Independent Approach to Molecular Dynamics Model Using DNA Encoded Amino Acids with Goal of Simulating Short Peptides to Proteins

Eric A. Chapman

November 22, 2021

Declaration

All figures and plots presented in this paper were produced by myself using the python code in the appendix or presentation making software. Of course; unless any figures are cited, referenced or stated otherwise.

- ***Eric A. Chapman***
November 22, 2021

Abstract

to be done at some point

Contents

List of Figures **3**

List of Tables **3**

1 Introduction - **4**

2 Mathematics - **4**

2.1 Geometry - 4

2.1.1 Line Intersecting an 'n'-dimensional sphere - 4

3 The Natural Amino Acids - **6**

3.1 What We Will be Looking At - 6

4 Turning Structural Formulae into Tables - **7**

4.1 Rules to Turn into Table - 7

4.1.1 Tetrahedral - 7

4.1.2 Trigonal Planar - 7

4.2 Preserving Correct Chirality - 8

A Structural, Skeletal Formulae & Structure Tables - **9**

A.1 Alanine - 9

A.2 Arginine - 9

A.3 Asparagine - 10

A.4 Aspartic Acid - 10

A.5 Cysteine - 10

A.6 Glutamine - 11

A.7 Glutamic Acid - 11

A.8 Glycine - 12

A.9 Histidine - 12

A.10 Isoleucine - 12

A.11 Leucine - 13

A.12 Lysine - 13

A.13 Methionine - 14

A.14 Phenylalanine - 14

A.15 Proline - 15

A.16 Serine - 15

A.17 Threonine - 16

A.18 Tryptophan - 16

A.19 Tyrosine - . . . 17
A.20 Valine - . . . 17

References **19**

List of Figures

1 Encoding of Amino Acids to RNA Bases . . . 6

List of Tables

1 The 20 Natural Amino Acids with Their 3 & 1 Letter Codes . . . 7
2 Example of Tetrahedral Table Row . . . 8
3 Example of Trigonal Table Row . . . 8
4 Alanine Structure Table . . . 9
5 Arginine Structure Table . . . 9
6 Asparagine Structure Table . . . 10
7 Aspartic Acid Structure Table . . . 10
8 Cysteine Structure Table . . . 11
9 Glutamine Structure Table . . . 11
10 Glutamic Acid Structure Table . . . 11
11 Glycine Structure Table . . . 12
12 Histidine Structure Table . . . 12
13 Isoleucine Structure Table . . . 13
14 Leucine Structure Table . . . 13
15 Lysine Structure Table . . . 14
16 Methionine Structure Table . . . 14
17 Phenylalanine Structure Table . . . 15
18 Proline Structure Table . . . 15
19 Serine Structure Table . . . 16
20 Threonine Structure Table . . . 16
21 Tryptophan Structure Table . . . 17
22 Tyrosine Structure Table . . . 17
23 Valine Structure Table . . . 18

1 Introduction -

2 Mathematics -

2.1 Geometry -

2.1.1 Line Intersecting an 'n'-dimensional sphere -

In Cartesian coordinates the generalised equation for a circle with an arbitrary origin is as such:

$$(x - x_0)^2 + (y - y_0)^2 = \rho^2 \tag{1}$$

But for generality we cannot use the x_0 notation used above to denote the origin coordinates. Thus, we shall say these are elements of a wider vector of n length/dimension, and each element counts as a coordinate. We shall therefore call this vector $\underline{\omega}$. We shall also treat x and y and so on in the same way by saying these are part of a vector which we shall label $\underline{r}$. Then for the inside of the brackets above to match up with new vector notation, we need simply subtract $\underline{\omega}$ from $\underline{r}$. Then take the dot product with itself, which looks as such:

$$\rho^2 = (\underline{r} - \underline{\omega}) \cdot (\underline{r} - \underline{\omega}) \tag{2}$$

Which for convenience we shall change the above notation be like so:

$$\rho^2 = (\underline{r} - \underline{\omega})^{\cdot 2} \tag{3}$$

Next, instead of our nice and simple two dimensional line of $y = mx + c$, we need to use a parametric form to have a line in higher dimensions. Like the $+c$ being an offset for the 2D line, we generalise the offsets for each axis in a vector $\underline{\delta}$. We will also assume that the x, y, z... components are a function of some parameter t with each dimension having some gradient m_i for the ith direction. Which we shall also keep in a vector labelled $\underline{m}$. Thus for each dimension, we have equations of the form:

$$x_i = m_i t - \delta_i \quad \Leftrightarrow \quad t = \frac{x_i + \delta_i}{m_i} \tag{4}$$

Then for the last bit of housekeeping, the above x_i will be the elements of the $\underline{r}$ vector that corresponds to the points of the line.

We can then substitute this into the previous equation that we have for a generalised sphere, which gives us:

$$\rho^2 = (\underline{m}t - \underline{\delta} - \underline{\omega})^{\cdot 2} \tag{5}$$

$$\rho^2 = \underline{m}^{\cdot 2} t^2 - 2\underline{m} \cdot (\underline{\delta} + \underline{\omega})t + \underline{\delta}^{\cdot 2} + \underline{\omega}^{\cdot 2} + 2\underline{\delta} \cdot \underline{\omega} \tag{6}$$

Then letting:

$$\varphi_a = \underline{m}^{\cdot 2} \tag{7}$$

$$\varphi_b = \underline{m} \cdot (\underline{\delta} + \underline{\omega}) \tag{8}$$

$$\varphi_c = \underline{\delta}^{\cdot 2} + \underline{\omega}^{\cdot 2} + 2\underline{\delta} \cdot \underline{\omega} - \rho^2 \tag{9}$$

$$\Rightarrow \quad 0 = \varphi_a t^2 - 2\varphi_b t + \varphi_c \tag{10}$$

The above can then be used to gain the two values of t via use of the quadratic formula, which keeping the '-2' in the equation above makes the formula somewhat nicer to use.

$$t = \frac{1}{\varphi_a}\left(\varphi_b \pm \sqrt{\varphi_b^2 - \varphi_a \varphi_c}\right) \tag{11}$$

$$\Rightarrow \quad t = t^{\pm}$$

3 The Natural Amino Acids -

3.1 What We Will be Looking At -

In total, there are twenty such amino acids that are produced from DNA encoding. These are encoded as such:

Universal Genetic Code Chart

Messenger RNA Codons and Amino Acids for Which They Code

First base	Second base: U	Second base: C	Second base: A	Second base: G	Third base
U	UUU, UUC: PHE	UCU, UCC, UCA, UCG: SER	UAU, UAC: TYR	UGU, UGC: CYS	U
	UUA, UUG: LEU		UAA, UAG: STOP	UGA: STOP	C
				UGG: TRP	A
					G
C	CUU, CUC, CUA, CUG: LEU	CCU, CCC, CCA, CCG: PRO	CAU, CAC: HIS	CGU, CGC, CGA, CGG: ARG	U
			CAA, CAG: GLN		C
					A
					G
A	AUU, AUC, AUA: ILE	ACU, ACC, ACA, ACG: THR	AAU, AAC: ASN	AGU, AGC: SER	U
	AUG: MET or START		AAA, AAG: LYS	AGA, AGG: ARG	C
					A
					G
G	GUU, GUC, GUA, GUG: VAL	GCU, GCC, GCA, GCG: ALA	GAU, GAC: ASP	GGU, GGC, GGA, GGG: GLY	U
			GAA, GAG: GLU		C
					A
					G

Figure 1: Encoding of Amino Acids to RNA Bases

One will notice that the above corresponds to RNA encoding (UCAG) which uses *uracil*, *cytosine*, *adenine* and *guanine*. For the above to use the nucleotides (or bases) that DNA uses, all one would have to do is swap uracil for *thymine*, i.e. 'U' for 'T'. Thus also to give names to the three letter codes that one will see in the chart above, do have a look at **Table 22 [1]**.

We shall then -in the proceeding subsection- go through and see the two 2-dimensional structures that the represent these molecules. These two representations being the structural and skeletal formulae. After that we shall discuss how to turn these in to some form of data structure that preserves chirality, implicit structure and gives what could be the start of a mapping in coordinate space.

	Amino Acid	3 Letter Code	1 Letter Code
1	Alanine	Ala	A
2	Arginine	Arg	R
3	Asparagine	Asn	N
4	Aspartic Acid	Asp	D
5	Cysteine	Cys	C
6	Glutamine	Gln	Q
7	Glutamic Acid	Glu	E
8	Glycine	Gly	G
9	Histidine	His	H
10	Isoleucine	Ile	I
11	Leucine	Leu	L
12	Lysine	Lys	K
13	Methionine	Met	M
14	Phenylalanine	Phe	F
15	Proline	Pro	P
16	Serine	Ser	S
17	Threonine	Thr	T
18	Tryptophan	Trp	W
19	Tyrosine	Tyr	Y
20	Valine	Val	V

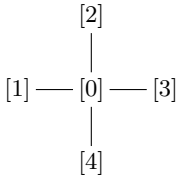
Table 1: The 20 Natural Amino Acids with Their 3 & 1 Letter Codes

4 Turning Structural Formulae into Tables -

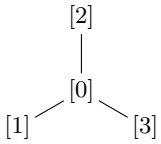
4.1 Rules to Turn into Table -

Right, so now we shall endeavour to build some data structure or framework where we can represent the structural formulae (as that shows all elements present in the molecule) as some form of table. We shall first suppose the two bond structures that we will most certainly see. Those being tetrahedral and trigonal which look like so:

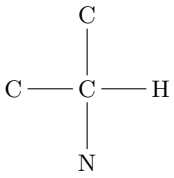
4.1.1 Tetrahedral -

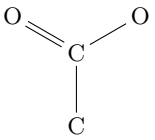


4.1.2 Trigonal Planar -



Now from looking at the above, there are at most five elements in and around the central element (depicted as [0]) we can therefore say that the table will have five columns. Suppose we have the following two cases:





Where one may recognise the right hand example to be similar to the carbon found amongst a carboxylic acid group and left to what we will soon refer to as the 'central' carbon of a given amino acid. So how would the above translate to their allocated 'row' within a table. Well, the tetrahedral case would look as such:

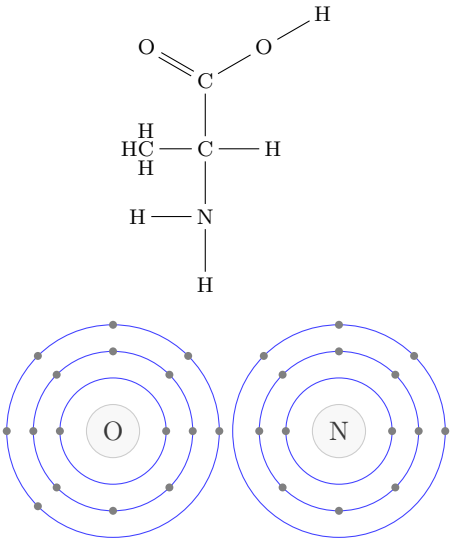
Path/Bond Index	0	1	2	3	4
0	C	-C	-C	-H	-N

Table 2: Example of Tetrahedral Table Row

Path/Bond Index	0	1	2	3	4
0	C	=O	-O	-C	X

Table 3: Example of Trigonal Table Row

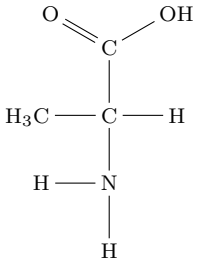
Here we can see that to interpret the table, one must accept a *rule of end-points*. In other words, we need to use 'X' as the identifier of a trigonal bonding around the atom in question of that row and is therefore an end point. Plus, in terms of what an atom can be bonded to, we must add '=O' and '-H' to the list of end points. It is best now to do this, not just for one row of the table but for an entire molecule, and so we shall construct the table for alanine. Recalling the structure that was presented earlier, to make the table, we will first need to present its structure in a more fuller fashion. i.e.

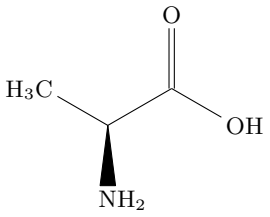


4.2 Preserving Correct Chirality -

A Structural, Skeletal Formulae & Structure Tables -

A.1 Alanine -

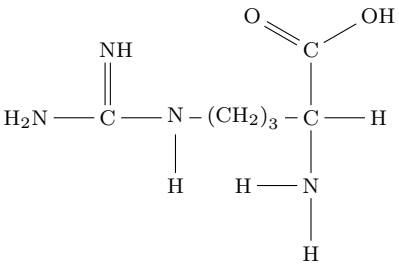


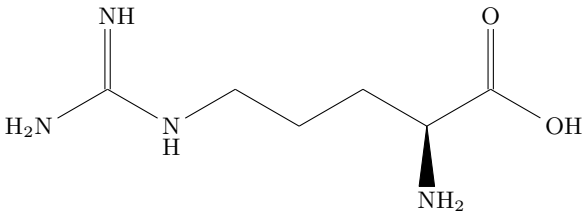


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-H	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
022	O	-H	ee	ee	-C

Table 4: Alanine Structure Table

A.2 Arginine -

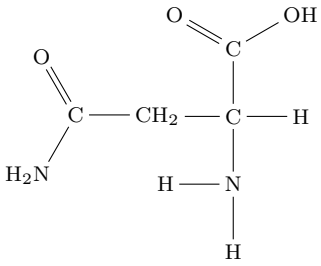


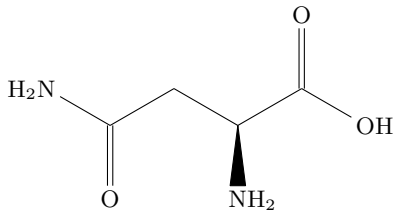


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	-H	-C	-H
022	O	-H	ee	ee	-C
0111	C	-N	-H	-C	-H
01111	N	-H	ee	-C	-H
011111	C	-N	=N	-N	X
0111111	N	-H	ee	-C	-H
0111112	N	-H	ee	=C	X

Table 5: Arginine Structure Table

A.3 Asparagine -

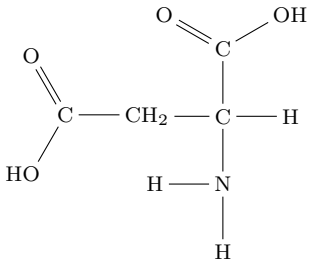


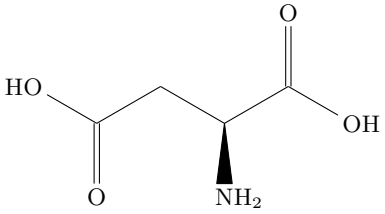


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-N	=O	-C	X
022	O	-H	ee	ee	-C
0111	N	-H	ee	-C	-H

Table 6: Asparagine Structure Table

A.4 Aspartic Acid -

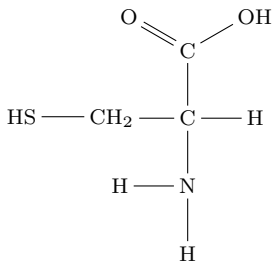


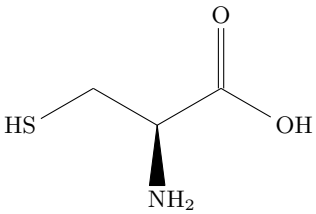


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-O	=O	-C	X
022	O	-H	ee	ee	-C
0111	O	-H	-C	ee	ee

Table 7: Aspartic Acid Structure Table

A.5 Cysteine -

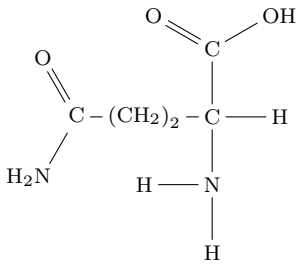


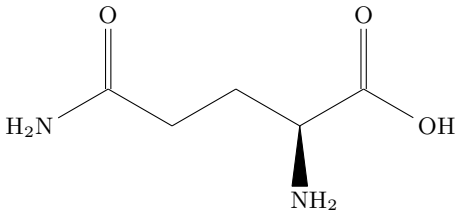


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-S	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	S	-H	ee	-C	ee
022	O	-H	ee	ee	-C

Table 8: Cysteine Structure Table

A.6 Glutamine -

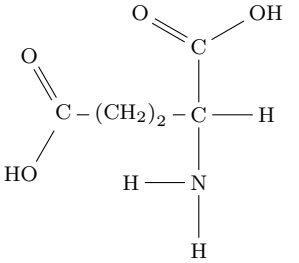


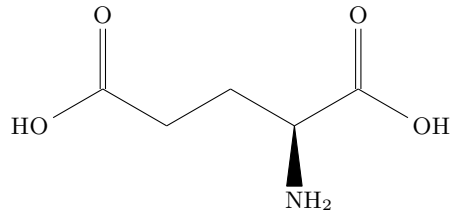


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	-H	-C	-H
022	O	-H	ee	ee	-C
0111	C	-N	=O	-C	X
01111	N	-H	ee	-C	-H

Table 9: Glutamine Structure Table

A.7 Glutamic Acid -

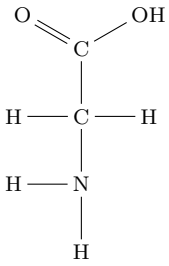


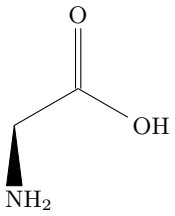


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	-H	-C	-H
022	O	-H	ee	ee	-C
0111	C	-O	=O	-C	X
01111	O	-H	-C	ee	ee

Table 10: Glutamic Acid Structure Table

A.8 Glycine -

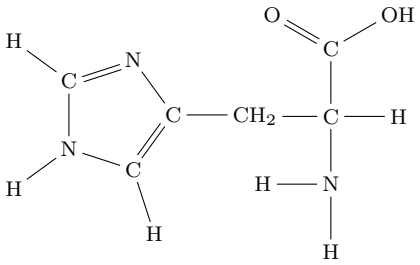


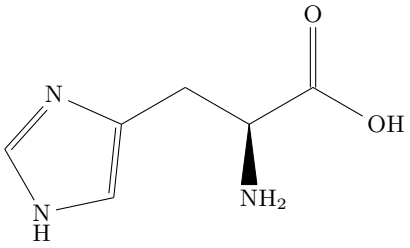


	[0]	[1]	[2]	[3]	[4]
0	C	-H	-C	-H	-N
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
022	O	-H	ee	ee	-C

Table 11: Glycine Structure Table

A.9 Histidine -

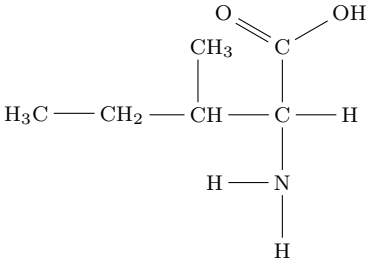


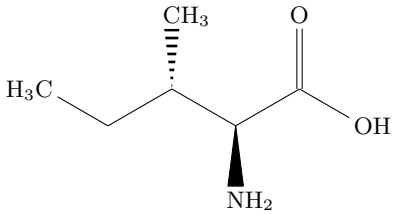


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	=C	-N	-C	X
022	O	-H	ee	ee	-C
0111	C	-N	=C	-H	X
0112	N	=C	ee	-C	X
01111	N	ee	-H	-C	-C
01121	C	-H	=N	-N	X

Table 12: Histidine Structure Table

A.10 Isoleucine -

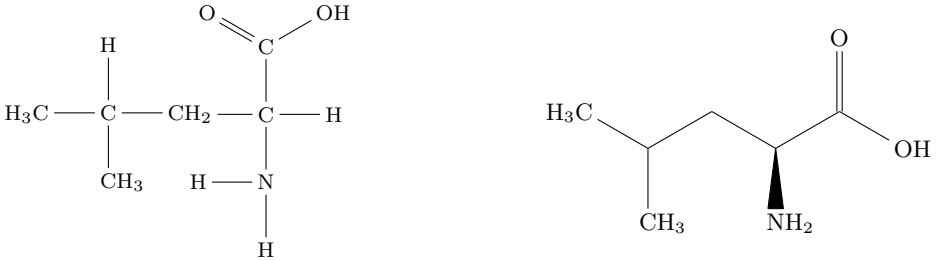




	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-C	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	-H	-C	-H
012	C	-H	-H	-H	-C
022	O	-H	ee	ee	-C
0111	C	-H	-H	-C	-H

Table 13: Isoleucine Structure Table

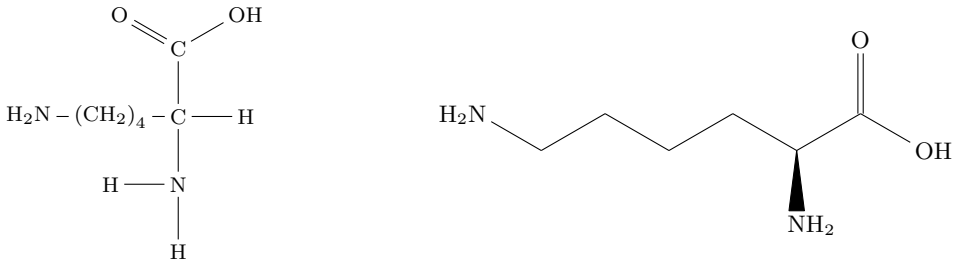
A.11 Leucine -



	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	-H	-C	-C
022	O	-H	ee	ee	-C
0111	C	-H	-H	-C	-H
0114	C	-H	-C	-H	-H

Table 14: Leucine Structure Table

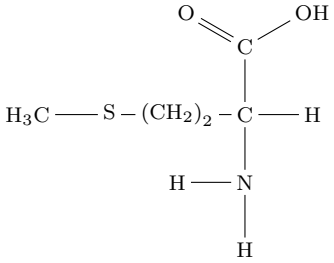
A.12 Lysine -

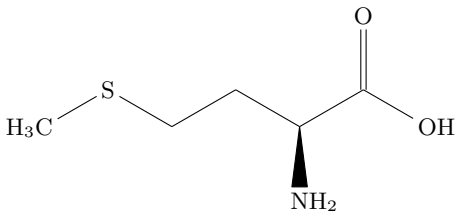


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	-H	-C	-H
022	O	-H	ee	ee	-C
0111	C	-C	-H	-C	-H
01111	C	-N	-H	-C	-H
011111	N	-H	ee	-C	-H

Table 15: Lysine Structure Table

A.13 Methionine -

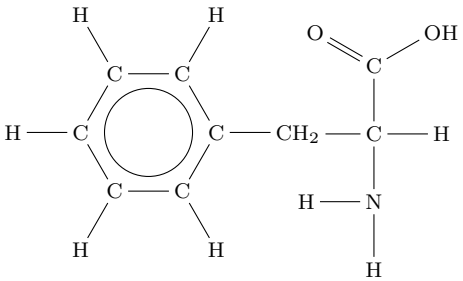


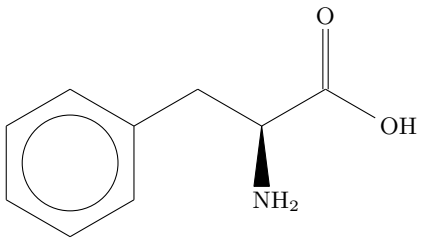


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-S	-H	-C	-H
022	O	-H	ee	ee	-C
0111	S	-C	ee	-C	ee
01111	C	-H	-H	-C	-H

Table 16: Methionine Structure Table

A.14 Phenylalanine -

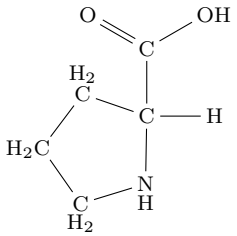


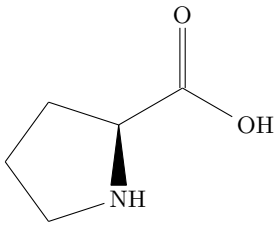


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	=C	-C	X
022	O	-H	ee	ee	-C
0111	C	=C	-C	-H	X
0112	C	-C	-H	=C	X
01111	C	-H	-C	=C	X
01121	C	=C	-H	-C	X
011112	C	-H	=C	-C	X

Table 17: Phenylalanine Structure Table

A.15 Proline -

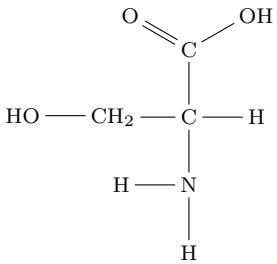


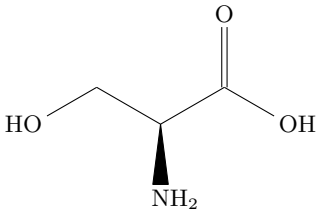


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-H	-H	-C	-C
02	C	=O	-O	-C	X
04	N	-C	-C	ee	-H
014	C	-H	-H	-C	-C
022	O	-H	ee	ee	-C
041	C	-H	-C	-N	-H
0144	C	-H	-C	-C	-H

Table 18: Proline Structure Table

A.16 Serine -

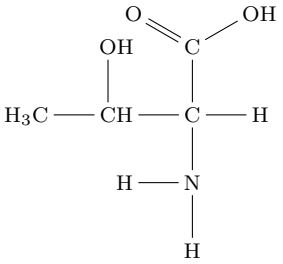


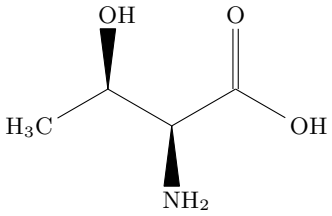


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-O	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	O	-H	ee	-C	ee
022	O	-H	ee	ee	-C

Table 19: Serine Structure Table

A.17 Threonine -

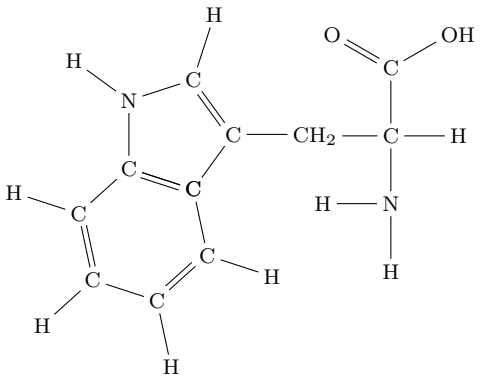


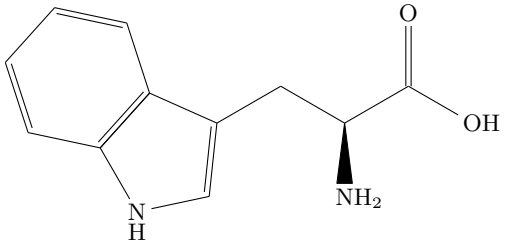


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-O	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-H	-H	-C	-H
012	O	ee	-H	ee	-C
022	O	-H	ee	ee	-C

Table 20: Threonine Structure Table

A.18 Tryptophan -

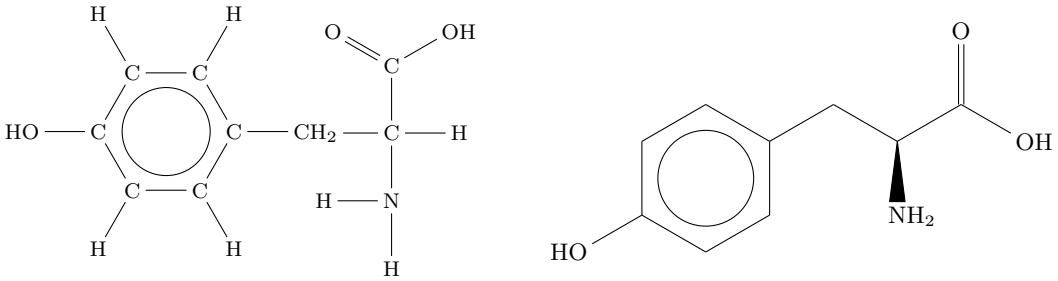




	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	=C	-C	X
022	O	-H	ee	ee	-C
0111	C	=C	-C	-C	X
0112	C	-N	-H	=C	X
01121	N	-H	ee	-C	-C
01111	C	-C	-N	=C	X
01113	C	=C	-C	-H	X
011111	C	-H	-C	=C	X
011131	C	-C	=C	-H	X
0111113	C	-H	=C	-C	X

Table 21: Tryptophan Structure Table

A.19 Tyrosine -

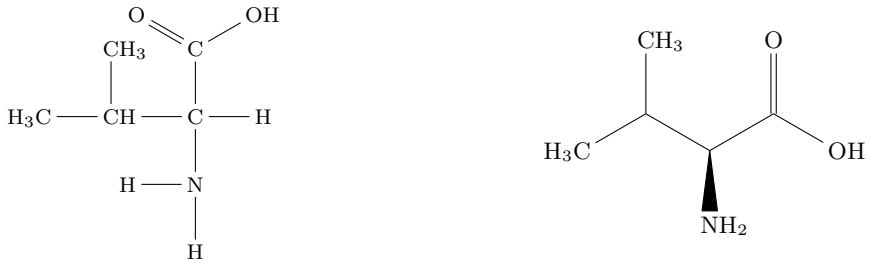


	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-H	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-C	=C	-C	X
022	O	-H	ee	ee	-C
0111	C	=C	-C	-H	X
0112	C	-C	-H	=C	X
01111	C	-H	-C	=C	X
01121	C	=C	-H	-C	X
011112	C	-O	=C	-C	X
0111121	O	-H	ee	-C	ee

Table 22: Tyrosine Structure Table

tyr

A.20 Valine -



	[0]	[1]	[2]	[3]	[4]
0	C	-C	-C	-H	-N
01	C	-C	-C	-C	-H
02	C	=O	-O	-C	X
04	N	-H	-C	ee	-H
011	C	-H	-H	-C	-H
012	C	-H	-H	-H	-C
022	O	-H	ee	ee	-C

Table 23: Valine Structure Table

References

[1] "Transcription vs Translation." Diffen.com. Diffen LLC, n.d. Web. 28 May 2021. < https://www.diffen.com/difference/Transcription_vs_Translation >

[2] Authors. Year of Publish. *Title*. edition/printing. City of publish: printing office/press.