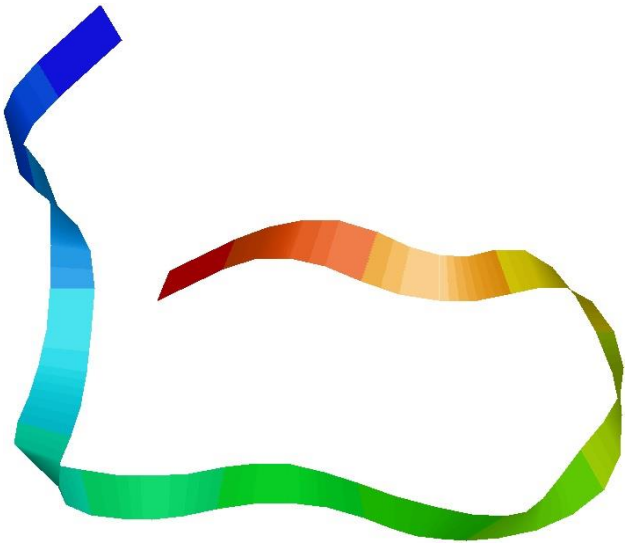
Wade Research Foundation Reports (2019) 8(1)

THE PEPTIDE WORKBOOK



David Wade, Ph.D.
Wade Research Foundation
(wade-research@hotmail.com)

Wade Research Foundation Reports (2019) 8(1)

This work is dedicated to the memory of R. Bruce Merrifield,
1984 Nobel Laureate in Chemistry.

TABLE OF CONTENTS:

PROTEINS, PEPTIDES, AND AMINO ACIDS (A.A.s)

Proteins are a major component of cells. Estimates for a typical human cell are shown in Table 1 [1]:

Table 1. Chemical components of human cells.

Molecule	Water	**Protein**	Lipids	RNA	DNA
% of Cell Mass	65	**20**	12	1	0.1

Proteins and peptides are “polymers” of amino acids (A.A.s) linked by chemical bonds [2]. They can be thought of as beads on a string (Figure 1), where the beads are A.A.s and the string is chemical bonds, called “amide” or “peptide” bonds, that link successive A.A.s. The size (length) of an A.A. polymer determines if it is a peptide or protein:

“peptide” < 100 A.A.s > “protein”.

Figure 1. Peptides and proteins are analogous to beads on a string.

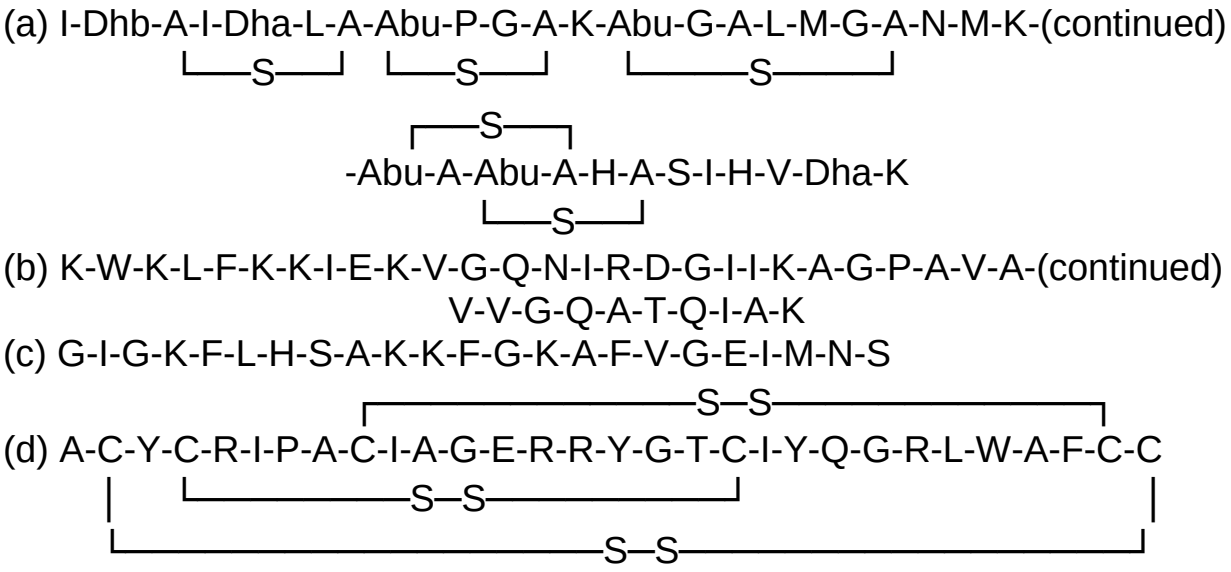
Wade Research Foundation Reports (2019) 8(1)

IMPORTANCE OF PEPTIDES AND PROTEINS

Peptides and proteins are ubiquitous in nature. All living organisms contain these types of molecules, where they have essential roles in life processes. At least 187,788,528 (as of 2/9/19) different peptides and proteins have been found in nature [3].

A major problem in 21st century medicine is the fact that many medically important microorganisms have become resistant to the antibiotics that have been used to treat microbial diseases [4]. Consequently, scientists are seeking to find new types of antibiotics, and peptides may help fulfill this need (Figure 2).

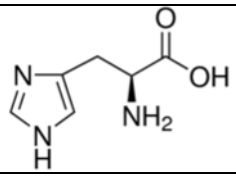
Figure 2. Examples of some peptides that kill microorganisms. The A.A. sequences of (a) Nisin A, a peptide that contains 34 A.A.s and is produced by the bacterium, *Streptococcus (Lactococcus) lactis* [4-6], (b) Cecropin A, a 37 A.A. peptide produced by larva of the silkmoth, *Hyalophora cecropia* [7], (c) Magainin 2, a 23 A.A. peptide produced by the skin of the African clawed frog, *Xenopus laevis* [8] , and (d) Defensin HNP-1, a 30 A.A. peptide produced by human neutrophils [9-11]. Single letters are symbols for the names of A.A.s, that will be described in the next section. Other abbreviations are: "-", peptide bond; Dhb, dehydrobutyrine; Dha, dehydroalanine; Abu, aminobutyric acid; A------A, lanthionine; Abu-----A, β-methyllanthionine, , " ┗─S─┛ ", thioether bond (where S = sulfur atom) and " ┗─S-S─┛ ", disulfide bond (where S = sulfur atom).



I.U.P.A.C.-I.U.B.M.B., J.C.B.N. SYMBOLS FOR THE NAMES OF A.A.s

There are 21 different types of A.A.s that occur in natural (gene encoded) proteins [2]. Each type of A.A. has been assigned a name by the International Union of Pure and Applied Chemistry–International Union of Biochemistry and Molecular Biology, Joint Commission on Biochemical Nomenclature (I.U.P.A.C.-I.U.B.M.B., J.C.B.N.) [12, 13]. Formal (systematic) or trivial names for A.A.s are long and complex. To make it easier for scientists to work with these names, the IUPAC-IUBMB, JCBN has also assigned symbols to represent the names of the A.A.s. The symbols are of two types: 3-letter symbols and 1-letter symbols. The 1-letter symbols are letters of the English alphabet. An example is shown in Table 2, and the full list is shown in Table 3 (next page).

Table 2. Example of the chemical structure and names for an A.A. [13].

	Names:		Symbols:	
Chemical structure:	Systematic:	Trivial:	3-Letter:	1-Letter:
	2-Amino-3-(1*H*-imidazol-4-yl) -propanoic acid	Histidine	His	H

Five letters of the English alphabet are not used in this workbook because they have ambiguous assignments (i.e., the letters “B”, “X”, and “Z” represent more than one A.A. each), or they have no assignments (i.e., letters “J” and “O” have not been assigned to any A.A.s) (Table 4).

Table 4. Comparison of the English alphabet and the IUPAC-IUBMB, JCBN single letter symbols for the names of A.A.s. The single letter symbols, B, J, O, X, and Z, (dark gray shading) are not included in the letters used to design peptides based on names, words, or phrases.

A	B	C	D	E
F	G	H	I	J
K	L	M	N	O
P	Q	R	S	T
U	V	W	X	Y
Z				

Table 3. Names and IUPAC-IUBMB, JCBN symbols for the 21 A.A.s [13]:

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

The IUPAC-IUBMB, JCBN single letter symbolism is used by chemists, biochemists, and molecular biologists throughout the world. It is the standard method of representing A.A. sequences in the chemical, biochemical, and molecular biology literature.

The U.S. National Center for Biotechnology Information (N.C.B.I.) has a computer database containing the A.A. sequences of hundreds of millions of proteins 187,417,574 (as of 1/30/19) [3]. These A.A. sequences are stored in N.C.B.I.'s computers as IUPAC-IUBMB, JCBN single letter symbols. In later sections of this workbook, you will design a peptide and then use this protein database to determine if the peptide has been found in nature.

Table 5 (next page) shows a list the names of States and territories of the United States. An examination of the letters in these names reveals that only 38-39% of the names on this list are composed entirely of letters that are compatible with the IUPAC-IUBMB, JCBN single letter symbols for the names of A.A.s.

Table 5. Names of US States and territories [14]. Those containing letters that are all compatible with IUPAC-IUBMB, JCBN single letter symbols for the names of A.A.s are shaded gray. Incompatible letters are B, J, O, X, Z.

STATE (19/50 = 38%):	Letter(s)	STATE (19/50 = 38%):	Letter(s)
Alabama	B	Ohio	O
Alaska		Oklahoma	O
Arizona	Z, O	Oregon	O
Arkansas		Pennsylvania	
California	O	Rhode Island	O
Colorado	O	South Carolina	O
Connecticut	O	South Dakota	O
Delaware		Tennessee	
Florida	O	Texas	X
Georgia	O	Utah	
Hawaii		Vermont	O
Idaho	O	Virginia	
Illinois	O	Washington	O
Indiana		West Virginia	
Iowa	O	Wisconsin	O
Kansas		Wyoming	O
Kentucky		FEDERAL DISTRICT (0/1=0%):	
Louisiana	O	District of Columbia	O, B
Maine			
Maryland		TERRITORY (5/13=39%):	
Massachusetts		American Samoa	O
Michigan		Guam	
Minnesota	O	Northern Mariana Islands	O
Mississippi		Puerto Rico	O
Missouri	O	U.S. Virgin Islands	
Montana	O	Baker Island	B
Nebraska	B	Howland Island	O
Nevada		Jarvis Island	J
New Hampshire		Johnston Atoll	J, O
New Jersey	J	Kingman Reef	
New Mexico	X, O	Midway Atoll	O
New York	O	Navassa Island	
North Carolina	O	Palmyra Atoll	O
North Dakota	O	Wake Island	

Wade Research Foundation Reports (2019) 8(1)

PERIODIC TABLE OF THE CHEMICAL ELEMENTS

Figure 3 shows the atomic symbols for chemical elements [15], and Table 6 shows the element types found in the 21 A.A.s to be shown in the next section. Note that four of the atomic symbols for elements [H (hydrogen), C (carbon), N (nitrogen), and S (sulfur)] are the same as the IUPAC-IUBMB, JCBN single letter symbols for the names of four A.A.s [H (Histidine), C (Cysteine), N (Asparagine), and S (Serine)] (Tables 2-4). In order to avoid confusion, an effort has been made to label each type of symbol (element or A.A.) when both are used in a description.

Figure 3. Chemical elements used to construct A.A.s [15]

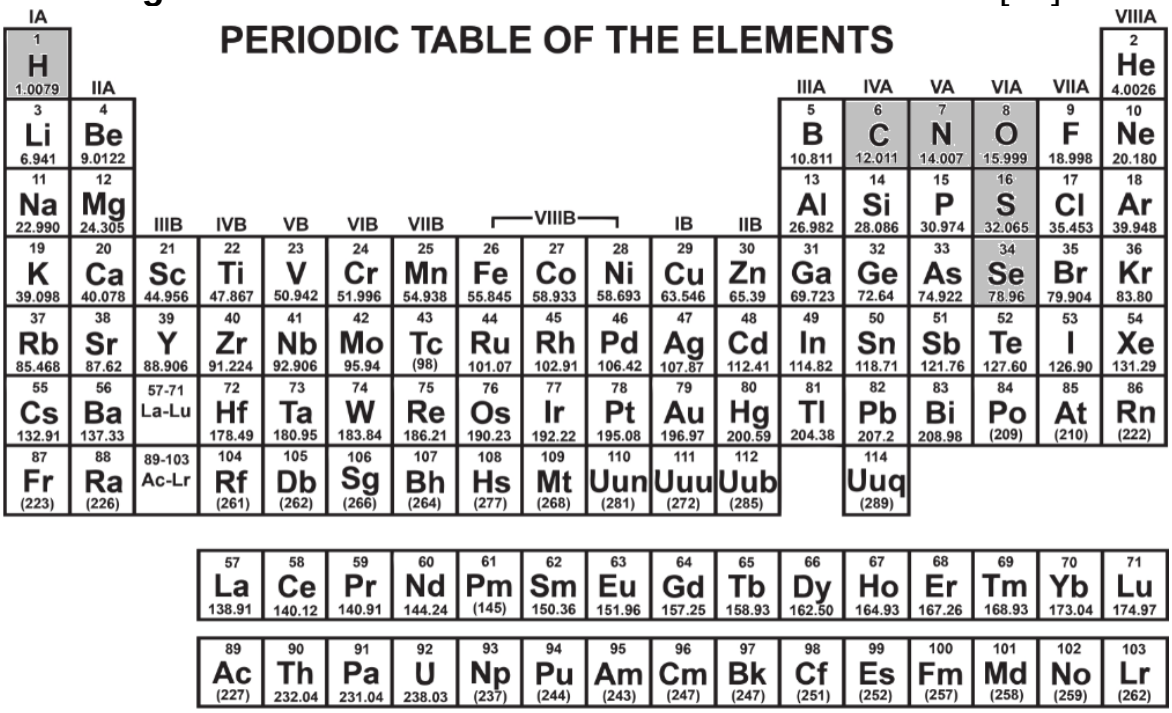


Table 6. Chemical elements found in the 21 A.A.s [2].

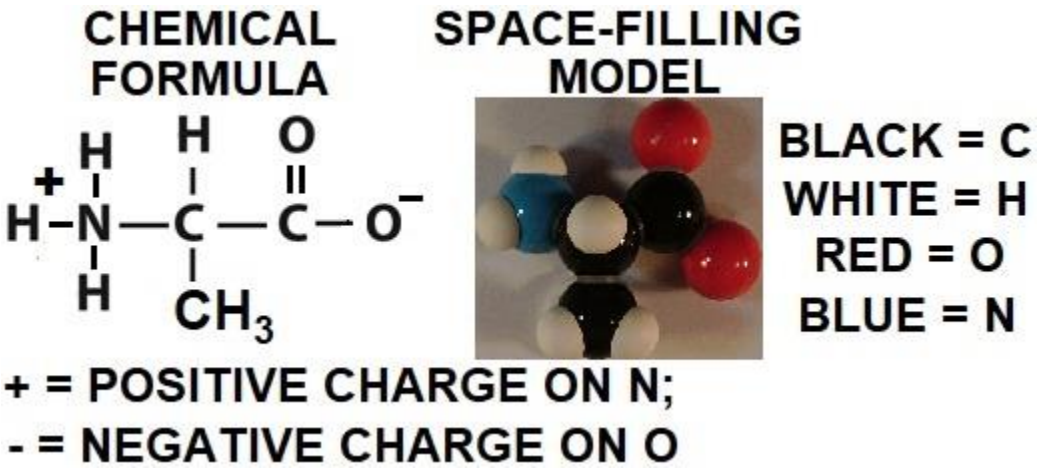
Element	Atomic Symbol*	Model Color	Atomic Number
Hydrogen	**H**	H	1
Carbon	**C**	**C**	6
Nitrogen	**N**	**N**	7
Oxygen	**O**	**O**	8
Sulfur	**S**	**S**	16
Selenium	**Se**	**Se**	34

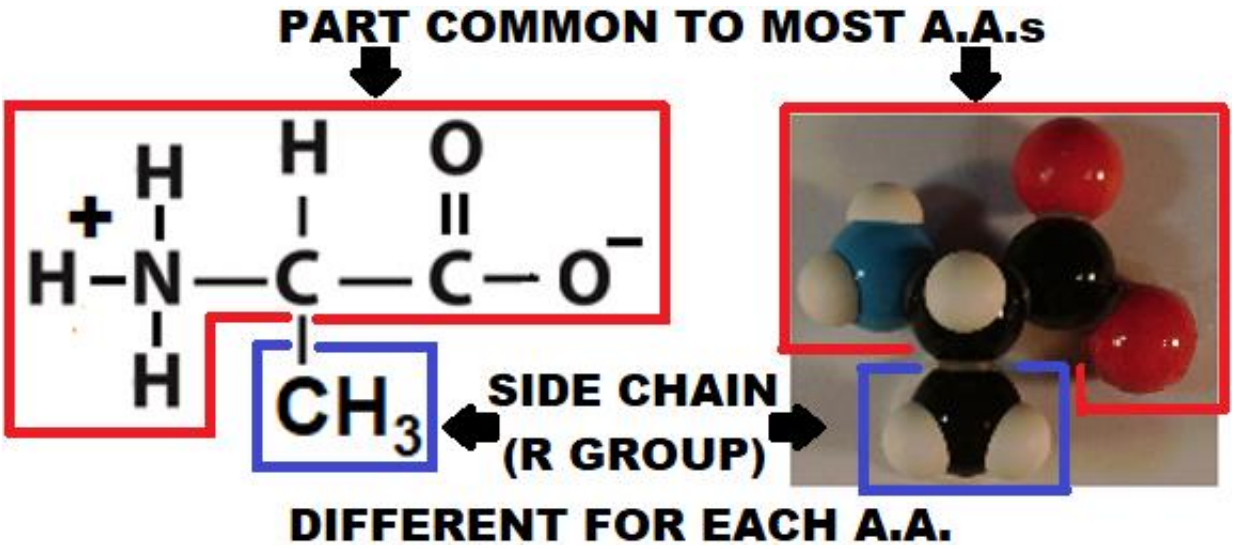
*Atomic (not IUPAC-IUBMB, JCBN A.A.) symbol.

Wade Research Foundation Reports (2019) 8(1)

21 A.A.s

Figure 4. The general features of the 21 A.A.s shown in Figures 5-25, on the following pages, are illustrated for an example A.A., Alanine (A) [2]. The letters in the figure represent atomic symbols for elements.



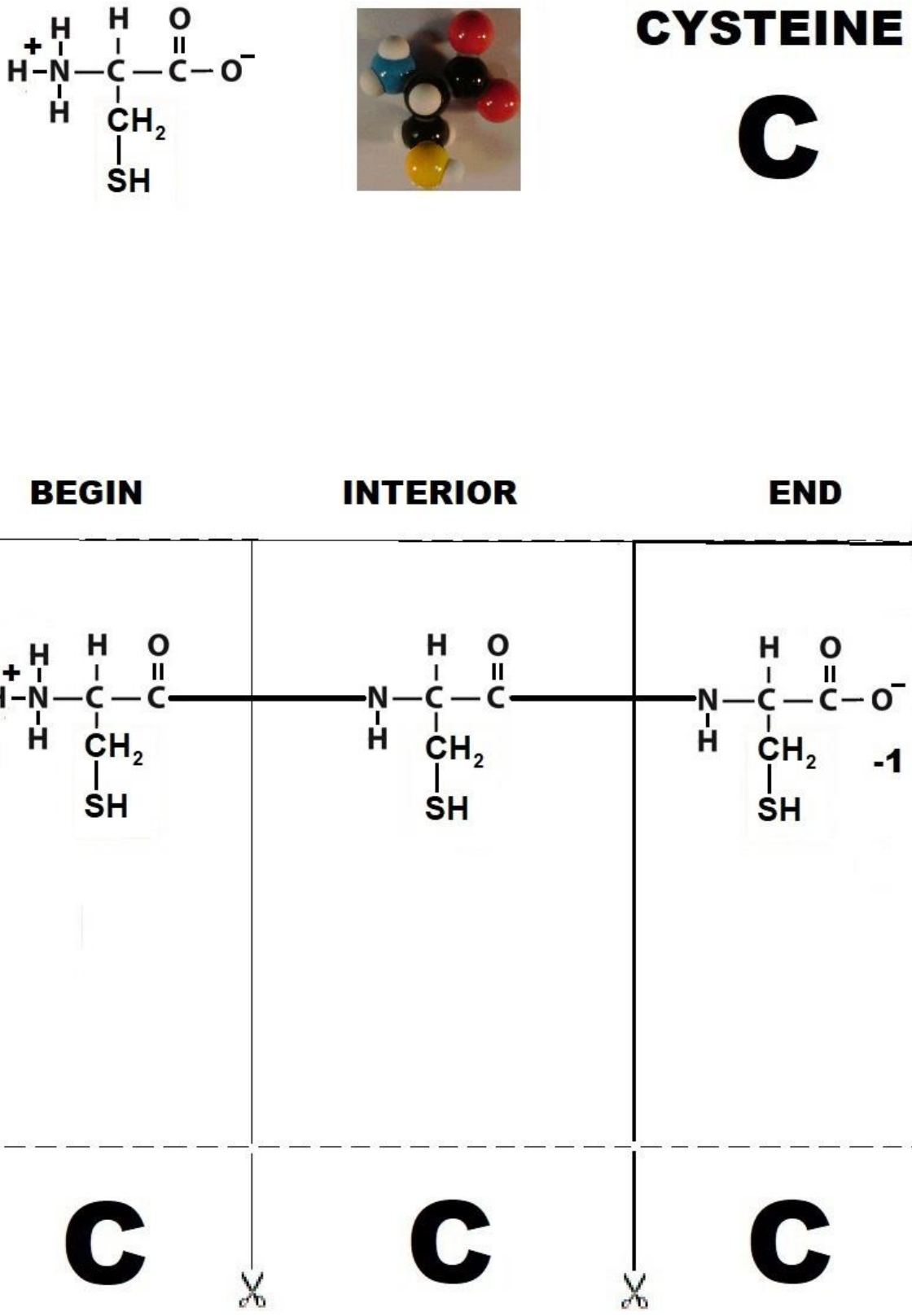


Wade Research Foundation Reports (2019) 8(1)

Figure 5.

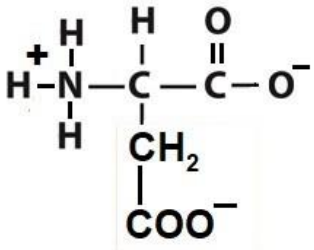
Wade Research Foundation Reports (2019) 8(1)

Figure 6.



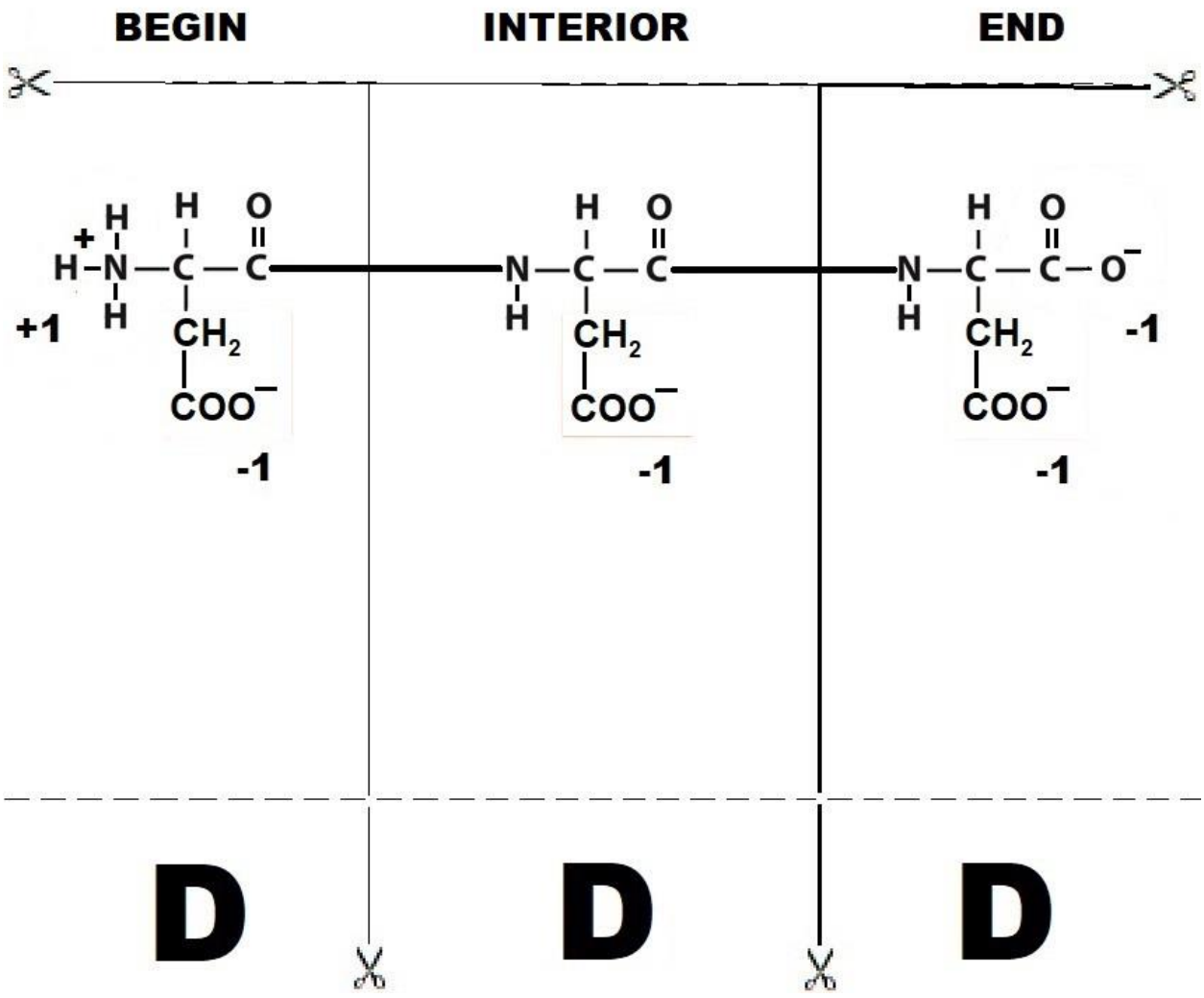
Wade Research Foundation Reports (2019) 8(1)

Figure 7.



ASPARTIC ACID

D

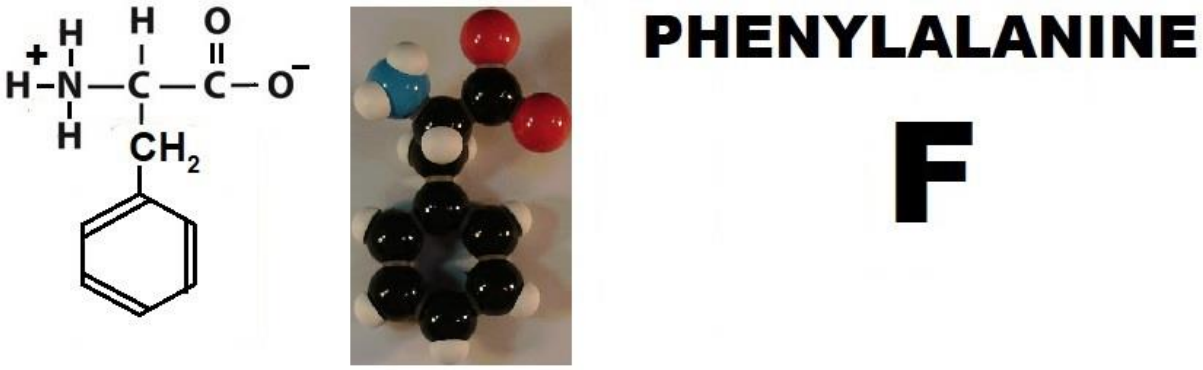


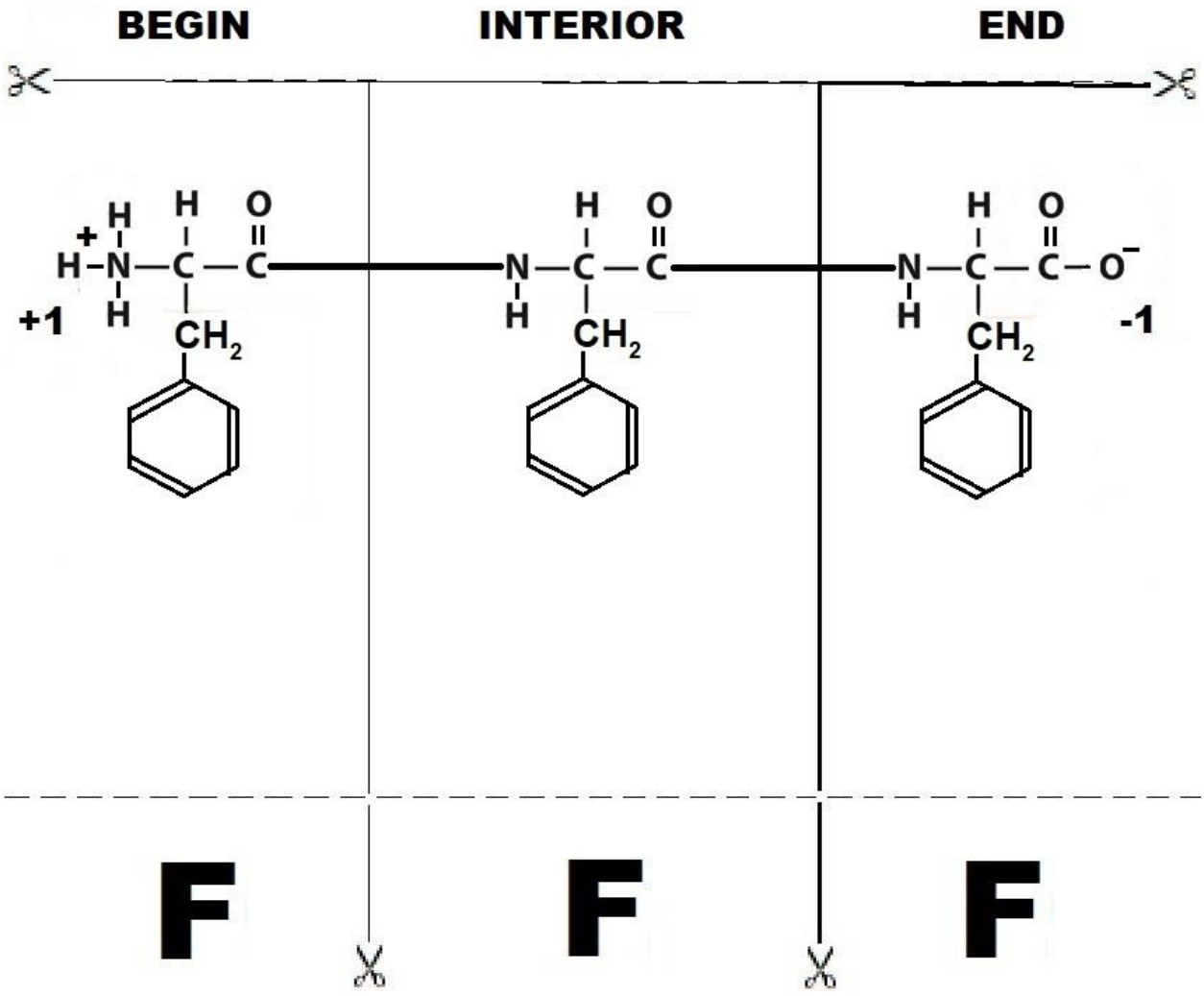
Wade Research Foundation Reports (2019) 8(1)

Figure 8.

Wade Research Foundation Reports (2019) 8(1)

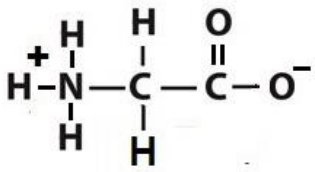
Figure 9.





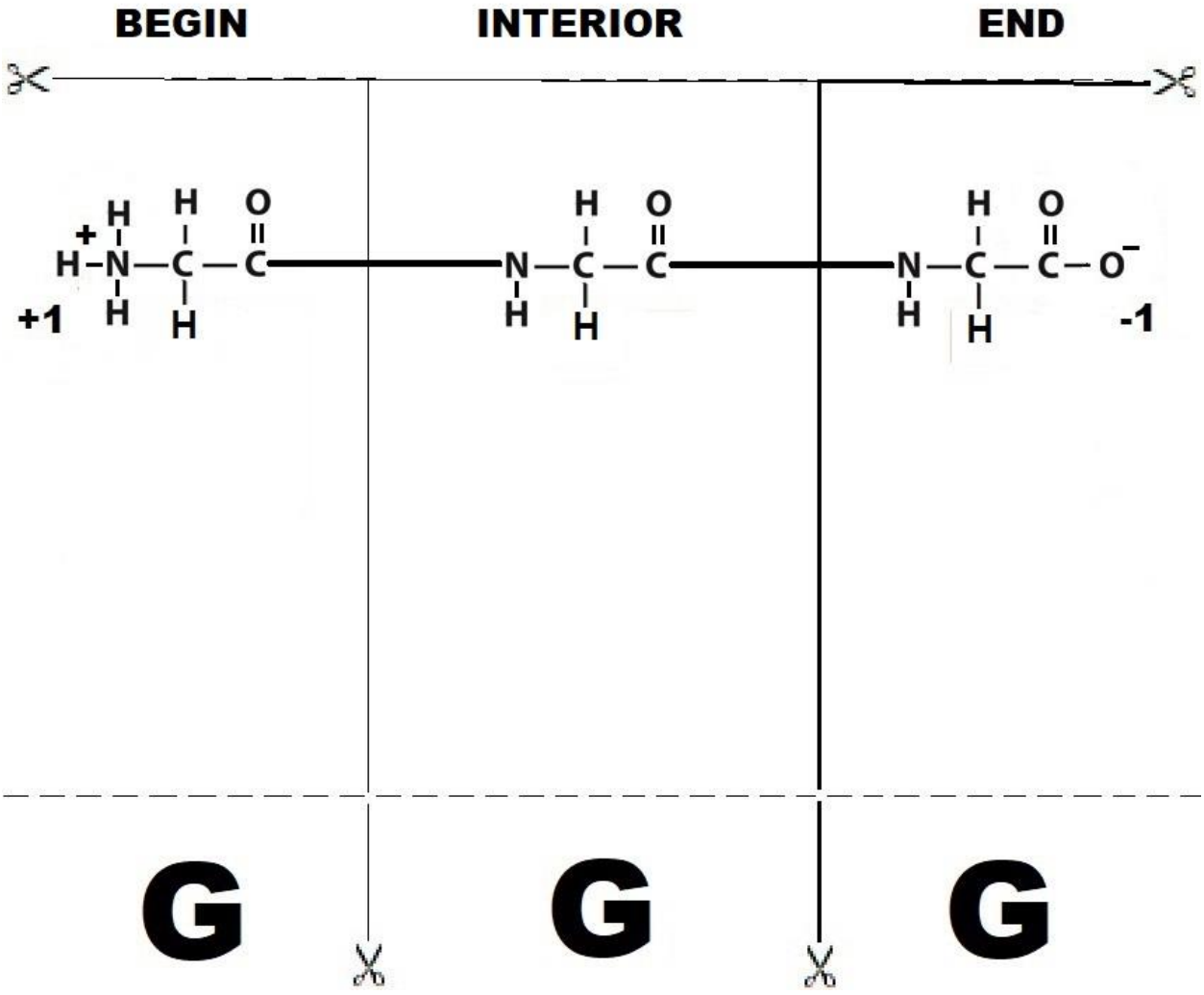
Wade Research Foundation Reports (2019) 8(1)

Figure 10.



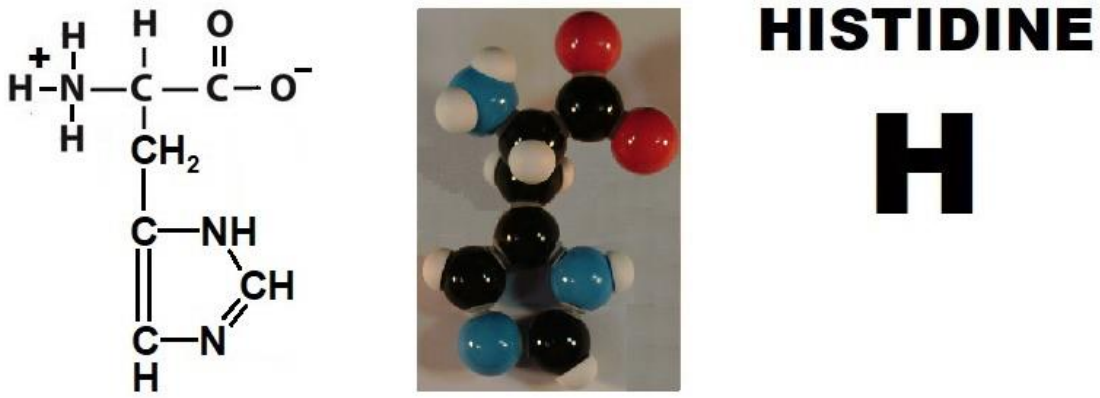
GLYCINE

G



Wade Research Foundation Reports (2019) 8(1)

Figure 11.

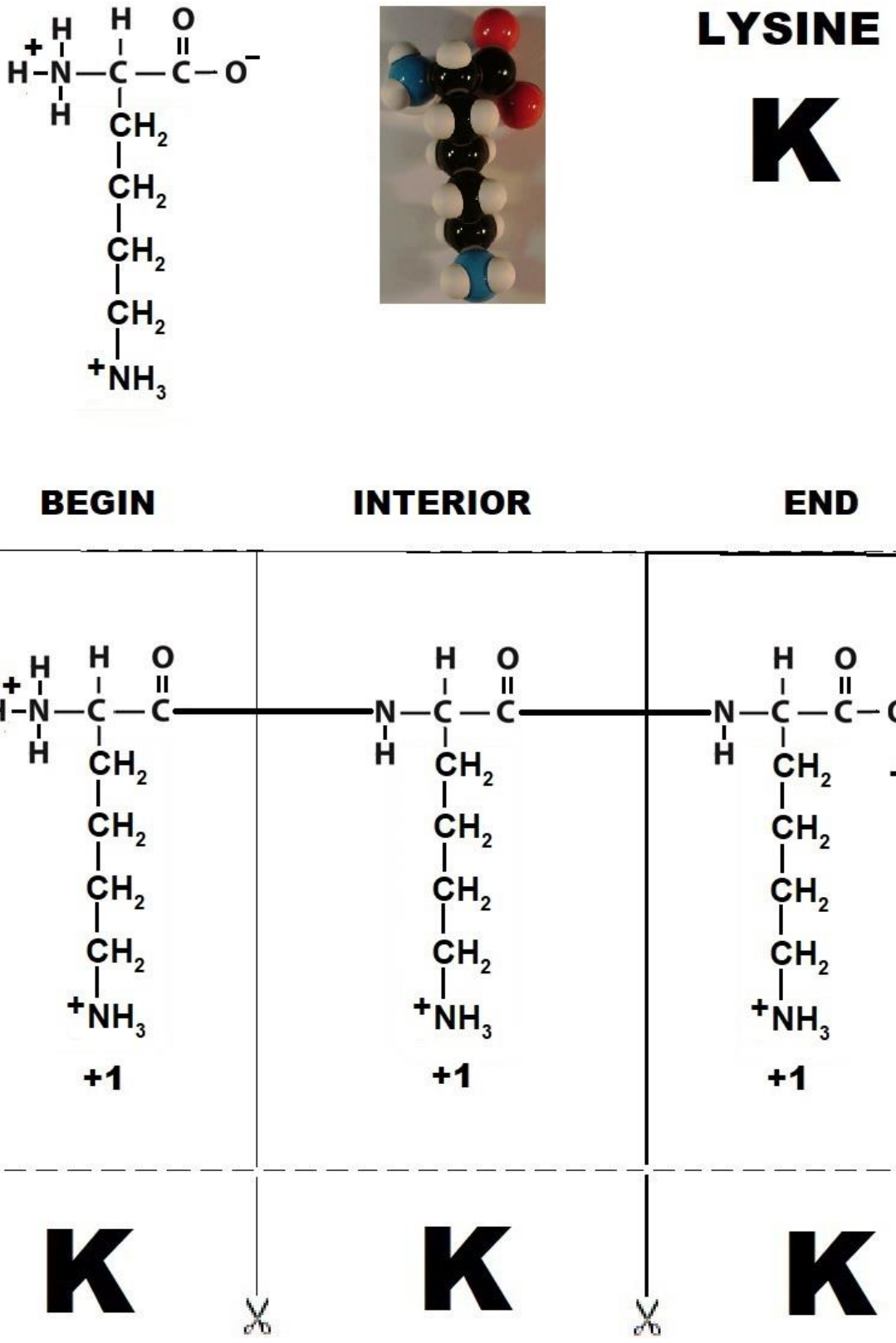


Wade Research Foundation Reports (2019) 8(1)

Figure 12.

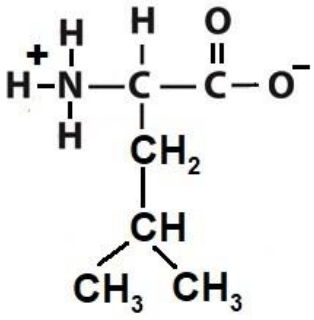
Wade Research Foundation Reports (2019) 8(1)

Figure 13.



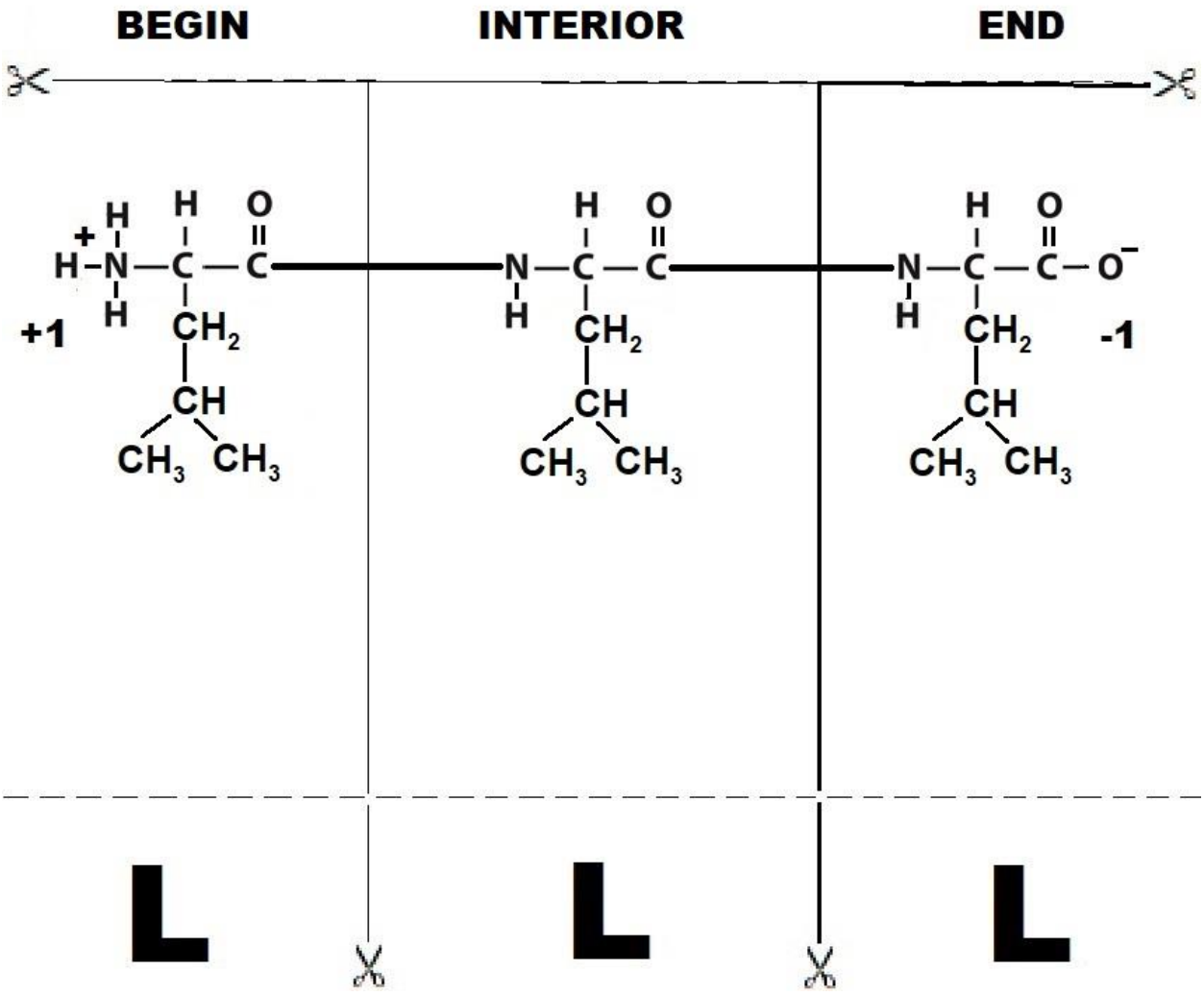
Wade Research Foundation Reports (2019) 8(1)

Figure 14.



LEUCINE

L

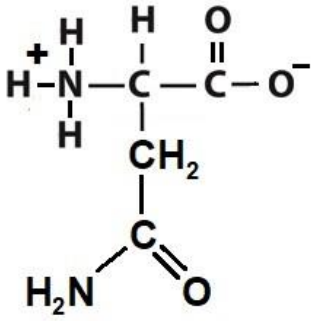


Wade Research Foundation Reports (2019) 8(1)

Figure 15.

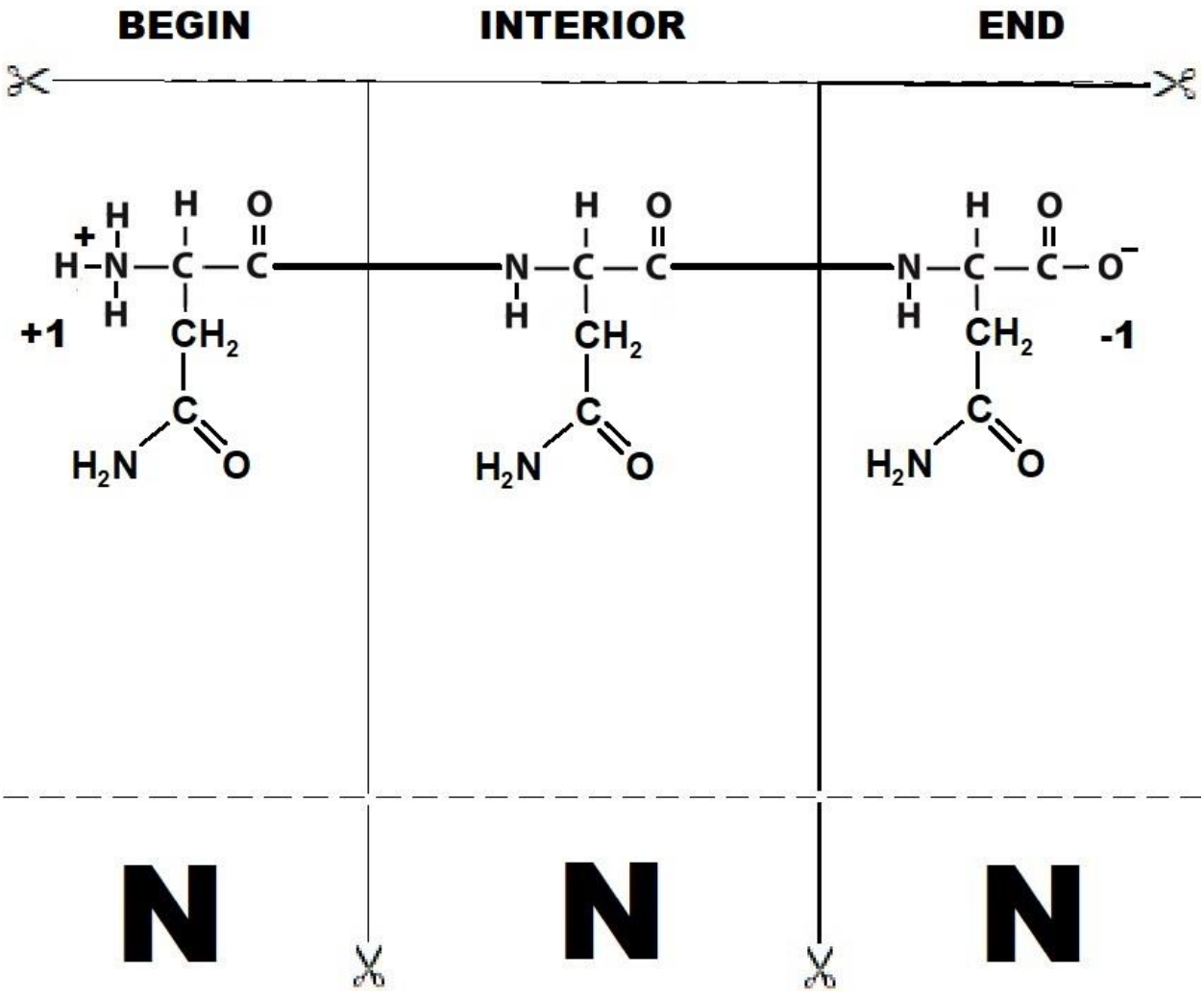
Wade Research Foundation Reports (2019) 8(1)

Figure 16.



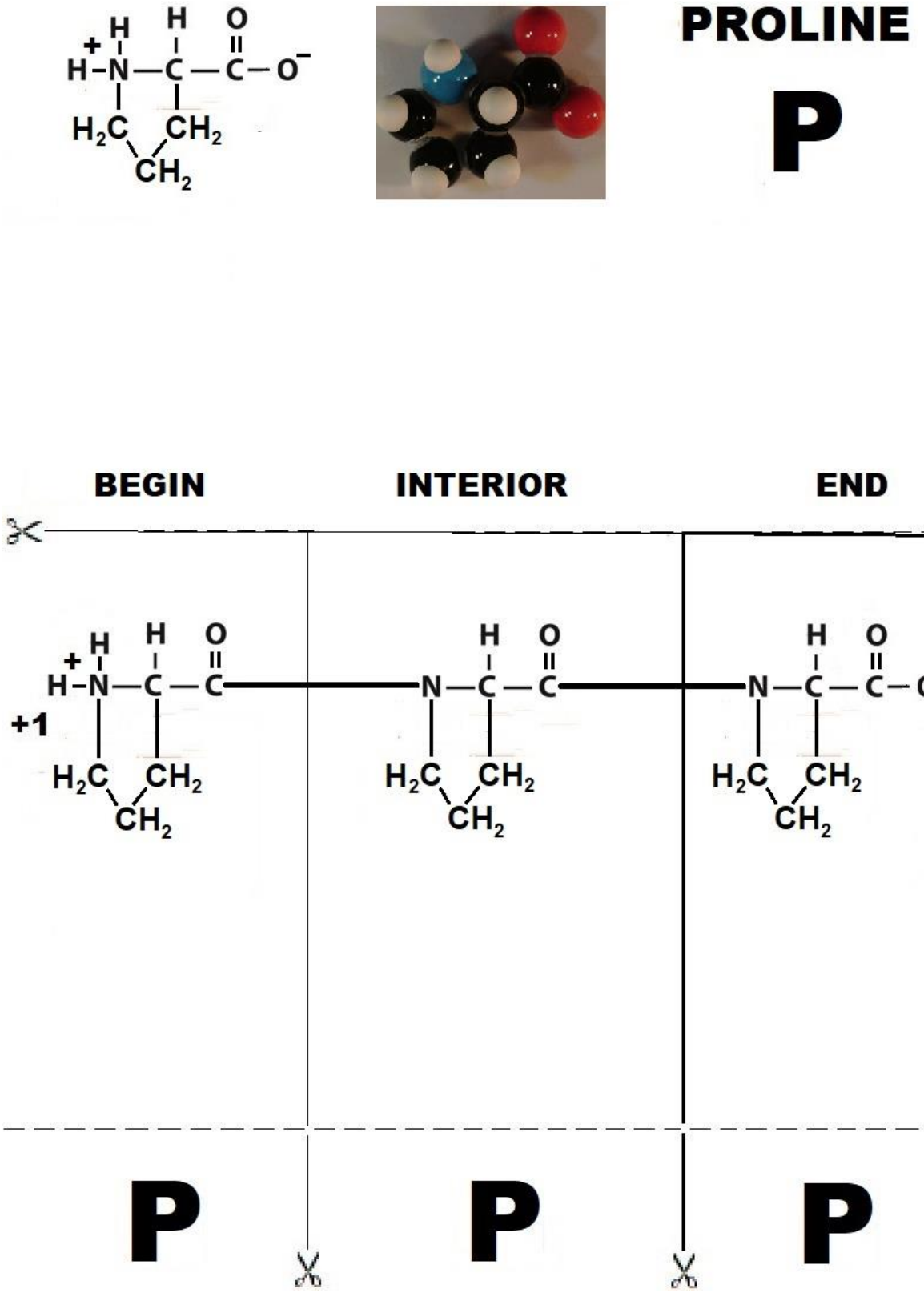
ASPARAGINE

N



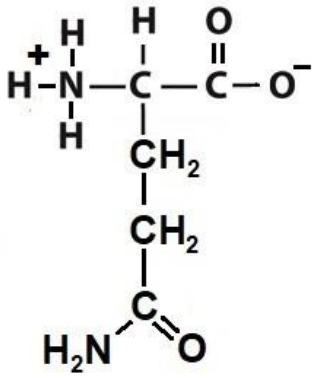
Wade Research Foundation Reports (2019) 8(1)

Figure 17.



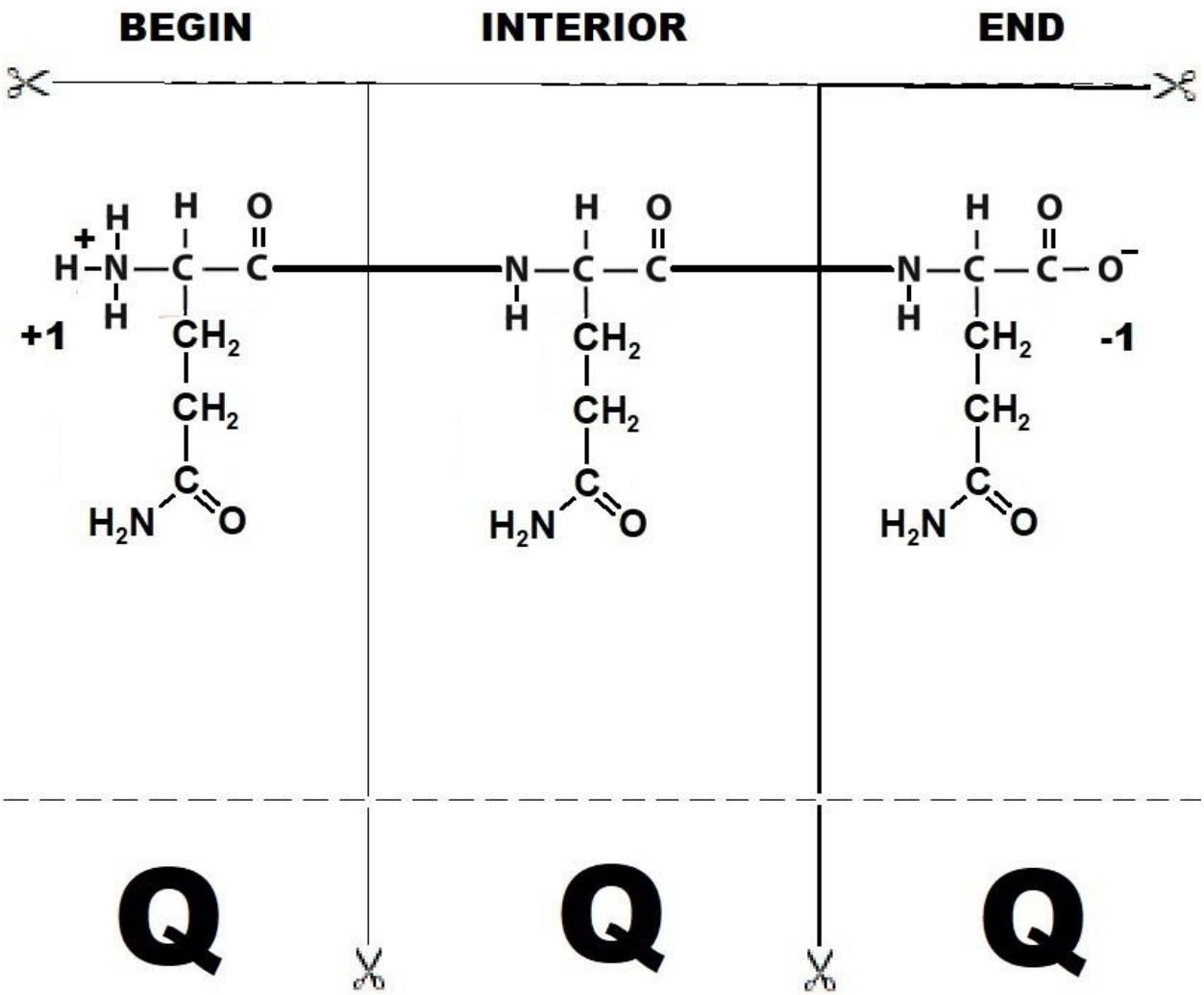
Wade Research Foundation Reports (2019) 8(1)

Figure 18.



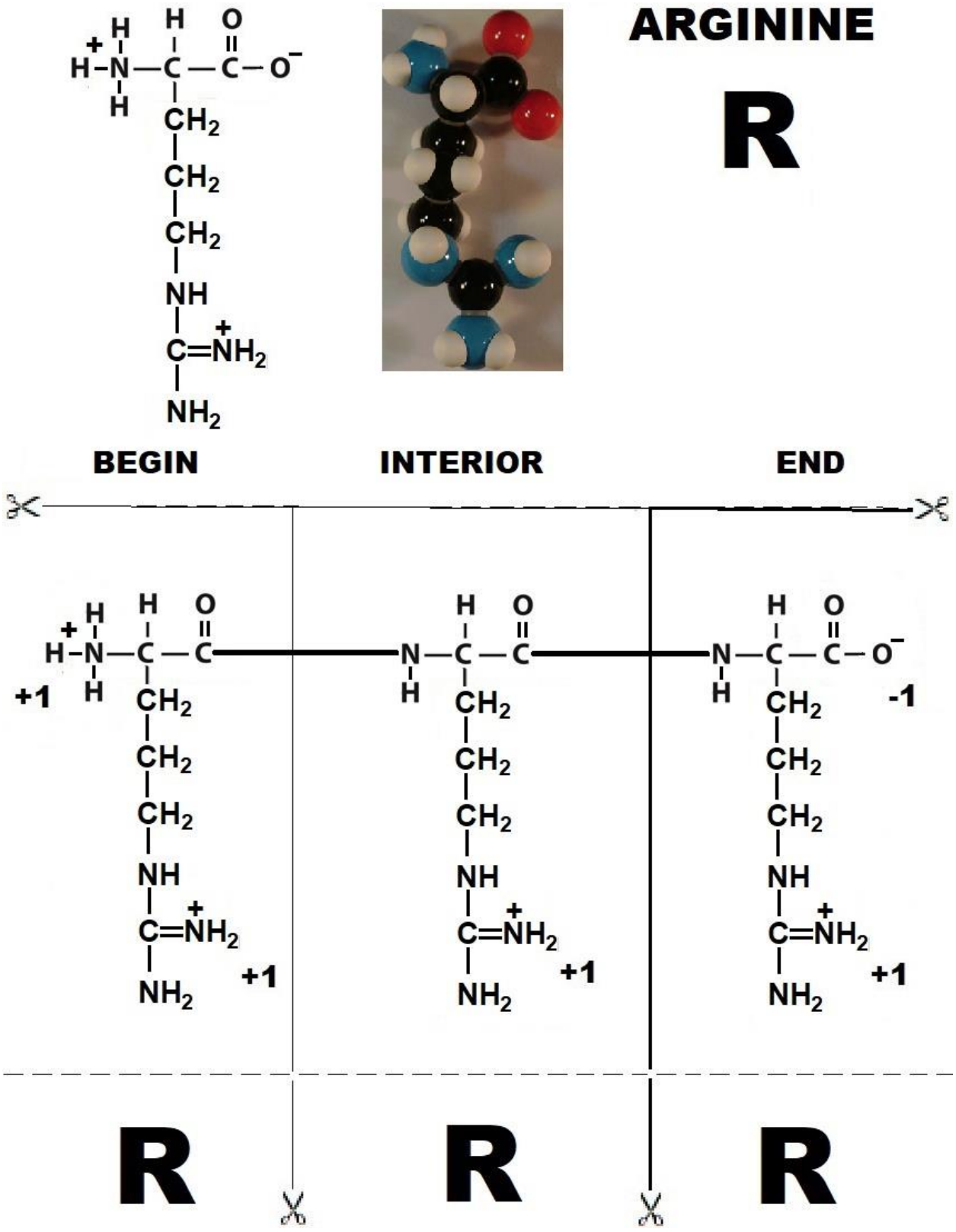
GLUTAMINE

Q



Wade Research Foundation Reports (2019) 8(1)

Figure 19.



Wade Research Foundation Reports (2019) 8(1)

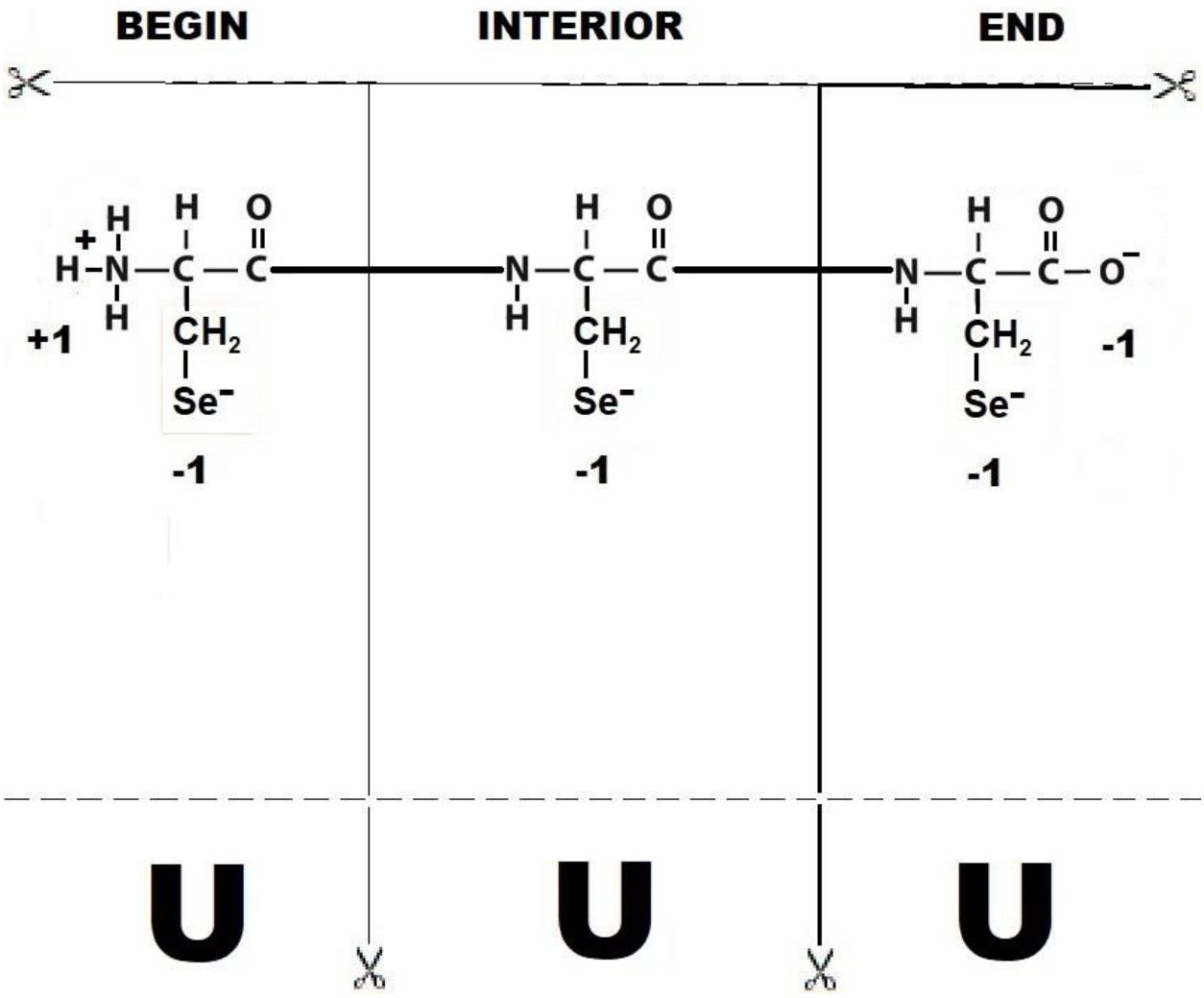
Figure 20.

Wade Research Foundation Reports (2019) 8(1)

Figure 21.

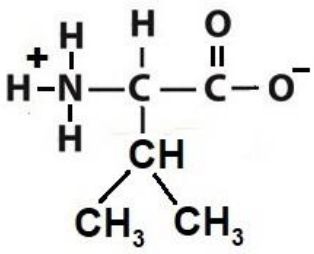
Wade Research Foundation Reports (2019) 8(1)

Figure 22.



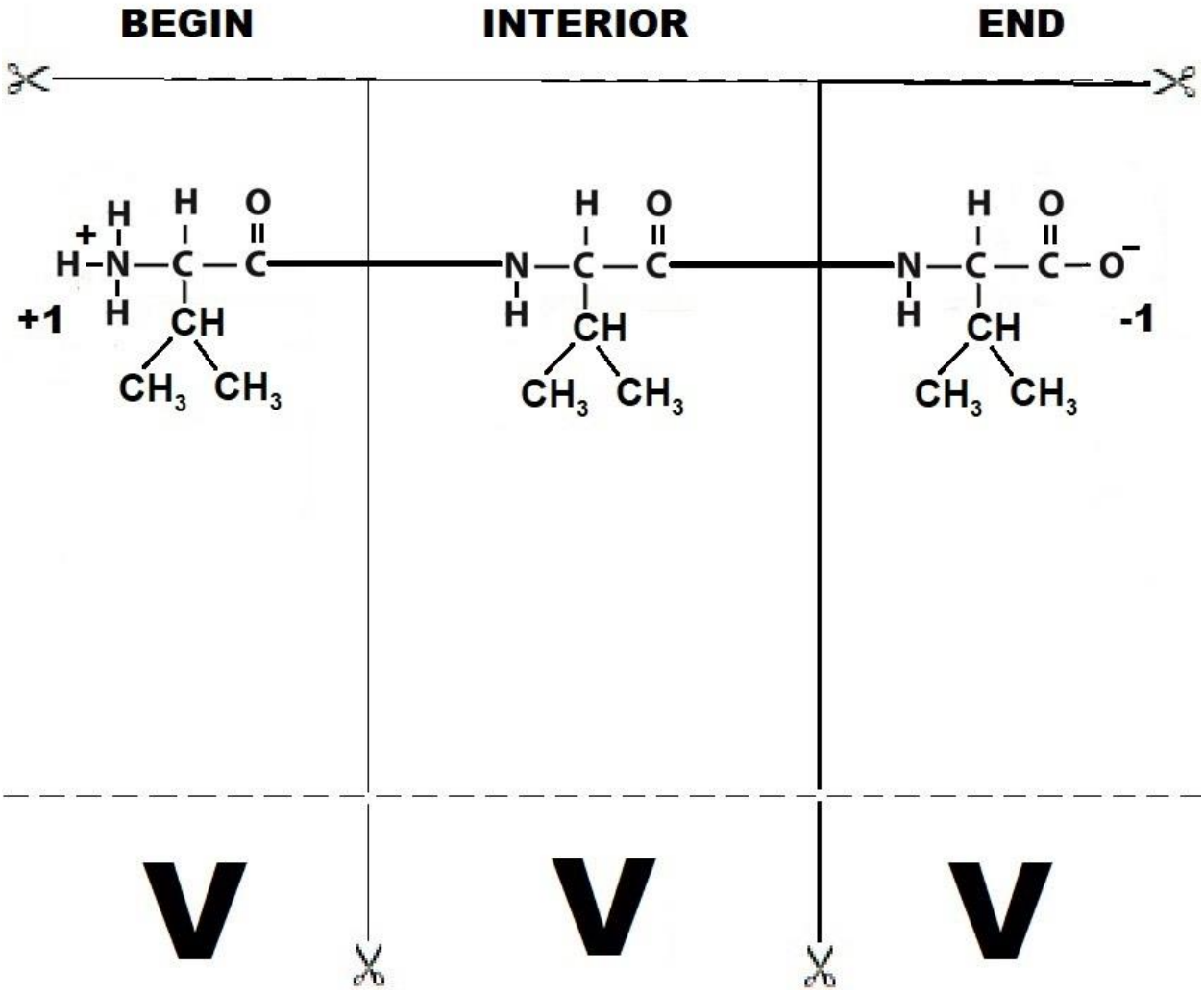
Wade Research Foundation Reports (2019) 8(1)

Figure 23.



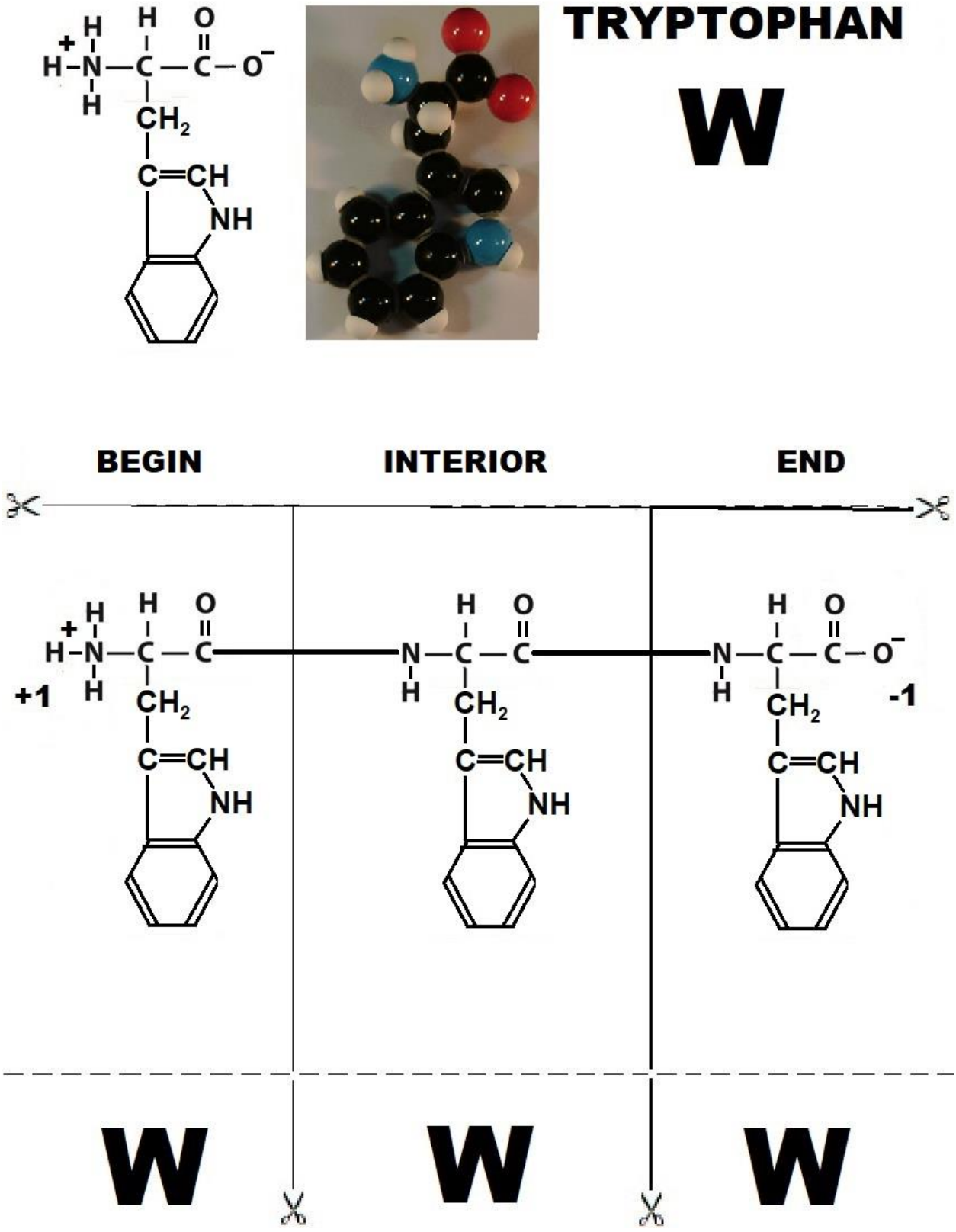
VALINE

V



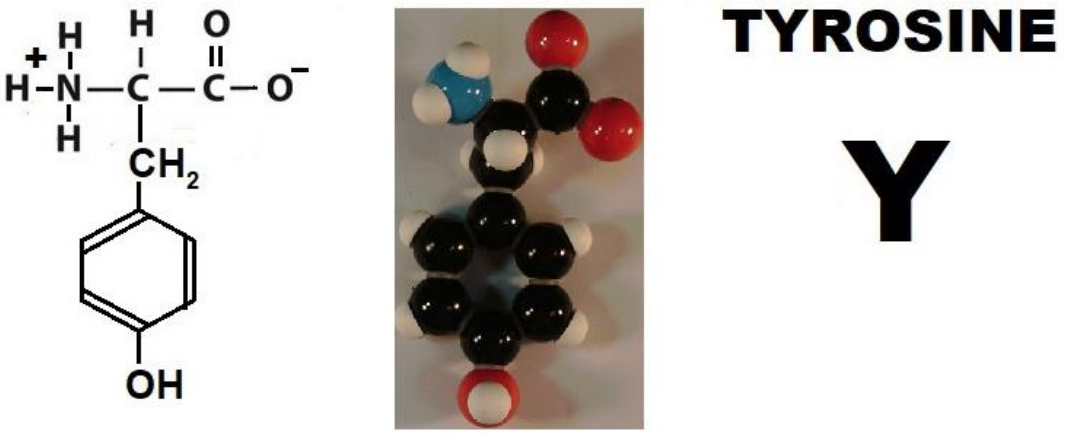
Wade Research Foundation Reports (2019) 8(1)

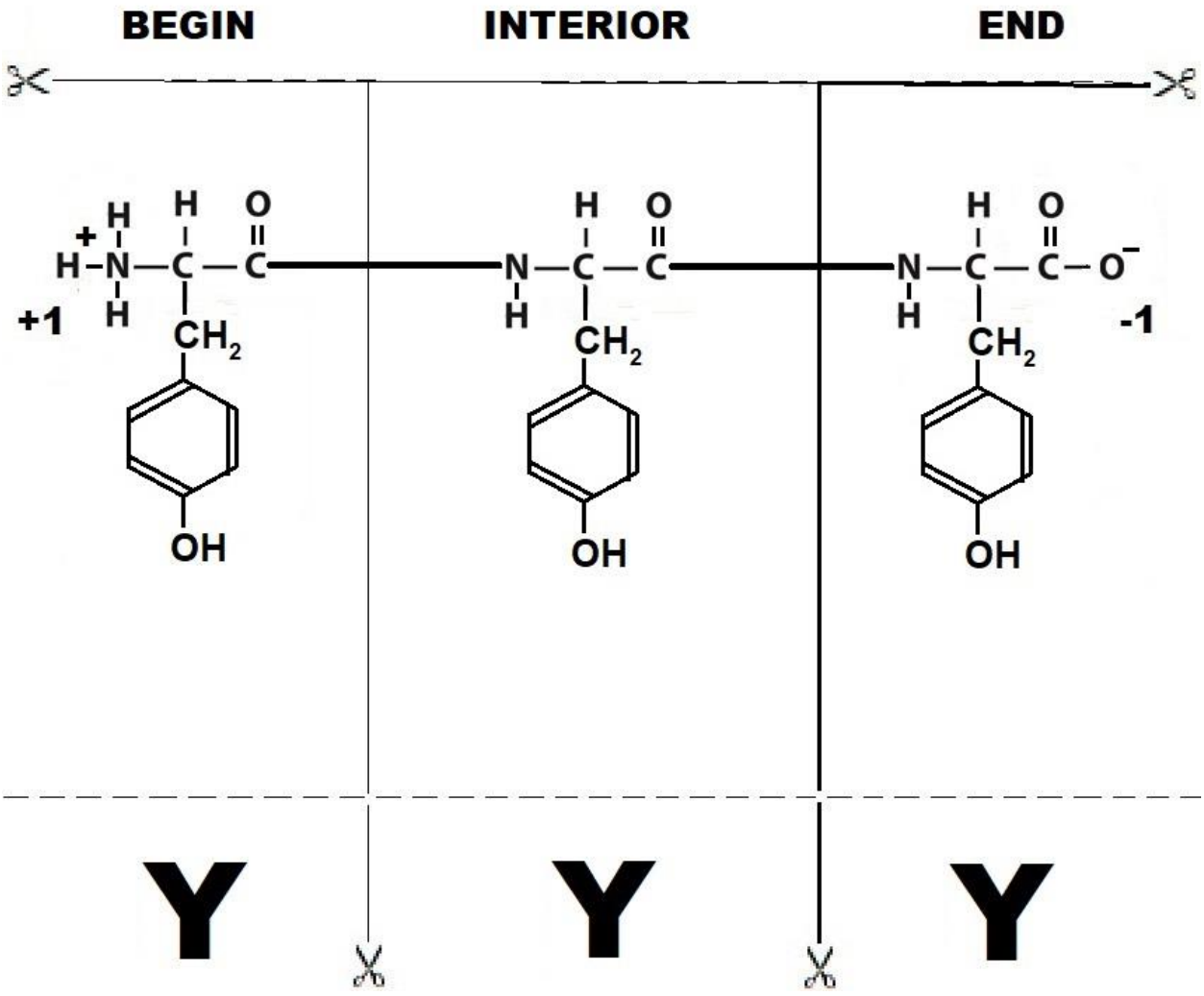
Figure 24.



Wade Research Foundation Reports (2019) 8(1)

Figure 25.





ADVANCED TOPIC: CYSTEINE AND SELENOCYSTEINE:

When either, or both, of the A.A.s, Cysteine (C, or the 3-letter symbol, Cys) and Selenocysteine (U, or the 3-letter symbol, Sec), are present in an A.A. sequence, their side chains will have the ability to make chemical bonds with the side chains of other Cys or Sec in the same, or other, peptides [i.e., to form the disulfide bond of Cystine (Cys-S-S-Cys; where S is a sulfur atom), the diselenide bond of Selenocystine (Sec-Se-Se-Sec, where Se is a selenium atom) or a Selenosulfide (Sec-Se-S-Cys) bond] [16]. This results in many possible bonding combinations for peptides that contain these A.A.s. (e.g., Table 7). One of these combinations (the box with gray shading in Table 7) will be seen in the next section of the workbook.

Table 7. Cysteine (C) side chain bonding combinations for peptides containing 1 or 2 C. Subscripts "B", "I", and "E" refer to the "Beginning", "Interior", and "End" locations of the peptide, respectively. Horizontal dashed lines indicate chains of A.A.s, and vertical plus horizontal solid lines that link Cs indicate disulfide bonds (i.e., Cys-S-S-Cys, where S = sulfur atom).

One C in a peptide; interpeptide bonding:		Two Cs in a peptide; intrapeptide bonding:
C_B---------- | C_B----------	-----C_I----- | -----C_I-----	C_B-----C_I----- └─────┘
C_B---------- | -----C_I-----	-----C_I----- | ----------C_E	C_B----------C_E └──────────┘
C_B---------- | ----------C_E	----------C_E | ----------C_E	-----C_I-----C_E └─────┘

DESIGN A PAPER PEPTIDE (WITH SCISSORS AND TAPE):

1) Choose a name, word, or phrase that is composed of letters of the English alphabet [17-22]. For example, choose the name, “Captain America”. (Note: Captain America is a fictional superhero appearing in American comic books published by Marvel Comics [23].)
2) Compare the chosen name, word, or phrase with the IUPAC-IUBMB, JCBN single letters symbols for the names of A.A.s (Tables 3 and 4) to ensure that all letters of the chosen name, word, or phrase are compatible with the single letter symbol nomenclature. In “Captain America”, all letters are compatible with the IUPAC-IUBMB, JCBN nomenclature.
3) Combine the letters of the name, word, or phrase to produce a continuous string of letters. For example:

   Captain America ➔ CaptainAmerica or CAPTAINAMERICA
4) In this sequence of letters, the first letter, “C”, will be considered the “Begin” A.A., letters “APTAINAMERIC” will be considered “Interior” A.A.s, and the final letter, “A”, will be considered the “End” A.A. Using the “21 A.A.s” section of the workbook, print out the pages corresponding to the letters in CAPTAINAMERICA. You will have to print extra copies (or photocopy) the pages for A and I. Using scissors, cut out the “BEGIN” part of one page for Cysteine (C), the “INTERIOR” parts of 3 pages for Alanine (letter “A”), one page for Proline (letter “P”), one page for Threonine (letter “T”), two pages for Isoleucine (letter “I”), one page for Asparagine (letter “N”), one page for Methionine (letter “M”), one page for Glutamic acid (letter “E”), and one page for Arginine (letter “R”). Then cut out the “END” part of one page for Alanine (letter “A”).
5) Assemble the paper pieces in the proper order (Figure 26).
6) Tape the pieces of paper together to form a continuous, linear sequence of letters (C-A-P-T-A-I-N-A-M-E-R-I-C-A) (Figure 27). At this point, the peptide would be complete, if it did not contain two Cs. As described in the previous section, peptides that contain 2 Cs, 2 Us, or a C and a U, can form intrapeptide chemical bonds, called disulfide bonds (for 2 Cs), diselenide bonds (for 2 Us), or selenosulfide bonds for one C and one U.
7) Using a pencil or pen, draw a line linking the sulfur atom (S) in the side chain of each C. This will represent the disulfide bond. The peptide is now complete (see below and Figure 27). Figure 28 shows some terminology that is used for describing peptide structures [2].

C-A-P-T-A-I-N-A-M-E-R-I-**C**-A

Wade Research Foundation Reports (2019) 8(1)

Figure 26. The individual pieces of A.A. pages arranged in correct order for peptide CAPTAINAMERICA.

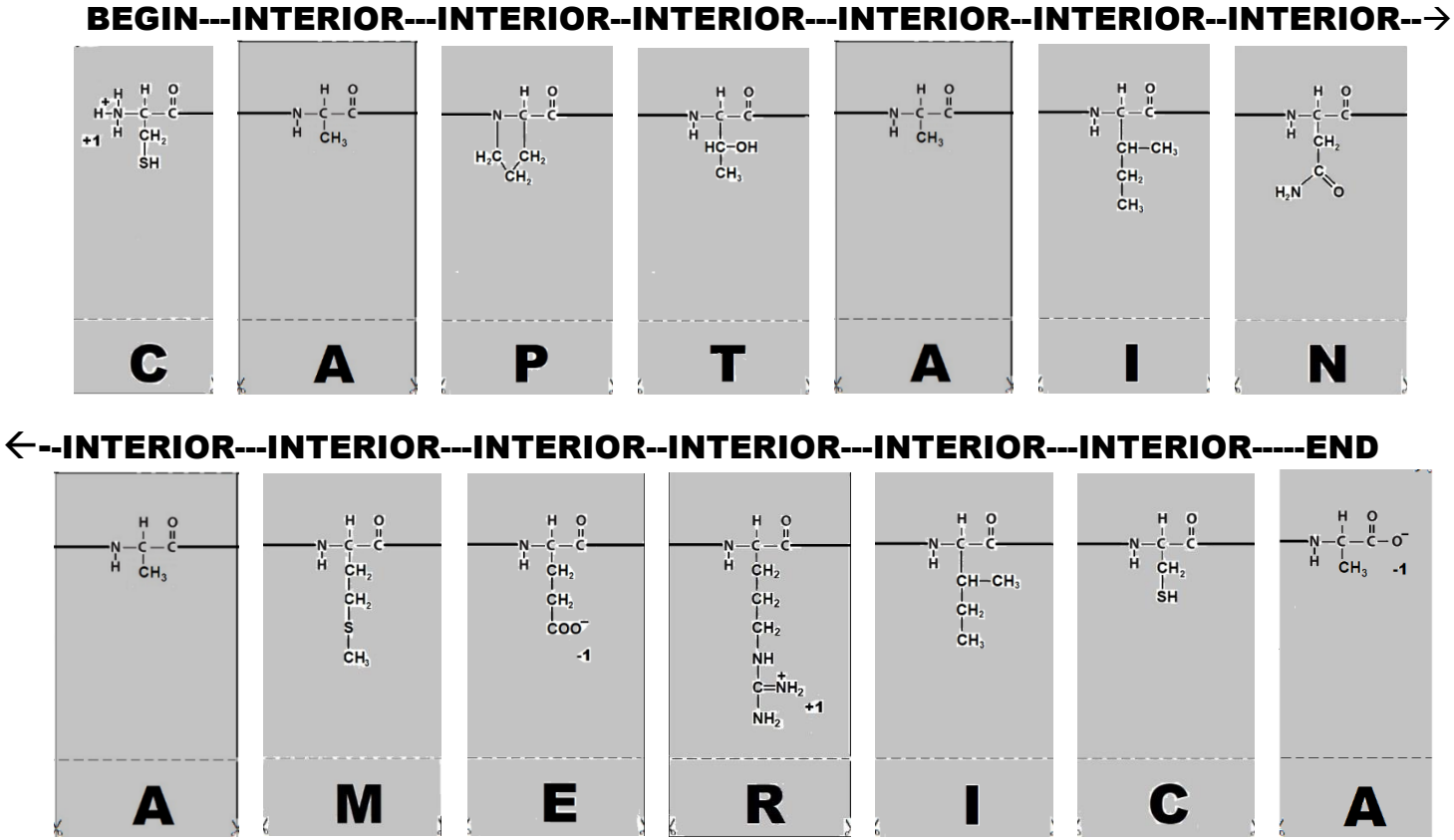
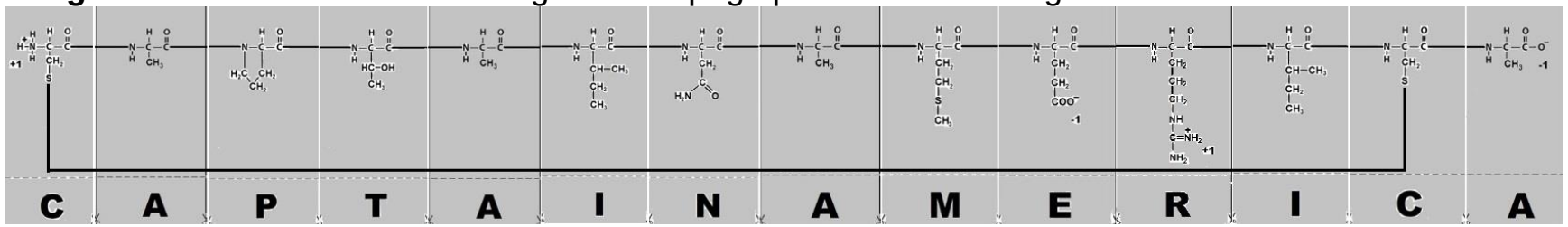
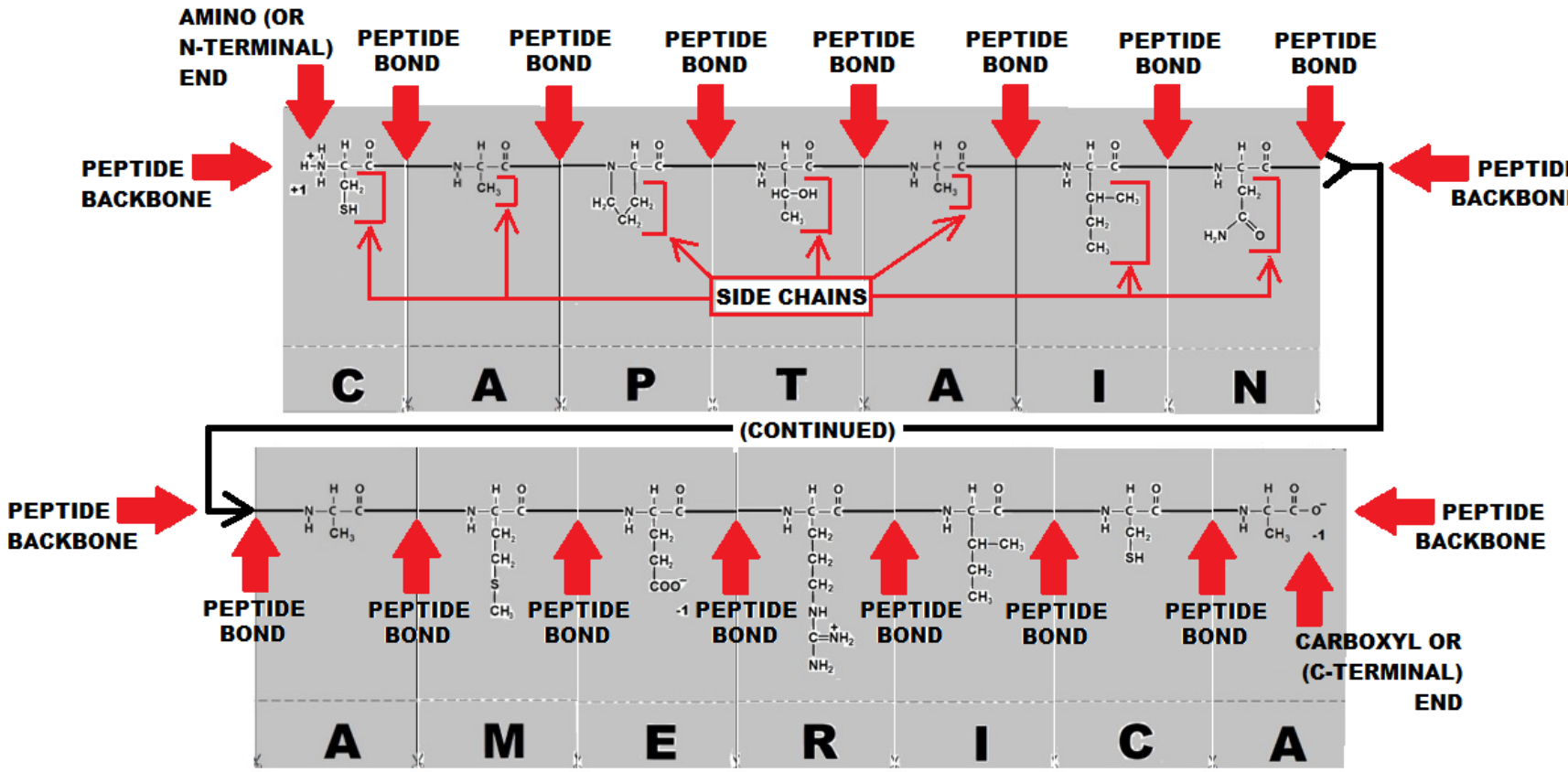


Figure 27. Result after connecting the A.A. page pieces and drawing the disulfide bond between Cs.



Wade Research Foundation Reports (2019) 8(1)

Figure 28. Some peptide terminology [2]:



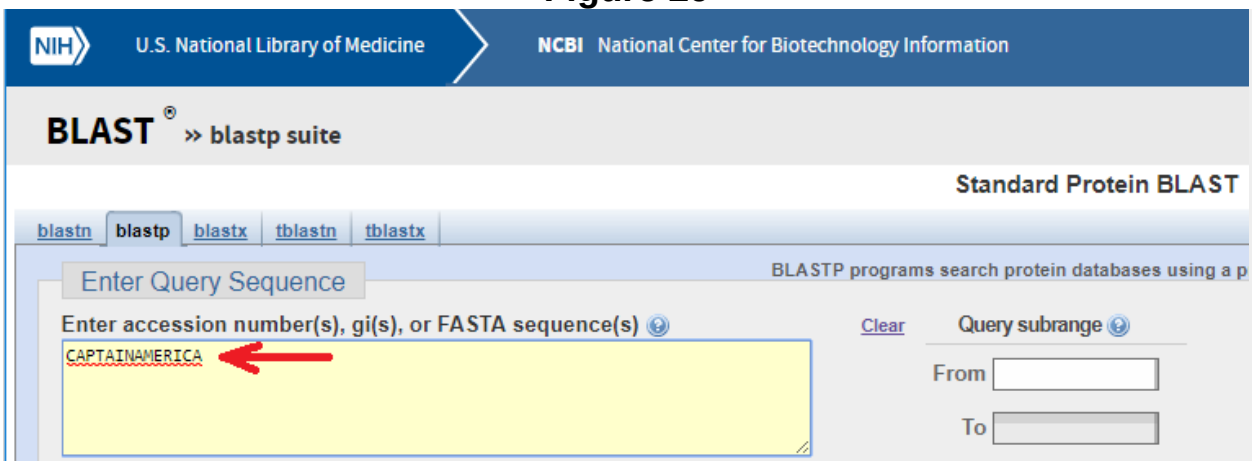
Wade Research Foundation Reports (2019) 8(1)

HAS YOUR PEPTIDE BEEN FOUND IN NATURE?

This section will enable you to determine if the peptide that you designed (e.g., CAPTAINAMERICA) has been found in nature. Since the instructions for each step are accompanied by figures, individual figures captions have been omitted to save space.

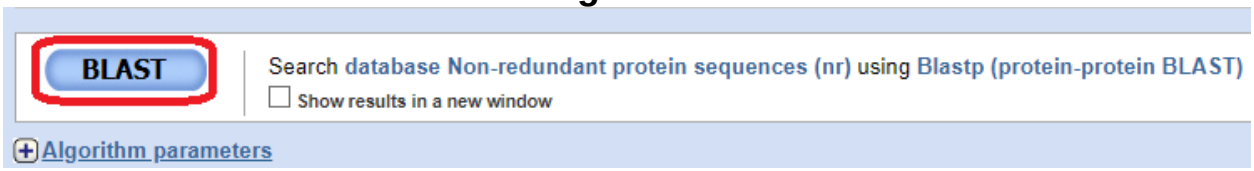
1) Consider "CAPTAINAMERICA" as the example (hypothetical) peptide. Perform a standard protein "BLAST" (Basic Local Alignment Search Tool) search of the N.C.B.I.'s protein database for "CAPTAINAMERICA" [24].
Go to: https://blast.ncbi.nlm.nih.gov/blast.cgi?page=proteins
Enter "CAPTAINAMERICA" (without the parentheses) into the box labeled, "Enter accession number(s), gi(s), or FASTA sequence(s)".

Figure 29



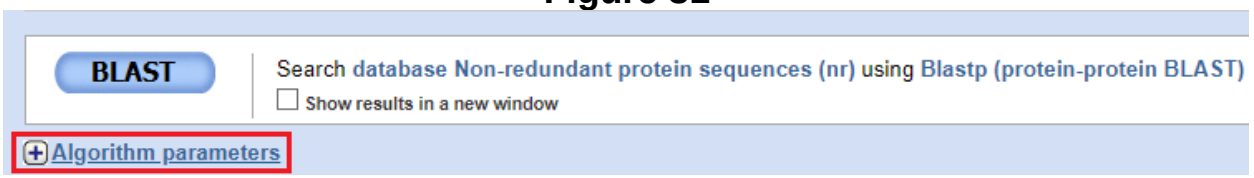
2) Scroll down the page and click on "BLAST".

Figure 30



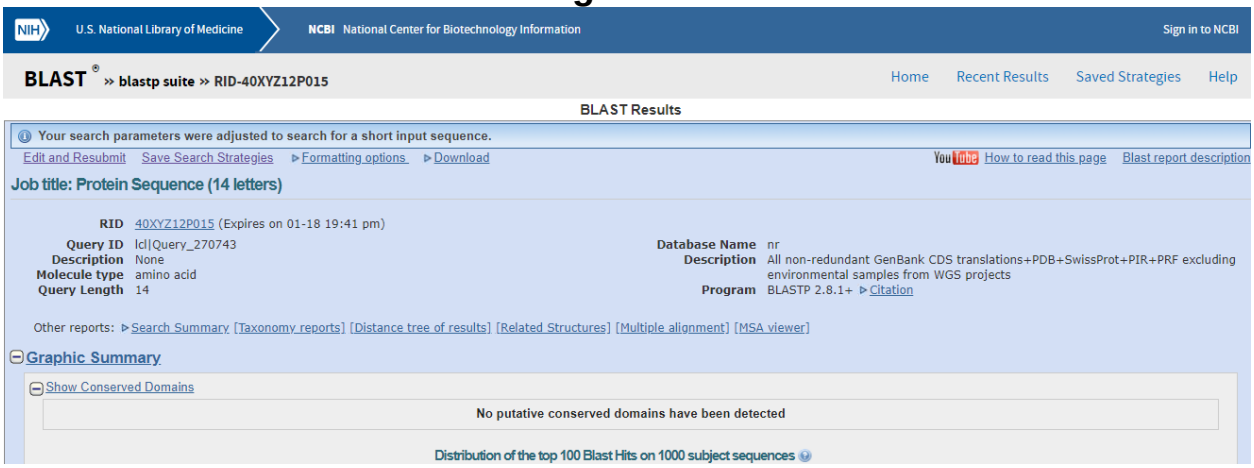
The "BLAST" algorithm is automatically set to find 100 "hits" (protein A.A. sequences that contain all, or parts, of the query sequence), but it can be adjusted to find from 10 to 20,000 "hits". In this example, it was adjusted to 1,000 "hits".

Figure 31



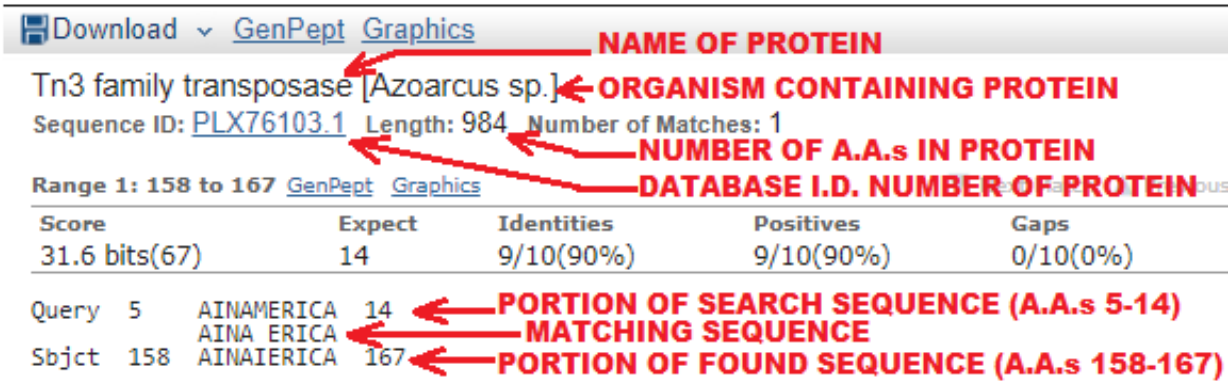
3) Wait for the N.C.B.I. computer to search the database. At the time that this example was searched (1/11/19), the protein database contained 184,243,125 protein A.A. sequences, and the search took one minute. The top of the results page is shown below.

Figure 32.



4) Scroll down the page to locate the sequence alignments, and any matches with the search ("query") sequence. The best matches will occur at the top of these alignments.

Figure 33.



The best result is shown above. It indicates that the entire "CAPTAINAMERICA" sequence has not yet been found in nature. There are 14 A.A.s in this hypothetical peptide, and 9 of those A.A.s (or 64% of the sequence) occur in a protein, Tn3 family transposase [25], of the organism *Azoarcus* sp., a genus of nitrogen-fixing bacteria [26]. An examination of the other "hits" results would indicate that other portions of the "CAPTAINAMERICA" sequence occur in other proteins.

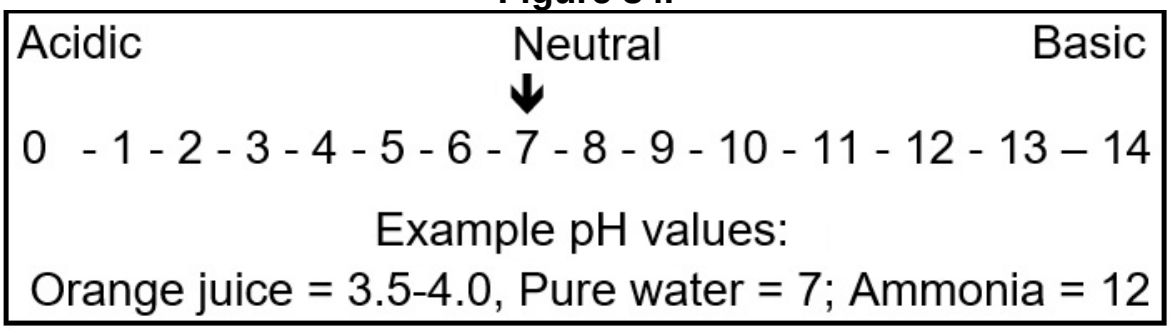
Occasionally, a match will occur with a protein whose three-dimensional (3D) structure has been determined. Unless you have experience using the

Protein Data Bank (PDB) [27], you may not recognize such a match from its Sequence, or database, ID. However, if you do discover such a match, you would then be able to go to the PDB website and view the protein (and your peptide) in 3D. This type of analysis can sometimes provide information about the function of your peptide within the protein.

PREDICT BIOLOGICAL AND OTHER PROPERTIES OF YOUR PEPTIDE.

1) Determine the "net charge" on the peptide at pH 7. "pH" refers to the acidic or basic character of a peptide [2]. The "pH" scale extends from "0" (most acidic) to "14" (most basic), with pH 7 being "neutral".

Figure 34.



The A.A.s in this workbook (Figures 5-25) have charges associated with them, and the charges shown are the those that would occur on the A.A.s in the "CAPTAINAMERICA" peptide at pH 7 [2]. Such charges can affect the physical, chemical, and biological properties of peptides. Examine peptide "CAPTAINAMERICA" (Figures 26-27) and locate the charges (+1 or -1). Add the charges to obtain the "net charge" for "CAPTAINAMERICA". It is "0".

Table 8. Calculation of the net charge on CAPTAINAMERICA at pH 7.

A.A.:	Symbol:	Location in peptide:	Charge:
Cysteine	C	Beginning	+1
Glutamic acid	E	Middle	-1
Arginine	R	Middle	+1
Alanine	A	End	-1
		Total charges =	(+1-1+1-1 = 0)

2) All peptides will exhibit biological properties. Sometimes these properties can be predicted by use of various online prediction programs, such as:

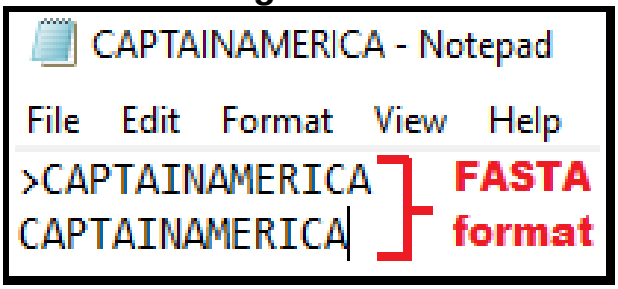
a) ExPASy Protparam tool [28] which allows the computation of various physical and chemical parameters for a user entered protein sequence (e.g., molecular weight = 1463.75; theoretical pI = 5.99).
b) Antimicrobial and anticancer properties:
   b(i) APD3: Antimicrobial peptide calculator and predictor [29]. The CAPTAINAMERICA sequence resembles the sequences of defensins, and it may exhibit antimicrobial activity.
   b(ii) IACP: a sequence-based tool for identifying anticancer peptides [30]. CAPTAINAMERICA is predicted to have anticancer activity (Probability: Anticancer = 0.789621; Non-anticancer = 0.210379).

MAKE COMPUTER MODELS OF YOUR PEPTIDE:

There are many computer-based, molecular modeling programs, but a free and user-friendly program is the Deep View/Swiss-PdbViewer program [31]. One problem with this program is that it does not have the ability to incorporate the A.A. Selenocysteine (i.e., the symbol, “U”) into peptides. However, since the properties of Selenocysteine are very similar to those of Cysteine [16], one can simply substitute Cysteine (i.e., the symbol, “C”) for Selenocysteine in the model (and remember that it really represents U). As in the previous section of the workbook, each step is accompanied by a figure, and figure captions have been omitted in order to save space.

1) Download the Deep View/Swiss-PdbViewer program on your computer, install it, and open it.
2) Create a “text document” and enter the peptide’s A.A. sequence in FASTA format (Figure 35).

Figure 35:



3) Load your A.A. sequence, in FASTA format, into the Deep View/Swiss-PdbViewer program (Figure 36, next page). Select the text document that contains your peptide’s A.A. sequence, and it will be uploaded into the

program. It will appear as a wireframe structure in an alpha helical conformation (i.e., a coiled, cylindrical structure) (Figure 37).

Figure 36.

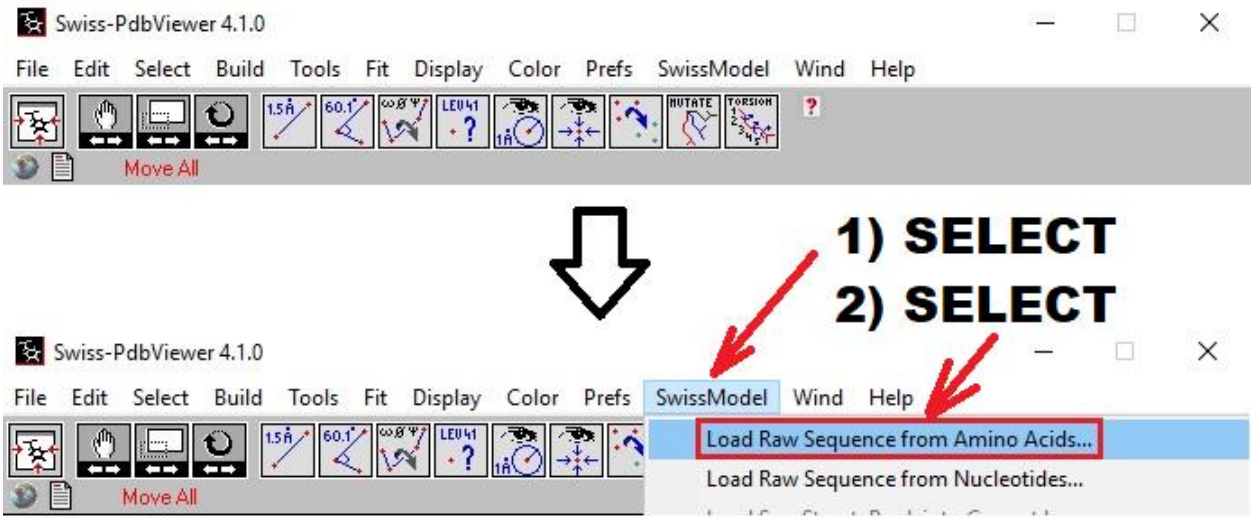
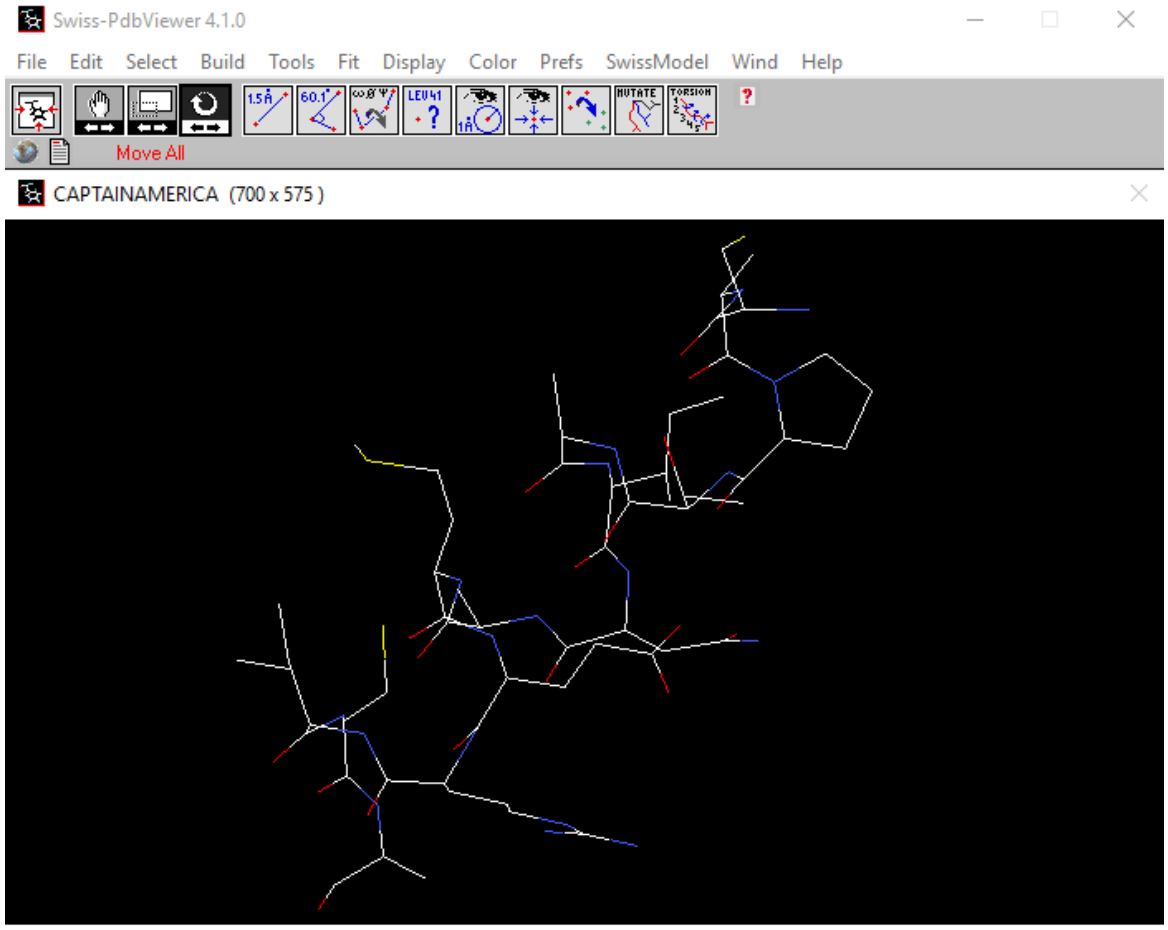


Figure 37.



4) You can rotate the structure by selecting the circular arrow on the tools panel at the top of the screen, and then left clicking on the computer mouse and moving the mouse (Figure 38). Figure 39 shows two views of the structure that were obtained by rotating it horizontally by 90 degrees.

Figure 38.

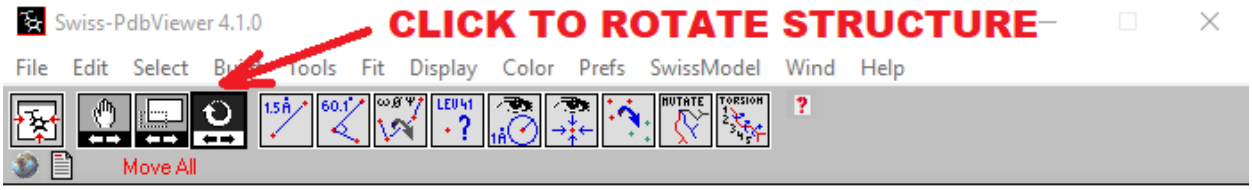
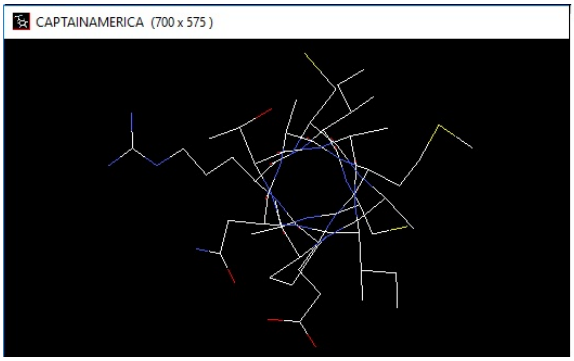
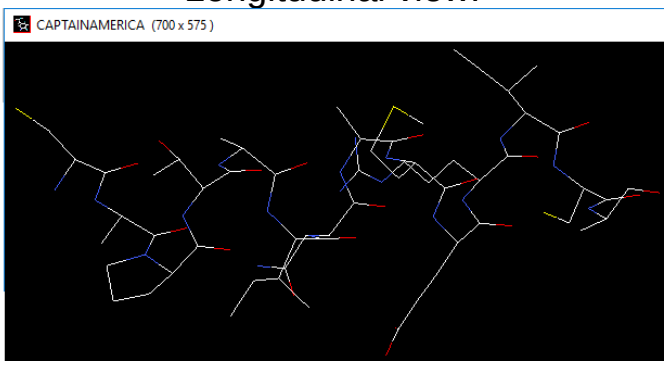


Figure 39.

Axial view:



Longitudinal view:



5) To obtain a less complicated view of the peptide, convert the alpha helix structure to a beta strand (i.e., a flattened ribbon) structure by following the steps shown in Figure 40.

Figure 40.

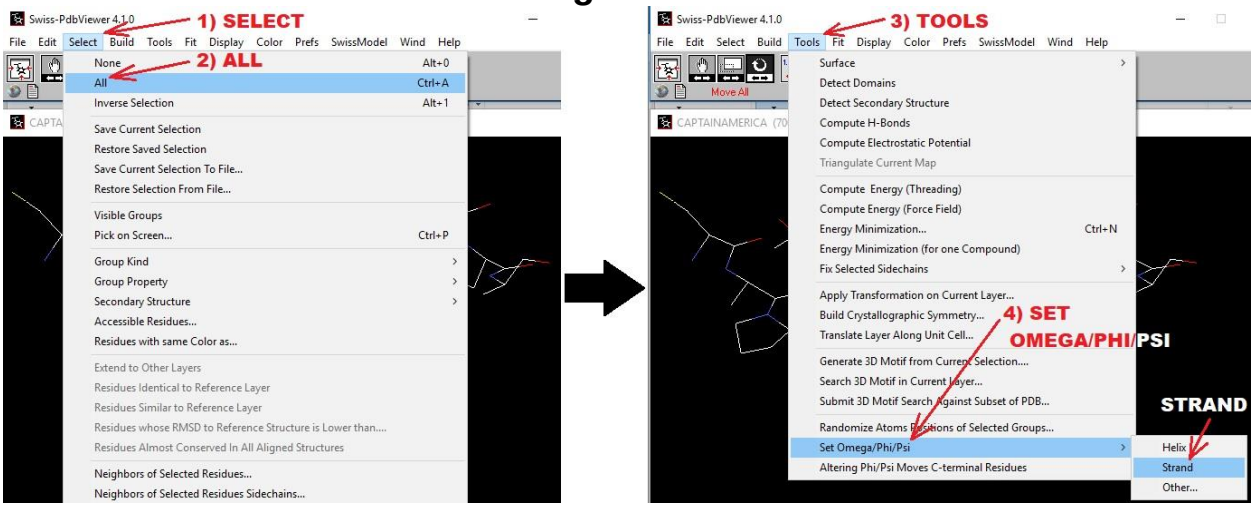
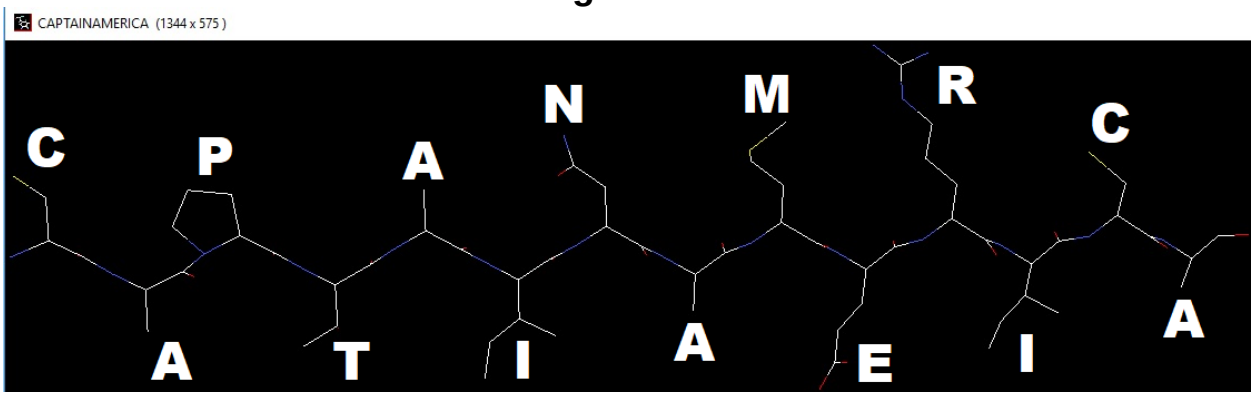


Figure 41 shows the results of modifying the alpha helical structure to a beta strand structure, with letters added to identify the component A.A.s.

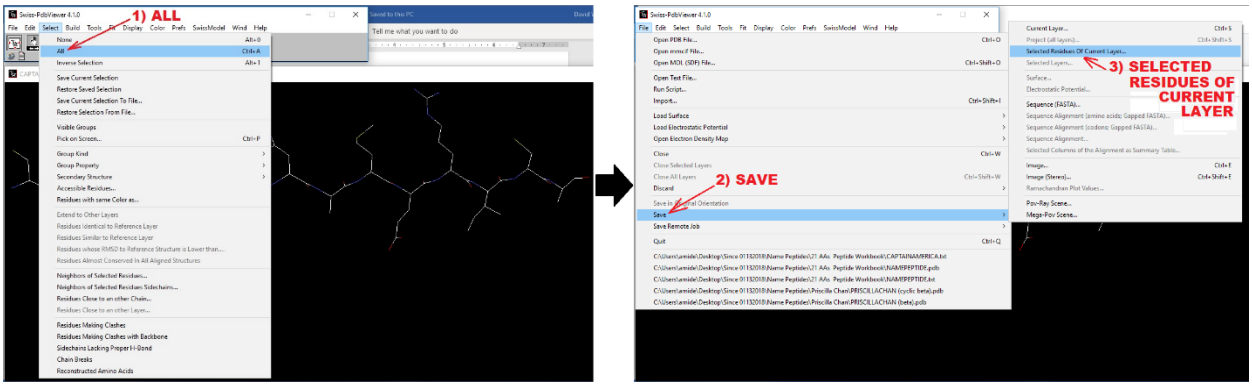
Figure 41.



The Deep View/Swiss-PdbViewer program has many features that enable the user to manipulate structure, to change A.A.s within the sequence, to add or remove A.A.s, to display various properties of the peptide, etc.

6) Save the 3D structure as a Protein Data Bank (pdb) file (Figure 42).

Figure 42.



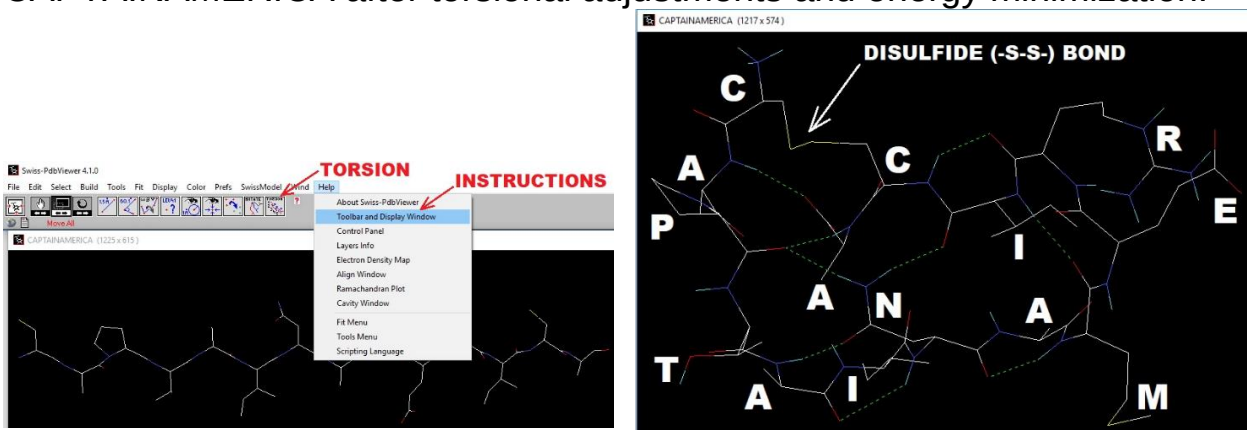
Give the new file the name of the peptide and be certain that it is saved as a ".pdb" file (e.g., CAPTAINAMERICA.pdb). A ".pdb" file can be used by other molecular modeling programs, such as RasMol (next section), to display the peptide in other formats.

As mentioned previously, peptides that contain one or more Cysteine (C), Selenocysteine (U), or both Cysteine and Selenocysteine A.A.s, can form chemical bonds between sulfur, selenium, or sulfur and selenium. This enables two molecules of peptide to link together, if there is only one C or one U in the peptide, or to form cyclic structures if the peptide contains two Cs or two Us, or a C and a U. As noted previously, CAPTAINAMERICA contains two Cs and could form a cyclic structure with a disulfide bond.

In order to create a cyclic structure, starting with the beta strand (flattened ribbon), bond angles within the beta strand structure must be

modified. This can be done using the torsion tool of the Deep View/Swiss-PdbViewer program (Figure 43), but the process will not be described here. Instructions on how to accomplish such adjustments can be obtained by clicking on “Help” (Figure 43). The cyclic structure shown in Figure 43 was created by starting with a beta strand structure, modifying bond angles until a disulfide bond formed (yellow lines in the figure), and then energy minimizing the structure (100,000 steps of steepest descent; final energy of -430 kJ/mol) to remove steric problems. In addition to the disulfide bond, it is stabilized by 9 internal hydrogen bonds (dashed green lines)]. The cyclic structure was then saved as “CAPTAINAMERICA(cyclic).pdb”.

Figure 43. (Left) CAPTAINAMERICA as a beta strand. (Right) Cyclic CAPTAINAMERICA after torsional adjustments and energy minimization.



Displaying Molecular Models with RasMol:

1) Download the RasMol program [32] and install it on your computer.

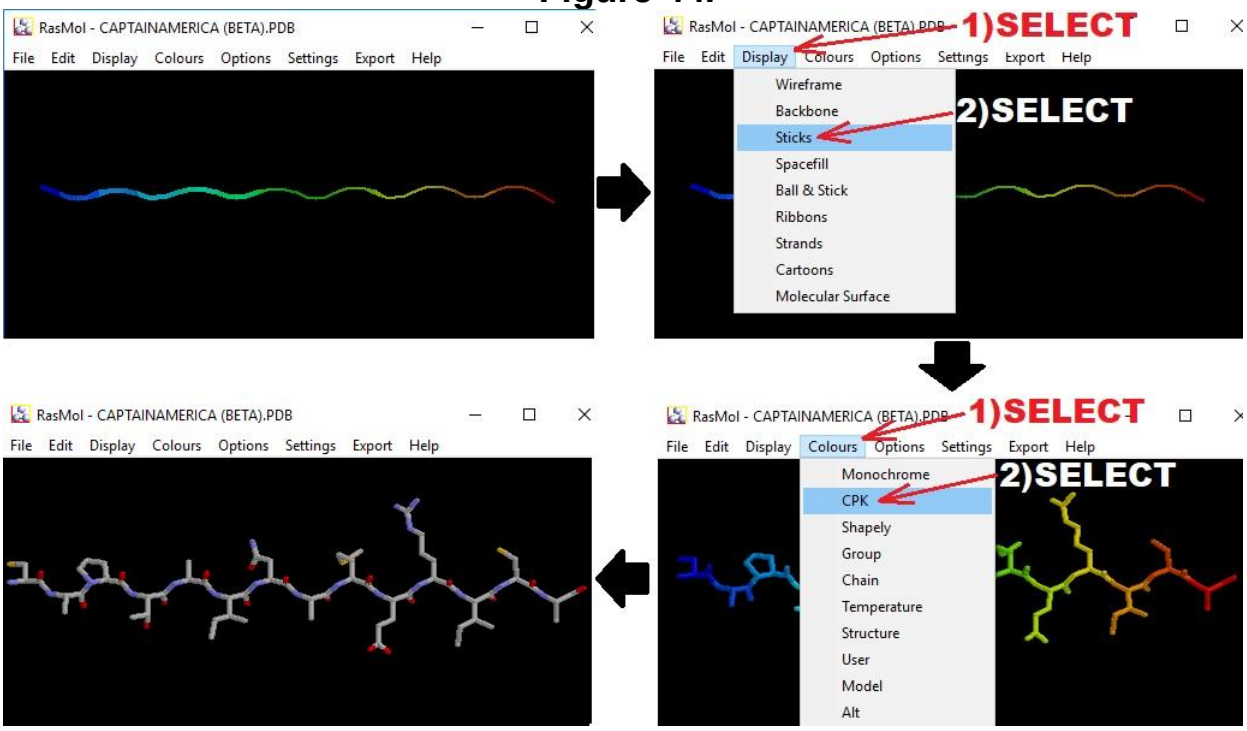
2) After the program is installed, you will be able to open it by simply double clicking on the file (e.g., CAPTAINAMERICA.pdb). A multicolored, ribbon model of your peptide will appear (Figure 44).

3) To better visualize the model, convert it to a stick figure (select “Display” and then “Sticks”) and then give it a “CPK” color scheme (select “Colours” and then “CPK”). The color scheme for atoms will now be gray = Carbon, white = Hydrogen (not shown in figure below), Red = oxygen, and Blue = nitrogen. In Figure 44, letters were added to indicate the positions of A.A.s.

4) Adjust the position of the stick figure model to get the best view, by left-clicking and holding down the mouse button and moving the mouse. You can translate (move up or down, and left or right) the model by right-clicking on the mouse button and holding it down while moving the mouse. You can

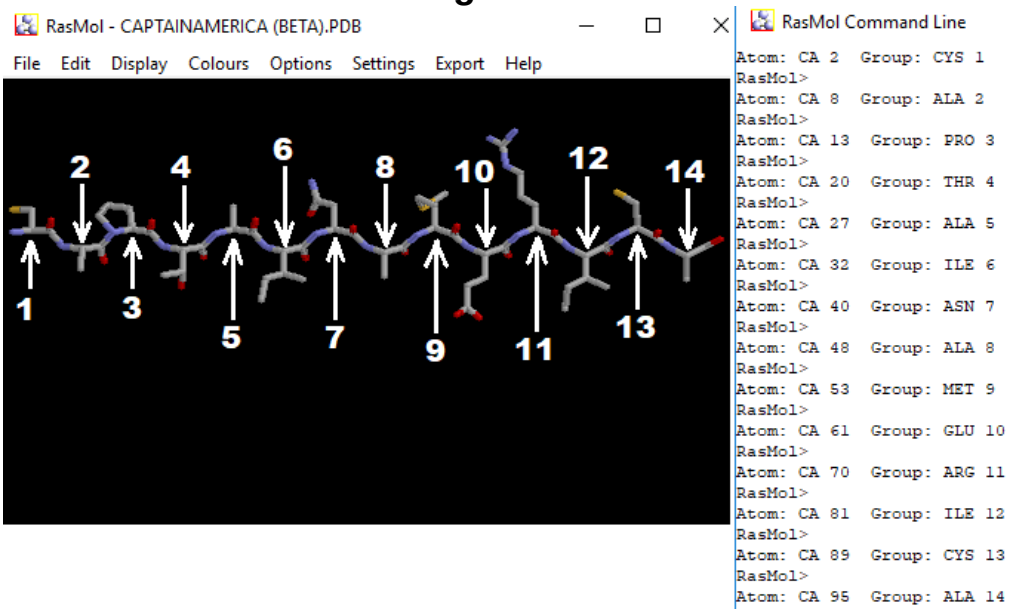
increase or decrease the size of the model by simultaneously holding down the “Shift” key on the keyboard while left clicking on and moving the mouse.

Figure 44.



5) You can identify each A.A. in the peptide, by clicking on any part of it in the model, and then referring to the RasMol command line screen (i.e., the second screen that always appears when viewing structures with RasMol) (Figure 45).

Figure 45.



6) You can convert the stick figure model to a space-filling model by clicking on “Display” and then “Spacefill” (Figure 46).

7) If you have the MS Windows Paint program on your computer, you can convert the model on your computer screen to a picture, and then modify the picture (e.g., by adding letters to it). Copy the screen by clicking the “PrtScr” key on the keyboard. Then open the MS Paint program on your computer, and “Paste” the copied screen into the Paint program. Then use the features of the Paint program to add letters and numbers (e.g., Figures 36-49).

Figure 46.

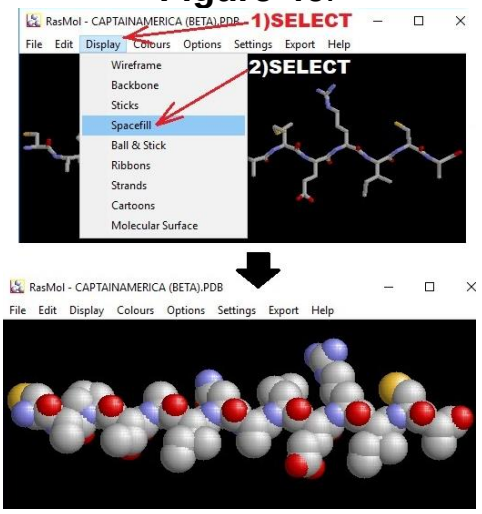
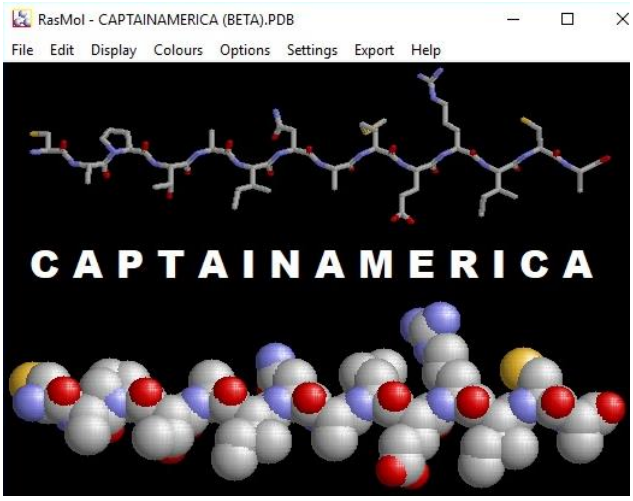


Figure 47.



9) The CAPTAINAMERICA(cyclic).pdb file was displayed with RasMol, as a “Sticks” diagram, using “CPK” “Colours”, and copied into the MS Paint program for the addition of lettering, an arrow, and charge (+/-) symbols (Figure 48).

Figure 48.

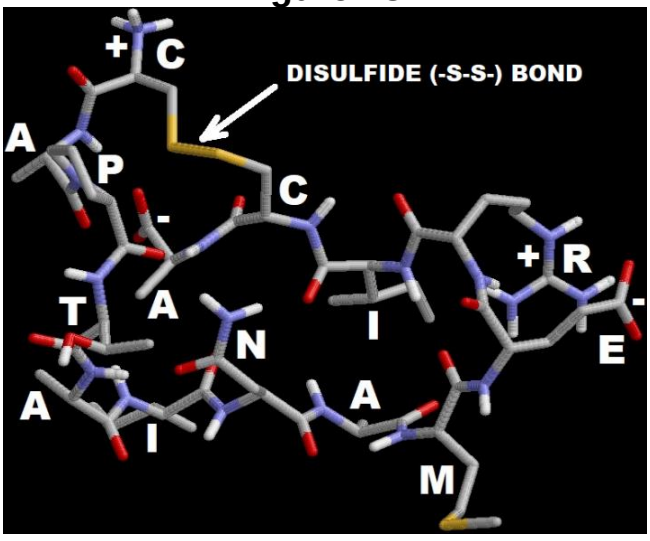
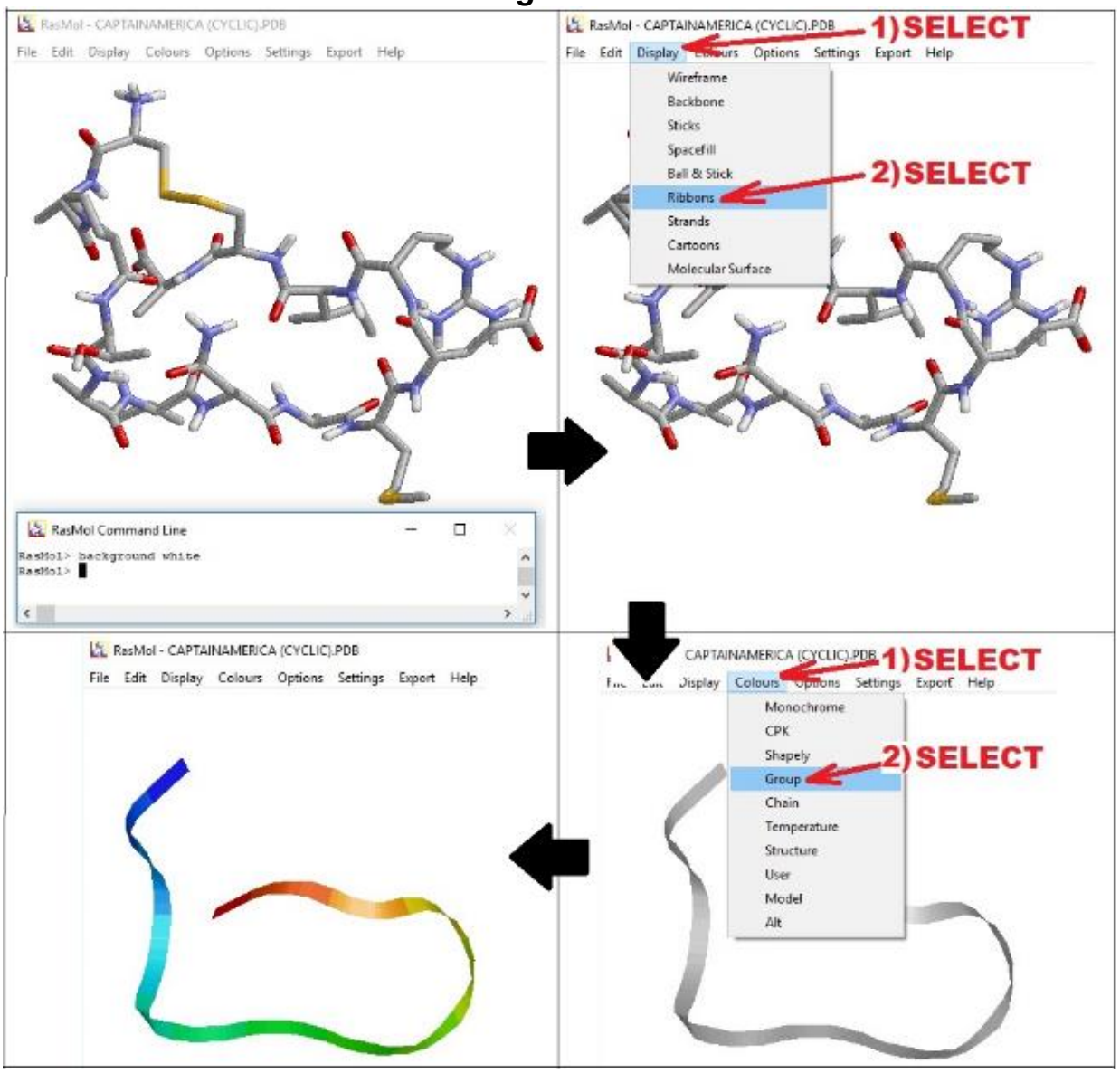


Figure 49

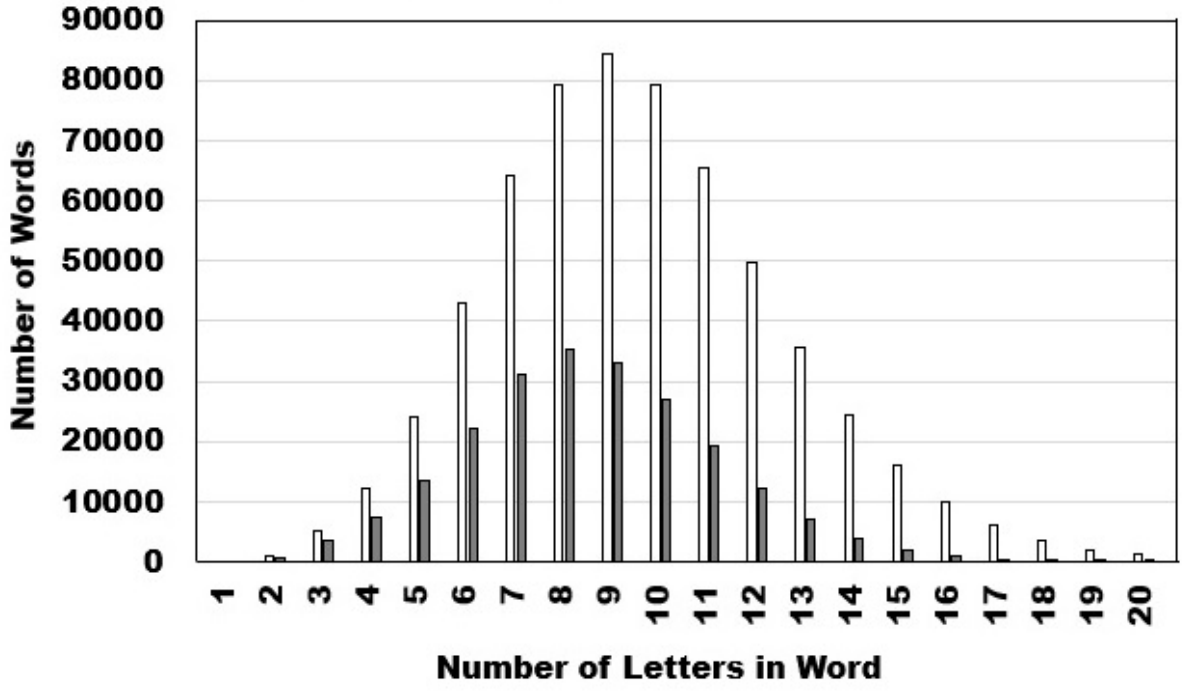


Modified views of cyclic CAPTAINAMERICA were prepared with the RasMol program (Figure 49). (Upper left) The background color of the figure on the previous page was changed from black to white by inserting the command "background white" into the RasMol Command Line screen. (Upper right ➔ lower right) The "Sticks" diagram was modified to a "Ribbons" diagram. (Lower right ➔ lower left). The "Ribbons" diagram was given a "Group" "Colour" to distinguish the ends of the peptide. In the final diagram (lower left), the purple end of the peptide is the beginning of the A.A. sequence, and the red end is the end of the sequence.

CONCLUDING REMARKS

This workbook, and the references therein [17-22], illustrates how it is possible to use the English language as a source for designing novel peptides. Since not all the letters of the English alphabet are used (i.e., B, J, O, X, and Z are omitted), the source is finite but still substantial in size. For example, the Lots of Words database currently (2/9/19) contains 608,641 words, of which 220,430 (36%) could be used to design peptides. In addition, the lengths of most of these words ranges from 2-16 letters (Figure 50), which would correspond to peptide lengths of 2-16 A.A.s, well within the range for synthesis by chemical methods [33]. As mentioned earlier, all peptides will exhibit some form of biological activity, and some peptides that are designed using the methods described in this workbook may be found to be useful in medicine and other fields.

Figure 50. Lengths of words, and their abundance, in the Lots of Words database (white bars) [17], and the subset of words that could be used to design peptides using the methods described in this workbook (gray bars).



ACKNOWLEDGEMENT

The author thanks a Ph.D. biochemist, who shall remain anonymous, for reviewing the manuscript and providing helpful suggestions.

REFERENCES:

1. Chemical composition of the human body. (https://en.wikipedia.org/wiki/Composition_of_the_human_body)
2. Nelson, D.L. and Cox, M.M., Lehninger Principles of Biochemistry, 6th ed., W.H. Freeman and Co., New York, pp. 58-69, 75-89, 102-104, 115-125.
3. National Center for Biotechnology Information (N.C.B.I.). (https://www.ncbi.nlm.nih.gov/protein/)
4. Hassan, M., Kjos, M., Nes, I.F., Diep, D.B., and Lotfipour, F. Natural antimicrobial peptides from bacteria: characteristics and potential applications to fight against antibiotic resistance. Journal of Applied Microbiology (2012) 113(4): 723-36.
5. Gross, E. and Morell, J.L. The structure of nisin. Journal of the American Chemical Society (1971) 93(18): 4634-4635.
6. Kaletta, C. and Entian, K.D. Nisin, a peptide antibiotic: cloning and sequencing of the nisA gene and posttranslational processing of its peptide product. Journal of Bacteriology (1989) 171(3): 1597-1601.
7. Steiner, H., Hultmark, D., Engström, A., Bennich, H., and Boman, H.G. Sequence and specificity of two antibacterial proteins involved in insect immunity. Nature (1981) 292(5820): 246-248.
8. Zasloff, M. Magainins, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. Proceedings of the National Academy of Sciences USA (1987) 84(15): 5449-5453.
9. Ganz, T., Selsted, M.E., Szklarek, D., Harwig, S.S., Daher, K., Bainton, D.F., and Lehrer, R.I. Defensins. Natural peptide antibiotics of human neutrophils. Journal of Clinical Investigation (1985) 76(4): 1427-1435.
10. Selsted, M.E., Harwig, S.S., Ganz, T., Schilling, J.W., and Lehrer, R.I. Primary structures of three human neutrophil defensins. Journal of Clinical Investigation (1985) 76(4): 1436-1439.
11. Zhang, Y., Doherty, T., Li, J., Lu, W., Barinka, C., Lubkowski, J., and Hong, M. Resonance assignment and three-dimensional structure determination of a human alpha-defensin, HNP-1, by solid-state NMR. Journal of Molecular Biology (2010) 397: 408-422.
12. International Union of Pure and Applied Chemistry–International Union of Biochemistry and Molecular Biology, Joint Commission on Biochemical Nomenclature (I.U.P.A.C.-I.U.B.M.B., J.C.B.N.). (https://iupac.org)

13. IUPAC-IUBMB, JCBN nomenclature and symbolism for amino acids. (https://www.qmul.ac.uk/sbcs/iupac/AminoAcid/)
14. US states and territories. (https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States)
15. Periodic table of the chemical elements. (https://www.vectorstock.com/royalty-free-vector/periodic-table-of-elements-vector-984320)
16. Reich, H.J. and Hondal, R.J. Why nature chose selenium. ACS Chemical Biology (2016) 11(4): 821-841.
17. One source of words is the Lot of Words website. The word list can be searched to find words that do not contain the letters, B, J, O, X, or Z (220558 words) (https://lotsofwords.com). There are also many other databases of names and phrases that can be found on the internet.
18. 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.
19. Wade, D. and Wade, S. The name game: use of words composed of letters of the English alphabet as a source of novel bioactive peptides. Biopolymers (2003) 71 (3): 322 (Abstract P082).
20. 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.
21. 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.
22. Wade, D., and Lea, M.A. The WALMART peptide. Wade Research Foundation Reports (2015) 7(1): 1-13.
23. Captain America (https://en.wikipedia.org/wiki/Captain_America)
24. N.C.B.I. Basic Local Alignment Search Tool (BLAST). (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins)
25. Transposase (https://en.wikipedia.org/wiki/Transposase)
26. *Azoarcus* sp. (https://en.wikipedia.org/wiki/Azoarcus)
27. Protein Data Bank (https://www.rcsb.org)
28. ExPASy Protparam tool (https://web.expasy.org/protparam/)
29. APD3: Antimicrobial peptide calculator and predictor (http://aps.unmc.edu/ap/prediction/prediction_main.php)

30. IACP: a sequence-based tool for identifying anticancer peptides (http://lin-group.cn/server/iacp)
31. Deep View/Swiss-PdbViewer program (https://spdbv.vital-it.ch)
32. RasMol (http://www.openrasmol.org)
33. 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. Proceedings of the National Academy of Sciences USA (1990) 87: 4761-65.

Originally published online January 27, 2019.
Revised versions published online February 5 and 9, 2019.