Wade Research Foundation Reports (2010) 5 (2)

Four Billionaire Name Peptides That Contain the CendR Motif

David Wade
Wade Research Foundation, P.O. Box 257, Princeton, New Jersey 08542-0257, USA

Abstract

The name-to-peptide concept uses the International Union of Pure and Applied Chemistry-International Union of Biochemistry and Molecular Biology, Joint Commission on Biochemical Nomenclature (IUPAC-IUB, JCBN) system for abbreviating the names of amino acids (AAs) with single letters of the English alphabet, and considers the strings of letters in personal and other names as strings of AAs, or peptides. The method can be used to design and create novel peptides of potential medical usefulness. This article reports the results of a theoretical analysis of hundreds of personal names extracted from the 2010 Forbes List of the World's Billionaires with the goal of finding names containing letter compositions corresponding to AA sequences containing the CendR motif (i.e., R/K-X-X-R/K at the carboxyl terminal end of the peptide, where R is Arginine, K is Lysine, and X can be any AA). The names of US citizens were separated from the Forbes List, and searched for the presence of the CendR motif. Only 10 of the 403 names were found to contain it, and of these only 4 had given and family names composed entirely of letters corresponding to IUPAC-IUB, JCBN abbreviations for gene-encoded AAs: Ty Warner, Stewart Rahr, Randal Kirk, and Nancy Lerner. The given and family names of these four were combined to form letter sequences corresponding to name peptides. Protein databases were searched for the occurrence of these peptides in nature, and portions of each peptide were found in a variety of natural proteins, including proteins of known three dimensional structure. Each name peptide was modeled as two commonly found structures in proteins, the α-helix and the β-strand, and potential properties of these hypothetical peptides are discussed.

Introduction

Amino acids (AAs), peptides and proteins

Peptides are polymers of AAs that are often compared to beads on a string, where the beads are AAs and the string is covalent chemical bonds, called peptide or amide bonds, that link successive AAs (Figure 1) [1]. AA polymers containing less than 100 AAs are peptides, and those containing 100 or more AAs are proteins. Peptides are ubiquitous in nature, where they perform functions essential for life. An example of a well known peptide is the hormone, insulin, a polymer containing 51 AAs and that is associated with the disease, diabetes.

AA nomenclature and the name-to-peptide method

Several decades ago, the International Union of Pure and Applied Chemistry-International Union of Biochemistry and Molecular Biology, Joint Commission on Biochemical Nomenclature (IUPAC-IUB, JCBN) officially adopted a system whereby the formal chemical names of AAs are abbreviated by single letters of the English alphabet, and this system is in common use throughout the world (Table 1) [2]. In 2003, Wade proposed a novel method for creating biologically active peptides in which the strings of letters in personal and other names were considered as strings of IUPAC-IUB, JCBN single letter

(Text continued on bottom of next page.)

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

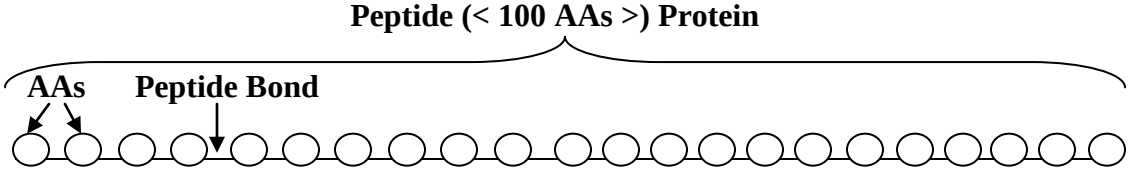


Table 1. The IUPAC-IUB, JCBN one- and three-letter abbreviations for amino acids (AAs) [2]. The letters B, J, O, X, and Z, marked with asterisks, are either unassigned, or have ambiguous assignments.

1-letter symbol	3-letter symbol	AAs	1-letter symbol	3-letter symbol	AAs
A	Ala	Alanine	N	Asn	Asparagine
B*	Asx	Aspartic acid or Asparagine	**O***	(None)	(None)
C	Cys	Cysteine	P	Pro	Proline
D	Asp	Aspartic acid	Q	Gln	Glutamine
E	Glu	Glutamic acid	R	Arg	Arginine
F	Phe	Phenylalanine	S	Ser	Serine
G	Gly	Glycine	T	Thr	Threonine
H	His	Histidine	U	Sec	Selenocysteine
I	Ile	Isoleucine	V	Val	Valine
J*	(None)	(None)	W	Trp	Tryptophan
K	Lys	Lysine	**X***	Xaa	Unknown or 'other' amino acid
L	Leu	Leucine	Y	Tyr	Tyrosine
M	Met	Methionine	**Z***	Glx	Glutamic acid or Glutamine

Figure 2. Frequency of occurrence of letters of the English alphabet in English language text [6]. The black bars indicate those letters (O, B, J, X, Z) for which there is no unambiguous assignment in the IUPAC-IUB, JCBN system for abbreviating the names of AAs with single letters of the English alphabet.

(Continued from previous page.)

abbreviations for AAs, and used to design peptides [3-5]. The major deficiency of the name-to-peptide concept is that not all letters of the English alphabet have official IUPAC-IUB, JCBN AA assignments (Table 1 and Figure 2). Unassigned letters, or those with ambiguous assignments, are B, J, O, X, and Z.

(Text continued on next page.)

Figure 3. Structures of the AAs, (left) Lysine (K) and (right) Ornithine (O). Carbon and hydrogen atoms are not shown. The figures were created using Symyx ® Draw, version 3.1.2, Symyx Technologies, Inc.

(Continued from previous page.)

Of these five letters, O occurs most frequently (7.5%), B occurs much less frequently (1.5%), and the remaining three, J, X and Z occur with almost insignificant frequencies (0.1%) in text [6]. It would be useful for the name-to-peptide method to have an AA assignment for at least the letter, O. Sometimes this letter is used as an abbreviation for the AA, Ornithine, in the scientific literature [3-5]. Ornithine is not a gene encoded AA, and, therefore, not included in the IUPAC-IUB, JCBN nomenclature system for abbreviating the names of AAs. However, Ornithine does occur as a component in the urea cycle of humans, and it is found in some naturally occurring proteins [7]. Ornithine (O) is structurally very similar to Lysine (K), with the only difference between the two AAs being that Ornithine has one less methylene group ($-CH_2$) in its side chain (Figure 3). The two AAs have nearly identical pK values for their side chain amino ($-NH_3^+$) groups [7]. The name-to-peptide method was validated in 2004 by Wade, Yang, and Lea using the name of former US Secretary of State, Colin Powell, as the basis for creating the peptide, COLINPOWELL, where O represented Ornithine [7]. Upon testing, this peptide exhibited anticancer and immune stimulating properties. The name-to-peptide method differs from rational drug design, in that it does not use nature as a starting point, and the resulting peptide may, or may not, be found in nature. However, the method has the ability to generate peptides of potential medical usefulness, and, due to the fact that names are important in all cultures, it also has the potential benefit of increasing interest in peptide science among the general public by facilitating a better understanding of this field of research.

Cend Rule and CendR peptides

A class of peptides has recently been described that induces cell internalization and tissue penetration, and that may be useful in targeted drug delivery [8-11]. The active forms of these peptides have an exposed R/K-X-X-R/K motif at their carboxyl (C-) terminal ends, where R is Arginine, K is Lysine, and X can be any AA. This motif has been termed, the C-end rule or CendR, and peptides containing the motif will be referred to as CendR peptides herein. Peptides that contain the motif internally, and not at the C-terminal end of the peptide, are inactive, but can be activated by proteolytic cleavage that results in the movement of the CendR motif to the C-terminal end of the peptide.

The Forbes List of the World's Billionaires

The Forbes list of the World's Billionaires is an annual compilation by Forbes magazine of the names and fortunes of hundreds of the world's wealthiest people [12]. The list for 2010 contains 1,011 names, and the total wealth of these individuals is about $3.6 trillion ($3,567,800,000,000). Among the world's billionaires, are 403 US citizens (40% of the total), and their total wealth is about $1.3 trillion ($1,344,200,000,000). It is possible to consider the names of these individuals as sequences of single letter, IUPAC abbreviations for polymers of AAs (i.e., as peptides).

Methods and Results

Search of the Forbes List of the World's Billionaires

The Forbes List of the World's Billionaires was searched for the names of US citizens, and 403

(Text continued on next page.)

Table 2. Names on the 2010 Forbes List of the World's Billionaires whose constituent letters correspond to gene-encoded AAs in the IUPAC-IUB, JCBN nomenclature system (Table 1), and that contain the CendR motif (K/R-X-X-K/R, where K is Lysine, R is Arginine, and X can be any AA). Letters in the name that corresponded to the CendR motif are in bold and underlined. The letter, J, shown in parentheses in one name, was omitted from consideration. The estimated net charge (NC) on each peptide at pH 7 was determined from the pK values of the α-amino and α-carboxyl groups of AAs (-NH_3^+; pK ≈ 9-11; -COOH; pK ≈ 2), and the side chain groups of Arginine [-NH-(C=NH_2^+)-NH_2, pK ≈ 12], Aspartic acid (-COOH, pK ≈ 3), Cysteine (-SH, pK ≈ 8), Glutamic acid (-COOH, pK ≈ 4), Histidine (imidazole, pK = 6), Lysine (-NH_3^+; pK ≈ 10), and Tyrosine (-OH, pK = 10) [1].

Name on Forbes List	Rank on Forbes List: World	US	AA sequence of name peptide	Number of AAs	Est. NC at pH 7
Ty Wa**r**ne**r**	342	50	TYWA**R**NE**R**	8	+1
Stewart **R**ah**r**	536	60	STEWART**R**AH**R**	11	+2
Randal (J.) **K**ir**k**	655	64	RANDAL**K**IR**K**	10	+3
Nancy Le**r**ne**r**	937	69	NANCYLE**R**NE**R**	11	0

(Continued from previous page:)

were found. This subset of names was searched for names containing letter sequences corresponding to the CendR motif, and 10 names were found to contain it. However, only 4 of the 10 have given and family names composed entirely of letters that correspond to gene-encoded AAs: Ty Warner, Stewart Rahr, Randal Kirk, and Nancy Lerner (Table 2).

Name peptides and BLAST searches

The given and family names of these four individuals were combined to form sequences of letters that were considered to be name peptides. It would be of interest to determine if the name peptides occur in nature, either as self contained peptides or as AA segments within larger natural proteins, because such information might yield clues to potential biological functions of the peptides. The occurrence of the four name peptides in naturally occurring proteins was investigated by using their AA sequences to do Basic Local Alignment Search Tool (BLAST) searches of the National Center for Biotechnology Information (NCBI) protein databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [13]. The BLAST program was also used to perform alignments of peptide AA sequences with the AA sequences of proteins of known three dimensional (3D) structure in the Protein Data Bank (PDB; http://www.pdb.org/pdb/home/home.do). The results of BLAST searches are shown in Tables 3-6, and information about the organisms containing these AA sequences is shown in Table 7. One search sequence, TYWARNER, was found in its entirety in the protein databases, and there were numerous partial sequence matches with all four of the search sequences. In addition, partial matches with all four search sequences were found in proteins with known 3D structures. Two types of structures commonly found in the 3D structures of proteins are the α-helix, which has a coiled, cylindrical shape, and the β-strand, which has an extended flattened shape (Figure 4) [1]. Those portions of the name peptides that occurred in proteins with known 3D structure were found in both α-helices and β-strands (Figures 5-8).

Modeling of name peptides

In addition to determining if the name peptides occur in natural proteins, and the 3D structures of the peptides in those proteins, it also would be of interest to examine the potential structures of the name peptides in isolation as this might also provide clues to possible biological functions of the peptides. This would be of particular interest in the event that the peptides were synthesized for testing purposes.

Three dimensional wireframe models were created with the Swiss PdbViewer program (v4.0.1)

(Text continued on Page 10)

Table 3. BLAST search results for peptide TYWARNER. The number of protein sequences searched was 10,568,660.

Search sequence:	Sequence found:	AAs in Protein:	% Identity with search:	Protein & Chain	Database ID:
TYWARNER	TYWARNER	203-210	100	Hypothetical protein HEAR0157 [*Herminiimonas arsenicoxydans*]	ref|YP_001098521.1
"	_YWARNER	220-226	87	Hypothetical protein mma_0185 [*Janthinobacterium* sp. Marseille]	ref|YP_001351875.1
"	TYW_RNER	497-504	87	Hypothetical protein Vvad_PD1757 [*Victivallis vadensis* ATCC BAA-548]	ref|ZP_06242878.1
"	TYWARN_R	853-860	87	Polyketide synthase [*Chondromyces crocatus*]	emb|CAQ18838.1
"	_YWAR_ER	6-12	75	Hypothetical protein SORBIDRAFT_02g021210 [*Sorghum bicolor*]	ref|XP_002462191.1
"	_YW_RNER	122-128	75	Cell cycle protein [*Roseiflexus castenholzii* DSM 13941]	ref|YP_001433203.1
"	_YWAR_ER	86-92	75	Aminotransferase, class V [*Aeromicrobium marinum* DSM 15272]	ref|ZP_06047029.1
"	_YWARNE_	112-117	75	Hypothetical protein THA_850 [*Thermosipho africanus* TCF52B]	ref|YP_002334652.1
"	TYWARN	41-46	75	Glycoside hydrolase family 3 domain protein [*Hyphomicrobium denitrificans* ATCC 51888]	ref|ZP_05374821.1
"	TYWAR_E_	85-91	75	RagB/SusD domain protein [*Rhodothermus marinus* DSM 4252]	ref|YP_003290756.1
"	TYWARN_ _	48-53	75	AGAP005971-PA [*Anopheles gambiae* str. PEST]	ref|XP_316013.4
"	_ _ WARNER	129-134	75	Aminotransferase [*Streptomyces ghanaensis* ATCC 14672]	ref|ZP_04689321.1
"	_ _WARNE_	103-107	62	Obelin [*Obelia longissima*]	PDB 1QV0_A
"	T_W_RNE	348-354	62	Glycerol phosphate lipoteichoic acid synthase [*Bacillus subtilis*]	PDB 2W8D_A

Wade Research Foundation Reports (2010) 5 (2)

Table 4. BLAST search results for peptide STEWARTRAHR. The number of protein sequences searched was 10,391,716.

Search sequence:	Sequence found:	AAs in Protein:	% Identity with search:	Protein & Chain	Database ID:
STEWARTRAHR	S _ EWART _ AHR	150-160	82	Major facilitator superfamily MFS_1 [*Burkholderia* sp. CCGE1001]	ref|ZP_06294835.1
"	STE _ A _ TRAHR	703-713	82	Valyl-tRNA synthetase [*Halomicrobium mukohataei* DSM 12286]	ref|YP_003177358.1
	_ _ _ WARTRAHR	269-276	73	Urease accessory protein D [*Synechococcus* sp. BL107]	ref|ZP_01469140.1
"	_ TEWARTR _ _ _	570-576	64	Hypothetical protein BRADO0472 [*Bradyrhizobium* sp. ORS278]	ref|YP_001202658.1
"	_ TEW _ RTRA _ _	161-168	64	HAD family hydrolase [*Salinispora arenicola* CNS-205]	ref|YP_001535720.1
"	_ _ EWARTRA _ _	12-18	64	Predicted protein [*Ostreococcus lucimarinus* CCE9901]	ref|XP_001422032.1
"	_ _ EW _ RT _ AHR	398-406	64	Cytochrome P450 [*Streptomyces ghanaensis* ATCC 14672]	ref|ZP_04685351.1
"	_ _ EW _ RTRA _ R	141-149	64	DNA polymerase III, delta subunit [*Sphaerobacter thermophilus* DSM 20745]	ref|YP_003319471.1
"	_ TEWAR _ _ _ _ _	288-292 (426-430 in PDB model)	45	High affinity cAMP-specific 3',5'-cyclic phosphodiesterase 7A, catalytic domain residues 139-456 [*Homo sapiens*]	pdb|3G3N|A
"	STEW _ _ _ _ _ _ _	167-170	36	Cholesterol oxidase [*Brevibacterium sterolicum*]	pdb|1COY|A
"	_ _ _ WAR _ RA _ R	386-393 (377-384 in PDB model)	54	1-Deoxy-D-xylulose 5-phosphate reductoisomerase [*Mycobacterium tuberculosis*]	pdb|2JCV|A
"	_ _ EWAR _ R _ _ _	139-144	45	TTHA1264, a putative M16-family zinc peptidase [*Thermus thermophilus*]	pdb|3EOQ|A
"	_ TEWA _ T _ _ _ _	333-338	45	1-Deoxy-D-xylulose 5-phosphate synthase [*Deinococcus radiodurans*]	pdb|2O1X|A

Wade Research Foundation Reports (2010) 5 (2)

Table 5. BLAST search results for peptide RANDALKIRK. The number of protein sequences searched was 10,421,070.

Search sequence:	Sequence found:	AAs in Protein:	% Identity with search:	Protein & Chain	Database ID:
RANDALKIRK	RANDALK_R_	132-140	80	Predicted protein [*Coprinopsis cinerea okayama7*#130]	ref|XP_001831246.1
”	_ _NDALKI_K	255-262	70	N-methyltryptophan oxidase	pdb|2UZZ|A
“	RA_DAL_ _RK	84-93	64	Hypothetical protein Pmar_PMAR022537 [*Perkinsus marinus* ATCC 50983]	gb|EER14004.1
“	R_ _DALKIR_	344-352	64	Amino acid adenylation [*Burkholderia oklahomensis* C6786]	ref|ZP_02367647.1
“	RA_ _ALKIR_	298-306	64	Type VI secretion protein IcmF [*Ralstonia pickettii* 12D]	ref|YP_002981628.1
“	_ _ _DALKIRK	627-633	64	Unnamed protein product [*Vitis vinifera*]	emb|CBI39537.1
“	_AND_LK_RK	86-94	64	Phospholipid-translocating P-type ATPase, flippase family protein [*Tetrahymena thermophila*]	ref|XP_001020378.1
“	RA_D_LKIR_	558-566	64	Phosphoenolpyruvate-protein phosphotransferase [*Sideroxydans lithotrophicus* ES-1]	ref|ZP_05338765.1
“	_ _NDALKIR_	428-434	64	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase family protein [*Acidithiobacillus caldus* ATCC 51756]	ref|ZP_05291862.1
“	_AND_LKIR_	85-92	64	Proliferating cell nuclear antigen [*Symbiodinium goreaui*]	gb|ABO40129.1
“	_AN_ALKIR_	213-220	64	Hypothetical protein GB2207_09796 [marine gamma proteobacterium HTCC2207]	ref|ZP_01223535.1
“	_ANDALK_R_	200-207	64	Hypothetical protein [*Oryza sativa* Japonica Group]	gb|AAK13110.1|AC080019_2
“	RANDAL_I_ _	1113-1120	64	NAD-specific glutamate dehydrogenase large form [*Vibrio furnissii* CIP 102972]	ref|ZP_05877392.1

Table 6. BLAST search results for peptide NANCYLERNER. The number of protein sequences searched was 10,421,070.

Search sequence:	Sequence found:	AAs in Protein:	% Identity with search:	Protein & Chain	Database ID:
NANCYLERNER	NA_CYL_RNE_	155-164	73	TU4 isoform C [*Mus musculus*]	gb|ACC61063.1
"	NA_ _YLERNE_	100-109	73	Hypothetical protein [*Entamoeba histolytica* HM-1:IMSS]	ref|XP_649983.2
"	_ _NCYLER_ER	145-153	73	RNA (guanine-9-) methyltransferase domain-containing protein [*Toxoplasma gondii* GT1]	gb|EEE25686.1
"	_ANCYLER_ _ _	257-263	64	Putative pleiotropic regulatory protein [*Synechococcus* sp. BL107]	ref|ZP_01468834.1
"	_ANCYL_ _NE_	34-42	64	Conserved hypothetical protein [*Raphidiopsis brookii* D9]	ref|ZP_06303801.1
"	_ _NCYL_RNE_	164-171	64	Predicted protein [*Naegleria gruberi*]	gb|EFC44168.1
"	_ _NCYLERN_ _	2195-2201	64	Hypothetical protein Fisuc_0824 [*Fibrobacter succinogenes* subsp. *succinogenes* S85]	ref|YP_003248916.1
"	_ _ _ _YLERNER	119-125	64	Hypothetical protein NRI_0709 [*Neorickettsia risticii* str. Illinois]	ref|YP_003081920.1
"	A_ _YLER_ER	208-217	64	Extend3d SID of MAD1 bound to the PAH2 domain of MSIN3B [*Mus musculus*]	pdb|1PD7|B
"	_ _NCYLE_N_ _	93-99	54	Yeast fatty acid synthase, subunit alpha [*Saccharomyces cerevisiae*]	pdb|2PFF|B
"	_ _NCYLE_ _E_	132-139	54	Hypothetical protein EUBIFOR_01996 [*Eubacterium biforme* DSM 3989]	ref|ZP_03489407.1
"	_ _NCY_ _RNE_	67-74	54	Protoheme IX farnesyltransferase [*Kocuria rhizophila* DC2201]	ref|YP_001855073.1
"	_ _ _CYL_RN_R	152-159	54	Hypothetical protein SORBIDRAFT_03g003695 [*Sorghum bicolor*]	ref|XP_002457230.1
"	_ _ _CYL_ _NER	94-101	54	Hypothetical protein CdifQCD-2_14696 [*Clostridium difficile* QCD-23m63]	ref|ZP_05402311.1
"	_ _ _CYL_RNE_	136-142	54	TPR repeat-containing protein [*Desulfovibrio salexigens* DSM 2638]	ref|YP_002992528.1

Table 7. Organisms cited in Tables 3-6. Descriptions were obtained by internet searches.

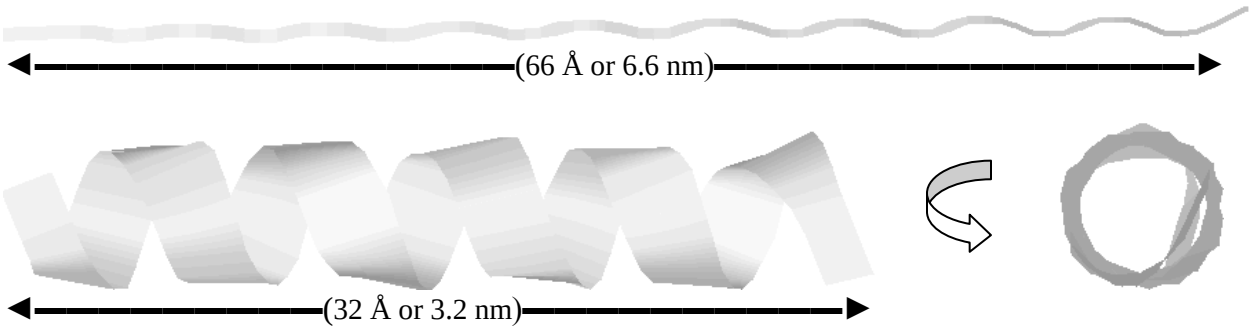
Latin name:	English description:
Acidithiobacillus caldus	Gammaproteobacterium
Aeromicrobium marinum	Pelagic bacterium
Anopheles gambiae	Mosquito vector of malaria
Bacillus subtilis	Gram-positive bacterium commonly found in soil
Bradyrhizobium	Gram-negative, nitrogen fixing, soil bacteria
Burkholderia	Gram negative proteobacteria,some of which are pathogenic
Burkholderia oklahomensis	Gram negative, possibly pathogenic, proteobacteria
Chondromyces crocatus	Myxobacterium
Clostridium difficile	Bacteria that causes diarrhea and other intestinal disease
Coprinopsis cinerea okayama7	Higher fungus with a typical mushroom form
Deinococcus radiodurans	Gram positive, extremophilic, highly radioresistant, bacterium
Desulfovibrio salexigens	Gram-negative, anaerobic, marine, sulfate reducing bacterium
Entamoeba histolytica	Parasitic protozoan that infects humans
Eubacterium biforme	Gram positive bacterium
Fibrobacter succinogenes	Gram negative bacterium present in the rumen of cattle
Halomicrobium mukohataei	Halophilic archaeon
Herminiimonas arsenicoxydans	Metalloresistant bacterium
Homo sapiens	Humans
Hyphomicrobium denitrificans	Methylotrophic denitrifying bacterium
Janthinobacterium	Purple bacterium
Kocuria rhizophila	Gram positive soil bacterium
Mus musculus	House mouse
Mycobacterium tuberculosis	Pathogenic bacterium, causative agent of tuberculosis
Naegleria gruberi	Free-living, soil and freshwater amoeboflagellate
Neorickettsia risticii	Gram negative bacterium; causative agent of Potomac horse fever
Obelia longissima	Marine invertebrate, Hydrozoan
Oryza sativa	Rice
Ostreococcus lucimarinus	Single-celled alga, abundant in oceans
Perkinsus marinus	Protozoan, pathogen of oysters
Ralstonia pickettii	Gram-negative soil bacterium
Raphidiopsis brookii	Cyanobacteria
Rhodothermus marinus	Thermohalophilic bacterium
Roseiflexus castenholzii	Filamentous photosynthetic bacterium
Saccharomyces cerevisiae	Baker's yeast
Salinispora arenicola	Marine bacterium
Sideroxydans lithotrophicus	Proteobacteria
Sorghum bicolor	Sorghum, a species of plant
Sphaerobacter thermophilus	Gram positive bacterium
Streptomyces ghanaensis	Gram positive bacterium
Symbiodinium goreaui	Alga
Synechococcus	Marine cyanobacterium
Tetrahymena thermophila	Ciliated protozoan
Thermosipho africanus	Thermophilic eubacterium
Thermus thermophilus	Gram negative eubacterium
Toxoplasma gondii	Parasitic protozoa; cause of Toxoplasmosis
Vibrio furnissii	Gram negative bacterium

(Continued on next page.)

Table 7 (Con't.). Organisms cited in Tables 3-6. Descriptions were obtained by internet searches.

Latin name:	English description:
Vibrio furnissii	Gram negative bacterium
Victivallis vadensis	Sugar-fermenting anaerobic bacteria
Vitis vinifera	Plant, grape vine

Figure 4. Ribbon diagrams of two types of structure commonly found in proteins [1]: the β-strand which has a flattened, extended structure (top) and the α-helix, which has a coiled, cylindrical shape [bottom left (longitudinal view) and bottom right (cross sectional view)]. Ribbon diagrams only show the shape of the peptide backbone [i.e., successive amide bonds (-CO-NH-) within the peptide], and do not show AA side chains. Both structures shown below were modeled from a peptide containing 20 Alanine residues. In the β-strand conformation, the peptide is 66 Å long whereas in the α-helical conformation, it is only 32 Å long, a 51% reduction in length. (Note: The scale of the two objects shown is not the same, but the measured dimensions are correct)



(Continued from page 4:)

using omega, phi, and psi angles corresponding to a β–strand (an extended, nearly flat structure) or α-helix (a coiled, cylindrical structure). Each model was subjected to one round of energy minimization (5,000 steps of steepest descent), and the energy of each resulting structure was recorded. Electrostatic potential models were also made with the Swiss PdbViewer program using default parameters. After creation and energy minimization, the models were transferred to the RasWin Molecular Graphics program (Windows version 2.6-ucb), and modeled as stick figures using the CPK color scheme. Molecular measurements were also made with the RasWin program. The stick figures were then transferred to the Microsoft ® Paint program (version 5.1) for the addition of letters and charges (+/-).

Figures 9 and 10 show the results of modeling all 4 peptides in β-strand and α-helical conformations, respectively. Table 8 lists the dimensions and energies of all 4 peptides in both conformations. The range of lengths was 17-32 Å, or 1.7-3.2 nanometers (nm), which would place all of these structures and any functions that they are eventually found to exhibit, within the realm of nanotechnology. Converting the extended, β-strand structures into coiled, cylindrical and more compact α-helical structures resulted in an average 22% reduction in energy, and an average 47% reduction in length.

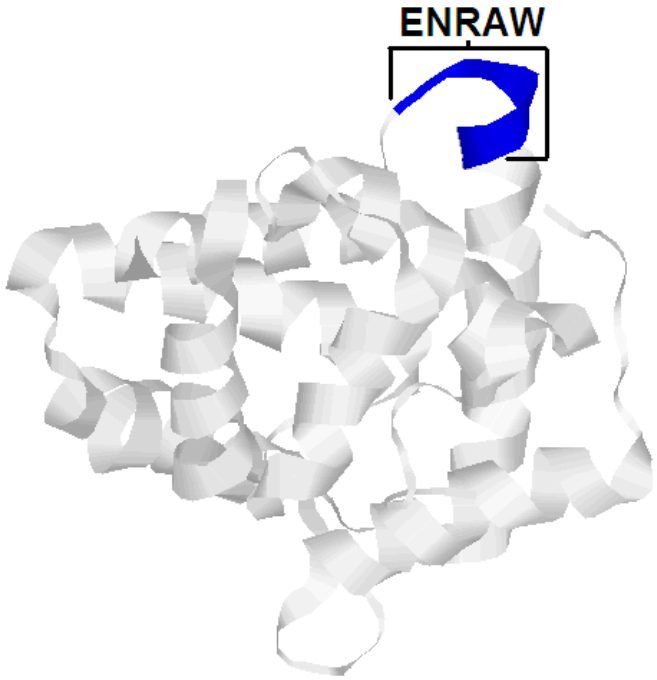
Figures 11 and 12 are electrostatic potential (EP) diagrams for each of the structures in Figures 9 and 10. All EP diagrams were made with the Swiss PdbViewer program, using default parameters which placed potentials around only charged AA residues.

It should be noted that the net charges at pH 7 of all four peptides could be increased by +1 (i.e., TYWARNER, +1 → +2; STEWARTRAHR, +2 → +3; RANDALKIRK, +3 → +4; NANCYLERNER, 0 → +1) simply by chemically synthesizing the peptides as C-terminal amides (i.e., -COOH → -$CONH_2$). This modification removes the negative charge from the C-terminal end of the peptide, and would

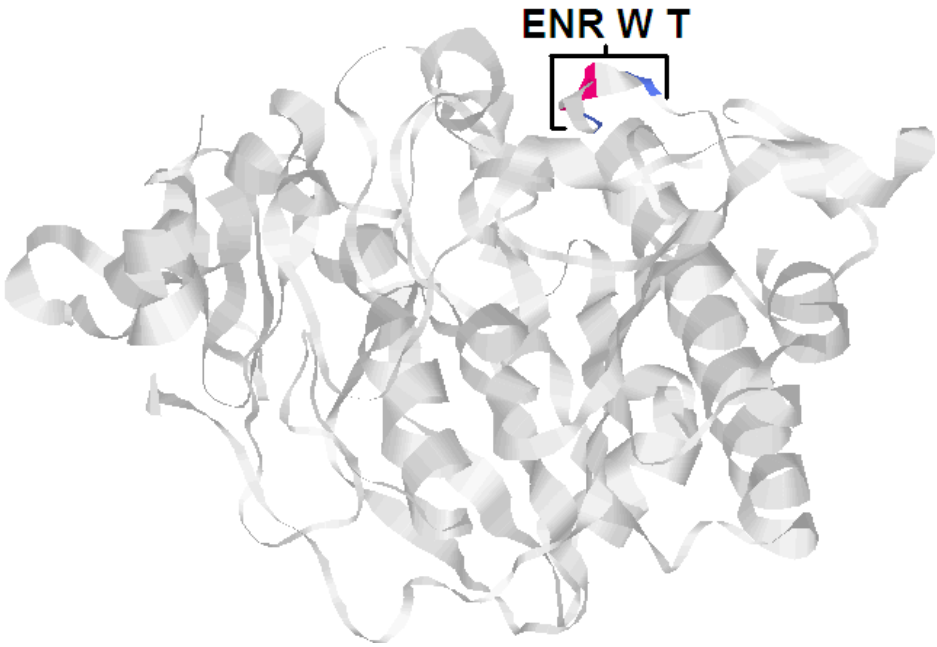
(Text continued on page 15.)

Wade Research Foundation Reports (2010) 5 (2)

Figure 5. Locations and 3D structures of portions of peptide TYWARNER in the 3D structures of two proteins. The peptide segments are shown in color (blue and purple), and the position of each protein has been oriented for maximum visibility of the sequences of interest, which results in the reversed sequences shown below. (Top) AA sequence Trp103-Glu107 (WARNE) in obelin from *Obelia longissima* (PDB 1QV0, chain A). (Bottom) AA sequence Thr348-Glu354 (T_W_RNE) in glycerol phosphate lipoteichoic acid synthase from *Bacillus subtilis* (PDB 2W8D, chain A). The AA sequences of interest occur at the surface of the proteins, and at the end of an α-helix and in a loop in 1QV0, and in a β-strand and a loop in 2W8D.



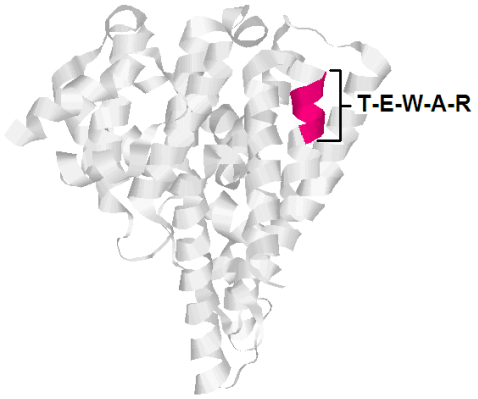
PDB 1QV0



PDB 2W8D

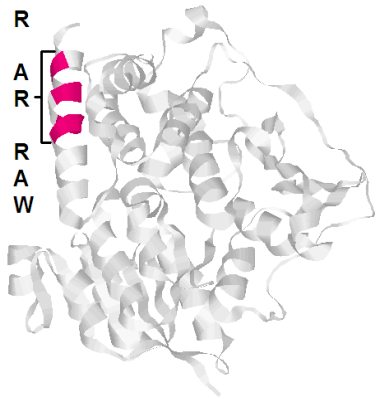
Wade Research Foundation Reports (2010) 5 (2)

Figure 6. Locations and 3D structures of portions of peptide STEWARTRAHR in the 3D structures of five proteins. The peptide segments are shown in color (purple and green). (Upper left) AA sequence, Thr426-Arg430 (TEWAR) in the 3D structure of high affinity cAMP-specific 3',5'-cyclic phosphodiesterase 7A, catalytic domain residues 139-456, of *Homo sapiens* (PDB 3G3N). (Upper center) AA sequence, Ser167-Trp170 (STEW; sequence appears reversed in figure) in the 3D structure of cholesterol oxidase of *Brevibacterium sterolicum* (PDB 1COY). (Upper right) AA sequence, Trp377-Arg384 (WAR _ RA - R) in the 3D structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase of *Mycobacterium tuberculosis* (PDB 2JCV). (Lower left) AA sequence, Glu139-Arg144 (EWAR _ R) in the 3D structure of TTHA1264, a putative M16-family zinc peptidase from *Thermus thermophilus* (PDB 3EOQ). (Lower right). AA sequence, Thr333-Thr338 (TEWA _ T), in the 3D structure of 1-deoxy-D-xylulose 5-phosphate synthase from *Deinococcus radiodurans* (PDB 2O1X). In all structures, except the upper center one (PDB 1COY), the entire peptide AA sequence occurs within an α-helix. In the upper center structure, one end of the peptide AA sequence occurs at the end of an α-helix, and the other end occurs as part of a β-strand.

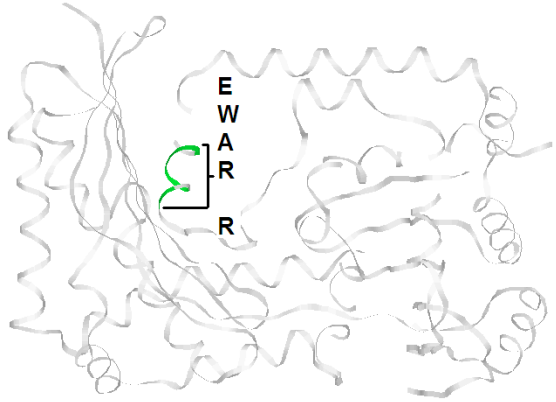


PDB 3G3N

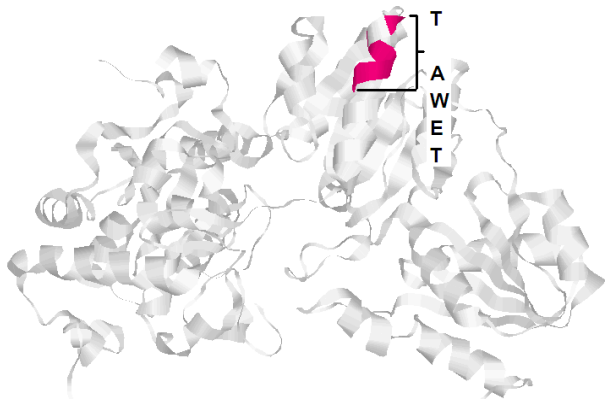
PDB 1COY



PDB 2JCV



PDB 3EOQ

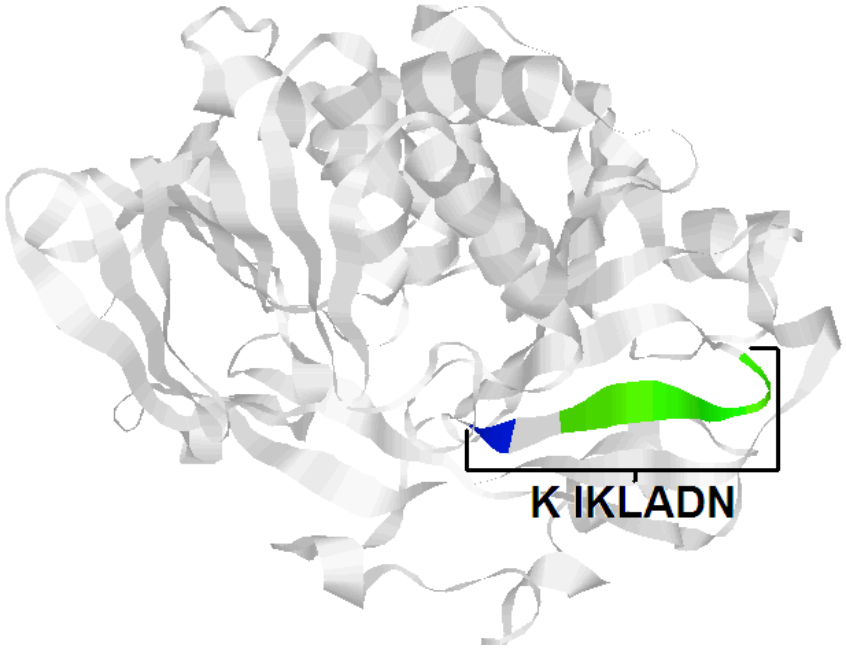


PDB 2O1X

Wade Research Foundation Reports (2010) 5 (2)

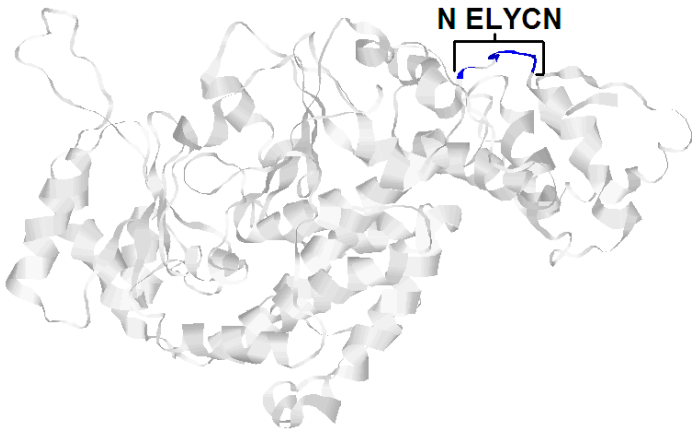
Figure 7. Locations and 3D structures of portions of peptide RANDALKIRK in the 3D structures of two proteins. The peptide segments are shown in color (purple, green, and blue). (Top) AA sequence, Asn318-Ile323 (NDALKI) in Bna3p, a putative kynurenine aminotransferase from *Saccharomyces cerevisiae* (PDB 3B46, chain A). In this structure, the peptide segment occurs within an α-helix. (Bottom) AA sequence, Asn255-Lys262 (NDALKI _ K) in N-methyltryptophan oxidase (PDB 2UZZ, chain A). The protein has been oriented for maximum visibility of the AA sequence of interest which results in the sequence appearing to be reversed. In this structure, the peptide segment occurs within a β-strand.

PDB 3B46

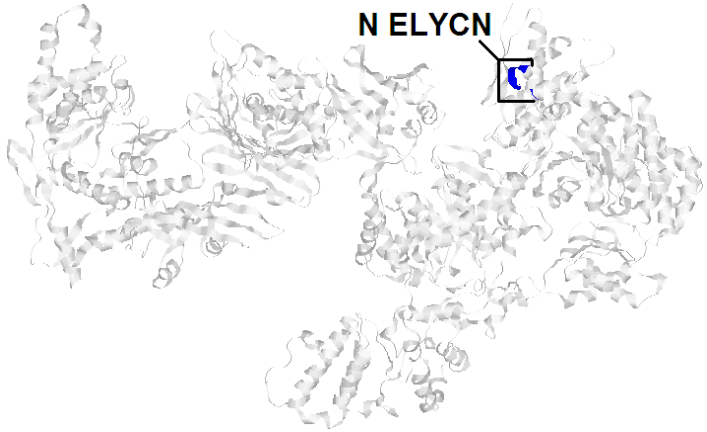


PDB 2UZZ

Figure 8. Locations and 3D structures of portions of peptide NANCYLERNER in the 3D structures of four proteins. The peptide segments are shown in color (blue and purple). (Upper left and upper right) AA sequence, Asn93-Asn99 (NCYLE N), in yeast fatty acid synthase, subunit alpha (PDB 2PFF) and subunit G (PDB 2UV8). In PDB 2PFF, the peptide sequence occurs in an unstructured region of the protein, and in PDB 2UV8 a portion of the sequence occurs in the end of an α-helix and a portion occurs in a β-strand. (Lower left) AA sequence, Ala208-Arg217 (A_ _YLER_ER), in extended SID of MAD1 bound to the PAH2 domain of MSIN3B (PDB 1PD7, chain B). (Lower right) AA sequence, Ala194-Arg200 (ANC_LER), STAL, a glycopeptides antibiotic sulfotransferase from *Streptomyces toyocaensis* (PDB 2OV8). In the two lower structures the peptide sequences occur within an α-helix. The proteins are oriented for maximum visibility of the sequence of interest, and, as a result, the sequences appear reversed in all proteins, except 2OV8 (lower right).

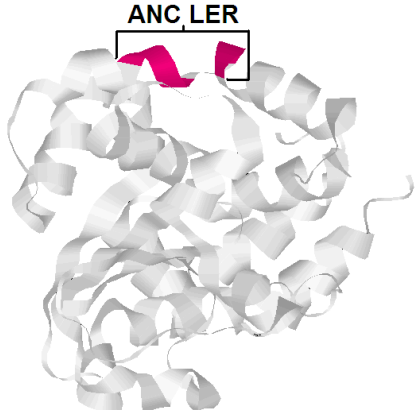


PDB 2PFF



PDB 2UV8

PDB 1PD7



PDB 2OV8

Table 8. Energy and length of each peptide structure after one round of energy minimization with the Swiss PdbViewer program.

	Energy (kJ/mol)		% Reduction In energy	Length (Å)		% Reduction in length
Peptide	β-Strand	α-Helix	β → α	β-Strand	α-Helix	β → α
TYWARNER	-548	-646	18	25	12	52
STEWARTRAHR	-670	-836	25	35	19	46
RANDALKIRK	-544	-683	25	33	17	48
NANCYLERNER	-964	-1157	20	36	20	44
Avg. ± Std. Dev.	-681 ± 197	-830 ± 233	22 ± 4	32 ± 5	17 ± 4	47 ± 3

(Continued from page 10.)

eliminate the electronegative (i.e., red colored) regions from the C-terminal ends of all peptides in the electrostatic potential diagrams shown in Figures 11 and 12. Such a modification could drastically change the biological properties of a peptide (e.g., enhancing antimicrobial properties) [7].

Discussion

This article examined four hypothetical peptides that have AA sequences that are based upon the letter compositions of four personal names. The names were considered as sequences of single letter, IUPAC-IUB, JCBN abbreviations for AAs. The resulting name peptides are 8-11 AAs in length, have net charges at pH 7 of 0 - +3, and contain the CendR motif which can be used to induce cell internalization and tissue penetration. One peptide was found to exist in its entirety in a naturally occurring protein, and fragments of all peptide AA sequences, covering all parts of all name peptide AA sequences, were found to occur in many naturally occurring proteins, including some proteins of known 3D structure. Those sequence fragments found in known 3D structures were found in α-helical, β-strand, loops and bend conformations within the overall protein 3D structures, demonstrating the conformational flexibility of these peptides.

Increasing the number of CendR name peptides

Ornithine (O) has not yet been tested to determine if it can replace Lysine (K) in the CendR motif [14]. CendR peptide sequences are thought to function by binding to neuropilin-1 (NRP1), a transmembrane receptor on the surfaces of cells [8-11]. Receptor-ligand interactions are usually very specific, and that might exclude the use of Ornithine as a substitute for Lysine, even though Ornithine is only one methylene group ($-CH_2$) shorter than Lysine. However, if experiments show that Ornithine can be used to replace Lysine, then the number of names on the Forbes List containing the CendR motif could be expanded from 4 to 7 (Table 9) if O is used as an abbreviation for Ornithine.

Use of billionaire name peptides as carrier molecules

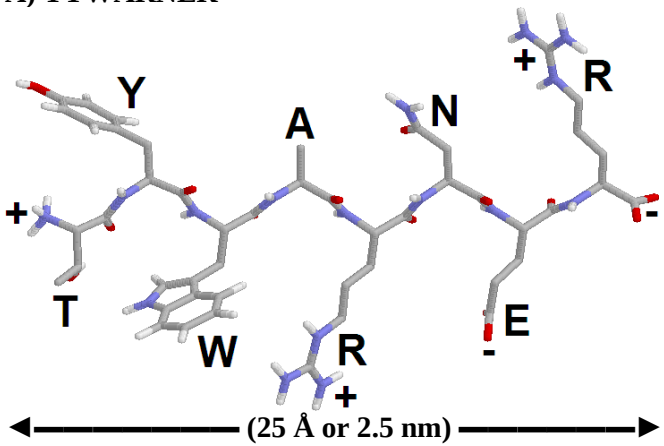
In addition to possessing the CendR motif, the four billionaire name peptides contain functional groups that would enable the chemical linkage of other molecules, such as drug molecules, to the peptides. This would enable the name peptides to be used as carriers to deliver these additional molecules into cells. If the functional groups located within the four AAs comprising the CendR motif are excluded, other functional groups available for derivatization include: the amine group ($-NH_2$) present at the amino terminal end of each peptide; the guanidino group [$-NH-C(NH)-NH_2$] in the side chain of Arginine (R) in RANDALKIRK and STEWARTRAHR; the hydroxyl group (-OH) in the side chains of Tyrosine (Y), Serine (S), and Threonine (T) of TYWARNER, STEWARTRAHR, and NANCYLERNER; indole ring nitrogens in the side chain of Tryptophan (W) in TYWARNER and STEWARTRAHR; the carboxyl group (-COOH) in the side chain of Glutamic acid (E), in STEWARTRAHR and NANCYLERNER, and

(Continued on page 22.)

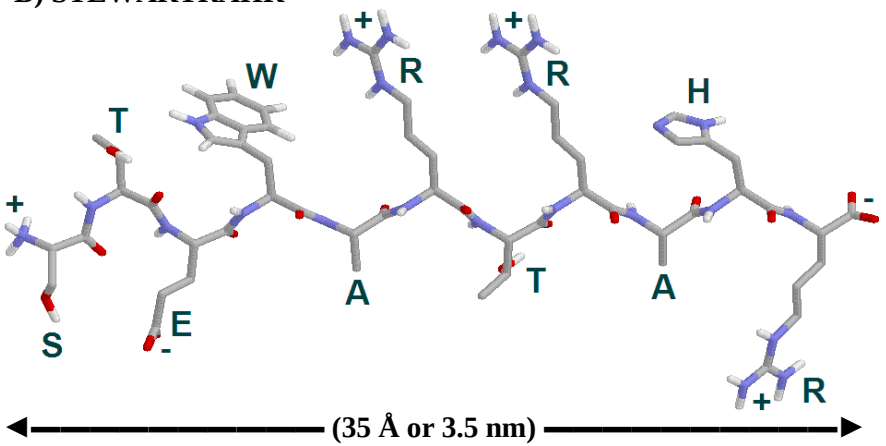
Wade Research Foundation Reports (2010) 5 (2)

Figure 9. Stick figure models of the β-strand (flat, extended) structures of the four name peptides containing the CendR motif. The models were created with the Swiss PdbViewer program which does not display the hydrogens of the methyl group (-CH_3) in Alanine group or the hydrogens of methylene (-CH_2-) groups in other AAs. The color scheme is gray for carbon, blue for nitrogen, red for oxygen, yellow for sulfur, and white for hydrogen. Charges (+/-) on various moieties at pH 7 are shown. Single letter abbreviations for AAS are shown adjacent to the AA. The average length of these structures is 32 Å, or 3.2 nm (Table 8).

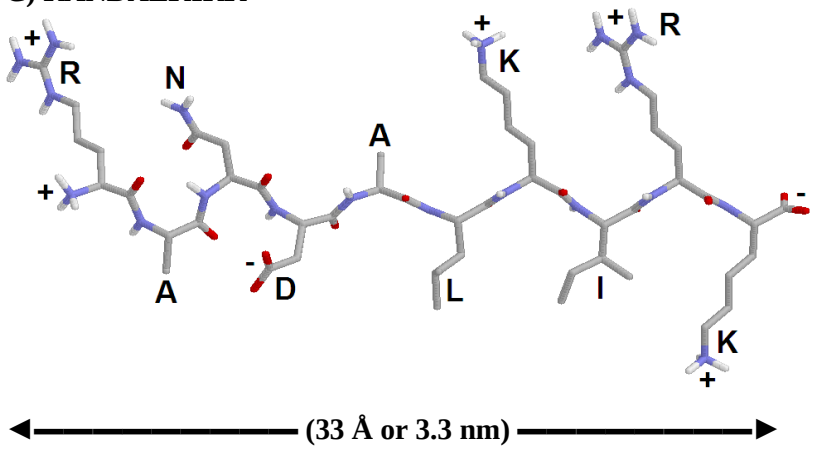
A) TYWARNER



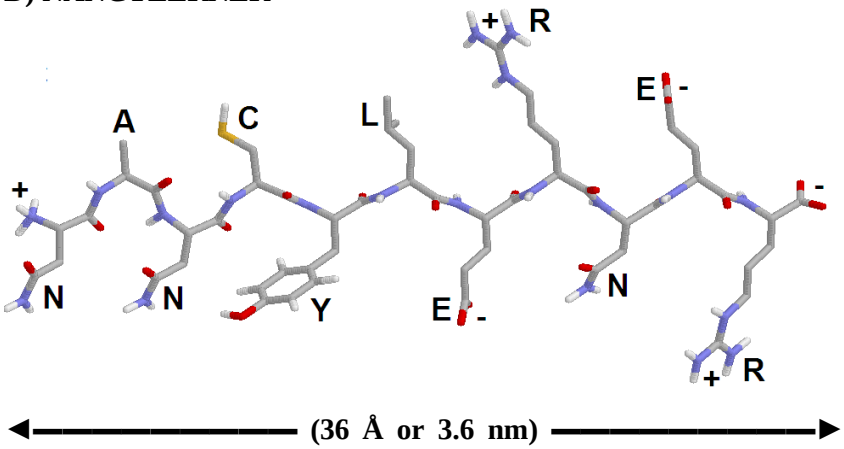
B) STEWARTRAHR



C) RANDALKIRK



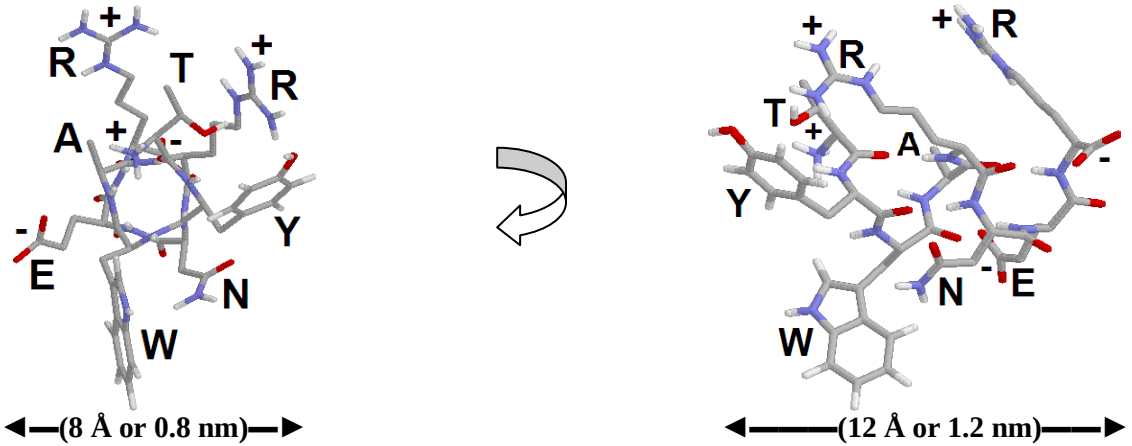
D) NANCYLERNER



Wade Research Foundation Reports (2010) 5 (2)

Figure 10. Stick figure models of the α-helical (cylindrical) structures of the four name peptides containing the CendR motif. This type of structure is held together by hydrogen (H-) bonds between peptide amide groups (C=O - H-N) at i+4 intervals within the peptide [1], and is thermodynamically favorable when the peptide interacts with other molecules, such as those comprising the lipid bilayers membranes of cells. (Left) Cross sectional view along the helix axis with the N-terminal end of the peptide closest to the viewer. (Right) View obtained by rotating the cross sectional view by 90° about the vertical axis. The N-terminal and C-terminal ends of the peptide are on the left and right ends of the models, respectively. Conversion of the β-strand structures of Figure 9 to the α-helical structures of this figure resulted in an average reduction in the overall length of peptides by 47%, and an average decrease in energy of 22% (Table 8).

A) TYWARNER:



B) STEWARTRAHR:

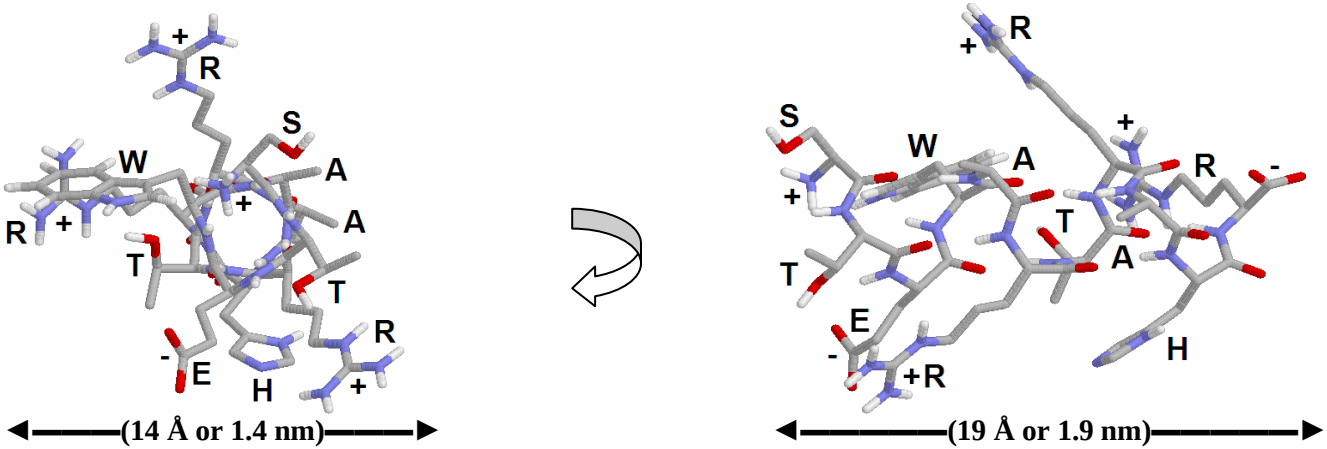
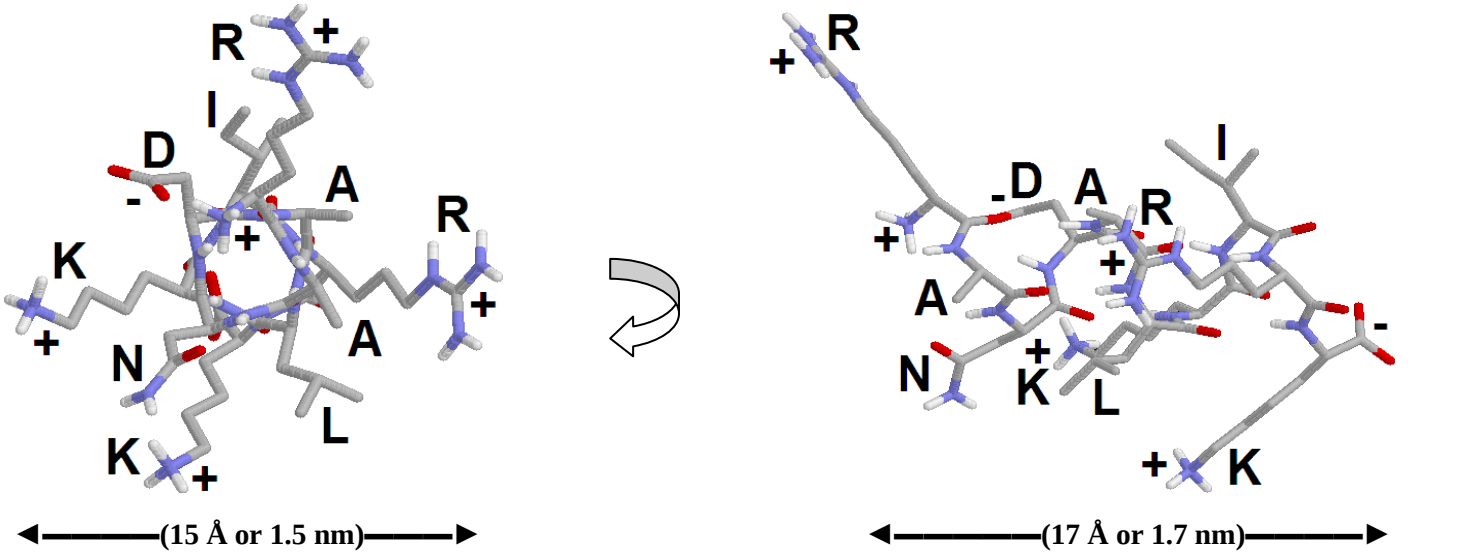


Figure 10 (Continued):

C) RANDALKIRK:



D) NANCYLERNER:

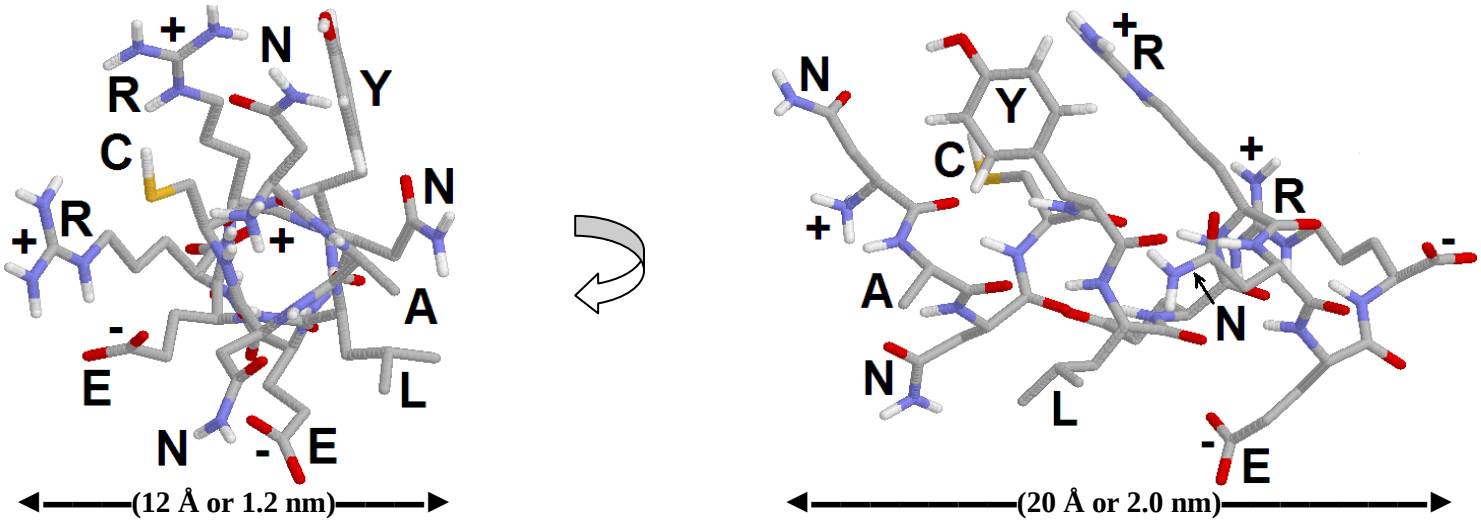
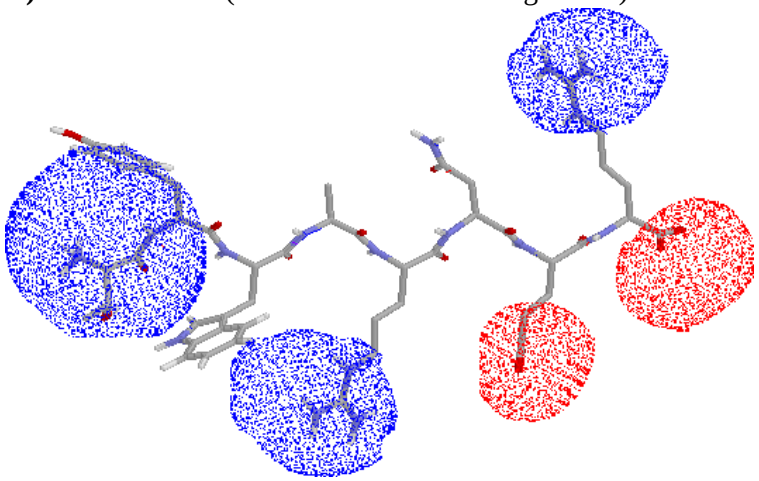
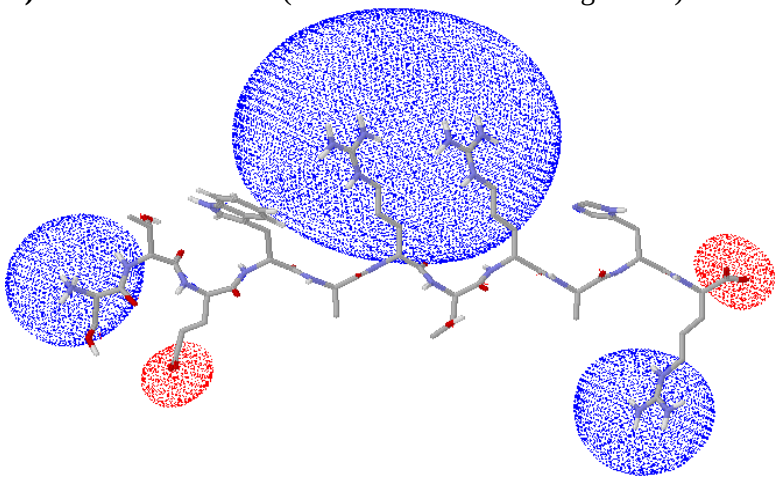


Figure 11. Electrostatic potential models of the four name peptides in the β-strand conformation, and in the same orientations shown in Figure 9. The blue and red areas represent regions of positive and negative electrostatic potential, respectively.

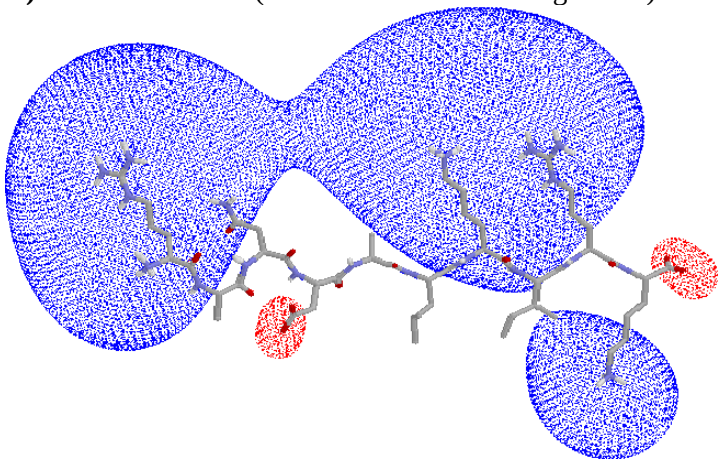
A) TYWARNER (same orientation as in Figure 9A).



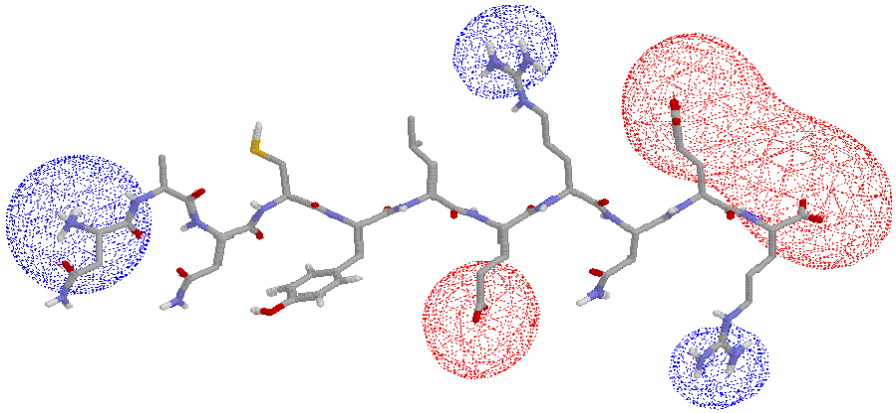
B) STEWARTRAHR (same orientation as in Figure 9B).



C) RANDALKIRK (same orientation as in Figure 9C).



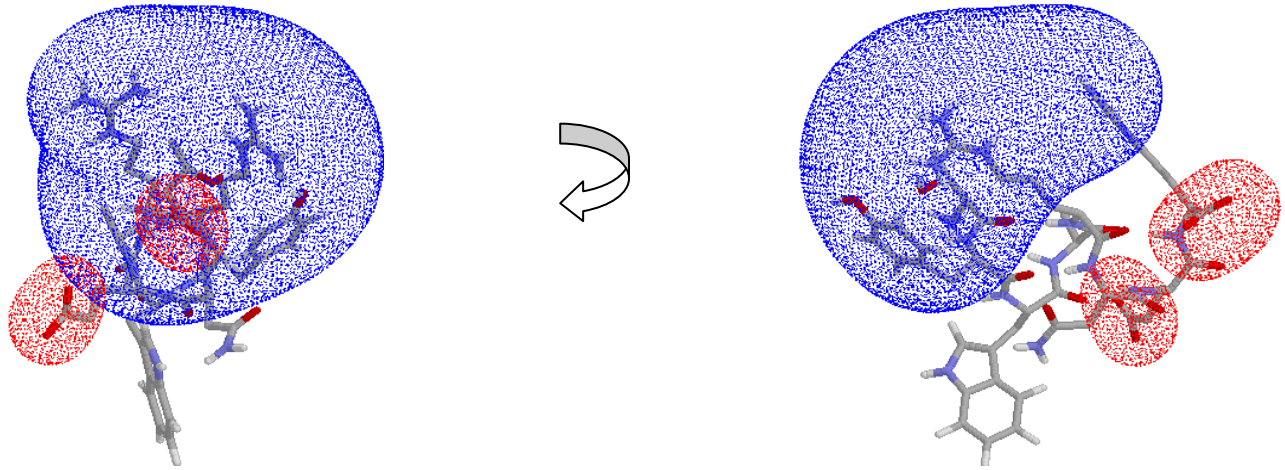
D) NANCYLERNER (same orientation as in Figure 9D).



Wade Research Foundation Reports (2010) 5 (2)

Figure 12. Electrostatic potential models of the four name peptides in α-helical conformations, and in the same orientations shown in Figure 10. The blue and red areas represent regions of positive and negative electrostatic potential, respectively.

A) TYWARNER (same orientations as in Figure 10A).



B) STEWARTRAHR (same orientations as in Figure 10B).

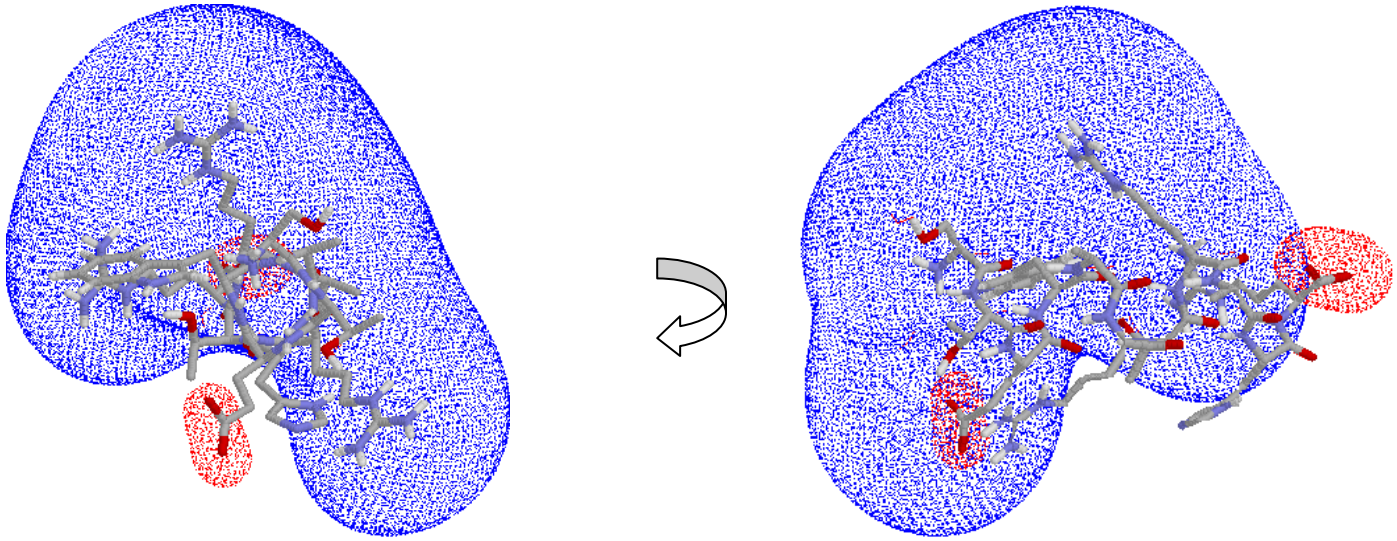
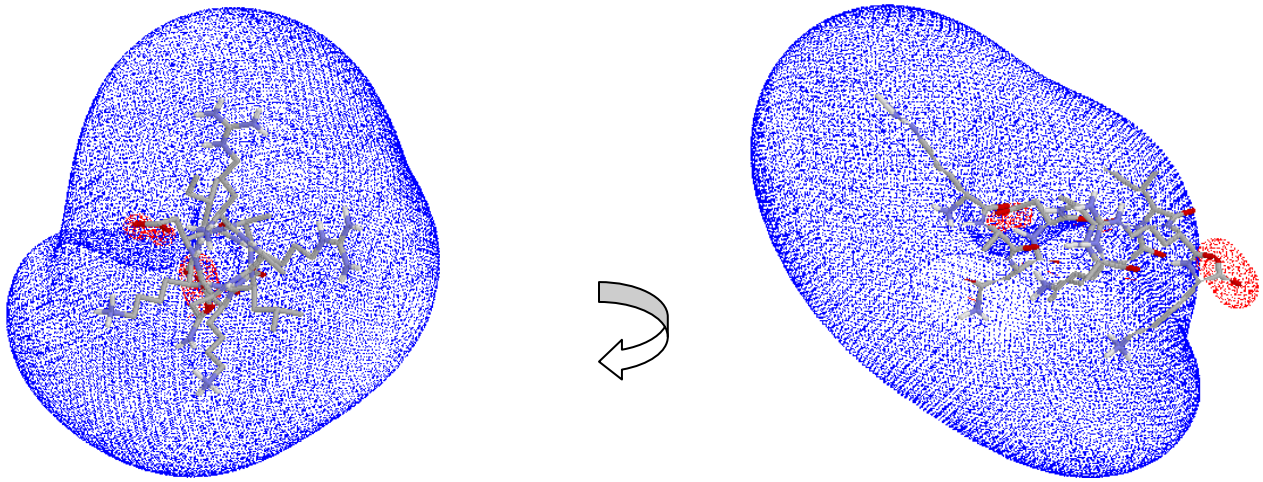


Figure 12 (Continued).

C) RANDALKIRK (same orientations as in Figure 10C).



D) NANCYLERNER (same orientations as in Figure 10D).

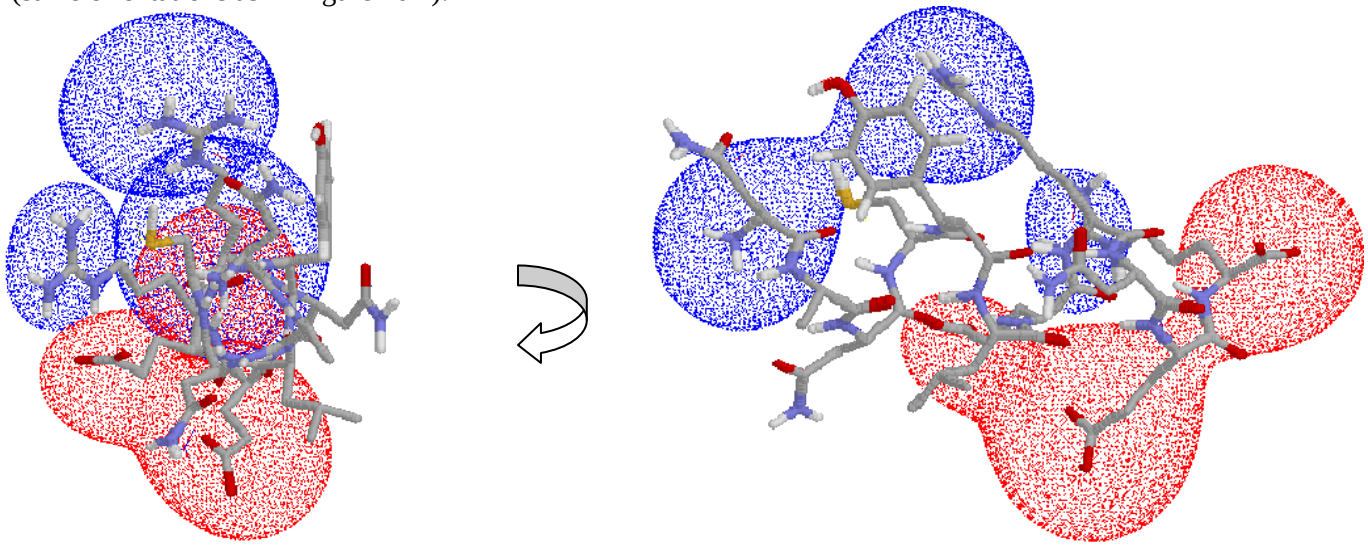


Table 9. Additional names on the Forbes List of the World's Billionaires that contain potential CendR motifs. These names contain only gene encoded AAs, except for the additional presence of Ornithine. AA abbreviations are the same as those used in Table 1, except O represents Ornithine.

Name on Forbes List	Rank on 2010 List		AA sequence of name peptide	Number of AAs:	Est. NC at pH 7
	World	US			
Theodore Le**r**ne**r**	316	49	THEODORELE**R**NE**R**	14	0
Arturo Mo**r**en**o**	937	69	ARTUROMO**R**EN**O**	12	+5
Norma Le**r**ne**r**	937	69	NORMALE**R**NE**R**	11	+2
Randolph Le**r**ne**r**	937	69	RANDOLPHLE**R**NE**R**	14	+1

Table 10. Functional groups of name peptides that can be chemically modified to add molecules to the name peptides. The CendR portions of the peptides were excluded from consideration, and the types and numbers of each functional group available for derivatization are indicated by Xs.

Name Peptide:		Functional Group:					
Other Part	CendR	N-term. $-NH_2$	$-NH-C(NH)-NH_2$	-OH	Indole N	-COOH	-SH
TYWA-	**R**NE**R**	X		XX	X		
STEWART-	**R**AH**R**	X	X	XXX	X	X	
RANDAL-	**K**IR**K**	X	X			X	
NANCYLE-	**R**NE**R**	X		X		X	X

Table 11. Peptide NANCYLERNER has the ability to form dimeric compounds through covalent, disulfide (-S-S-) bonds with other molecules containing Cysteine (C), such as homodimeric compounds composed of two NANCYLERNER, peptides, or other molecules containing thiol groups (-SH). If that other molecule were, for example, a drug, then the peptide NANCYLERNER could act as a carrier to deliver the drug into cells. In the table below, covalent, disulfide bonds are represented by vertical lines, and "Etc." indicates some molecule other than NANCYLERNER that contains a thiol group (e.g., a drug).

Peptide	Homodimer	Mixed Dimer
NANCYLERNER	NANCYLERNER | NANCYLERNER	NANCYLERNER | Etc.

(Continued from page 15.)

in the side chain of Aspartic acid (D), in RANDALKIRK; the thiol/sulfhydryl group (-SH) in the side chain of Cysteine (C) of NANCYLERNER (Tables 10). An example of the derivatization of peptide NANCYLERNER, as a homo- or mixed dimer, is shown in Table 11.

Acknowledgement

The author thanks a reviewer, who shall remain anonymous, for reading the article and providing constructive comments.

References

Note: Wade Research Foundation Reports are indexed in the EBSCO Academic Search™ Complete database (http://www.ebscohost.com/titleLists/a9h-journals.htm), and are available for a fee from EBSCO. They are also available free of charge on the website of the Wade Research Foundation

(www.wade-research.com), or by sending an e-mail request to: wade-research(at)hotmail.com. Please include "Reprint Request" in the subject line.

1. Nelson, D.L., and Cox, M.M., Lehninger Principles of Biochemistry, 5th edition, W.H. Freeman and Co., New York, 2008, pp. 71-140, 682-684, 878-881.
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, Wade S. The name game: use of words composed of letters of the English alphabet as a source of novel bioactive peptides. Biopolymers Peptide Science (2003) 71: 322 (abstract P082).
5. 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 Chemical Society, Cardiff, CA, USA, 2004, pp. 580-581.
6. Lewand, R.E. (2000). Cryptological Mathematics. The Mathematical Association of America, 2000, p. 36.
7. Wade, D., Yang, D., and Lea, M.A. Biological and structural properties of COLINPOWELL, a synthetic peptide amide. Wade Research Foundation Reports (2004) 1: 2-35. (http://www.wade-research.com/images/COLINPOWELL_10-25-04_.pdf)
8. Teesalu, T., Sugahara, K.N., Kotamraju, V.R., and Ruoslahti, E. C-end rule peptides mediate neuropilin-1-depenent cell, vascular, and tissue penetration. Proceedings of the National Academy of Sciences USA (2009) 106: 16157–16162.
9. Sugahara, K.N., Teesalu, T., Karmali, P.P., Kotamraju, V.R., Agemy, L., Girard, O.M., Hanahan, D., Mattrey, R.F., Ruoslahti, E. Tissue-penetrating delivery of compounds and nanoparticles into tumors. Cancer Cell (2009) 16: 510–520.
10. Niculescu-Duvaz, D., Springer, C., and Marais, R. C-end your drugs using peptide tags. Pigment Cell Melanoma Research (2010) in press.
11. Feron, O. Tumor-penetrating peptides: a shift from magic bullets to magic guns. Science Translational Medicine (2010) 2(34): 1-5.
12. Kroll, L., and Miller. M. The world's billionaires. Forbes.com, March 10, 2010. (http://www.forbes.com/2010/03/09/worlds-richest-people-slim-gates-buffett-billionaires-2010-intro.html)
13. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. Basic local alignment search tool. Journal of Molecular Biology (1990) 215 (3): 403–410.
14. Personal communication with E. Ruoslahti, on January 26, 2010.

Published June 22, 2010