Wade Research Foundation Reports (2007) 4 (1)

Peptides MICROSOFT, GOOGLE, YAHOO, COCACOLA and PEPSICOLA

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

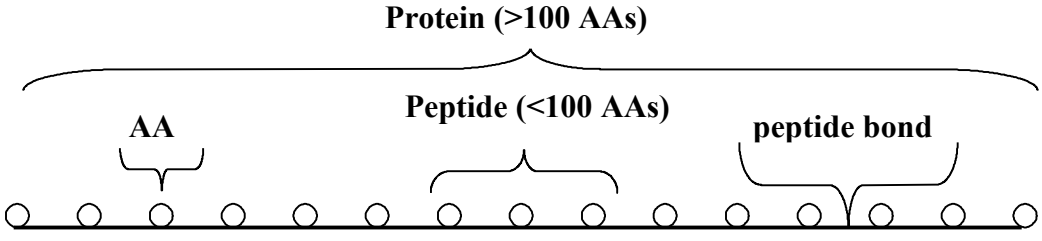
Abstract

Three of the most recognizable corporate names are Microsoft, a manufacturer of computer software, and Google and Yahoo, two internet search providers. Two of the most popular soft drink brand names are Coca-Cola and Pepsi-Cola. This article proposes to use the sequences of English alphabet letters in the names of these companies and products as the basis for creating five new peptides, MICROSOFT, GOOGLE, YAHOO, COCACOLA, and PEPSICOLA. This can be done by using a modified version of the International Union of Pure and Applied Chemistry's one letter abbreviations for the names of amino acids. In this modification, the unassigned letter, O, is used to represent the amino acid, Ornithine. Precedent for this assignment already exists in the chemical literature. Ornithine does not occur in natural proteins, and, therefore, the letter, O, does not occur among the amino acid sequences of any protein database. However, Ornithine is structurally and chemically very similar to the amino acid, Lysine, which does occur in natural proteins and which is represented by the one letter abbreviation, K, in the IUPAC system. A search of the National Center for Biotechnology Information protein databases for the letter sequences, MICRKSKFT, GKKGLE, YAHKK, CKCACKLA and PEPSICKLA, where the one letter abbreviation, K, for Lysine, replaced the letter, O, returned numerous exact (GKKGLE and YAHKK) and partial (MICRKSKFT, CKCACKLA and PEPSICKLA) sequence matches. This result indicates that all of the search sequences are common in proteins. There are no synthetic barriers to the creation of these Ornithine containing peptides. Consequently, it is predicted that the synthetic peptides, MICROSOFT, GOOGLE, YAHOO, COCACOLA, and PEPSICOLA, will all exhibit biological activities.

Introduction

Among the most important biomolecules of life are proteins, polymers of amino acids (AAs) that are held together by chemical bonds, called peptide bonds [1]. 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 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 [1].

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



Wade Research Foundation Reports (2007) 4 (1)

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 International Union of Pure and Applied Chemistry (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 [2].

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 does not occur in natural proteins. However, Ornithine, is important to human life. The body makes it from another AA, Arginine (R), and then uses the Ornithine to detoxify ammonia that also is made by the body [1]. Ornithine is structurally and chemically almost identical to the AA, Lysine (K) (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 sequences in the company names, Microsoft, Google and Yahoo, and product names, Coca-Cola and Pepsi-Cola, to create peptides (Figures 3 and 4).

Due to technological advances developed by R.B. Merrifield (1984 Nobel Prize in 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, 6]. For example, peptides corresponding to the names, MICROSOFT, GOOGLE, YAHOO, COCACOLA and PEPSICOLA could be synthesized in less than a day.

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

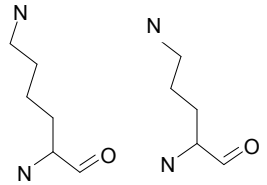


Figure 3. AA sequences of the peptides, MICROSOFT, GOOGLE, YAHOO, COCACOLA and PEPSICOLA. Hyphens represent peptide bonds.

Methionine (**M**)-Isoleucine (**I**)-Cysteine (**C**)-Arginine (**R**)-Ornithine (**O**)-
Serine (**S**)-Ornithine (**O**)-Phenylalanine (**F**)-Threonine (**T**)

Glycine (**G**)-Ornithine (**O**)-Ornithine (**O**)-Glycine (**G**)-Leucine (**L**)-Glutamic Acid (**E**)

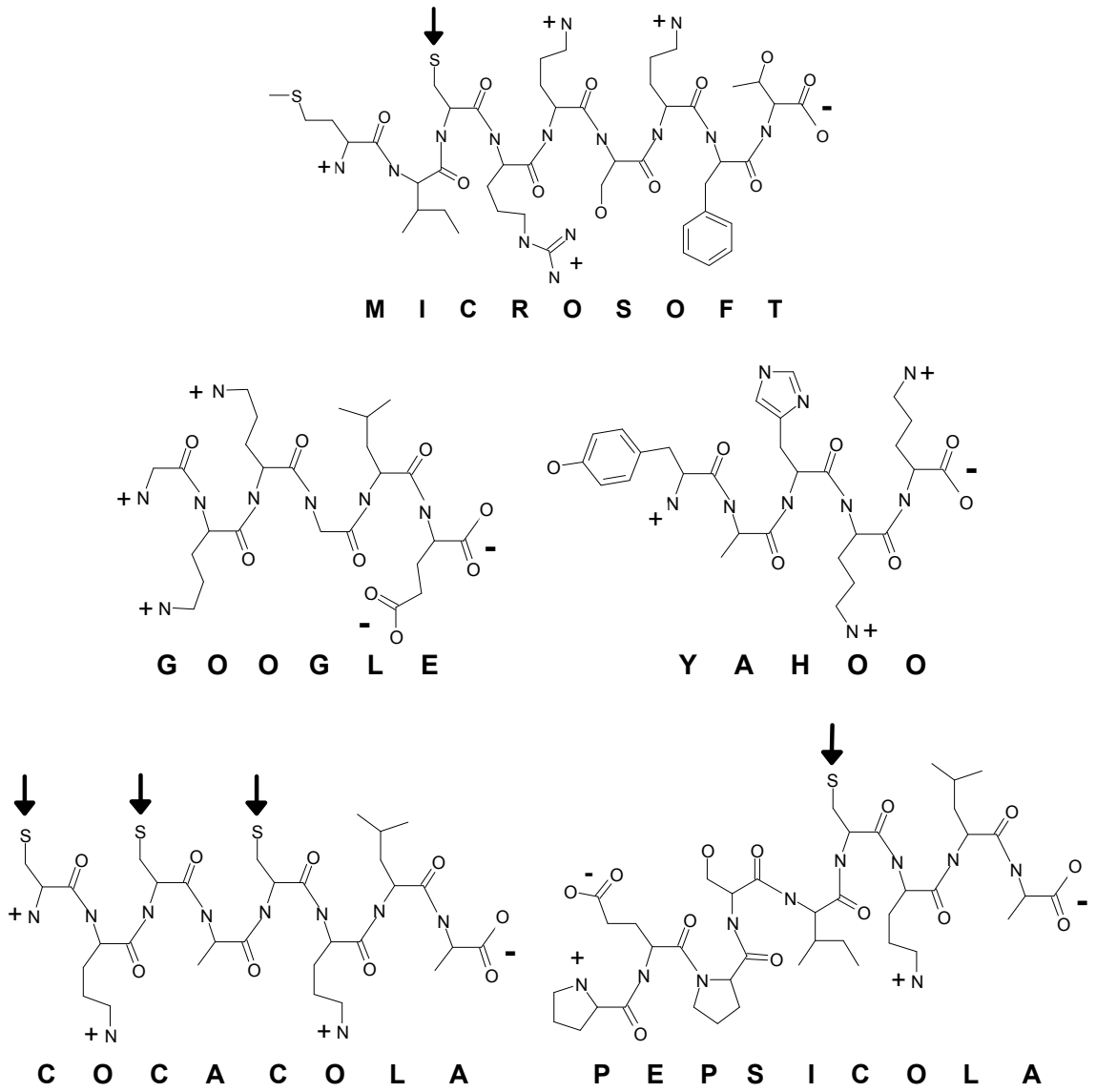
Tyrosine (**Y**)-Alanine (**A**)-Histidine (**H**)-Ornithine (**O**)-Ornithine (**O**)

Cysteine (**C**)-Ornithine (**O**)-Cysteine (**C**)-Alanine (**A**)-
Cysteine (**C**)-Ornithine (**O**)-Leucine (**L**)-Alanine (**A**)

Proline (**P**)-Glutamic Acid (**E**)-Proline (**P**)-Serine (**S**)-Isoleucine (**I**)-
Cysteine (**C**)-Ornithine (**O**)-Leucine (**L**)-Alanine (**A**)

Wade Research Foundation Reports (2007) 4 (1)

Figure 4. The chemical structures of peptides, MICROSOFT, GOOGLE, YAHOO, COCACOLA and PEPSICOLA. Single letter abbreviations for AAs are shown directly beneath the position of each AA in each peptide. The arrows indicate the location of sulfhydryl (-SH) groups that can participate in the formation of disulfide bonds. Charges (+/-) on AAs at pH 7 are shown, and net charges on the peptides at pH 7 are +3 for MICROSOFT, +1 for GOOGLE, +2 for YAHOO, +2 for COCACOLA, and 0 for PEPSICOLA.



In addition to the presence of Ornithine in all peptides, three also contain the AA, Cysteine (Figures 3 and 4). The side chain sulfhydryl (-SH) group of Cysteine can oxidize to form a disulfide bond (-S-S-). The sulfhydryl groups of MICROSOFT and PEPSICOLA could each form one intermolecular disulfide bond, whereas the sulfhydryl groups of COCACOLA could form both intra- and intermolecular disulfide bonds. For MICROSOFT and PEPSICOLA, the largest possible polymers would be dimers containing two peptide molecules (monomers) each. In contrast, the three sulfhydryl groups of COCACOLA would permit the formation of a great variety of disulfide bonding arrangements and a vast range of peptide polymers, from dimers to, in theory, polymers containing an infinite number of peptide monomers (Table 1). In reality, the numbers and sizes of polymers formed from COCACOLA would be determined by the probabilities of interactions of sulfhydryl groups within and between peptides, and would most likely produce dimers in which each monomer contained one intra- and one intermolecular disulfide bond.

Methods

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/) [7].

Two dimensional models of peptides were made with the ISIS™/Draw 2.4 program (MDL Information Systems, Inc.). Three dimensional (3D) models of proteins were obtained from the RSCB Protein Data Bank (PDB; http://www.pdb.org/pdb/home/home.do). Figures of proteins and peptides were made using the RasWin Molecular Graphics, Windows version 2.6-ucb program (http://mc2.cchem.berkeley.edu/Rasmol/v2.6/) [8], 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 (http://www.expasy.org/spdbv/), and the Microsoft Paint version 5.1 program.

Results and Discussion

The occurrence of MICROSOFT, GOOGLE, YAHOO, COCACOLA and PEPSICOLA sequence analogs in proteins

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 have biological activities. A BLAST search of the AA sequences in the NCBI protein databases will provide such information. However, due to the fact that the MICROSOFT, GOOGLE, YAHOO, COCACOLA, and PEPSICOLA sequences contain the letter, O, which does not occur among the AA sequences in protein databases, the BLAST search algorithm will interpret these sequences as MICR _ S _ FT, G _ _ GLE, YAH _ _, C _ CAC _ LA, and PEPSIC _ LA. 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 MICROSOFT, GOOGLE, YAHOO, COCACOLA, and PEPSICOLA with the letter, K, will form the new sequence analogs, MICRKSKFT, GKKGLE, YAHKK, CKCACKLA, and PEPSICKLA. When the sequence analogs were used for BLAST searches of short, nearly exact matches among the 4.5 million AA sequences of the NCBI protein databases, numerous exact (GKKGLE and YAHKK) and partial (MICRKSKFT, CKCACKLA, and PEPSICKLA) matches were found for each sequence (Tables 2-6), including matches within proteins of known 3D structure (Figures 5-14). This result indicated that all of the search

sequences occur commonly among proteins within the databases. Many of the proteins containing matching sequences have known biological functions. The BLAST search results could be interpreted as indicating that the MICRKSKFT, GKKGLE, YAHKK, CKCACKLA, and PEPSICKLA sequences, and by implication the MICROSOFT, GOOGLE, YAHOO, COCACOLA, and PEPSICOLA sequences, might also exhibit biological activities.

Creating MICROSOFT, GOOGLE, YAHOO, COCACOLA and PEPSICOLA peptides

There are no synthetic barriers to the creation of peptides, MICROSOFT, GOOGLE, YAHOO, COCACOLA and PEPSICOLA, 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, the creation of peptides, MICROSOFT, GOOGLE, YAHOO, COCACOLA and PEPSICOLA, is feasible, and they would have a high probability of exhibiting biological activities.

References

1. Rodwell, V.W., and Kennelly, P.J., Structures & Functions of Proteins & Enzymes: Amino Acids & Peptides, 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, 449-450.
2. 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/)
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. 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)
7. 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.
8. Sayle, R.A. and Milner, E.J., RasMol: Biomolecular graphics for all. Trends Biochem. Sci. (1995) 20: 374-376.

Wade Research Foundation Reports (2007) 4 (1)

Table 1. Polymers that can result from disulfide bond formation within and between peptide molecules of MICROSOFT, COCACOLA and PEPSICOLA. MICROSOFT and PEPSICOLA form only dimers, whereas COCACOLA forms many different polymers (only 3 shown). The net charge on COCACOLA polymers increases by +2 with the addition of each peptide molecule.

Peptide:	Polymer Structure:[a]	Intra-molecular disulfide:	Inter-molecular disulfide:	Number of peptides in polymer:	Net charge on polymer at pH 7:
MICROSOFT	M-I-C_3-R-O-S-O-F-T | S | S | M-I-C_3-R-O-S-O-F-T	0	1	2	+6
COCACOLA	C_1-O-C_3-A-C_5-O-L-A | | | S——S S | S——S S | | | C_1-O-C_3-A-C_5-O-L-A	1	1	2	+4
"	C_1-O-C_3-A--C_5-O-L-A | | | S S——S | S S——S | | | C_1-O-C_3-A--C_5-O-L-A	1	1	2	+4
"	S————S | | C_1-O-C_3-A-C_5-O-L-A | S | S | C_1-O-C_3-A-C_5-O-L-A | | S————S	1	1	2	+4
PEPSICOLA	P-E-P-S-I-C_6-O-L-A | S | S | P-E-P-S-I-C_6-O-L-A	0	1	2	0

Note: [a]Hyphens indicate peptide bonds. Disulfide bond symbolism: -C_n-S-S-C_m-, where C is Cysteine, n and m are the positions of Cysteine reidues, and S is sulfur.

Wade Research Foundation Reports (2007) 4 (1)

Table 2. Examples of the occurrence of **MICRKSKFT,** a structural analog of **MICROSOFT**, within proteins of the NCBI protein database.

Search Sequence:	Sequence Found:	Database Accession #:	Protein and Source:	Location in protein & total AAs in protein:
MICRKSKFT	**MICRK_ _FT**	XP_821891.1	Vacuolar ATP synthase subunit c (*Trypanosoma cruzi* strain CL Brener)	AAs 249-257; 381 AAs
"	**MICRK _ _ F**	ZP_00604122.1	Ribosomal-protein-alanine acetyltransferase (*Enterococcus faecium* DO)	AAs 1-8; 154 AAs
"	**M _ CR _ S _ F**	XP_814380.1	CLC-type chloride channel (*Trypanosoma cruzi* strain CL Brener)	AAs 695-702; 727 AAs
"	**M_CRKSK**	NP_507223.1	Serpentine receptor, class W family member (srw-44) (*Caenorhabditis elegans*)	AAs 357-363; 385 AAs
"	**MICRKS**	AAH67864.2	TMDCII protein (*Homo sapiens*)	AAs 168-173; 412 AAs
"	**MICRK**	BAF45923.1	Ubiquitin (*Solea senegalensis*)	AAs 94-98; 128 AAs
"	**ICRKSKF**	NP_219783.1	Na(+)-translocating NADH-quinone reductase subunit B (*Chlamydia trachomatis* D/UW-3/CX)	AAs 12-18; 503 AAs
"	**ICRK _ KF**	XP_828684.1	TatD related deoxyribonuclease (*Trypanosoma brucei* TREU927)	AAs 271-277; 334 AAs
"	**IC _ K _ KF**	PDB i.d. 2A20	Rim2 Zinc Finger Domain	AAs 14-20; 62 AAs
	ICRKSK	NP_001038675.1	Xeroderma pigmentosum, complementation group C (*Danio rerio*)	AAs 416-421; 879 AAs
"	**CRKSKF**	NP_296927.1	Na(+)-translocating NADH-quinone reductase subunit B (*Chlamydia muridarum* Nigg)	AAs 13-18; 503 AAs
"	**RKSKFT**	BAF45924.1	Fusicoccadiene synthase (*Phomopsis amygdali*)	AAs 26-31; 719 AAs
"	**RKSKFT**	CAM16480.1	LDL receptor-related protein 4 (*Mus musculus*)	AAs 1750-1755; 1905 AAs
"	**I _ RKSKFT**	ZP_01061430.1	P-II family protein (*Flavobacterium* sp. MED217)	AAs 7-14; 112 AAs

Figure 5. The location and 3D structure of peptide, **IC _ K _ KF**, from the AA sequence, **MICRKSKFT**, an analog of **MICROSOFT**, in a protein of known 3D structure, Rim2 zinc finger domain (PDB i.d. code 2A2O). The peptide backbone is shown as a ribbon and colored gray, and the **IC _ K _ KF** peptide is shown in space filling format and colored blue. The figures differ only by a 90º rotation about the vertical axis. The arrows point to the location of the Cysteine residue in this sequence.

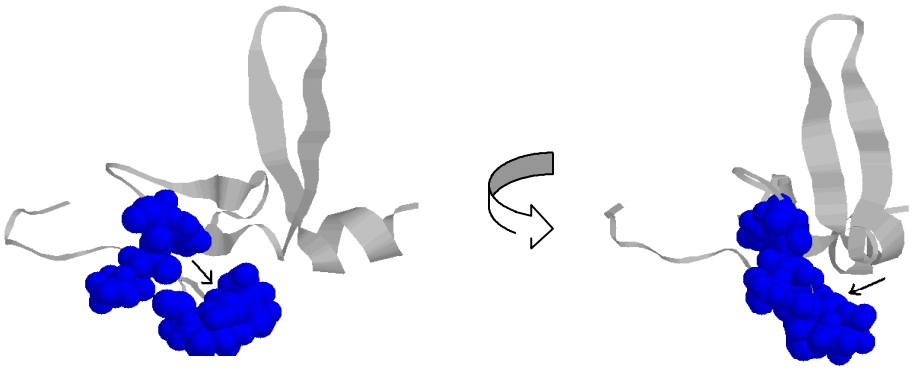
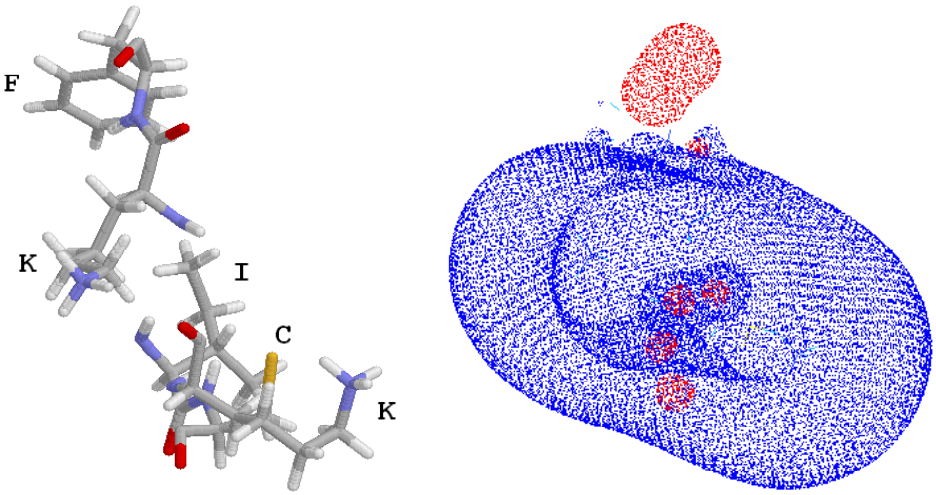


Figure 6. (Below left) A stick figure model of the isolated **IC _ K _ KF** peptide as it occurs in the 3D structure of the protein of Figure 5 (above, left side). The position of each AA in the peptide is indicated by the adjacent single letter abbreviation for the AA. The color scheme is gray for carbon, blue for nitrogen, red for oxygen, and white for hydrogen (some hydrogens are not shown). (Below right) An electrostatic potential model of the **IC _ K _ KF** peptide in the same structural orientation shown in the stick figure to the left. Blue areas are regions of positive electrostatic potential and red areas are regions of negative electrostatic potential. The electrostatic potential would have a direct effect on the ability of the peptide to interact with other molecules.



Wade Research Foundation Reports (2007) 4 (1)

Table 3. Examples of the occurrence of **GKKGLE,** a structural analog of **GOOGLE**, within proteins of the NCBI protein database.

Search Sequence:	Sequence Found:	Database Accession #:	Protein and Source:	Location in protein & total AAs in protein:
GKKGLE	**GKKGLE**	EAX33049.1	Pyridoxal-dependent decarboxylase domain protein, (*Coxiella burnetii* 'MSU Goat Q177)	AAs 215-220; 324 AAs
"	"	EAW81552.1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 24, isoform CRA_c (Homo sapiens)	AAs 159-164; 840 AAs
"	"	YP_930201.1	Protein-L-isoaspartate O-methyltransferase (*Pyrobaculum islandicum* DSM 4184)	AAs 133-138; 207 AAs
"	"	XP_001264829.1	Transcription initiation factor TFIID subunit 13, putative (*Neosartorya fischeri* NRRL 181)	AAs 254-259; 298 AAs
"	"	ZP_01513102.1	Zinc metalloproteinase Mpr protein (*Burkholderia phytofirmans* PsJN)	AAs 223-228; 251 AAs
"	"	XP_317894.3	ENSANGP00000004936 (*Anopheles gambiae* str. PEST)	AAs 189-194; 605 AAs
"	"	CAJ69966.1	Tagatose-1,6-bisphosphate aldolase (*Clostridium difficile* 630)	AAs 259-264; 289 AAs
"	"	YP_692328.1	Glutamate N-acetyltransferase/amino-acidacetyltransferase (*Alcanivorax borkumensis* SK2)	AAs 88-93; 407 AAs
"	"	YP_707512.1	Short chain dehydrogenase (*Rhodococcus* sp. RHA1)	AAs 40-45; 340 AAs
"	"	ABD77287.1	Mitochondrial malate dehydrogenase 2, NAD (*Oryctolagus cuniculus*)	AAs 272-277; 297 AAs
"	"	ZP_01352971.1	Response regulator receiver:Transcriptional regulatory protein-like (*Clostridium phytofermentans* ISDg)	AAs 36-41; 236 AAs
"	"	ZP_01274398.1	Hydroxymethylglutaryl-coenzyme A synthase, prokaryotic (*Lactobacillus reuteri* 100-23)	AAs 240-245; 385 AAs
"	"	ZP_01254940.1	Homoserine O-acetyltransferase (*Psychroflexus torquis* ATCC 700755)	AAs 192-197; 336 AAs
"	"	PDB i.d. 1VQQ|A	Penicillin Binding Protein 2a from methicillin-resistant (*Staphylococcus aureus* strain 27r)	AAs 257-262; 646 AAs

Figure 7. The location and 3D structure of peptide, **GKKGLE**, an analog of **GOOGLE**, in a protein of known 3D structure, the penicillin binding protein 2a from methicillin-resistant *Staphylococcus aureus* strain 27r (PDB i.d. code, 1VQQ). The peptide backbone is shown as a ribbon and colored gray, and the **GKKGLE** peptide is shown in space filling format and colored, green. The two figures differ only by a 90° rotation about the vertical axis. Peptide **GKKGLE** occurs in the non-penicillin binding domain of the protein.

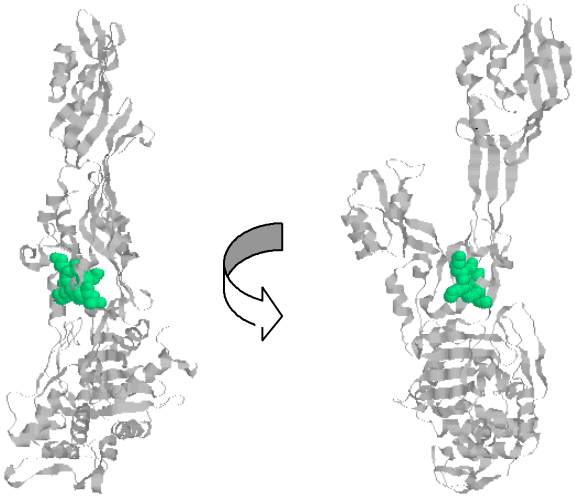
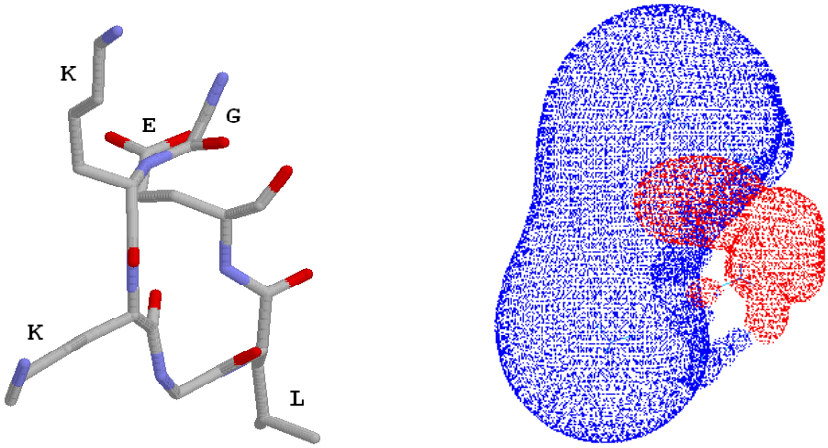


Figure 8. (Below left) A stick figure model of the isolated **GKKGLE** peptide as it occurs in the 3D structure of the protein of Figure 7 (above, left side). The position of each AA in the peptide is indicated by the adjacent single letter abbreviation for the AA. The color scheme is gray for carbon, blue for nitrogen, and red for oxygen. Hydrogens are not shown. (Below right) An electrostatic potential model of the **GKKGLE** peptide in the same structural orientation shown in the stick figure to the left. Blue areas are regions of positive electrostatic potential and red areas are regions of negative electrostatic potential. The electrostatic potential would have a direct effect on the ability of the peptide to interact with other molecules.



Wade Research Foundation Reports (2007) 4 (1)

Table 4. Examples of the occurrence of **YAHKK,** a structural analog of **YAHOO**, within proteins of the NCBI protein database.

Search Sequence:	Sequence Found:	Database Accession #:	Protein and Source:	Location in protein & total AAs in protein:
YAHKK	**YAHKK**	XP_001311174.1	Deoxyribonuclease II family protein (*Trichomonas vaginalis* G3)	AAs 105-105; 337 AAs
"	"	YP_948506.1	Citrate synthase I (*Arthrobacter aurescens* TC1)	AAs 219-223; 470 AAs
"	"	EAW57501.1	Zinc finger protein 541, isoform CRA_c (*Homo sapiens*)	AAs 93-97; 261 AAs
		NP_597731.2	Zinc finger protein 721 (*Homo sapiens*)	AAs 881-885; 923 AAs
"	"	XP_001267171.1	mRNA capping enzyme, beta subunit, putative (*Neosartorya fischeri* NRRL 181)	AAs 621-625; 799 AAs
"	"	YP_909645.1	Elongation factor Ts (*Bifidobacterium adolescentis* ATCC 15703)	AAs 156-160; 291 AAs
"	"	ZP_01575164.1	UvrD/REP helicase (*Clostridium cellulolyticum* H10)	AAs 355-359; 419 AAs
"	"	YP_796397.1	Type IV secretory pathway, VirD4 component (*Lactobacillus brevis* ATCC 367)	AAs 16-20; 543 AAs
"	"	YP_812278.1	D-alanine-D-alanine ligase related ATP-grasp enzyme (*Lactobacillus delbrueckii* subsp. bulgaricus ATCC BAA-365)	AAs 351-355; 362 AAs
"	"	ZP_01466116.1	Serine/threonine-protein kinase Pkn6 (*Stigmatella aurantiaca* DW4/3-1)	AAs 41-45; 583 AAs
"	"	ZP_01426506.1	Protein kinase (*Herpetosiphon aurantiacus* ATCC 23779)	AAs 130-134; 746 AAs
"	"	YP_891326.1	Iron ABC transporter, permease protein (*Campylobacter fetus* subsp. fetus 82-40)	AAs 509-513; 518 AAs
"	"	EAT36019.1	Transcription factor IIIA, putative (*Aedes aegypti*)	AAs 369-373; 395 AAs
"	"	YP_942709.1	Uracil-DNA glycosylase (*Psychromonas ingrahamii* 37)	AAs 163-167; 216 AAs
"	"	ZP_01272633.1	Dihydrouridine synthase TIM-barrel protein nifR3 (*Psychrobacter* sp. PRwf-1)	AAs 54-58; 330 AAs
"	"	AAI04183.1	FLJ37201 protein (*Homo sapiens*)	AAs 191-195; 206 AAs
"	"	PDB i.d. 1OKB|A	Uracil-DNA glycosylase from Atlantic cod (*Gadus Morhua*)	AAs 167-171; 223 AAs

Figure 9. The location and 3D structure of peptide, **YAHKK**, an analog of **YAHOO**, in a protein of known 3D structure, uracil-DNA glycosylase from Atlantic cod (*Gadus Morhua*) (PDB i.d. code, 1OKB). The peptide backbone is shown as a ribbon and colored gray, and the **YAHKK** peptide is shown in space filling format and colored, red. The two figures differ only by a 90° rotation about the horizontal axis.

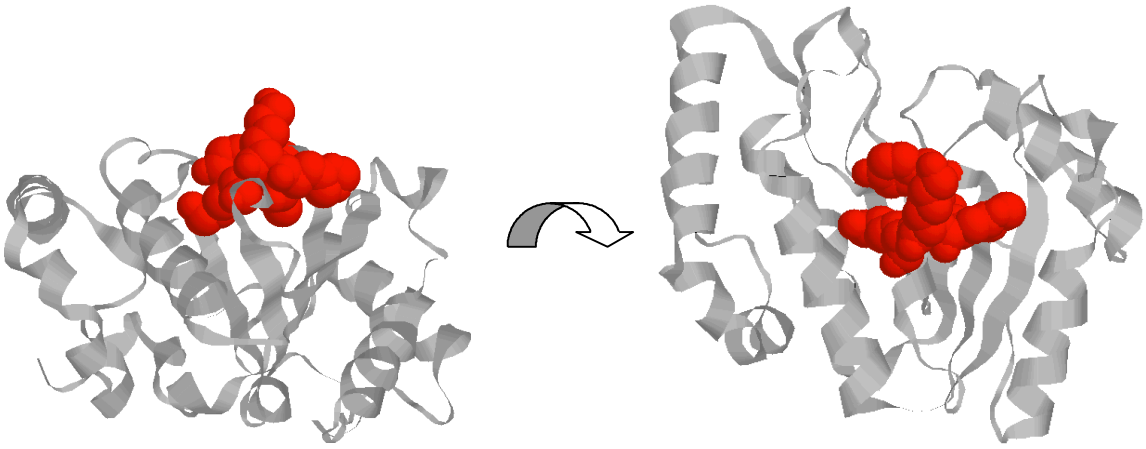
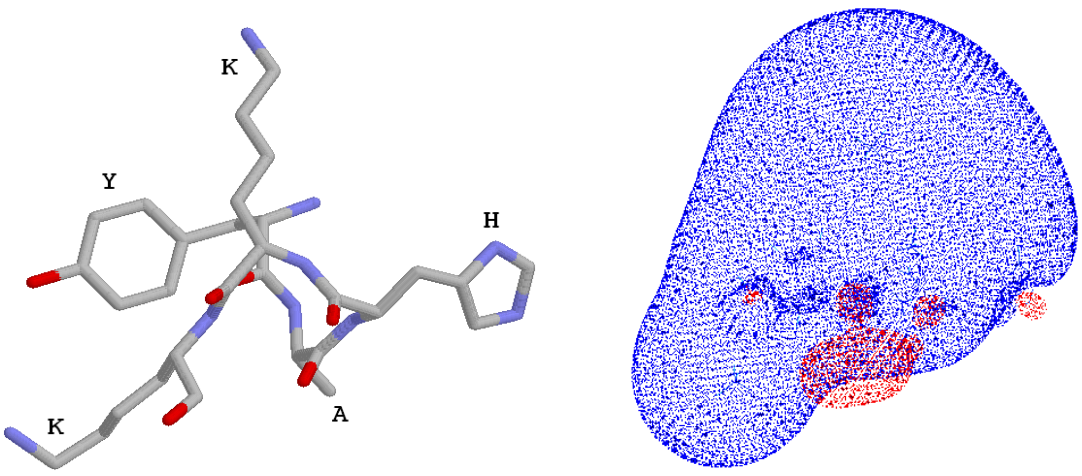


Figure 10. (Below left) A stick figure model of the isolated **YAHKK** peptide as it occurs in the 3D structure of the protein of Figure 9 (above, left side). The position of each AA in the peptide is indicated by the adjacent single letter abbreviation for the AA. The color scheme is gray for carbon, blue for nitrogen, and red for oxygen. Hydrogens are not shown. (Below right) An electrostatic potential model of the **YAHKK** peptide in the same orientation shown in the stick figure to the left. Blue areas are regions of positive electrostatic potential and red areas are regions of negative electrostatic potential. The electrostatic potential would have a direct effect on the ability of the peptide to interact with other molecules.



Wade Research Foundation Reports (2007) 4 (1)

Table 5. Examples of the occurrence of **CKCACKLA,** a structural analog of **COCACOLA**, within proteins of the NCBI protein database.

Search Sequence:	Sequence Found:	Database Accession #:	Protein and Source:	Location in protein & total AAs in protein:
CKCACKLA	**CKC _ C _ LA**	BAC85938.1	Unnamed protein product (*Homo sapiens*)	AAs 27-34; 161 AAs
"	**CKC _ CKLA** and **CKC _ CK**	AAA99803	185 kDa silk protein (*Chironomus pallidivittatus*)	AAs 973-978, 1020-1027; 1698 AAs
"	**CK _ ACKL**	NP_632817.1	Endonuclease III (*Methanosarcina mazei* Go1)	AAs 93-99; 145 AAs
"	**CKC _ CKL**	ZP_00907866.1	α-Glucosidase (*Clostridium beijerincki* NCIMB 8052)	AAs 775-781; 785 AAs
"	**CKC _ CK**	PDB i.d. 2VGH	Heparin-binding domain from vascular endothelial growth factor (*Homo sapiens*)	AAs 25-30; 55 AAs
"	**CKC _ CK**	CAM28207.1	Vascular endothelial growth factor (*Homo sapiens*)	AAs 133-138; 163 AAs
"	**CKC _ CK**	AAV73800.1	UL133 (Human herpesvirus 5)	AAs 91-96; 258 AAs
"	**CKCAC**	AAX86006.1	Metallothionein 1 (*Anopheles gambiae*)	AAs 26-30; 45 AAs
"	**C _ CACK**	EAL30651.1	GA20701-PA (*Drosophila pseudoobscura*)	AAs 20-26; 83 AAs
"	**KCACK**	PDB i.d. 1XM7	Hypothetical protein Aq_1665 (*Aquifex aeolicus*)	AAs 156-160; 195 AAs
"	**KC _ CKL**	XP_977126.1	Neurohypophysial hormones, N-terminal domain containing protein (*Tetrahymena thermophila* SB210)	AAs 296-301; 1725 AAs
"	**CACKLA**	ZP_00859856.1	Cation efflux protein (*Bradyrhizobium* sp. BTAi1)	AAs 310-315; 320 AAs
"	**KC _ CKLA**	XP_711996.1	Hypothetical protein CaO19.653 (*Candida albicans* SC5314)	AAs 14-20; 108 AAs

Figure 11. The location and 3D structure of peptide, **CKC _ CK**, from the AA sequence, **CKCACKLA**, an analog of **COCACOLA**, in a protein of known 3D structure, the heparin-binding domain from human vascular endothelial growth factor (PDB i.d. 2VGH). The peptide backbone is shown as a ribbon and colored gray, and the **CKC _ CK** peptide is shown as a space filling model and colored white. The two figures differ only by a 90º rotation about the vertical axis. Arrows indicate the locations of Cysteine residues in the sequence.

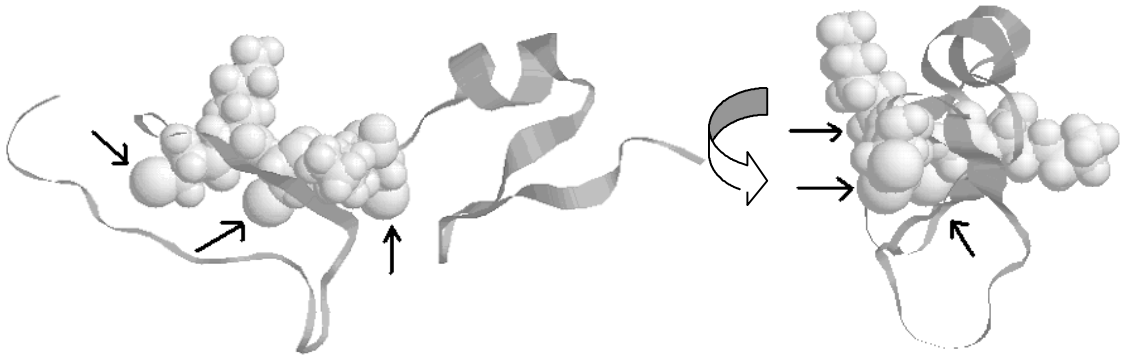
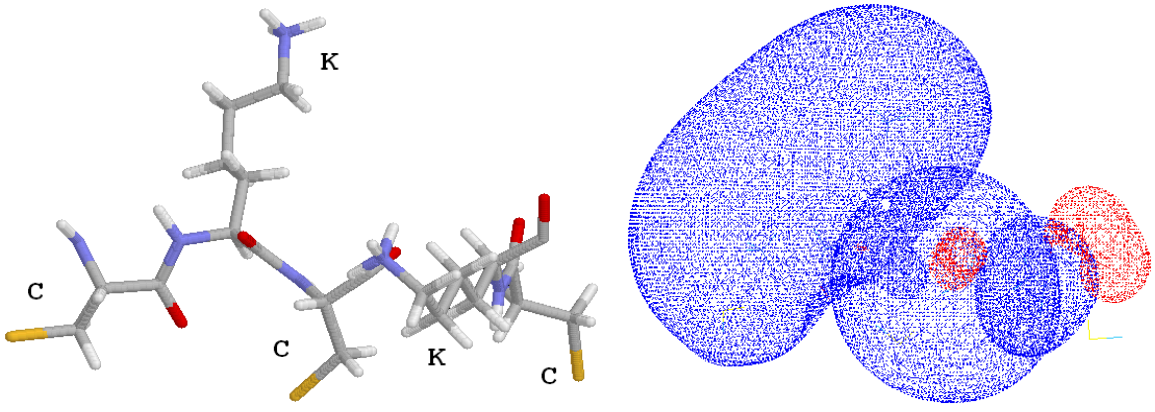


Figure 12. (Below left) A stick figure model of the isolated **CKC _ CK** peptide as it occurs in the 3D structure of the protein of Figure 11 (above, left side). The position of each AA in the peptide is indicated by the adjacent single letter abbreviation for the AA. The color scheme is gray for carbon, blue for nitrogen, red for oxygen, and white for hydrogen (some hydrogens are not shown). (Below right) An electrostatic potential model of the **CKC _ CK** peptide in the same orientation as the stick figure to the left. Blue areas are regions of positive electrostatic potential and red areas are regions of negative electrostatic potential. The electrostatic potential would have a direct effect on the ability of the peptide to interact with other molecules.



Wade Research Foundation Reports (2007) 4 (1)

Table 6. Examples of the occurrence of **PEPSICKLA,** a structural analog of **PEPSICOLA**, within proteins of the NCBI protein database.

Search Sequence:	Sequence Found:	Database Accession #:	Protein and Source:	Location in protein & total AAs in protein:
PEPSICKLA	**PEP _ IC _ LA**	AAI29966.1	Angiopoietin 4 (*Mus musculus*)	AAs 49-57; 509 AAs
"	**P _ PSIC _ LA**	AAL13364.1	SD07783p (*Drosophila melanogaster*)	AAs 234-242; 298 AAs
"	**P _ PSICK**	ZP_00910616.1	Xylan 1,4-beta-xylosidase (*Clostridium beijerincki* NCIMB 8052)	AAs 14-20; 535 AAs
"	**P _ PSIC**	PDB i.d. 1YI7	β-xylosidase (*Clostridium acetobutylicum*)	AAs 14-19; 542 AAs
"	**PEPSIC**	Q14B48|CC129_MOUSE	Coiled-coil domain-containing protein 129 (*Mus musculus*)	AAs 801-806; 1034 AAs
"	**PEPSI**	PDB i.d. 1YM7	Bovine G protein-coupled receptor kinase 2 (Grk2)	AAs 35-39; 689 AAs
"	**EPSICK**	ABG75862.1	Telomerase reverse transcriptase (*Anas platyrhynchos*)	AAs 601-606; 613 AAs
"	**PSICKL**	AAM28916.1	NBS/LRR (*Pinus taeda*)	AAs 383-388; 479 AAs
"	**PSICKL**	NP_176918.1	ATP binding / kinase/ protein serine/ threonine kinase (*Arabidopsis thaliana*)	AAs 138-143; 719 AAs
"	**SICKLA**	YP_591072.1	Type II and III secretion system protein (*Acidobacteria bacterium* Ellin345)	AAs 173-178; 764 AAs
"	**EPSIC _ LA**	NP_001027733.1	Fibroblast growth factor 11/12/13/14 (*Ciona intestinalis*)	AAs 158-165; 182 AAs
"	**EPSI _ KLA**	YP_411451.1	Translation elongation factor Tu (*Nitrosospira multiformis* ATCC 25196)	AAs 188-195; 396 AAs
"	**EPSI _ KLA**	ZP_01643637.1	Proton-translocating NADH-quinone oxidoreductase, chain M (*Stenotrophomonas maltophilia* R551-3)	AAs 486-493; 502 AAs

Figure 13. The location and 3D structure of peptide, **PEPSI**, from the AA sequence, **PEPSICOLA**, in a protein of known 3D structure, bovine G protein-coupled receptor kinase 2 (PDB i.d. 1YM7). The peptide backbone of the protein shown as a ribbon and colored gray, and the **PEPSI** segment is shown as a space filling model and colored, blue. The two figures differ only by a 90° rotation about the vertical axis.

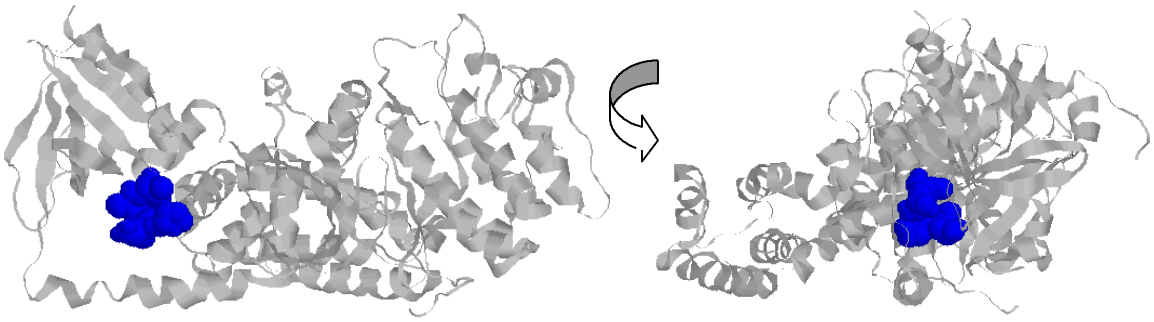


Figure 14. (Below left) A stick figure model of the isolated **PEPSI** peptide as it occurs in the 3D structure of the protein shown in Figure 13 (above, left side). The position of each AA in the peptide is indicated by the adjacent single letter abbreviation for the AA. The color scheme is gray for carbon, blue for nitrogen, and red for oxygen. Hydrogens are not shown. (Below right) An electrostatic potential model of the **PEPSI** peptide in the same orientation shown in the stick figure to the left. Red areas are regions of negative electrostatic potential. The electrostatic potential would have a direct effect on the ability of the peptide to interact with other molecules.