Electrical circuit models of protein structure

A framework for the analysis, classification, and synthesis of protein shapes

G. Sampath
Department of Bioinformatics and Computer Science
University of the Sciences in Philadelphia

Summary

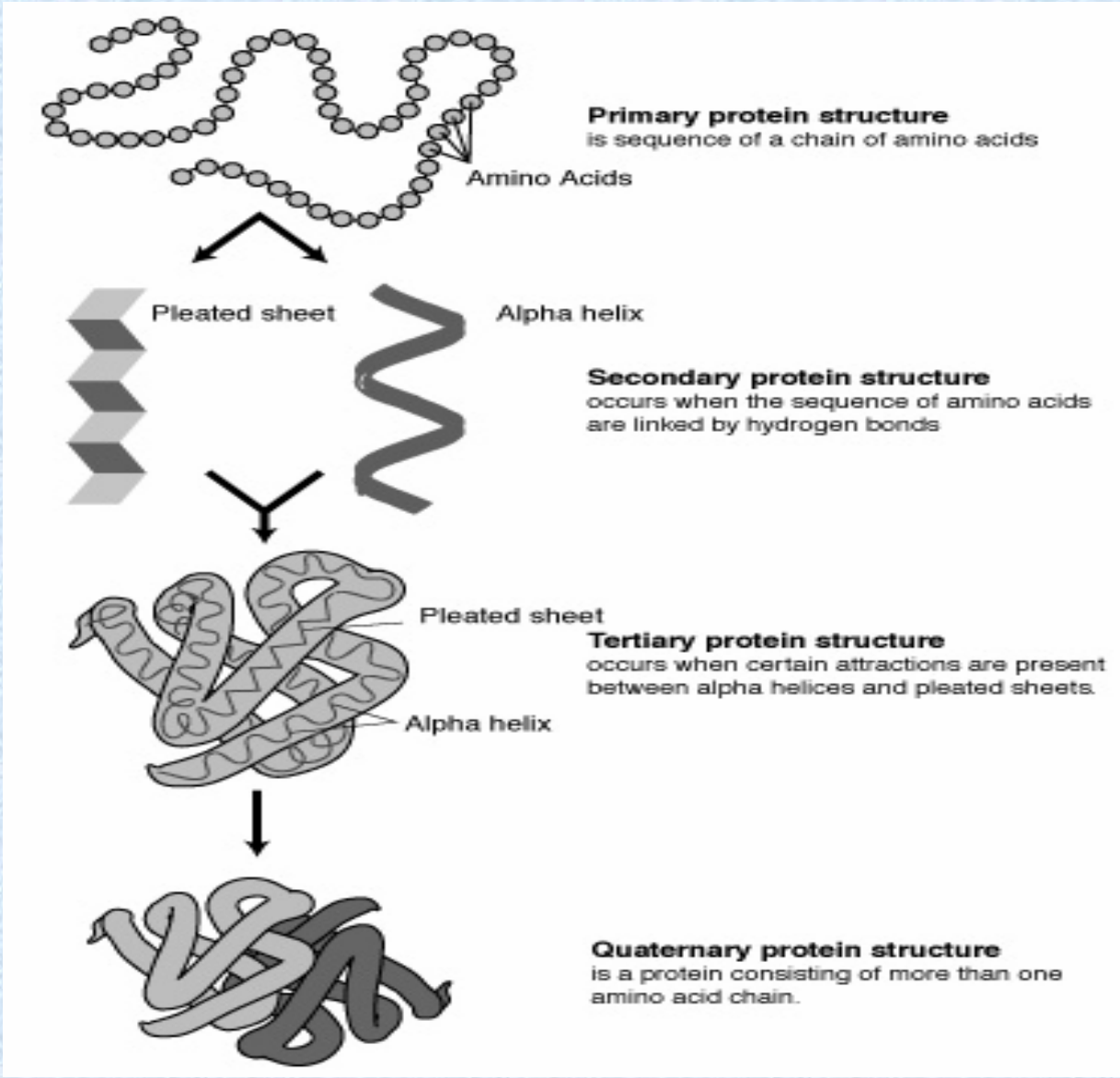
- Passive analog electrical circuits used to model protein structure
  - Secondary structure: α →L, β → C, γ → R
  - Tertiary structure: α-pair → M, distant folds → C
  - Constraints on protein structure based on circuit topology
- Protein shapes can be studied via their circuit analogues
  - Pole-zero maps for protein classes in the complex plane
- Poles and zeros can be used to synthesize a 'protein circuit'
  - May correspond to existing, non-existing or impossible protein
- Protein Circuit Analogue System: a framework for analysis and synthesis
  - Correlating real proteins to their circuit analogues
- Discussion
  - Do real proteins behave like electrical circuits in some frequency range?
  - Non-linear circuits
  - Computational considerations
  - Other issues

Modeling a complex system via its analogue in another domain

Examples

- Hodgkin-Huxley model
  - Nonlinear RC transmission line model of nerve signal transmission
- RLC circuit models in thermodynamics and quantum phenomena
  - Callen circuit
  - Quantum billiards
- Hopfield nets in computation
  - Neural nets for computation
  - Associative memory

Primary, secondary, tertiary, and quaternary structure of proteins



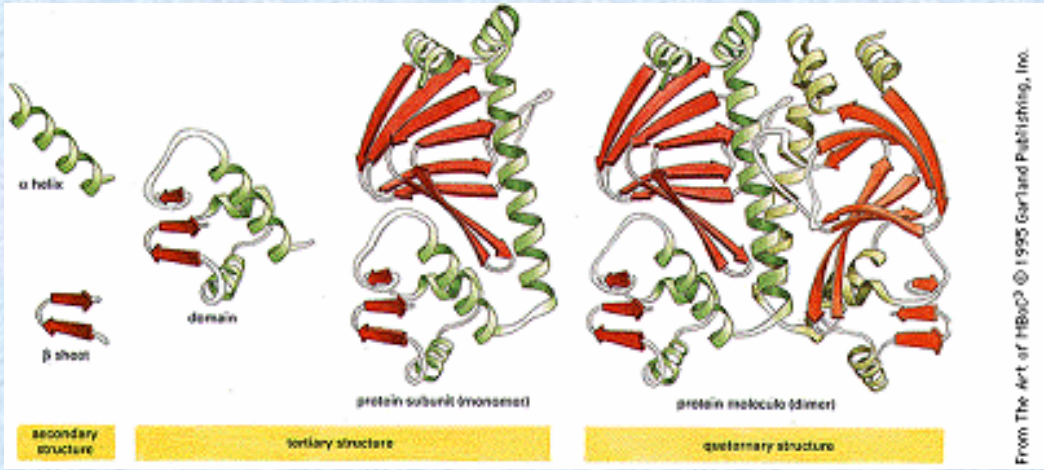
Secondary and tertiary structure building blocks

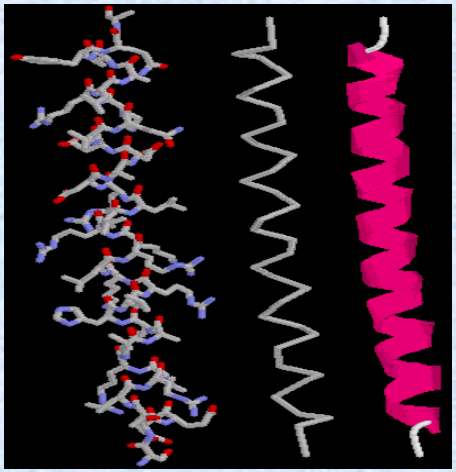
Secondary

- Helix
- Sheet
- Other (turns, loops)

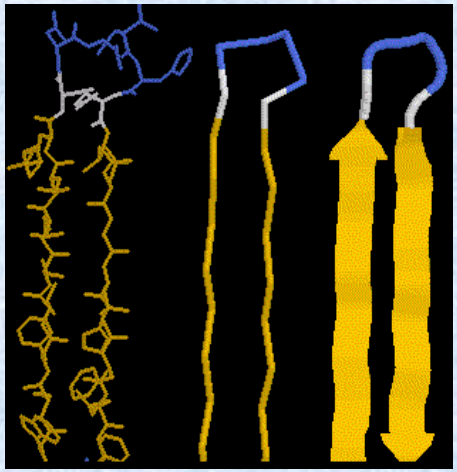
Tertiary

- Helix pair





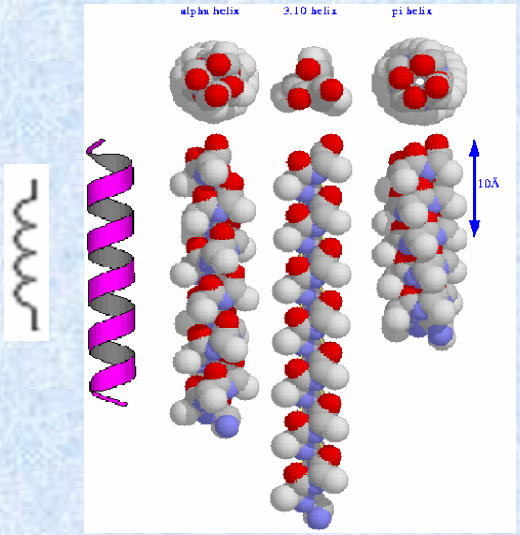
stick-ball backbone ribbon

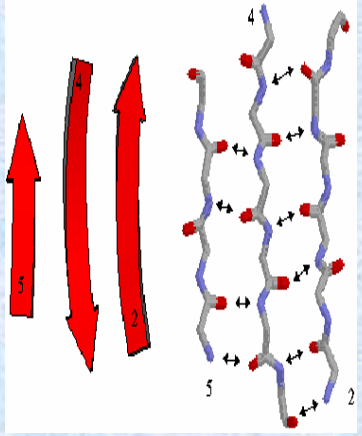


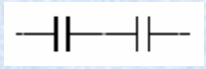
stick-ball backbone ribbon

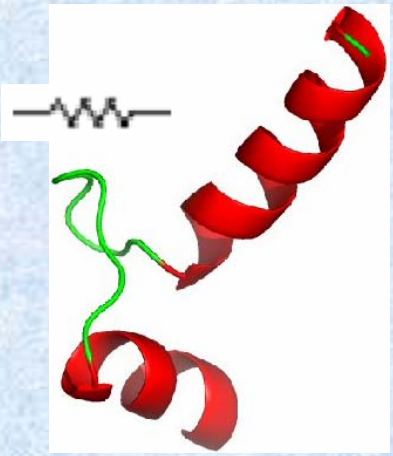
Protein structure and electrical circuits

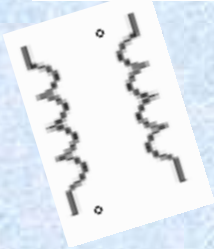
Similarities between protein structure units/motifs and circuit element types











Secondary unit	Circuit element
helix (H or α)	*inductor* (L)
strand/sheet (E or β)	*capacitor* (C)
turn (T or γ)	*resistor* (R)

Tertiary motif	Circuit element
Bond between distant residues	*Capacitive bridge*
Helix pair	*Transformer* *(mutual inductance M between coils)*

Protein structure circuit modeling parameters

Secondary structure	*Element property*	*Immittance (at 1 MHz)*
Helix (h turns)	$L(\alpha) = L_0\, h$	$2\pi 10^6 L_0\, h \rightarrow h$
Strand pair (b H bonds)	$C(\beta) = C_0\, b$	$2\pi 10^6 C_0\, b \rightarrow b$
Turn with t residues	$R(t) = R_0\, t$	$R_0\, t \rightarrow t$

Secondary structure modeling

→

p-RLC-s circuit

Tertiary motif	*Element property*	*Immittance (at 1 MHz)*
Helix pair: h_1, h_2 turns	$M = k\sqrt{(L_1 L_2)}$	$2\pi\, 10^6 L_0\, k\sqrt{(h_1 h_2)} \rightarrow k\sqrt{(h_1 h_2)}$
Capacitive bridge	$C_B = k_B$	$2\pi 10^6 k_B C_0 \rightarrow k_B$

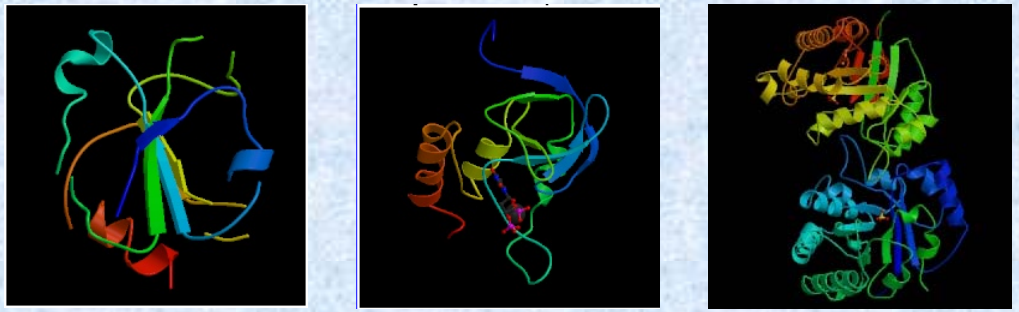
Tertiary structure modeling

→

p-RLC-t or p-RLCM-t circuit

Examples of ‘protein circuits’

- Thioredoxin
- Staphylococcus Nuclease
- Triose Phosphate Isomerase



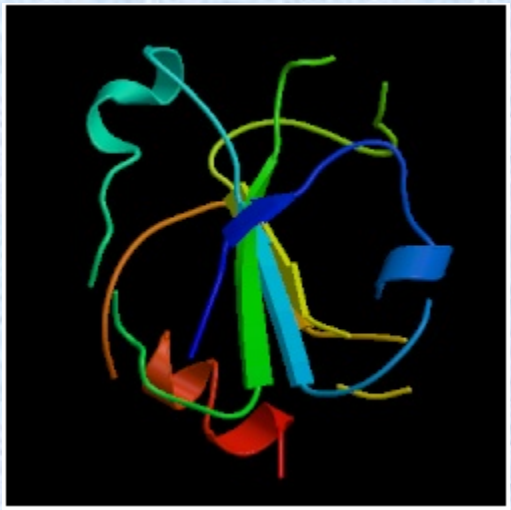
Protein structure circuit diagram as visualization schematic

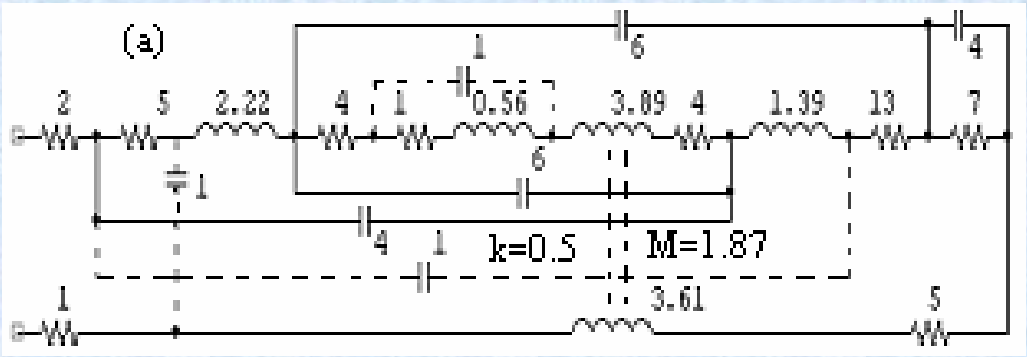
Compare with

Robinson schematics (ribbon diagrams) *Stick figures*

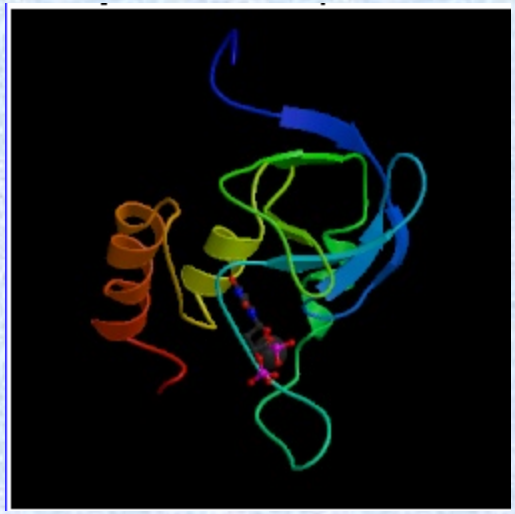
Skeletal structures *TOPS diagrams*

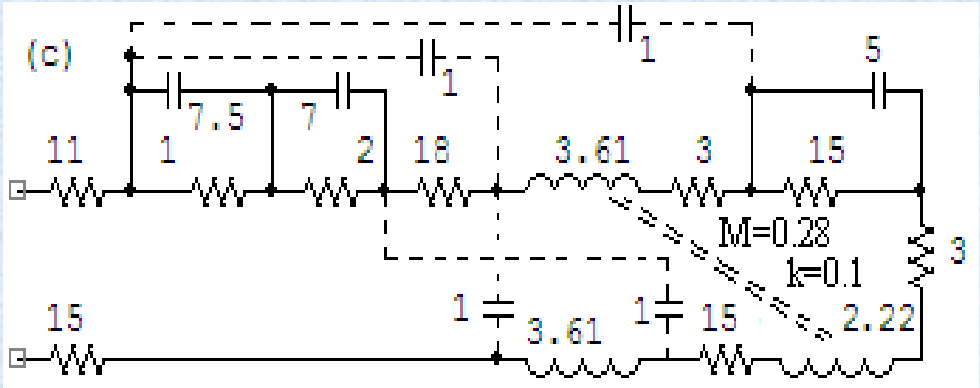
Thioredoxin (1SRX)



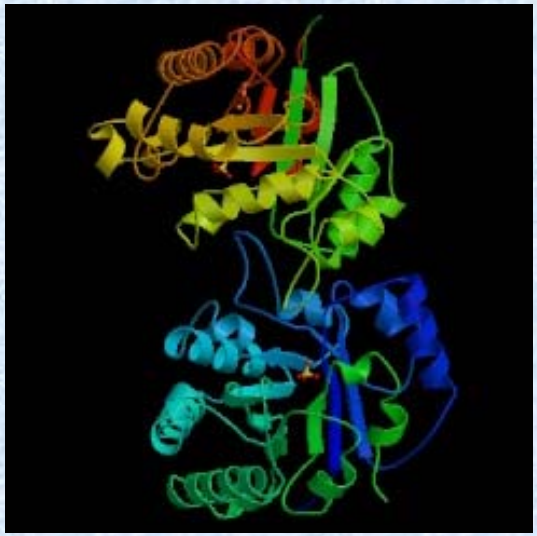


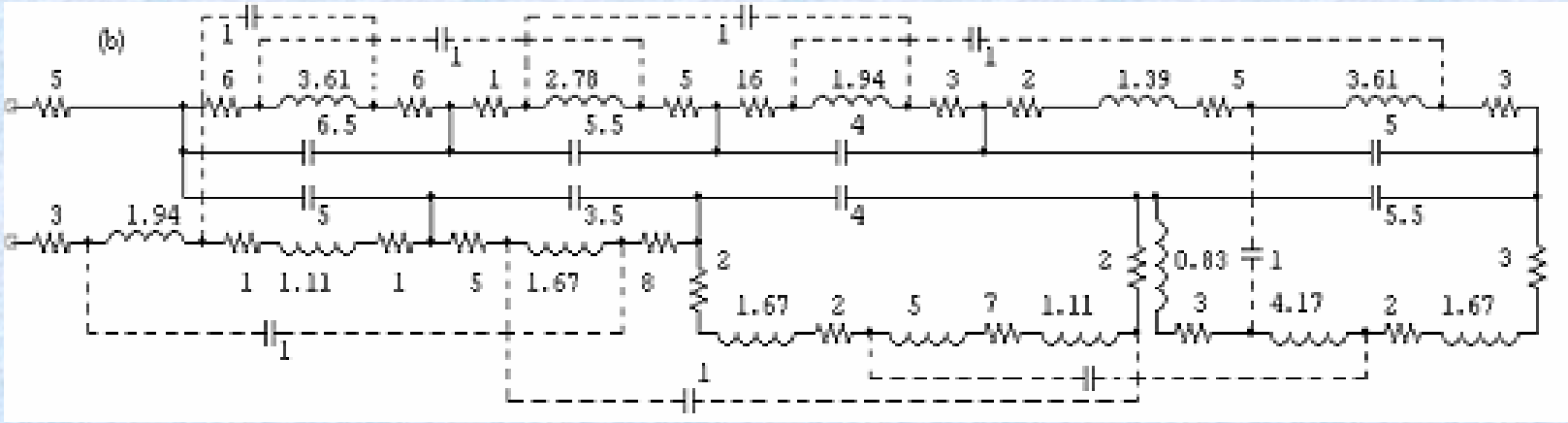
Staphylococcus Nuclease (2SNS)





Triose Phosphate Isomerase (8TIM)





From proteins to circuits and back

Structural properties/shape of protein $\overset{?}{\longleftrightarrow}$ *Properties of its p-RLC(M) circuit*

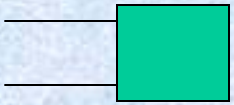
- Analytical approach
  - Search for and compare compatible characteristics in the two domains
  - Classification procedures based on those characteristics
- Synthetic approach
  - Search for protein structure and shape based on designed RLC(M) circuits
  - Consider circuits that are known and/or widely used in circuit applications

  Such a search may lead to three possibilities
  - a known protein structure,
  - one that has not been observed in nature
  - one that is perhaps physically or chemically impossible
- Folding studies
  - Circuit component sizes and distances used to compute possible 3-D shapes of a protein for a given RLC(M) circuit.

Circuit theory of analog linear passive circuits

Input impedance and transfer functions

- A circuit may be 1-terminal (1-port) or 2-terminal (2-port)



Circuit with linear passive elements R, L, C, M

Input voltage, input current, output voltage are all functions of complex frequency $s = \sigma + j\omega$

- 1-port fully specified by input impedance $Z(s) = V_{in}(s)/I_{in}(s)$

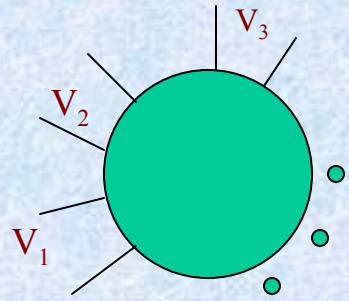
  → study protein structure

- 2-port fully specified by transfer function $T(s) = V_{out}(s) / V_{in}(s)$

  → study protein-protein interaction

- Multiports: specified by port-voltage matrix $\mathbf{V}(s)$

  → study protein complexes



Properties of positive real (p.r.) functions

Causality and stability require Z(s) = P(s)/Q(s) to be positive real

Necessary and sufficient conditions for Z(s) to be p.r.

1) Re s > 0 => Re Z(s) > 0
2) Re s >= 0 => Re Z(s) = 0

Usually not easy to prove this, so write Z(s) as

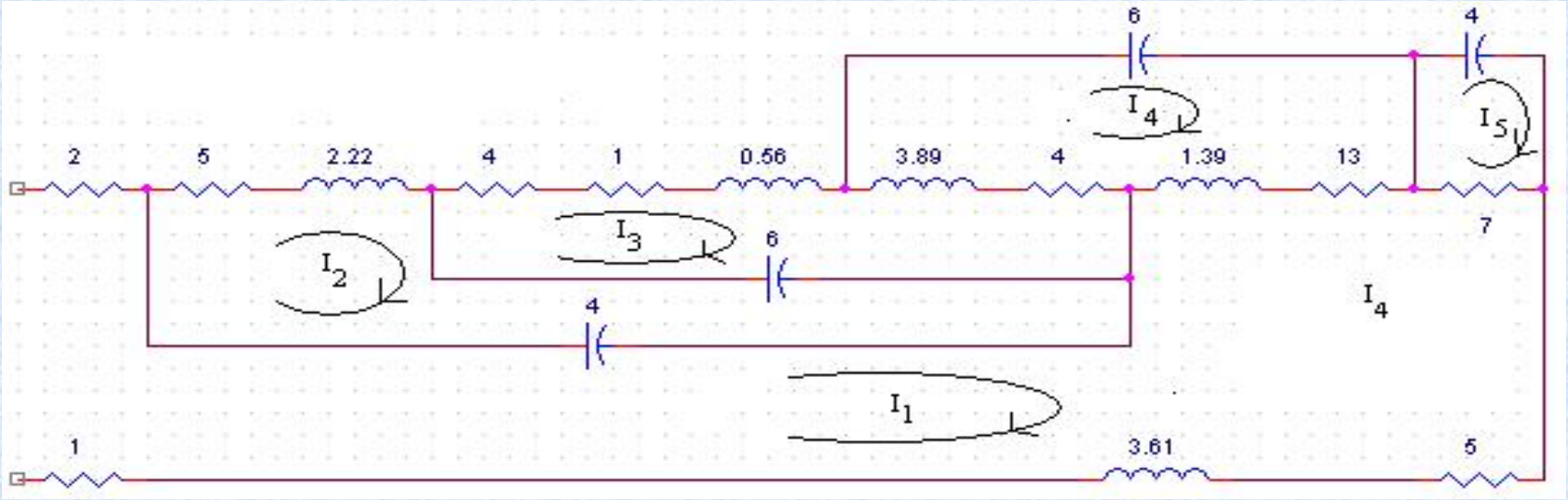
$$Z(s) = K \frac{(s-z_1) \ldots (s-z_n)}{(s-p_1) \ldots (s-p_m)}$$

Equivalent conditions for p.r. property

- K real positive
- |n - m| <= 1
- Poles and zeros real, appear in conjugate pairs
- Poles and zeros in left half of s plane (use Hurwitz test)
- Poles on imaginary axis must be single and in conjugate pairs, with positive real residues
- Re Z(jw) >= 0

Example of protein circuit analysis

Protein circuit for thioredoxin

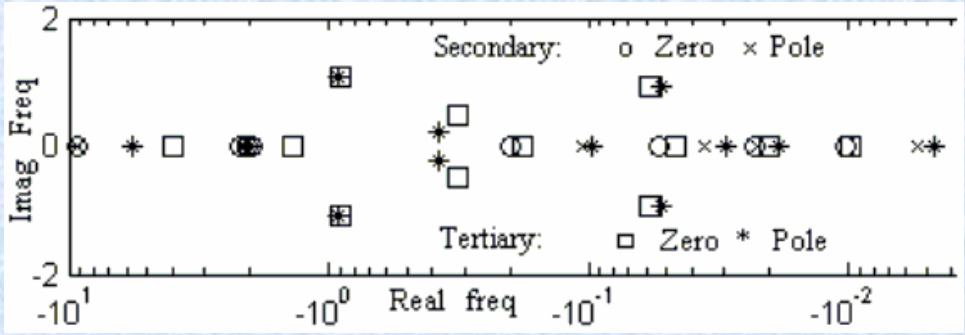


Loop analysis matrix

$$\begin{bmatrix} 5*s*s+28*s+1/4 & -1/4 & 0 & -1.39*s*s-13*s & -7*s \\ -1/4 & 2.22*s*s+5*s+1/6+1/4 & -1/6 & 0 & 0 \\ 0 & -1/6 & 4.45*s*s+9*s+1/6 & -4.45*s*s-9*s & 0 \\ -1.39*s*s-13*s & 0 & -4.45*s*s-9*s & 5.84*s*s+22*s+1/6 & 0 \\ -7*s & 0 & 0 & 0 & 7*s+1/4 \end{bmatrix}$$

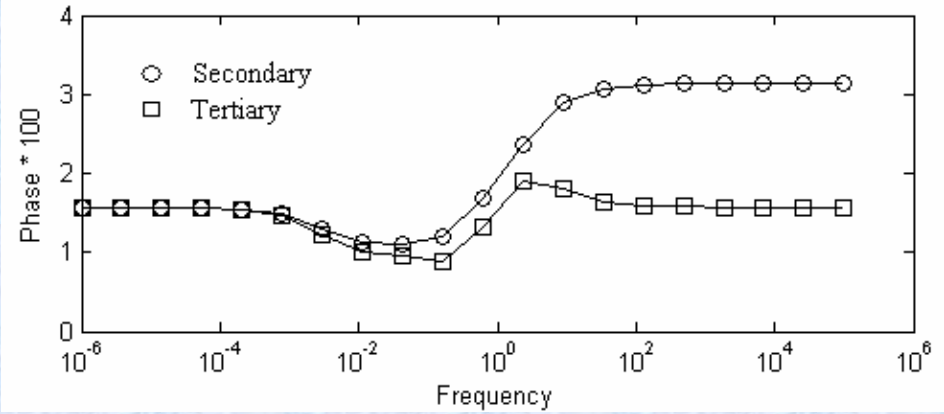
Input impedance of protein structure circuits: amplitude spectrum and pole-zero map of thioredoxin

- Amplitude response |Z(jω)| has predictable low-pass behavior, not shown. (Sometimes may have characteristic peak with damping or ringing.)



- Pole-zero map shows pattern of expansion going from all poles and zeros on negative real axis (secondary) to negative real and some complex poles and zeros (tertiary)

Input impedance of protein structure circuits: phase spectrum for thioredoxin



- Phase spectrum has distinguishing characteristics: for tertiary structure has a characteristic humped shape
- Properties of hump differ from one protein to another

Protein pole-zero maps

- One can consider mapping proteins in protein families to the corresponding poles and zeros at different levels: secondary, tertiary and quaternary.
- Conversely, one can look at the pole-zero distribution of a set of proteins and use the poles and zeros to group proteins.
- This leads to a classification scheme that can use training schemes similar to those used in neural-net-based methods.
- Conversely one could specify a set of poles and zeros from which it could be determined if there is a corresponding protein.

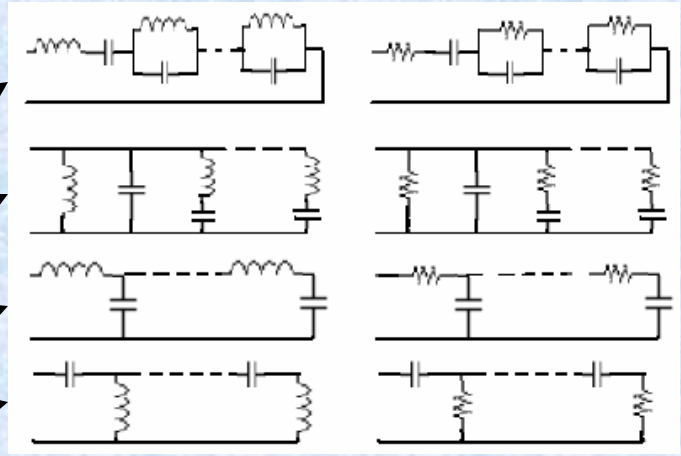
Circuit synthesis methods for impedance functions

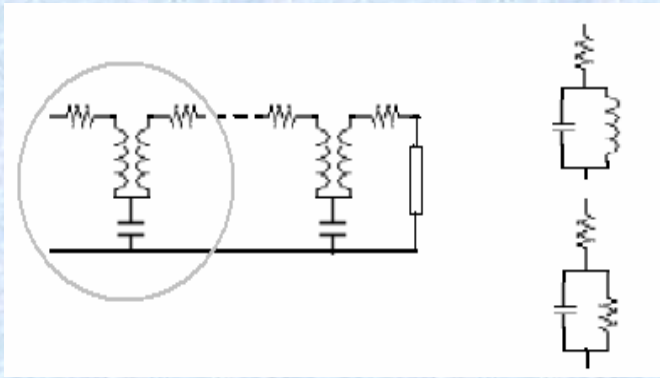
Standard methods

- Partial fractions
- Continued fractions

Standard methods lead to ladder networks. They may be canonical or non-canonical. Some cannot represent proteins.

- Foster I and II
- Cauer I and II
- Brune ladders
- Darlington synthesis
- Bott-Duffin synthesis
- Other methods - Miyata and Kuh
- Bridge networks - Seshu





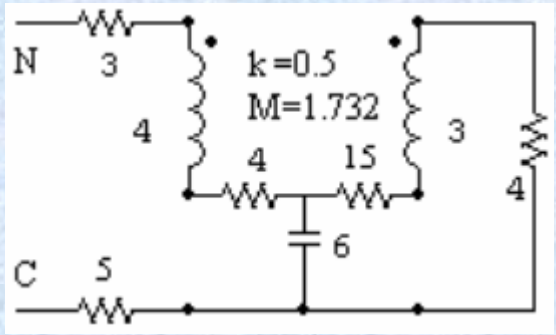
Example of protein shape synthesis

Protein circuit representing secondary + some tertiary structure

Impedance function

$$Z(s) = (18s^3 + 224s^2 + 457.17s + 10.33) / (s^2 + 4s + 0.056)$$

↓



Synthesized circuit (Brune synthesis)

→

Corresponding protein shape with secondary and partial tertiary structure

- 68 residues
- 2 α helices
- 1 sheet of two β strands
- 3 turn-loops (excluding the residues near the terminal ends)
- 1 helix pair (formed by the 2 α helices above)

Variations/restrictions on protein circuit topology

Secondary structure comes from linear sequence of amino acids

→

Elements in RLC(M) circuit cannot be arbitrarily connected

- p-RLC-s circuit can't be Foster II, Cauer II, bridge circuit, Bott-Duffin.
- If strands form sheet in primary sequence order (as in some barrels) then β-α- β motif can be represented by Foster I or Cauer I.
- When turns are present inductors may be replaced with lossy ones:

  L → series LR, RL or LRL R ≡ turn or loop
- Can cascade different forms in successive stages.
- A-type (but not B-type) Brune sections can be used for tertiary structure with helix pairs.
- In most cases, the dual network does not exist. In particular, non-planar circuits cannot have duals.
- In general, network realizing a given impedance function is not unique; some of these equivalent networks may or may not correspond to a protein structure.

Example: Foster synthesis

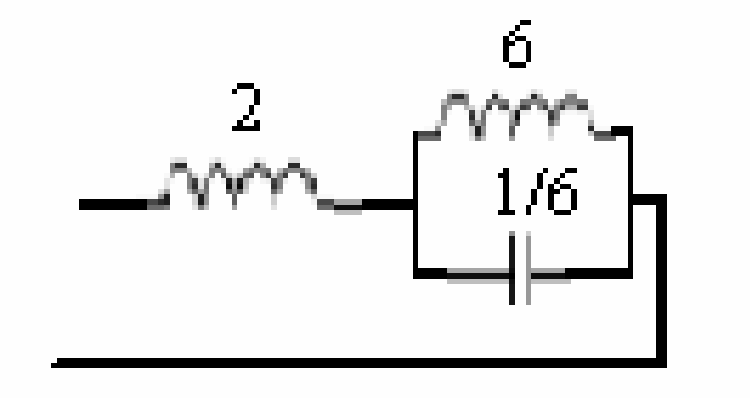
Partial function expansion example

Synthesis leads to protein circuit for secondary structure

Consider $Z(s) = (2s^3 + 8s)/(s^2 + 1)$

Partial fraction expansion leads to

$Z(s) = 2s + 1 / (s/6 + 1/6s)$



Example: Cauer synthesis

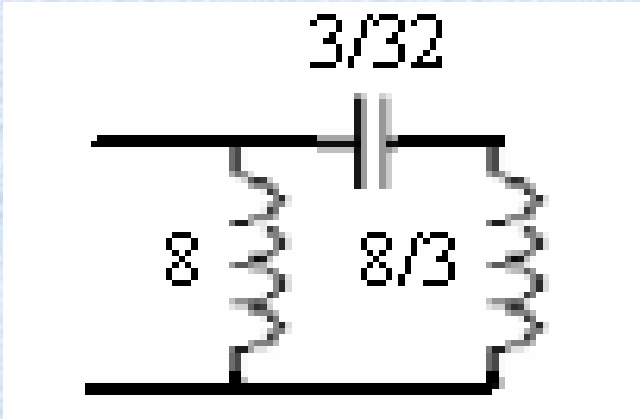
Continued fraction expansion example

Synthesis leads to protein circuit for secondary structure

Consider $Z(s) = (2s^3 + 8s)/(s^2 + 1)$

Continued fraction expansion of $Y(s) = 1 / Z(s) =$

$$\cfrac{1}{1/8s + \cfrac{1}{1/(3/32s) + \cfrac{1}{8/3s}}}$$

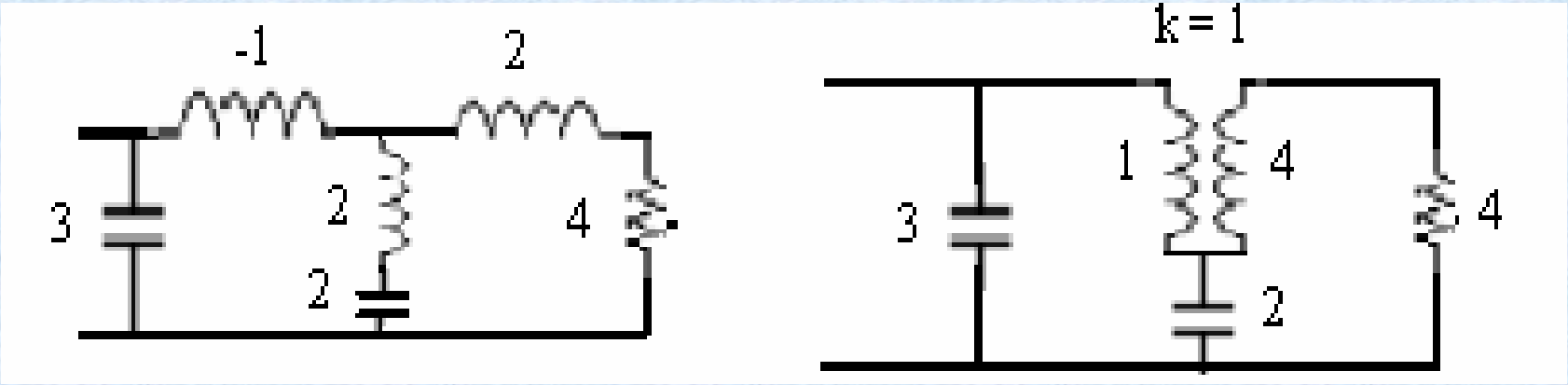


Example: Brune synthesis

Transformer-based synthesis example

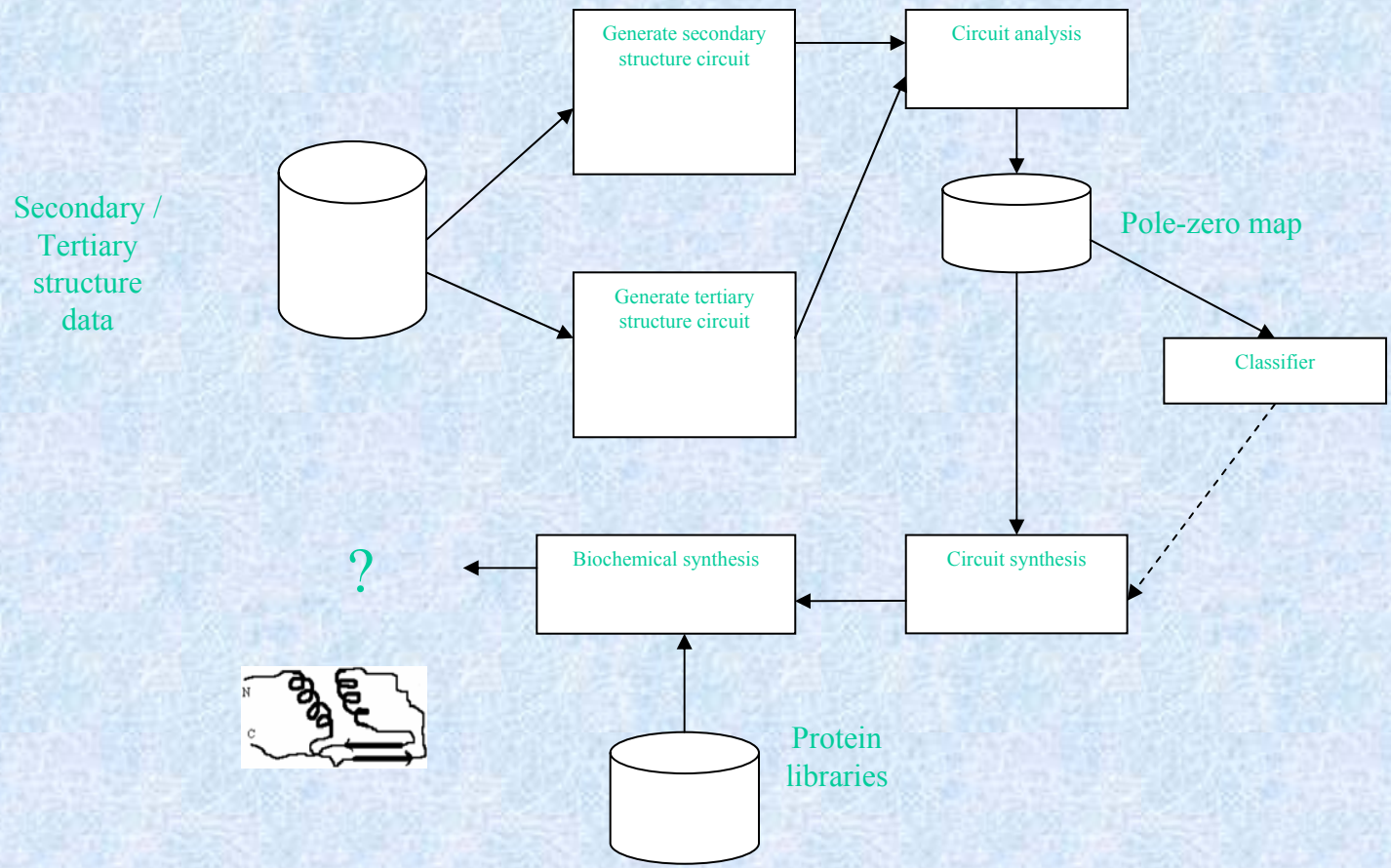
Leads to protein circuit for secondary structure and partial tertiary structure

Consider $Z(s) = (8s^2 + s + 4)/(24s^3 + 11s^2 + 20s + 1)$

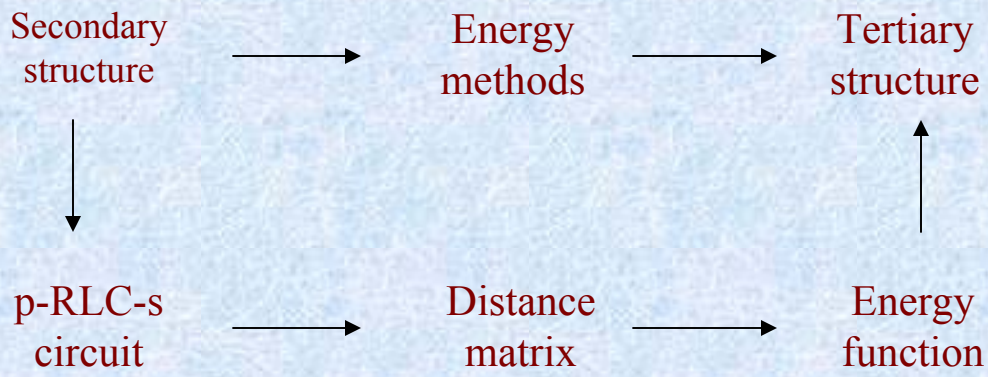


- 68 residues
- 2 α helices
- 1 sheet of two β strands
- 3 turn-loops (excluding the residues near the terminal ends)
- 1 helix pair (formed by the 2 α helices above)

Protein Circuit Analogue System



Application to protein folding



- The intermediate step of transforming protein secondary structure into the corresponding p-RLC-s circuit may lead to a considerable reduction in the size of the space over which folds need to be considered.
- Thus the number of possible combinations of the secondary structure elements to be considered in calculating the folded structure may be considerably reduced for many proteins.

Discussion

- Non-linear p-RLC(M) circuit → system of attractors
- Sensitivity analysis of Z(s) → how protein structure is affected by changes in component motifs
- Transfer function studies of protein interactions: multi-port methods
- Homology ↔ Properties of p-RLC(M) circuit
- Tertiary shape ↔ impedance function: use in drug discovery procedures?
- Measure electrical properties of proteins in secondary and native states (using methods used in study of electrical properties of DNA)
  - Compare with those of p-RLC(M) circuit
  - Amino acid polymers as passive nano-electrical circuits

Circuit theory bibliography

- C. A. Desoer and E. S. Kuh. *Basic Circuit Theory*. (1969)
- M. E. Van Valkenburg. *Modern Network Synthesis*. (1960)
- L. Weinberg. *Network Analysis and Synthesis*. (1962)
- G.C.Temes, J.W.LaPatra. *Intro to Circuit Synthesis and Design*. (1977)
- H.-W. Fink, C. Schönenberger. Nature **398**, 407-410. (1999)
- P. Dietz and M. Reif. PNAS **103** (5), 1244-7. (2006)

Credits

Several figures reproduced/adapted from various sources on the web