Generalized macro level models of amino acid sequences using passive electrical circuits

R. Marshall

Computer Science and Technology Department
Plymouth State University NH 03264 USA
rgmarshall@plymouth.edu

Abstract. In this paper, we discuss four types of electrical network circuits have been used for modeling amino acid sequences at the gross molecular level. The standard 20 amino acids have been partitioned into mutually exclusive partitions based on their chemical and structural properties and each partition is assigned a distinct electrical circuit. This enables the construction of electrical networks for representing any amino acid sequence. These networks have been analyzed to establish putative relationships between a network's electrical response and known behavior of protein sequences. To aid in the analysis of circuit response, we have also developed a unique visualization technique. Similar circuits have been used to model DNA/RNA sequences as well as non-biological strings which have applications in message encryption and computer security.

Keywords: Modeling and simulation, bioinformatics, passive electrical circuits

1 Introduction

Elaborate models of genes have been developed by various authors and organizations for trying to understand the intricate biological, chemical, and structural interactions between the atoms that constitute the gene. One of the motivating factors in developing these models has been to use simulation instead of conducting actual wet lab experiments and field studies necessary for testing the efficacy of new drugs, understanding drug interactions, studying the effect of mutations, etc. . These models, while comprehensive, nevertheless require a significant amount of computational time and effort. The network circuits for modeling gene sequences at the macro molecular level and the underlying methodology discussed in this paper represent an attempt to simplify the standard technique of modeling biological structures at the atomic level.

2 Mapping strings to electrical circuits - a generalized approach

Given a set of characters $c = \{c_1, c_2, \ldots c_k\}$, a set $d = \{d_1, d_2, \ldots d_l\}$ where d_i ($1 <= i <= l$) are mutually exclusive subsets of c, a set of electrical circuits $e = \{e_1, e_2, \ldots e_n\}$,

a set f = {f_1, f_2, … f_m} where f_i (1 <= *k* <= m) are mutually exclusive subsets of e, and a string s = $s_1 s_2 s_3$ … s_p where s_i is an element of set c (and, therefore, in exactly one set of d), we define the mapping of a string to a circuit as the set {g_1, g_2, .. g_q} where each gi is a tuple (u, v) where u is in d and v is in f. In order to form sets d and f, two independent set membership criteria are required, one for d and one for f. Set membership requirements are application dependent and since this issue has no immediate bearing, it will be addressed later in the paper.

Assume we are given an m-character string s = c_1 c_2 c_3 …… c_{k-2} c_{k-1} c_k …. c_m. Denote the impedance of the combined circuit representing the first k-2 characters by Z^{k-2} and the impedances of the single circuits representing the characters c_{k-1} and c_k by Z_{k-1} and Z_k, respectively. The exact makeup of the individual circuits is not relevant at this point but will be addressed section 3. In a fully serial connection (S), each new circuit representing the next character in the string s is connected serially to the overall circuit representing the previous characters. Likewise, in a fully parallel connection (P), each new circuit representing the next character in the string s is connected in parallel with the overall circuit representing the previous characters. In a serial-parallel (SP) connection scheme, the individual circuits representing the next two characters in the string s are connected as follows – the first one in series, the second one in parallel with the overall circuit for the string of length (k-2). Likewise, in a parallel–serial (PS) connection scheme, the circuits representing the next two characters in the string s are connected, the first one in parallel and the second one in series with the overall circuit for the string of length (k-2). In the four options which are depicted below, the underlined portions of the expressions are represented by a single superscripted Z followed by subscripted Z elements.

Serial (S) : $\underline{Z(c_1) + Z(c_2) + Z(c_3) \ \ldots. \ Z(c_{k-1})} + Z(c_k) + Z(c_{k+1}) \ \ldots. \ + \ Z(c_m)$

$= \qquad Z^{k-1} \qquad + \qquad Z_k \ + Z(c_{k+1}) \ + \ldots.$

Parallel (P) : $\underline{Z(c_1) \parallel Z(c_2) \parallel Z(c_3) \ \ldots. \ Z(c_{k-1})} \parallel Z(c_k) \parallel Z(c_{k+1}) \ \ldots. \parallel \ Z(c_m)$

$= \qquad Z^{k-1} \qquad + \qquad Z_k + Z(c_{k+1}) \ + \ldots..$

Serial-parallel (SP):

$(\ldots(((\underline{Z(c_1) + Z(c_2)) \parallel Z(c_3)}) + Z(c4)) \ \ldots. \ \underline{Z(c_{k-2})} + Z(c_{k-1})) \parallel Z(c_k)) \ \ldots. \ Z(c_m)$

$= \qquad Z_{k-2} \qquad + \qquad Z_{k-1} \quad \parallel \quad Z_k + \ldots.$

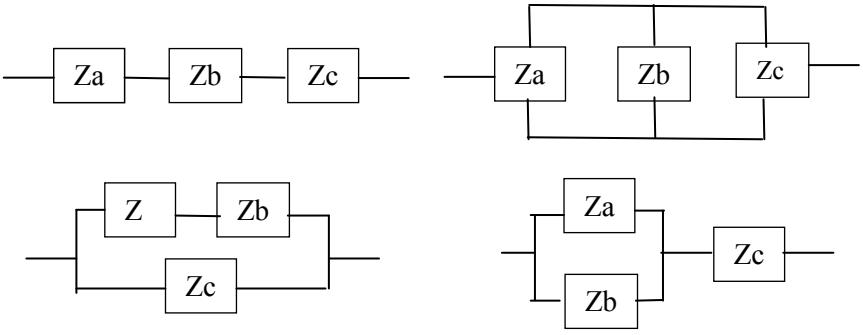
Parallel-serial (PS):

$(\ldots(((\underline{Z(c_1) \parallel Z(c_2)) + Z(c_3)}) \parallel Z(c4)) + \ \ldots. \ \underline{Z(c_{k-2})} \parallel Z(c_{k-1})) + Z(c_k)) \ \ldots. \ Z(c_m)$

$= \qquad Z^{k-2} \qquad \parallel \quad Z_{k-1} \quad + \quad Z_k + \ldots..$

The impedances of both the SP and PS versions depend on whether the message length is odd or even. To obtain the overall impedance of a string (or substring) of

however many characters is quite simple. The new impedance resulting from connecting the circuit representing the kth character to the existing circuit for the previous k-1 characters is given below for the S and P type connections. N and D stand for the numerator and denominator, respectively with appropriate subscripts or superscripts. A pictorial representation of the four models is shown below in the order - left top (S), right top (P), bottom left (SP) and bottom right (PS).



Serial:

$$Z^k = Z^{k-1} + Z_k$$

$$N^k / D^k = N^{k-1} / D^{k-1} + N_k / D_k$$

or, equivalently,

$$N^k = N^{k-1} D_k + D^{k-1} N_k \quad (1a)$$

$$D^k = D^{k-1} D_k \quad (1b)$$

Parallel:

$$Z^k = Z^{k-1} \parallel Z_k = Z^{k-1} Z_k / (Z^{k-1} + Z_k)$$

$$N^k / D^k = N^{k-1} N_k / (N^{k-1} Dk + D^{k-1} N_k)$$

or, equivalently,

$$N^k = N^{k-1} N_k \quad (2a)$$

$$D^k = N^{k-1} D_k + D^{k-1} N_k \quad \{\text{same as (1a)}\} \quad (2b)$$

The overall impedance of the circuit representing the string with k characters for the SP and PS models are shown below.

Serial-parallel:

$$Z^k = (Z^{k-2} + Z_{k-1}) \parallel Z_k = (Z^{k-2} + Z_{k-1}) Z_k / (Z^{k-2} + Z_{k-1} + Z_k)$$

$$= Z^{k-1} Z_k / (Z^{k-1} + Z_k) = N^{k-1} N_k / (N^{k-1} Dk + D^{k-1} N_k)$$

or, equivalently,

$$N^k = N^{k-1} N_k \quad (3a)$$

$$D^k = N^{k-1} D_k + D^{k-1} N_k \quad (3b)$$

Parallel-serial:

$$Z^k = (Z^{k-2} \parallel Z_{k-1}) + Z_k = Z^{k-2} Z_{k-1} / (Z^{k-2} + Z_{k-1}) + Z_k$$

$$= Z^{k-1} + Z_k = (N^{k-1} D_k + D^{k-1} N_k) / D^{k-1} D_k$$

or, equivalently,

$$N^k = N^{k-1} D_k + D^{k-1} N_k \quad (4a)$$

$$D^k = D^{k-1} D_k \quad (4b)$$

Note that the expressions for the four models are coupled recurrence equations. Even though expressions (1a), (1b) and (3a), (3b) are identical, the interpretations of N^{k-1} and D^{k-1} are quite different. A similar comment applies to expressions (2a), (2b) and (4a), (4b). We present a simple example to show why and how they are different.

For a string of length 1, the impedance expressions for S, P, SP and PS are all the same.

$$Z^1 = Z_1$$
$$N^1 = N_1 \text{ and } D^1 = D_1 \qquad (5a, b)$$

For strings of length 2, we have

S:

$$Z^2 = Z^1 + Z_2 \; ; \; N^2 / D^2 = N^1 / D^1 + N_2 / D_2$$
$$N^2 = N^1 D_2 + D^1 N_2 = N_1D_2 + D_1N_2 \qquad (6a)$$
$$D^2 = D^1 D_2 = D_1D_2 \qquad (6b)$$

P:

$$Z^2 = Z^1 \parallel Z_2 \; ; \; N^2 / D^2 = N^1 / D^1 \parallel N_2 / D_2 = N^1 N_2 / (N^1D_2 + D^1N_2)$$
$$N^2 = N^1 N_2 = N_1N_2 \qquad (7a)$$
$$D^2 = N^1D_2 + D^1N_2 = N_1D_2 + D_1N_2 \qquad (7b)$$

SP is the same as P and PS is the same as S.

For strings of length 3, we get

S:

$$Z^3 = Z^2 + Z_3 = N^2 / D^2 + N_3 / D_3 = (N^2 D_3 + D^2 N_3) / D^2 D_3$$
$$N^3 = N^2 D_3 + D^2 N_3 = (N_1D_2 + D_1N_2)D_3 + D_1D_2N_3 \qquad (8a)$$
$$D^3 = D^2 D_3 = D_1D_2D_3 \qquad (8b)$$

P:

$$Z^3 = Z^2 \parallel Z_3 = N^2 / D^2 \parallel N_3 / D_3 = N^2 N_3 / (N^2 D_3 + D^2 N_3)$$
$$N^3 = N^2 N_3 = N_1N_2 N_3 \qquad (9a)$$
$$D^3 = N^2 D_3 + D^2 N_3 = N_1N_2 D_3 + (N_1D_2 + D_1N_2)N_3 \qquad (9b)$$

SP:

$$Z^3 = (Z^1 + Z_2) \parallel Z_3 = Z^2 \parallel Z_3 = N^2 N_3 / (N^2D_3 + D^2N_3)$$
$$N^3 = N^2 N_3 = (N_1D_2 + D_1N_2)N_3 \qquad (10a)$$
$$D^3 = N^2 D_3 + D^2 N_3 = (N_1D_2 + D_1N_2)D_3 + D_1D_2N_3 \qquad (10b)$$

PS:

$$Z^3 = (Z^1 \parallel Z_2) + Z_3 = Z^2 + Z_3 = (N^2 D_3 + D^2 N_3) / D^2 D_3$$
$$N^3 = N^2 D_3 + D^2 N_3 = N_1N_2D_3 + (N_1D_2 + D_1N_2)N_3 \qquad (11a)$$
$$D^3 = D^2 D_3 = (N_1D_2 + D_1N_2)D_3 \qquad (11b)$$

The comments made earlier on the apparently equivalent impedances is borne out by looking at the detailed expressions for S (8a and 8b) and SP (10a and 10b); likewise, for P (9a and 9b) and PS (11a and 11b).

2.1 Closed form expressions for the overall impedance for the S and P models

For the S and P connection models, simple closed form expressions for the overall impedance in terms of magnitude and phase can be obtained for the entire message of m characters. However, this is not the case for the other two models. The closed form expression for the impedance magnitude and phase for the S model is obtained as follows.

$$Z = Z(c_1) + Z(c_2) + \ldots. Z(c_{m-1}) + Z(c_m) =$$

$$= [Re(Z(c_1)) + ... + Re(Z(c_m))] + j.[Im(Z(c_1)) + \ldots Im(Z(c_m))] = |Z| e^{-i\theta}$$

$$\text{where magnitude } |Z| = |\{ [\Sigma \text{ Real}(Z(c_k))]^2 + [\Sigma \text{ Imag}(Z(c_k))]^2 \}^{1/2} \quad (12a)$$

$$\text{and phase } \theta = \tan\text{-}1[([\Sigma \text{ Imag}(Z(c_k))]^2 / ([\Sigma \text{ Real}(Z(c_k))]^2)] \quad (12b)$$

Likewise, the closed form expression for the P model is

$$Z = Z(c_1) \,||\, Z(c_2) \,||\, \ldots. Z(c_{m-1}) \,||\, Z(c_m) = \pi_I Z_i / \Sigma_i \pi_{k\ k!=i} Z_k$$

$$|Z| = |\pi_I Z_i| / \Sigma_i \pi_{k\ k!=i} Z_k| \quad (13a)$$

$$\theta = \tan\text{-}1\ [\text{Imag}(|\pi_I Z_i) / \text{Real}(\pi_i Z_i) - \text{Imag}(|\Sigma_i \pi S_{k\ k!=i} Z_k) / \text{Real}(\Sigma_i \pi_{k\ k!=i} Z_k)] \quad (13b)$$

It should be mentioned in passing that (13a) is a continued fraction expansion in the complex domain.

2.2 Fibonacci sequences as a special case

If we have a message in which all characters in the message are the same or the characters are from the same tri-graph or character subset, then all the circuits associated with the message (assuming the tri-graphs or character subsets are mutually exclusive) will be the same. Consequently, the overall impedances of the four models reduce to simple expressions. Let m be the message length and Z the impedance of the circuit associated with the character(s).

S: $$Z^m = Z + Z + \ldots.. + Z = mZ \quad (14a)$$

P: $$Z^m = Z \,||\, Z \,||\, Z \ldots. \,||\, Z = Z / m \quad (14b)$$

SP: $$Z^m = (Z + Z) \,||\, Z) + \ldots\ldots\ldots Z =$$
$$F_m / F_{m+1} \text{ (m odd) or } F_{m+1} / F_m \text{ (m even)} \quad (14c)$$

PS: $$Z^m = (Z \,||\, Z) + Z) \,||\ldots\ldots Z = F_{m+1} / F_m \text{ (m even) or } F_m / F_{m+1} \text{ (m odd)} \quad (14d)$$

In the expressions for SP and PS above, F_n is the nth Fibonacci number. Of course, F_{m+1} / F_m and F_m / F_{m+1} for large m yield the golden ratio and its inverse. Later on in the results section of the paper we show the direct connection between Fibonacci fractions and the circuit models used in representing poly amino acids. Research articles that have explored aspects of the connection between Fibonacci numbers and the human genome include Perez [10] and Yamagishi and Shimabukuro [12].

3 Elementary RLC circuit models and associated impedance expressions

In the previous section, we ignored the compositional aspects of the different circuits and merely characterized the circuits by their impedances Z. Here we focus on a possible set of circuits which use only passive circuit elements – resistors, inductors and capacitors.

3.1 Elementary RLC circuits and interconnectivities

A simple passive electrical circuit falls into one of several categories – a) purely resistive or R circuit comprising of a single resistor (if more than one resistor is present and these are connected in series or in parallel or both, they can all be replaced by a single resistor), b) an RL circuit which consists of a single resistor connected either in series or in parallel with a single inductor (RL circuit) , b) a purely inductive or L circuit (if multiple inductors are present, they can all be replaced a single inductor), c) a purely capacitive or C circuit (similar comments as in (a) or (b) if multiple capacitors are present), d) an RL circuit in which a single resistor in connected either in series or in parallel with an inductor, e) an RC circuit in which a single resistor in connected either in series or in parallel with an capacitor, f) an RLC circuit in which the single R, single L and single C components may be connected in one of eight ways. The Laplace impedances for the different circuits that come under the six categories are shown in Table 1.

Table 1. Laplace impedance, magnitude and phase for circuit configurations base (R: resistor; L: inductor; C: capacitor; $s = i\omega$; $\omega = 2\pi f$ where f is the frequency)

Configuration		Laplace impedance	Impedance magnitude	Impedance phase
Serial (S)				
1	Z_R	R	R	0
2	Z_L	sL	ωL	$\pi / 2$
3	Z_C	$1 / sC$	$1 / \omega C$	$- \pi / 2$
4	$Z_R + Z_L$	sL + R	$R^2 + \omega^2 L^2$	$\tan^{-1}(\omega L / R)$
5	$Z_R + Z_C$	(sRC + 1) / sC	$(1 + \omega^2 R^2 C^2)^{1/2} / \omega C$	$\tan^{-1}(1 / \omega RC) - \pi / 2$

6	$Z_L + Z_C$	$(s^2LC + 1) / sC$	$(1 - \omega^2LC)^{1/2} / \omega C$	$-\pi / 2$
7	$Z_R + Z_L + Z_C$	$(s^2LC + sRC + 1) / sC$	$[\omega^2R^2C^2 + (1 - \omega^2LC)^2]^{1/2} / \omega c$	$\tan^{-1}((1 - \omega^2LC) / \omega RC) - \pi / 2$
Parallel (P)				
8	$Z_R \mid\mid Z_L$	$sRL / (sL + R)$	$\omega RL / (R^2 + \omega^2L^2)^{1/2}$	$-\tan^{-1}(\omega L/R)$
9	$Z_R \mid\mid Z_C$	$R / (1 + sRC)$	$R / (1 + \omega^2R^2C^2)^{1/2}$	$-\tan^{-1}(\omega RC)$
10	$Z_L \mid\mid Z_C$	$sL / (s^2LC + 1)$	$\omega L / (1 - \omega^2L^2C^2)$	$\pi / 2$
11	$Z_R \mid\mid Z_L \mid\mid Z_C$	$sLR / (s^2 RLC + sL + 1)$	$\omega RL / [\omega^2L^2 + (1 - \omega^2LC)^2]^{1/2}$	$\pi / 2 - \tan^{-1}(\omega L/ (1 - \omega^2LC))$
Serial-parallel (SP)				
12	$Z_R + (Z_L \mid\mid Z_C)$	$(s^2RLC + sL + R) / (s^2LC + 1)$	$[R^2(1 - \omega^2LC)^2 + \omega^2L^2]^{1/2} / (1 - \omega^2LC)$	$\tan^{-1}(\omega L / R(1 - \omega^2LC))$
13	$Z_L + (Z_R \mid\mid Z_C)$	$(s^2RLC + sL + R) / (sRC + 1)$	$[R(1 - \omega^2LC) + \omega^2L^2]^{1/2} / [(1 + \omega^2R^2C^2)]^{1/2}$	$\tan^{-1}(\omega L / R(1 - \omega^2LC)) - \tan^{-1}(\omega RC)$
14	$Z_C + (Z_R \mid\mid Z_L)$	$(s^2RLC + sL + R) / [sC(sL + R)]$	$[R(1 - \omega^2LC) + \omega^2L^2]^{1/2} / [\omega C(R^2 + \omega^2L^2)^{½}]$	$\tan^{-1}(\omega L / R(1 - \omega^2LC)) - \pi/2 - \tan^{-1}(\omega L/R)$
Parallel-serial (PS)				
15	$Z_R \mid\mid (Z_L + Z_C)$	$R(s^2LC + 1) / (s^2LC + sRC + 1)$	$R(1 - \omega^2LC) / [(1 - \omega^2LC)^2 + \omega^2R^2C^2]^{1/2}$	$-\tan^{-1}(RC / (1 - \omega^2LC))$
16	$Z_L \mid\mid (Z_R + Z_C)$	$sL (sRC + 1) / (s^2LC + sRC + 1)$	$\omega L(1 + \omega^2RC)^{1/2} / [(1 - \omega^2LC)^2 + \omega^2R^2C^2]^{1/2}$	$-\tan^{-1}(1/\omega RC) - \tan^{-1}(RC / (1 - \omega^2LC))$
17	$Z_C \mid\mid (Z_R + Z_L)$	$(sL + R) / (s^2LC + sRC + 1)$	$(R^2 + \omega^2L^2)^{1/2} / [(1 - \omega^2LC)^2 + \omega^2R^2C^2]^{1/2}$	$\tan^{-1}[\omega L / R] - \tan^{-1}(RC / (1 - \omega^2LC))$

Sensitivity expressions have also been obtained for the above circuits and these are used in associating specific circuits to specific subsets in a partition.

3.2 String – to – circuit mapping as applied to genetic sequences.

Amino acid / protein/gene sequences.

The models and modeling technique described in section 2 and 3 can be used to convert DNA, RNA and protein strings into 2-dimensional image representations. For DNA the basic letter set would A (adenine), C (cytosine), G (guanine) and T (thymine) but for RNA it would be A, C, G and U(uracil). The frequencies of occurrences of these elements in some genetic string would, of course, be different from letter frequencies associated with ordinary text. For amino acid / protein strings, the letter set would be the standard twenty 1-letter amino acid codes {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y}. See Appendix A for the correspondence between the 1-letter code and the amino acid. The emphasis of the amino acid models is at a macro level where a single RLC circuit is associated with an amino acid and no

consideration is given to whether there is any correlation between an amino acid's circuit representation and the acid's chemical composition. Elaborate models of amino acids and DNA/RNA which have some connection to their chemical properties and secondary structures (alpha helix, parallel and anti-parallel beta sheets, etc.) at the atomic level have appeared elsewhere. See Marshall [8], [9].

3.3 Partitioning of amino acids

The groupings of the amino acids into subsets whose membership is predicated on chemical and structural characteristics such as acidic/basic, hydrophilic/hydrophobic, polar/non-polar, amino acid substitution probabilities, alpha helix and beta sheet formers and breakers, etc. In this paper, the particular amino acid subsets that have been used are based on the hydropathy index of Kyte and Doolittle [7], point accepted mutations (PAM) of Dayhoff, Schwartz and Orcutt [1], block substitution matrix (BLOSUM) of Hennikoff and Hennikoff [4], and interface and octanol hydrophobicity scales of Wimley and White [11], [12]. Subset groupings based on other criteria are also possible as, for example, Kimura [6], Jukes and Cantor [2], Gonnet, Cohen and Benner [3] and JTT matrices of Jones, Taylor and Thornton [4].

3.4 Amino acid partition-to- circuit mapping procedure

The assignment of an amino acid partition to a circuit is based on circuit sensitivity considerations as well as the composition of the partition since amino acids are placed in a partition share one or more chemical/structural properties.

Circuit ordering procedure.

Of the eight sets of sensitivity expressions for circuits with all three circuit elements, the R || L || C circuit is the most sensitive with regard to changes in any of the four circuit parameters - R. L, C and frequency. The complete ordering of all seventeen circuits is shown below in Table 2.

Table 2. Ordering of circuits based on sensitivity to circuit parameter changes

Rank	1	2	3	4	5	6
Circuit	R || L || C	(R+ L) || C	(R + C) || L	(L + C) || R	(R || L) +C	(R || C) + L
Rank	7	8	9	10	11	12
Circuit	R + (L || C)	R || L	R || C	L || C	R + L + C	L + C
Rank	13	14	15	16	17	
Circuit	R + C	R + L	C	L	R	

Amino-acid partitions and ordering .

The partitions are also ordered based on some numerical value or values associated with the specific chemical or structural property upon which the partitions were created. The ordered partitions are given below.

Kyte - Doolittle Hydropathy index (based on extrapolated data)

Amino Acid	I	L	V	F	C	M	A	G	T	S
Hydropathy index	4.5	4.2	3.8	2.8	2.5	1.9	1.8	-0.4	-0.7	-0.8

Amino Acid	W	Y	P	H	N	D	E	Q	K	R
Hydropathy index	-0.9	-1.3	-1.6	-3.2	-3.5	-3.5	-3.5	-3.5	-3.9	-4.5

Kyte-Doolittle Partition:
{[I, L, V], [F, C], [M, A], [G], [T, S, W, Y], [P], [H, N, D, E, Q], [K, R]}

Items in each partition have been placed based on the hydropathy index of an amino acid falling in one of a predefined set of ranges. This set is shown below:
[x < -3.5, -3.5 <=x < -2.5; -2.5 <= x < -1.5; -1.5 <= x < -0.5; -0.5 <= x < 0.5;
0.5 <= x < 1.5; 1.5 <= x < 2.5; 2.5<= x < 3.5; 3.5 <= x]

Partitions based on Wimley and White hydrophobic scales (experimental data):
The values are in the range [-1.8, 2.1] for the Interface scale and in the range [-2.1, 3.6] for the Octanol scale
Interface Scale Partition:
{[W, F, Y], [L, I, C, M], [G], [V, S, T, A], [N, P, Q], [H, R, K, D, E]}
Octanol Scale Partition:
{[W, F, Y], [I, L], [C], [M], [G], [V, S, T, A], [N, P, Q], [H, R, K, D, E]}

In both these sets, the charged amino acids are in the extreme left sub-partition (most positive values) and the aromatic acids are in the rightmost sub-partition (most negative values).

Partitions based on Block Substitution Matrix (BLOSUM60):
Log odds alignment substitution frequencies. The matrix diagonal values are shown

S	A	V	I	L	T	E	Q	K	M
4	4	4	4	4	5	5	5	5	5
R	N	D	G	F	P	Y	H	C	W
5	6	6	6	6	7	7	8	9	11

The amino acids have been rearranged on the basis of increasing log odds values.
BLOSUM60 Partition:
{[S, A, V, I, L], [T, E, Q, K, M, R], [N, D, G, F], [P, Y], [H], [C], [W]}

Partitions based on Point Accepted Mutation (PAM25):
Log odds mutation frequencies matrix diagonal values are shown

S	A	N	T	V	D	E	Q	K	G

2	2	2	3	4	4	4	4	5	5
I	L	M	P	R	H	F	Y	C	W
5	6	6	6	6	6	9	10	12	17

The amino acids have been rearranged here as well on the basis of increasing log odds values.

PAM25 Partition:

{[S, A, N], [T],[V, D, E, Q],[K, G, I],[L, M, P, R, H], [F,Y],[C], [W]}

Partitions based on side chain property:

Side chain property	Amino acid
Acidic R group	D, E
Basic R group	K, H, R
Uncharged polar	G, S, T, Q, C, Y, N
Non-polar	A, V, I, P, M, L, F, W

Side chain property	Amino acid
Branched	I, L, V
Unbranched	A
Aliphatic	A, I, L, M, P, V
Aromatic	F, W, Y

Side Chain Partition:

{(D, E), (K, H, R), (G), (S, T, Q, Y, N), (I, L, V), (A, M, F, W), (P), (C)}

The sub-partitions are based on whether the amino acids are acidic, basic, uncharged polar, branched non-polar, or non-polar. Amino acids G, P and C have been placed in their own partitions since each of these acids is unique – G has the simplest side chain, P has its side chain incorporated into the backbone and C is the only amino acid that contains sulfur. The data for PAM250 and BLOSUM60 have been taken from the National Institutes of Health government web site (www.ncbi.nlm.nih.gov).

4 Results and discussion

We now show several examples employing the techniques described in the preceding sections for a variety of amino acid sequences and amino-acid circuit mappings. The generalized circuit expressions (1a, b) through (4a, b) for the four models have been computed for a set of input parameters. In generating the computations, we have used the following values for the various parameters:

Circuit elements R, L and C have each been assigned a value of 1. Unless otherwise indicated, the input frequency has been chosen to be 0.001 Hz. The amino acid sequences have been taken from the set {'P', 'PP', 'PPP', 'PPPP', 'PPPPP', 'ACDEF', 'LGLWL', 'TGLWL', 'LGTWL', and 'LGLWT'}. The partition to circuit mapping is based on the ordering shown in the table in section 4.3.1. Also, we did not

specify different R, L and C values for the different circuits even though this could be done. The special lengths and contents of the sequences have been chosen to bring out the differences between the four models and the utility of the models in particular contexts. After the overall impedance magnitude and phase of a model have been computed, we then compute the voltage of the circuit V(□t) = 1000 I(□t) Z(□t) where I(□t), the current, is a simple sine function over a 1-second time period. The voltage has been sampled at 900 uniformly divided time instants and each voltage value is taken to represent a pixel. These 900 pixel values have been rendered into a 30 by 30 color image using a thresholding technique based on the scheme given below:

Set all 900 pixels to white: pixel color = (red = 255, green = 255, blue = 255)

For each pixel, pixel color = pixel color + red $0 < V(t) \bmod 256 <= 85$

pixel color + green $85 < V(t) \bmod 256 <= 170$

pixel color + blue otherwise

It should be noted that at each time instant, only one of the primary colors is being modified. The magnitude and phase images generated by each of the models for a variety of cases are depicted in figures 1 through 4.

In Figures 1a and 1b, the magnitude and phase images for the amino acid sequence 'ACDEF' are shown. The number of circuits associated with string 'ACDEF' is dependent on which partition (Kyte-Doolittle, PAM25, etc.) and the sub-partition to which each of the five amino acids in the string belongs. For example, in the PAM25 partition, four circuits are involved since 'D' and 'E' lie in the same sub-partition whereas 'A', 'C', and 'F' lie in three different sub-partitions.

Partition model	Circuit model			
	Serial	Parallel	Serial-Parallel	Parallel-serial
Kyte-Doolittle				
Wimley-White Interface				
Wimley-White Octanol				
PAM-25				
BLOSUM-60				
Side Chain				

Figure 1a. Circuit model response to input sequence 'ACDEF' for the six amino acid partitions – Magnitude images.

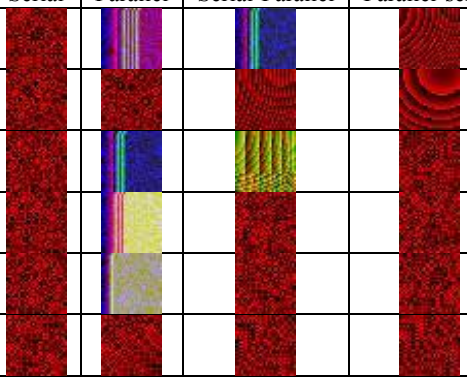
Partition model	Circuit model			
	Serial	Parallel	Serial-Parallel	Parallel-serial
Kyte-Doolittle				
Wimley-White Interface				
Wimley-White Octanol				
PAM-25				
BLOSUM-60				
Side Chain				

Figure 1b. Circuit model response to input sequence 'ACDEF' for the six amino acid partitions – Phase images.

In Figures 2a and 2b the images generated for the magnitude and phase responses of the four circuit models to purely alphabetic strings of lengths 1, 2, 3 and 4 are shown. For a string of length 1, the S, P, SP and PS models are the same since there is only one basic circuit associated with whatever character the string contains. For strings of length 2, as there are only two characters in the string, only the S and P models are germane, the other two are not. For strings greater than 2, all four models are possible. Since the first four message strings in Figure 1 are all comprised of the character 'P' and there is exactly one associated sub-partition and one associated circuit, the overall circuits' impedances are easily obtainable by using expressions (14a) thru (14d) and that the images depicted do indeed reflect the Fibonacci fractions in sequential order; the phase images will, of course, be all the same. Also, for much longer strings with the same character present throughout, the images converge and become indistinguishable from one image to the next (the 'golden ratio' effect).

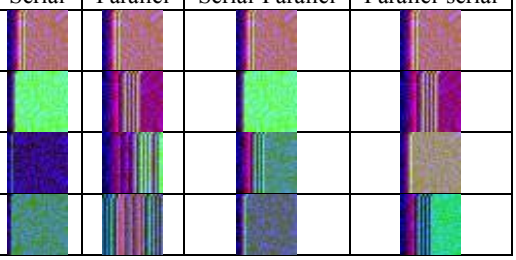
Input sequence	Circuit model			
	Serial	Parallel	Serial-Parallel	Parallel-serial
'P'				
'PP'				
'PPP'				
'PPPP'				

Figure 2a. The Fibonacci effect - Circuit response to repeat sequences for the Kyte-Doolittle partition – Magnitude images.

Input sequence	Circuit model			
	Serial	Parallel	Serial-Parallel	Parallel-serial
'P'				
'PP'				
'PPP'				
'PPPP'				

Figure 2b. The Fibonacci effect - Circuit response to repeat sequences for the Kyte-Doolittle partition – Phase images.

In figures 3a and 3b, the focus is on position sensitivity, .i.e., amino acid string pairs which differ in exactly one position. The results shown for the string 'LGLWL' are based on a single amino acid 'T' being substituted at specific positions in the string - at the beginning, in the middle and at the end.

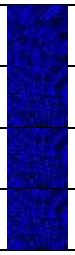
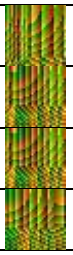
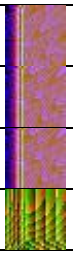
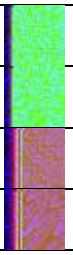
Input sequence	Circuit model			
	Serial	Parallel	Serial-Parallel	Parallel-serial
'LGLWL'				
'TGLWL'				
'LGTWL'				
'LGLWT'				

Figure 3a. Position sensitivity - Circuit response to mutated input sequences for the BLOSUM-60 partition – Magnitude images.

Input sequence	Circuit model			
	Serial	Parallel	Serial-Parallel	Parallel-serial
'LGLWL'				
'TGLWL'				
'LGTWL'				
'LGLWT'				

Figure 3b. Position sensitivity - BLOSUM-60 partition – Phase images.

In Figure 4, we have shown the 'difference' images, i.e. the images that result when an image for a base string is subtracted, pixel by corresponding pixel, from the base string which has been modified at specific positions. The image entries for the S model are identical and that should come as no surprise since the ordering of circuit

elements is irrelevant in a circuit which has multiple serially connected elements. A similar comment applies to the P model as well. Figures 4a and 4b shows the influence of frequency changes on the images that are generated while all other circuit parameters are held constant.

Circuit model	Message string = 'LGLWL'	
	Frequency = 0.001	Frequency = 0.0001
S		
P		
SP		
PS		

Figure 4a. Circuits' frequency dependence - magnitude

Circuit model	Message string = 'LGLWL'	
	Frequency = 0.001	Frequency = 0.0001
S		
P		
SP		
PS		

Figure 4b. Circuits' frequency dependence - phase

References

1. Dayhoff, M.O., R.M. Schwartz and B.C. Orcutt. "A model for evolutionary change in proteins". Atlas of Protein Sequence and Structure , Vol. 5. 1978. pp 345–352.
2. Jukes, T.H. and C.R. Cantor. "Evolution of protein molecules". In Munro, H.N. Mammalian protein metabolism. New York: Academic Press. 1969. pp. 21–123.
3. Gonnet, G.H. M.A. Cohen and S.A. Benner. "Exhaustive matching of the entire protein sequence database". Science, Volume 256, No. 5062, June 1992. pp. 1443–1445.
4. Henikoff, S. and J.G. Henikoff. "Amino Acid Substitution Matrices from Protein Blocks". Proceedings of the National Academy of Sciences, Vol. 89, No. 22. 1992. pp. 10915–10919.
5. Jones, D.T., W.R. Taylor and J.M. Thornton. "The rapid generation of mutation data matrices from protein sequences". Computer Applications in Biosciences, No. 8, 1992, pp. 275–282.
6. Kimura, M. "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences". Journal of Molecular Evolution, Vol. 16, No. 2, December 1980, pp. 111–20.

7. Kyte, J. and R.F. Doolittle. "A simple method for displaying the hydropathic character of a protein". Journal of Molecular Biology, Vol. 157, No. 1, 1982, pp. 105–32.
8. Marshall, R. Modeling DNA/RNA Strings Using Resistor—Capacitor (RC) Ladder Networks. The Computer Journal. Volume 53 Issue 6, July 2010. Oxford University Press Oxford. pp. 644-660.
9. Marshall, R. Modeling secondary structures and secondary structure linkages of protein sequences. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC), Istanbul, Turkey, Oct. 2010, pp. 2995-3000.
10. Perez, J.-C. "Codon populations in single-stranded whole human genome DNA are fractal and fine-tuned by the Golden Ratio 1.618". Interdisciplinary Sciences: Computational Life Sciences Volume 2, no. 3. September 2010. pp. 228-240.
11. Wimley, W.C., T.P. Creamer and S.H. White. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry, Volume 35, No. 16, April 1996, pp.5109-5124. Wimley, W.C. and S.H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nature Structural Biology Vol 3, October 1996, pp. 842-848.
12. Yamagishi, M. E.B., and A.I. Shimabukuro. Nucleotide frequencies in human genome and Fibonacci numbers. Bulletin of Mathematical Biology, Volume 70, no. 3, April 2008. pp 643-653.