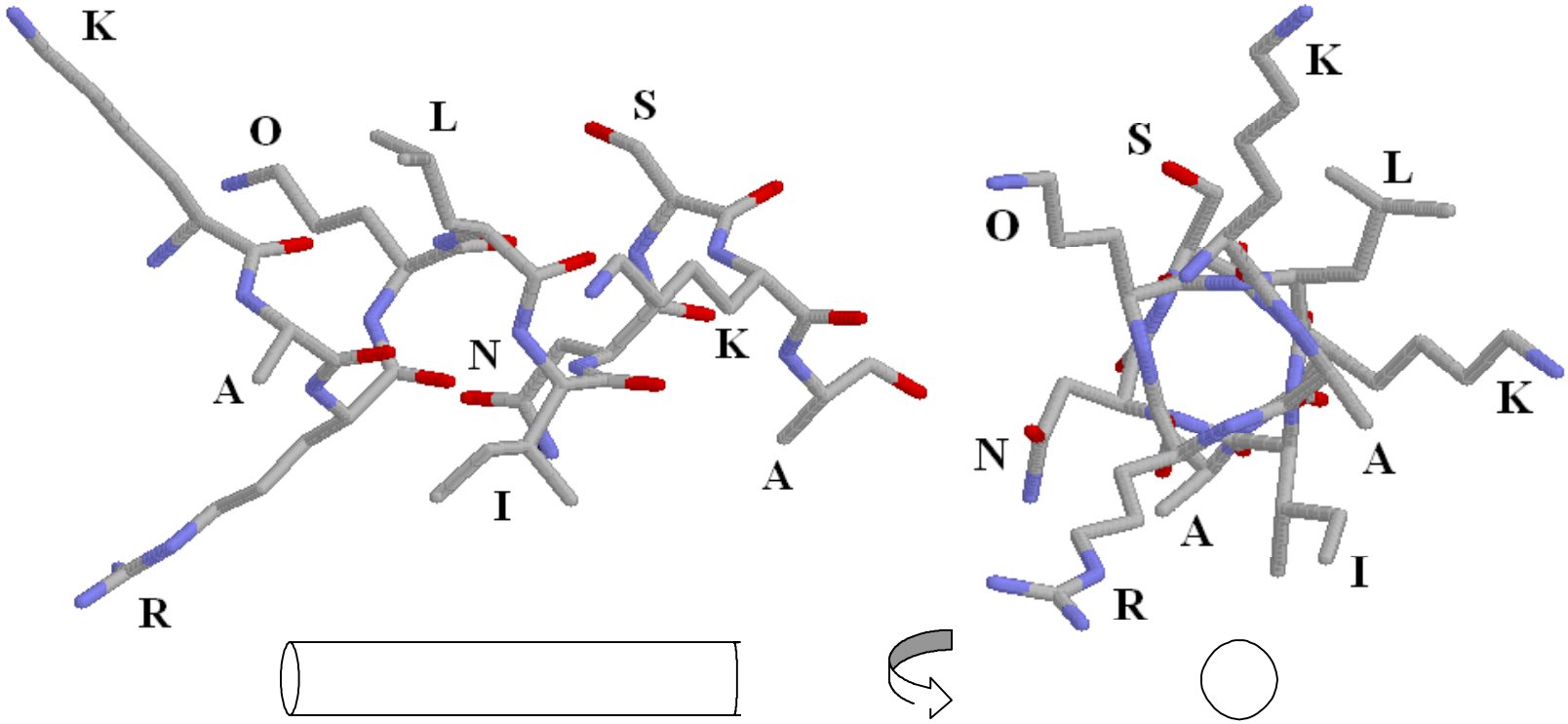
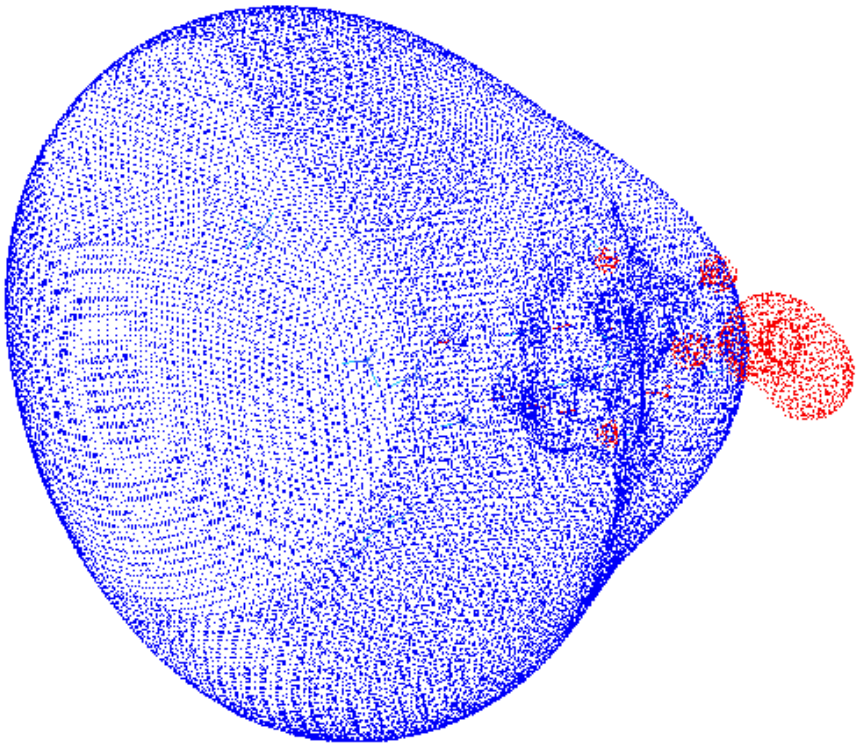


Figure 5. Electrostatic potential model of the KAROLINSKA peptide. The orientation of the peptide is similar to that of the longitudinal view of Figure 4, with the N-terminal end to the left and the C-terminal end to the right. Much of the carbon skeleton of the peptide is not visible since it is colored white, and, therefore, hidden by the background color. However, some nitrogens within the skeleton are visible as turquoise color. Blue areas represent regions of positive electrostatic potential, and red areas represent 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.)

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 KAROLINSKA sequence contains the letter, O, which does not occur among the AA sequences in protein databases, the BLAST search algorithm will interpret this sequence as KAR_LINSKA. 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, KAROLINSKA, with the letter, K, will form the new sequence analog, KARKLINSKA, that represents a peptide that is very similar in structure to the peptide of interest, KAROLINSKA (i.e., only one more $-CH_2-$ group, and only a 1.3% increase in molecular mass). When the KARKLINSKA 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 a variety of types of living organisms, most had 70% sequence identity with the search sequence,

Wade Research Foundation Reports (2007) 4 (5)

Table 1. Examples of the results of a BLAST search of the NCBI protein databases for KARKLINSKA, a close structural analog of KAROLINSKA.

Sequence Found:	Protein:	Source:	Accession Number:	% Identity:
KARKLIN	Transaldolase; AAs 114-120 of 333 AAs	*Prochlorococcus marinus* str. MIT 9301	YP_001090769.1	70%
ARKL _ NSK	Ribulose 1,5-bisphosphate carboxylase/oxygenase activase precursor; AAs 241-248 of 308 AAs	*Zea mays*	AAG22094.2	70%
ARKLINS	(Acyl-carrier-protein) phosphodiesterase; AAs 23-29 of 213 AAs	*Clostridium cellulolyticum* H10	ZP_01576526.1	70%
KAR _ LI _ SK	Myosin class II heavy chain (ISS); AAs 493-501 of 621 AAs	*Ostreococcus tauri*	CAL57983.1	70%
K _ _ K _ INSKA	Threonine dehydratase; AAs 214-223 of 403 AAs	*Campylobacter lari* RM2100	ZP_00368325.1	70%
RKLIN _ KA	Oxidoreductase-like; AAs 186-193 of 347 AAs	*Caldivirga maquilingensis* IC-167	ZP_01711608.1	70%
KARK _ INS	Type IV pilus assembly protein PilP; AAs 74-81 of 176 AAs	*Methylobacillus flagellatus* KT	YP_546564.1	70%
KLINSKA	RNA helicase; AAs 1230-1236 of 1381 AAs	*Plasmodium yoelii yoelii* str. 17XNL	XP_725673.1	70%
RKLI _ SKA	Sodium:alanine symporter family protein; AAs 504-511 of 511 AAs	*Gramella forsetii* KT0803	YP_862684.1	70%
AR _ LINS _ A	Tyrosyl-tRNA synthetase; AAs 357-365 of 400 AAs	*Campylobacter lari* RM2100	ZP_00369035.1	70%
KA _ KLI _ SK	ABC transporter, permease; AAs 271-279 of 485 AAs	*Oenococcus oeni* ATCC BAA-1163	ZP_01544940.1	70%
ARKL _ N _ KA	Mitochondrial glycosylase/lyase; AAs 197-205 of 376 AAs	*Saccharomyces cerevisiae*	NP_013651.1	70%
R _ LINSKA	Exodeoxyribonuclease V, beta subunit; AAs 796-803 of 1222 AAs	*alpha proteobacterium* HTCC2255	ZP_01450099.1	70%

Results and Discussion (Con't.)

KARKLINSKA, and they covered all 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, KAROLINSKA, would probably also exhibit biological activities.

There are no synthetic barriers to the creation of the KAROLINSKA 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 KAROLINSKA peptide is feasible, and it would have a high probability of exhibiting biological activities.

References

1. Ett medicinskt universitet. Website of the Karolinska Institute, Stockholm, Sweden, June 27, 2006. (http://ki.se/ki/jsp/polopoly.jsp?d=138&a=5151&l=sv)
2. Nobel Prize. Website of the Karolinska Institute, Stockholm, Sweden, April 10, 2007. (http://ki.se/ki/jsp/polopoly.jsp?d=1269&a=2197&l=en)
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)
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.