Wade Research Foundation Reports (2020) 9(1)

Election 2020 USA: Peptides MICHAELPENCE and KAMALAHARRIS

David Wade, Wade Research Foundation, Princeton, N.J., USA

Abstract

The 2020 USA Presidential election campaign features four major political party candidates for President and Vice President: Donald Trump and Michael Pence of the Republican party, and Joseph Biden and Kamala Harris of the Democratic party. Hypothetical peptides, MICHAELPENCE and KAMALAHARRIS, were designed and subjected to theoretical analyses. If these peptides were created by chemical synthesis or molecular biological techniques, they could be expected to exhibit biological activities.

Introduction

Peptides and proteins are polymers of amino acid (AA) subunits, and can be illustrated as "beads on a string", where the "beads" represent AAs and the "string" represents the chemical bonds that link successive AAs (Figure 1) [1]. If the number of "beads" (AAs) on the "string" is less than 100, then the polymer is called a "peptide". If the number of "beads" (AAs) on the "string" is more than 100, then the polymer is called a "protein".

Figure 1. "Beads on a string" representation of an amino acid (AA) polymer. The symbol, "O", represents an AA and "-" represents the chemical bond linking successive AAs.

etc.-O-O-O-O-O-O-O-O-O-O-O-O-O-O-O-O-O-O-O-O-O-etc.
Peptide < 100 AAs > Protein

There are 20-21 AAs that are commonly found in proteins (Table 1), and chemists have given them common or trivial names. The International Union of Pure and Applied Chemistry-International Union of Biochemistry and Molecular Biology (IUPAC-IUB), Joint Commission on Biochemical Nomenclature formalized this naming process by assigning two additional types of names to these AAs: systematic names that are long and complex, and 3- and 1-letter symbols. The 3- and 1-letter symbols are constructed entirely from letters of the 26-letter English alphabet. The 1-letter symbols include all letters of the English alphabet except J and O, which have not been assigned to any AA. The letters, B, X, and Z, each have multiple AA assignments (i.e., are ambiguous).

The AA sequences of peptides and proteins are commonly specified using the 1-letter symbols shown in Table 1. One reason for this choice is that it enables sequence information to be stored more efficiently. For example, the National Center for Biotechnology Information (NCBI) stores AA sequence information on 314,303,451 proteins (as of 9/10/20) in its computer database in the form of 1-letter symbols, and it can be searched using strings of 1-letter symbols, except "U" (e.g., searching with the BLAST algorithm [2]).

Since AA sequence information is stored as sequences of English alphabet letters, it is occasionally possible to find AA sequences whose single letter symbols correspond to names or words. For example, a search of the NCBI non-redundant protein database

Table 1. Names of the AAs incorporated into protein under mRNA direction. The letters J and O have no AA assignments, and the letters B, X, and Z have been omitted because they each have multiple AA assignments (i.e., are ambiguous) [1].

Trivial Name	Systematic name	3-Letter Symbol	**1-Letter Symbol**
Alanine	2-Aminopropanoic acid	Ala	**A**
Cysteine	2-Amino-3-mercaptopropanoic acid	Cys	**C**
Aspartic acid	2-Aminobutanedioic acid	Asp	**D**
Glutamic acid	2-Aminopentanedioic acid	Glu	**E**
Phenylalanine	2-Amino-3-phenylpropanoic acid	Phe	**F**
Glycine	Aminoethanoic acid	Gly	**G**
Histidine	2-Amino-3-(IH-imidazol-4-yl)propanoic acid	His	**H**
Isoleucine	2-Amino-3-methylpentanoic acid	Ile	**I**
Lysine	2,6-Diaminohexanoic acid	Lys	**K**
Leucine	2-Amino-4-methylpentanoic acid	Leu	**L**
Methionine	2-Amino-4-(methylthio)butanoic acid	Met	**M**
Asparagine	2-Amino-3-carbamoylpropanoic acid	Asn	**N**
Proline	Pyrrolidine-2-carboxylic acid	Pro	**P**
Glutamine	2-Amino-4-carbamoylbutanoic acid	Gln	**Q**
Arginine	2-Amino-5-guanidinopentanoic acid	Arg	**R**
Serine	2-Amino-3-hydroxypropanoic acid	Ser	**S**
Threonine	2-Amino-3-hydroxybutanoic acid	Thr	**T**
Selenocysteine	2-Amino-3-selanylpropanoic acid	Sec	**U**
Valine	2-Amino-3-methylbutanoic acid	Val	**V**
Tryptophan	2-Amino-3-(IH-indol-3-yl)propanoic acid	Trp	**W**
Tyrosine	2-Amino-3-(4-hydroxyphenyl)propanoic acid	Tyr	**Y**

for the AA sequence, Tryptophan-Alanine-Leucine-Methionine-Alanine-Arginine-Threonine, was done using the BLAST program, and the single letter symbols for the names of the AAs [3]:

AA sequence: Tryptophan-Alanine-Leucine-Methionine-Alanine-Arginine-Threonine

↓

1-letter symbols: W-A-L-M-A-R-T

↓

BLAST search sequence: WALMART

The BLAST search result shows that the AA sequence WALMART, which corresponds to the name of a well-known company, occurs in the NCBI database 438 times (as of 9/10/20) For example, it occurs in the AA sequence that comprises the enzyme, acyl-CoA dehydrogenase of the bacterium, *Pseudomonas* sp. AAs 71-77, of the 128 AA-long sequence (Sequence ID HCL40313.1), are etc.-WALMART-etc. The BLAST process can be used to find any sequence of English alphabet letters, except B, J, O, U, X, or Z, within the hundreds of millions of AA sequences in the protein databases.

Most peptides and proteins have been isolated from natural sources, but the IUPAC-IUB single letter symbol nomenclature for AAs, combined with the methodologies of chemical synthesis and molecular biology, make it possible to design and create novel peptides and proteins that correspond to almost any sequence of letters except those containing the letters B, J, O, X, and Z. For example, the peptide, WALMART, was created using chemical synthesis, tested in the laboratory, and found to exhibit both antimicrobial and anticancer properties [3].

The fact that sequences of English alphabet letters also correspond to sequences of AA symbols makes it easy to teach nonscientists about peptides and proteins. Most members of the general public (e.g., youngsters) have heard and/or seen the word, “protein”, and they may have some idea that protein is an important component of the diet, but they do not know exactly what proteins are. If it is shown to them that the sequence of letters in their name can correspond to the AA sequence of a peptide/protein, then the topic of proteins becomes more interesting to them, and they may be more receptive to learn more about proteins.

Methods and Results

Peptides designed using the names of 2020 US Vice Presidential election candidates:

The 2020 US Presidential election campaign features four major political party candidates for President and Vice President: Donald Trump and Michael Pence of the Republican party, and Joseph Biden and Kamala Harris of the Democratic party. If the letters in these names are considered as IUPAC-IUB single letter symbols for the names of AAs, it would not be possible to form a peptide from AAs symbolized by the sequences of letters in either of the names of the Presidential candidates, Donald Trump and Joseph Biden. The letters, O and J, have not been assigned to any AA by the IUPAC-IUB, and the letter, B, is an ambiguous symbol representing more than one AA. However, all of the letters comprising the names of both Vice Presidential candidates are unambiguous single letter symbols for the names of AAs in the IUPAC-IUB system, and it would be possible to form peptides from the AAs symbolized by the letters in the names of Michael Pence and Kamala Harris. Therefore, hypothetical peptides, MICHAELPENCE and KAMALAHARRIS, were designed and subjected to theoretical analyses.

Predicted properties of the name peptides, MICHAELPENCE and KAMALAHARRIS:

The ProtParam Tool, of the ExPASy molecular biology server [4] was used to predict several physical and biological properties of the hypothetical name peptides, MICHAELPENCE and KAMALAHARRIS. The results are shown in Table 2. Both peptides have the same lengths (12 AAs) and similar molecular weights (1,387-1,389 vs. 1,325). Interestingly, their net charges at pH 7 are identical in magnitude but opposite in sign (-3 vs. +3). The theoretical pI values, or the pH values at which the peptides have no net charge, are as expected. The aliphatic indices, the relative volume occupied by aliphatic side chains of the AAs, Alanine (A), Valine (V), Isoleucine (I), and Leucine (L) were somewhat similar (73.33 vs. 98.33). The grand average of hydropathicity values showed peptide MICHAELPENCE to be slightly less hydropathic (hydrophobic) than

Table 2. Predicted physical properties of the name peptides, MICHAELPENCE and KAMALAHARRIS.

Property	MICHAELPENCE	KAMALAHARRIS
1) Number of AAs	12	12
2) Molecular weight	1388.59 (open chain); 1386.59 (cyclic disulfide)	1324.61
3) Net charge at pH 7	-3	+3
4) Theoretical pI	4.24	12.01
5) Estimated half-life in: a) mammalian reticulocytes, *in vitro* b) yeast, *in vivo* c) *Escherichia coli*, *in vivo*	 30 hours >20 hours >10 hours	 1.3 hours 3 mins. 3 mins.
Instability index	64.47 (unstable)	60.11 (unstable)
Aliphatic index	73.33	98.33
Grand average of hydropathicity	-0.150	0.042

peptide KAMALAHARRIS (-0.150 vs. 0.042). Using the Kyte and Doolittle scale of AA hydropathicity (hydrophobicity) [5], the percent of hydrophobic AAs in MICHAELPENCE is 50% and in KAMALAHARRIS is 58%. Based on their predicted charge and hydrophobicity characteristics, both peptides are predicted to be water soluble.

Predictions of peptide stability, or half-life, are not consistent. The instability index, which provides an estimate of the stability of a protein in a test tube, suggests that both peptides are unstable. However, predictions of the half-life of both peptides, both *in vitro* and *in vivo*, suggest that peptide MICHAELPENCE would be more stable than peptide KAMALAHARRIS.

Do the name peptides, MICHAELPENCE and KAMALAHARRIS, occur in nature?

The NCBI non-redundant protein sequences database contains the AA sequences of 314,303,451 proteins (as of 9/10/20). This database was searched with the Basic Local Alignment Search Tool (BLAST) algorithm [2] for the occurrence of the AA sequences MICHAELPENCE and KAMALAHARRIS. The results are shown in Tables 3 and 4. Neither peptide was found in its entirety, but all parts of both peptides were found.

Molecular models of the name peptides, MICHAELPENCE and KAMALAHARRIS:

Molecular models of the hypothetical name peptides, MICHAELPENCE and KAMALAHARRIS, were created using the Deep View-Swiss Pdb Viewer, v.4.1 program [6]. Each peptide was modeled in two extremes of conformation that are commonly found in proteins, as a β-strand (a flat, ribbon-like structure) and as an α-helix (a coiled structure with an overall cylindrical shape) (Figure 2) [1].

In addition to β-strand and α-helical conformations, peptide MICHAELPENCE contains two Cysteine (C) AAs which enables it to cyclize via a bond between sulfur atoms in the two Cs (below and Figure 3). The cyclic structure is the most realistic form of the peptide in a water environment, at pH 7, and in an oxygen atmosphere.

MICHAELPENCE ➔ MICHAELPENCE

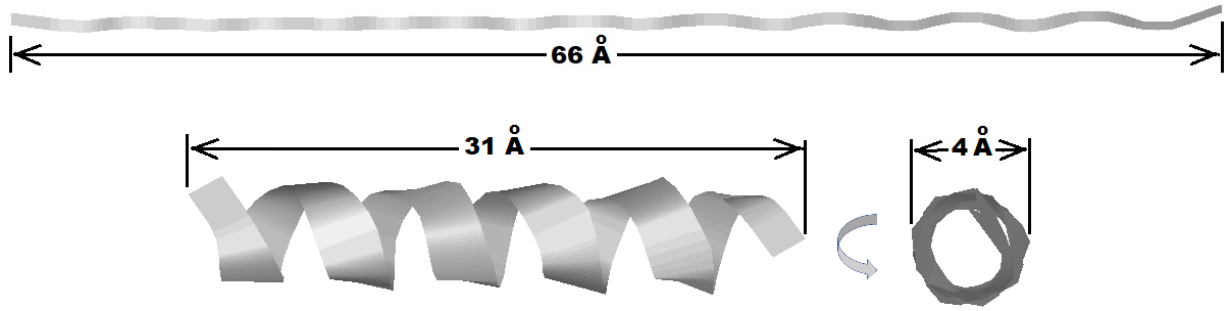
Table 3. Examples of the results of a BLAST search of the NCBI non-redundant protein database for the AA sequence, MICHAELPENCE.

Sequence Found:	Percent Identity:	Organism:	Protein:	Location in Protein:	Sequence ID:
MI_HAELPE__E	75%	*Nitrososphaeria archaeon*	Ubiquitin-conjugating enzyme E2	1-9/161	HHO64024.1
MICH_ELP____	58%	*Acidobacteria bacterium*	Lipoyl(octanoyl) transferase LipB	8-15/242	HHR98019.1
MICHA__PE___	58%	*Xanthobacter autotrophicus*	AMP-binding protein	133-141/ 556	WP_012114797.1
___H_E_PENCE	58%	*Oceanicoccus* sp. KOV_DT_Chl	Pyruvate dehydrogenase (acetyl-transferring), homodimeric type	692-700/ 882	WP_101757466.1
____A_LPENCE	58%	*Gammaproteobacteria bacterium*	CinA family nicotinamide mononucleotide deamidase-related protein	118-125/ 428	NNC98938.1

Table 4. Examples of the results of a BLAST search of the NCBI non-redundant protein database for the AA sequence, KAMALAHARRIS.

Sequence Found:	Percent Identity:	Organism:	Protein:	Location in Protein:	Sequence ID:
KAMALAHAR____	75%	*Betaproteobacteria bacterium*	SDR family oxidoreductase	167-175/ 267	TAK80783.1
__MALAHARR__	67%	*Armatimonadetes bacterium*	Pilus assembly protein PilM	326-333/ 535	NLC55439.1
_____AHARRIS	58%	*Sandarakinorhabdus limnophila*	Uroporphyrinogen-III C-methyltransferase	53-59/ 256	WP_022680757.1

Figure 2. Ribbon diagrams of the two extremes of conformation commonly found in proteins. The model shown is poly-Glycine, containing 20 Glycine (G) AAs. Top, β-strand. Bottom, α-helix. The α-helical structure is stabilized by several intrachain hydrogen bonds (16 for this particular model; not shown). Notice that the length of the α-helix is less than half of the length of the β-strand. [Note: Ångström (Å) = 1.0 x 10^{-10} meters.] The models were produced with the Deep View-Swiss Pdb Viewer, v.4.1, program [6], viewed with the RasMol program [7], and transferred to the MS Paint program for the addition of numbers and letters.



Peptide KAMALAHARRIS (Figure 4) would most likely adopt the α-helix structure in a water environment at pH 7, and when in contact with other macromolecules, as it is stabilized by several (9) intrachain hydrogen bonds [1].

A comparison of the molecular sizes of the MICHAELPENCE and KAMALAHARRIS peptide models, in the β-strand and α-helix conformations, shows that, like their molecular weights, their molecular sizes are similar. They are 39-40 Å in length as β-strands, and 21-22 Å in length as α-helices.

Electrostatic potential models of name peptides MICHAELPENCE and KAMALAHARRIS:

Electrostatic potential models of peptides provide information about charged regions within the peptide that may be important in influencing electrostatic interactions within the peptide and with other molecules. Oppositely charged regions would attract each other (+ ➔🡐-), and similarly charged regions would repel each other (+ 🡐➔ + and - 🡐➔ -). Figures 5 and 6 show electrostatic potential models of peptides MICHAELPENCE and KAMALAHARRIS.

Interestingly, the electrostatic potential model for peptide MICHAELPENCE is mostly a red color, indicating a high amount of negative charge. In US politics, the color, red, is associated with the Republican Party. Also, the electrostatic potential model for peptide KAMALAHARRIS is mostly a blue color, indicating a high amount of positive charge, and in US politics the color, blue, is associated with the Democratic Party. However, it is just coincidental that the electrostatic potential model colors correspond to the colors associated with the political parties of the candidates whose names were used as the basis for creating the two hypothetical peptides (Figure 7).

Figure 3. Stick figure models of peptide, MICHAELPENCE, in the β-strand (top), α-helical (middle), and cyclic, disulfide-bonded (bottom) conformations. The cyclic structure is the most realistic form of the peptide in a water environment, at pH 7, and in an oxygen atmosphere. Models were created with the Deep View-Swiss Pdb Viewer, v. 4.1, program, energy minimized with the same program, visualized with the RasMol program [7], and transferred to the MS Paint program for the addition of IUPAC-IUB letters, numbers, and charge symbols. The atom color scheme is gray for carbon, red for oxygen, blue for nitrogen, yellow for sulfur, and white for hydrogen. Methyl (-CH_3) and methylene (-CH_2-) hydrogens are not shown.

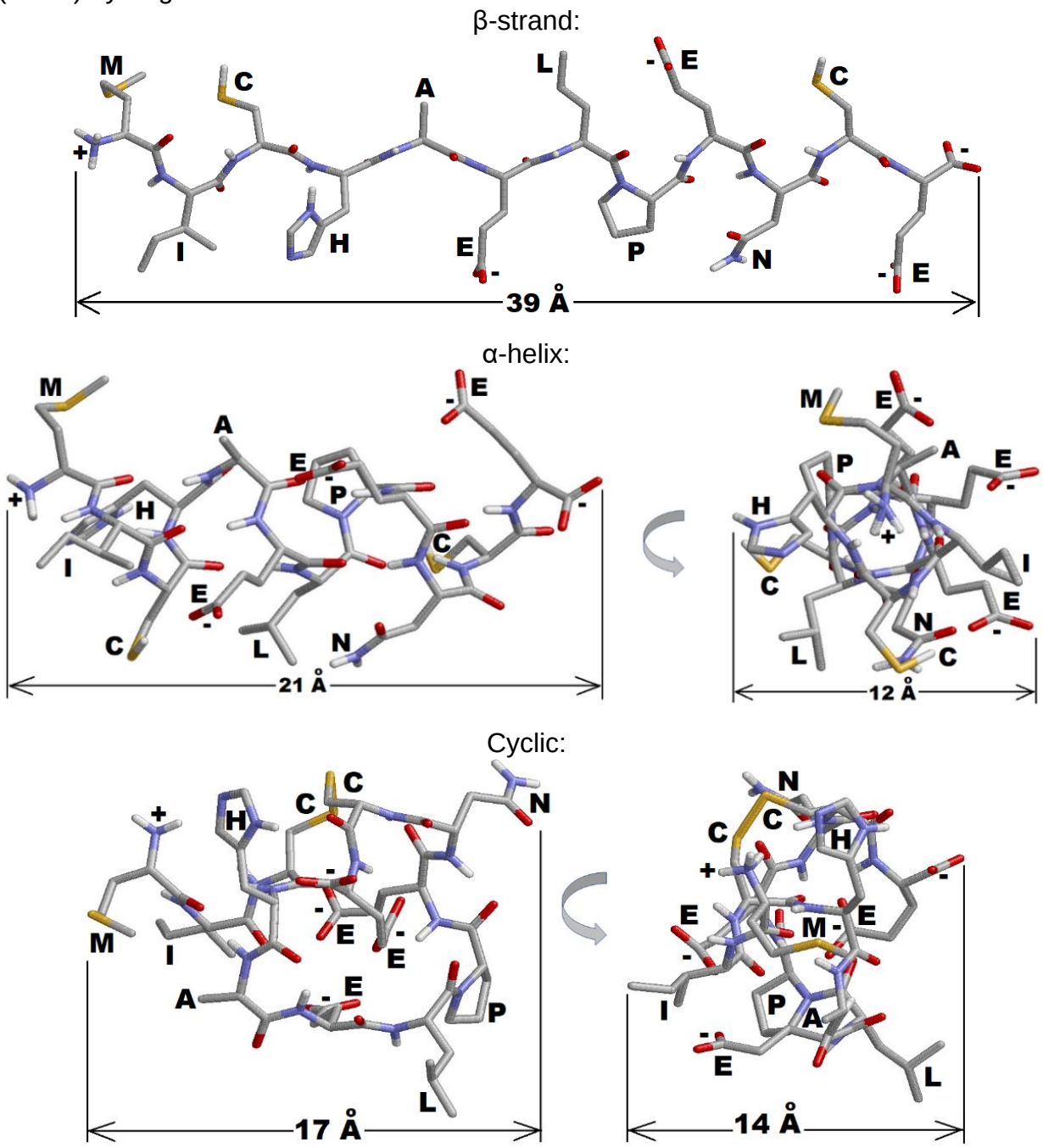
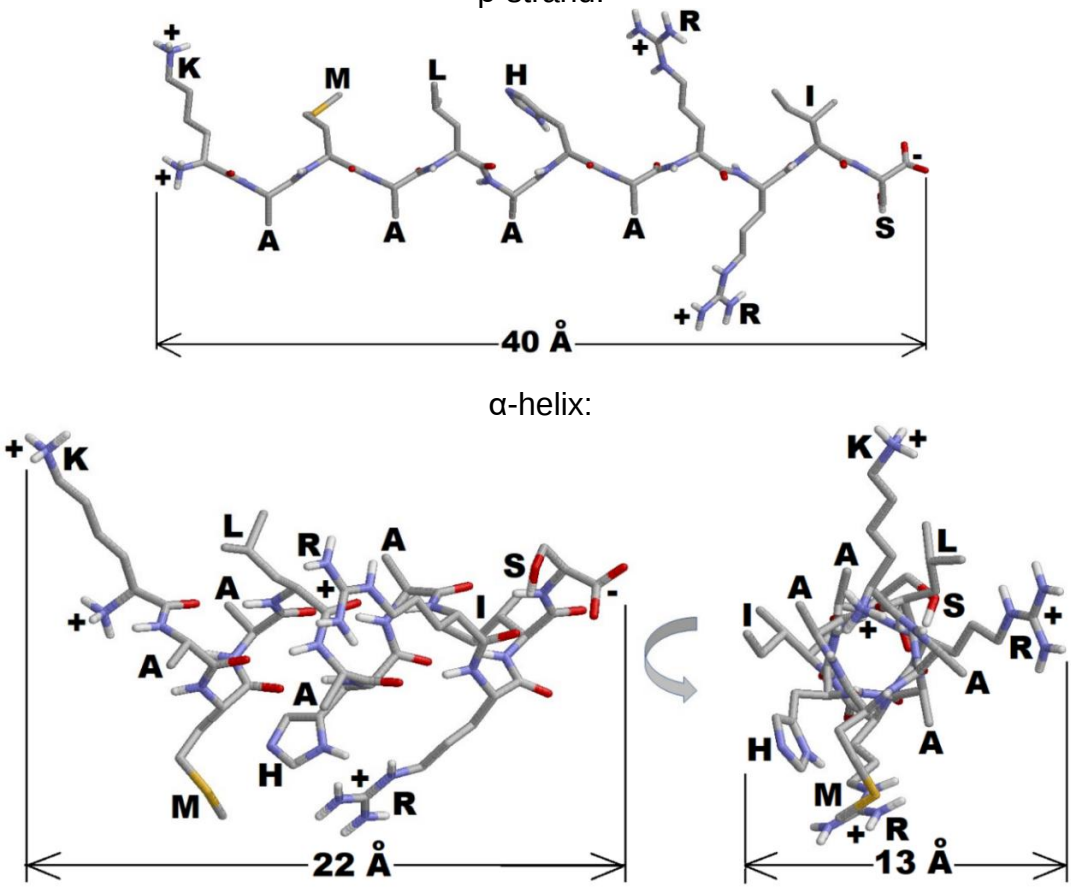


Figure 4. Stick figure models of peptide KAMALAHARRIS in the β-strand (top) and α-helical (bottom) conformations. The α-helical conformation is stabilized by 9 intrachain hydrogen bonds (not shown) and is the most likely structure in a water environment, at pH 7, and when in contact with other molecules. See the legend of Figure 3 for additional details.



Docking of the name peptides, MICHAELPENCE and KAMALAHARRIS:

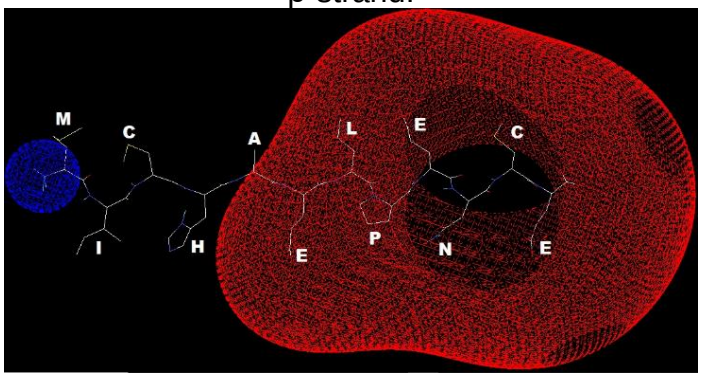
Since the name peptides MICHAELPENCE and KAMALAHARRIS are predicted to have equal and opposite net charges at pH 7, they would be expected to attract each other (+➔🡐-) when in proximity. The most probable structures of both peptides in an aqueous environment at pH 7, cyclic for peptide MICHAELPENCE and α-helix for KAMALAHARRIS, were docked with the ClusPro 2.0 docking server [8]. The server generates multiple models, and an example of a result is shown in Figure 8.

Target Prediction:

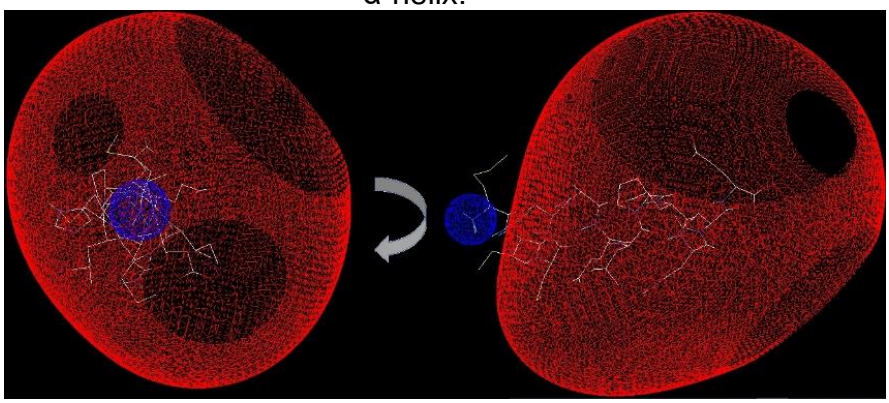
In addition to interacting with each other, it could be expected that peptides MICHAELPENCE and KAMALAHARRIS would interact with numerous other molecules, including cellular proteins. Both peptides were analyzed by a program, SwissTargetPrediction [9], that predicts potential cellular protein targets of molecules. The results for potential human targets are shown in Table 5, and there are differences in the human proteins predicted to be targets of the two peptides.

Figure 5. Electrostatic potential models of peptide MICHAELPENCE in the β-strand (top), α-helix (middle), and cyclic (bottom) conformations. The cyclic model is the most realistic structure of the peptide in a water environment, at pH 7, and in an oxygen atmosphere. The models were created as wireframe structures with the Deep View-Swiss Pdb Viewer, v.4.1, program, and viewed with the same program. Letters were added with the MS Paint program to indicate the location of each AA. Bonds between atoms are visible as thin white lines. Regions of positive and negative electrostatic potential are shown in blue and red, respectively. A black background is used to highlight both the bonds between atoms and the regions of negative and positive potential.

β-strand:



α-helix:



Cyclic:

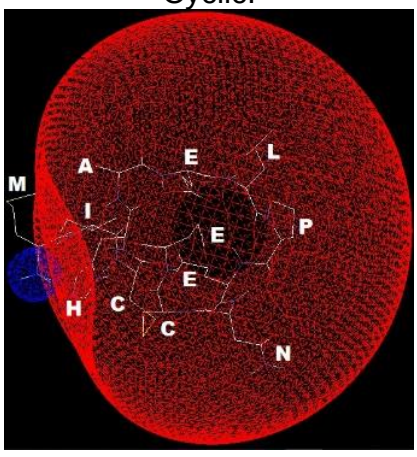
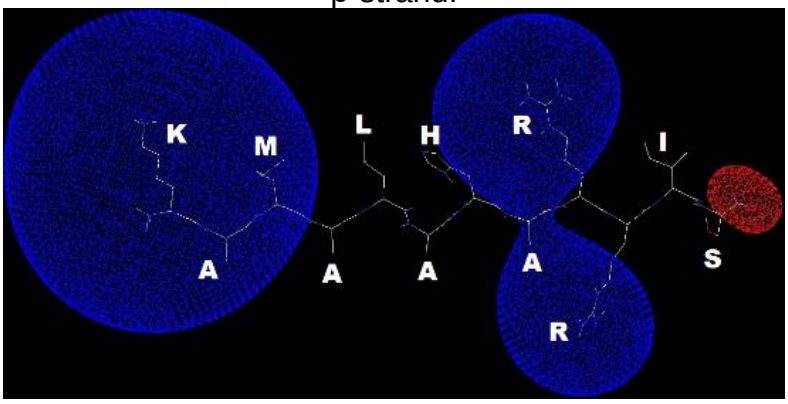


Figure 6. Electrostatic potential models of peptide KAMALAHARRIS in the β-strand (top) and α-helix (bottom) conformations. The α-helix is the most realistic structure when the peptide is in a water environment, at pH 7, and when in contact with other macromolecules. See the legend to Figure 5 for additional details.

β-strand:



α-helix:

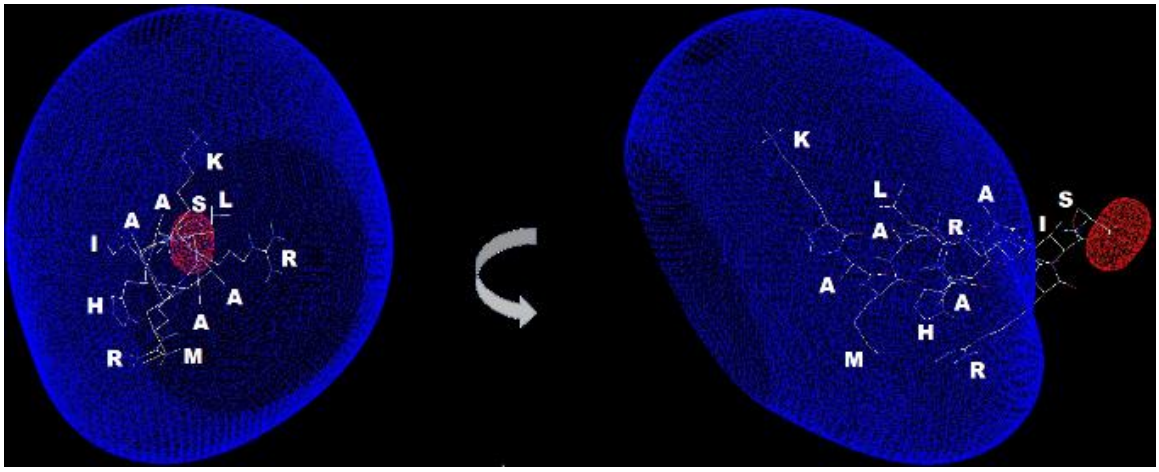


Figure 7. Images of the two Vice-Presidential candidates, Kamala Harris and Michael Pence, superimposed upon electrostatic potential (EP) diagrams for their "name peptides". Blue is used to indicate both positive EP and the Democratic Party, and red is used to indicate both negative EP and the Republican Party.

Figure 8. Example of a result of docking of the cyclic form of peptide MICHAELPENCE with the α-helical form of peptide KAMALAHARRIS using the ClusPro 2.0 server [8]. (Top) Stick figure model of the docked peptides, showing the location of each AA in both peptides. (Bottom) Electrostatic potential model of the two docked peptides in the same orientations shown in the stick figure models at the top part of the figure.

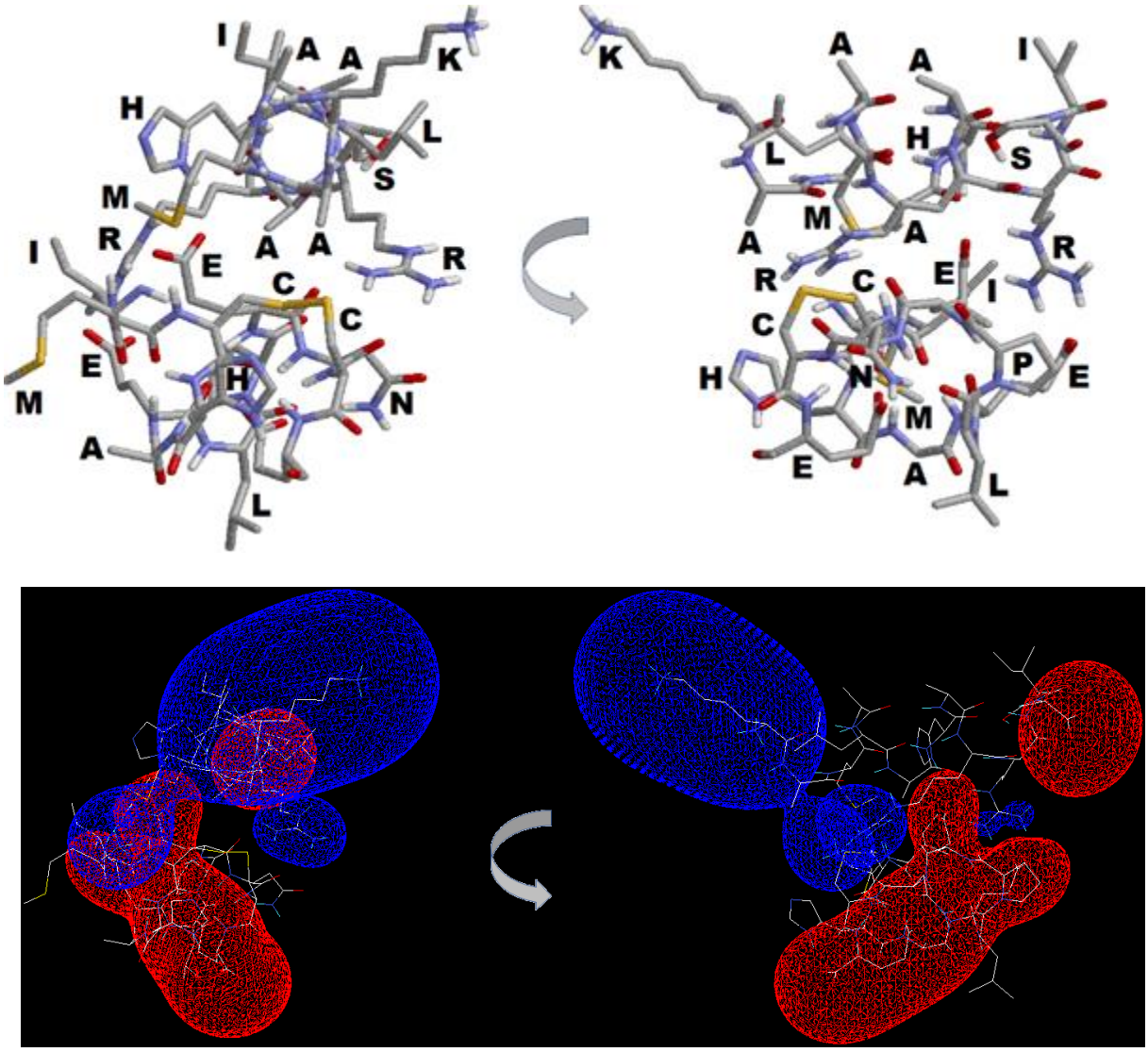


Table 5. Results of SwissTargetPrediction analyses of peptides MICHAELPENCE and KAMALAHARRIS [9]. The peptides are assumed to be bioactive, and the prediction is based on the similarity of the peptides with a library of compounds that are known to be active on human proteins. The result is expressed as a Percent Probability, which is the probability that the corresponding human protein is a target of the query peptide.

MICHAELPENCE		KAMALAHARRIS	
Target Classes	Percent Probability	Target Classes	Percent Probability
Protease	26.7	Family A G protein-coupled receptor	40.0
Membrane receptor	20.0	Eraser	20.0
Eraser	20.0	Protease	13.3
Surface antigen	13.3	Membrane receptor	6.7
Family A G protein-coupled receptor	6.7	Surface antigen	6.7
Kinase	6.7	Kinase	6.7
Unclassified protein	6.7	Unclassified protein	6.7
Total	100.1	Total	100.1

Summary and Conclusions

The names of the Vice-Presidential candidates of the two major US political parties were used as the basis for designing two hypothetical name peptides: MICHAELPENCE and KAMALAHARRIS. The peptides were then subjected to theoretical analyses and found to be similar in molecular weights and sizes. They also had net charges at pH 7 that were equal in magnitude, but opposite in sign. Based on their net charges at pH 7 and the hydrophobic characteristics of their constituent AAs, both hypothetical peptides were predicted to be water soluble.

Although the peptides were not found to occur in their entirety in a protein database, all parts of the peptides' AA sequences were found to occur within the database. Therefore, the peptides' AA sequences would not be incompatible with nature.

Molecular models of both peptides were constructed, and a docking program was used to illustrate how the oppositely charge peptides might interact. Finally, potential human molecular targets of both peptides were predicted. Based on the theoretical analyses of these two hypothetical peptides, it is predicted that they would exhibit biological properties, and that these properties may be of a useful nature.

References:

1. Wade, D. The Peptide Workbook, Wade Research Foundation Reports (2019) 8(1): 1-47. (http://www.wade-research.mynetworksolutions.com/images/Peptide_ Workbook_020919.pdf)
2. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. (1990) 215: 403-410. (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins)

3. Wade, D., and Lea, M.A. The WALMART peptide. Wade Res. Found. Rep. (2015) 7(1): 1-13. (http://www.wade-research.mynetworksolutions.com/images/WRF_Reports__2015_7_1_. pdf)
4. ProtParam Tool of ExPASY, the Swiss Institute of Bioinformatics Resource portal (https://web.expasy.org/protparam/)
5. Kyte, J. and Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. (1982) 157(1): 105-132.
6. Guex, N. and Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis (1997) 18: 2714-2723. (https://spdbv.vital-it.ch)
7. Sayle, R. and Milner-White, E.J. RasMol: Biomolecular graphics for all. Trends in Biochemical Sciences (1995) 20 (9): 374-376. (http://www.openrasmol.org)
8. Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., and Vajda, S. The ClusPro web server for protein-protein docking. Nature Protocols (2017) 12(2): 255-278. (https://cluspro.org/login.php)
9. Gfeller, D., Michielin, O., and Zoete, V. Shaping the interaction landscape of bioactive molecules. Bioinformatics (2013) 29(23): 3073-3079. (http://www.swisstargetprediction.ch)

(Published online September 15, 2020.)