Improved Hot-Spot Location Technique for Proteins Using a Bandpass Notch Digital Filter

Parameswaran Ramachandran, Wu-Sheng Lu, and Andreas Antoniou

ISCAS, Seattle
May 21, 2008

**Department of Electrical Engineering,
University of Victoria, BC, Canada.**

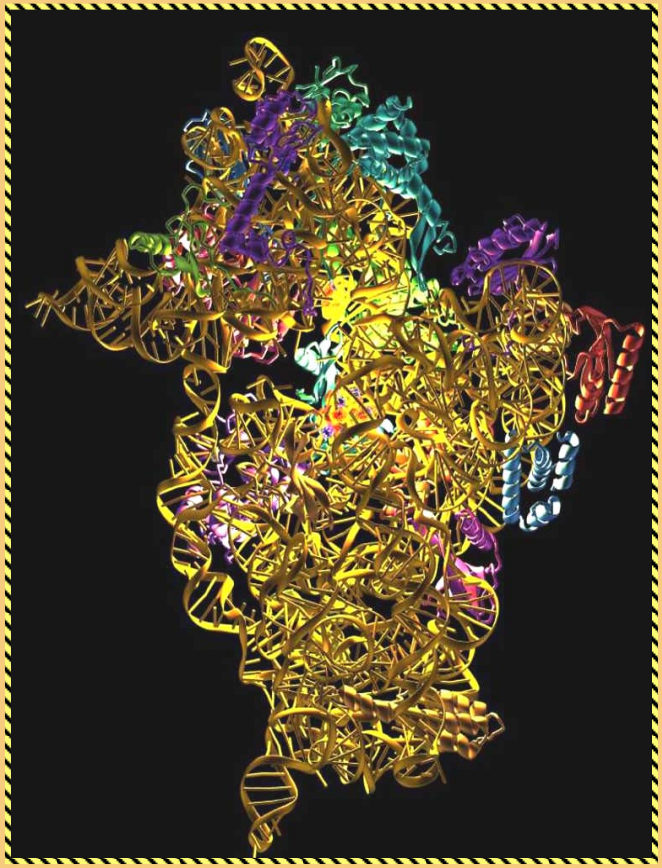
Proteins

- Proteins are the building blocks of cells.
- Proteins
  - form the structural components (e.g., skin proteins)
  - catalyze chemical reactions (e.g., enzymes)
  - transport and store materials (e.g., hemoglobin)
  - regulate cell processes (e.g., hormones)
  - protect the organism from foreign invasion (e.g., antibodies)
- Proteins are long polymers of subunits called **amino acids.** There are 20 different amino acids that form proteins.

Proteins (cont'd)

- Proteins perform their functions by folding into unique **three-dimensional (3-D)** structures.

Image Credit: Catholic University of Brussels, Biotechnology



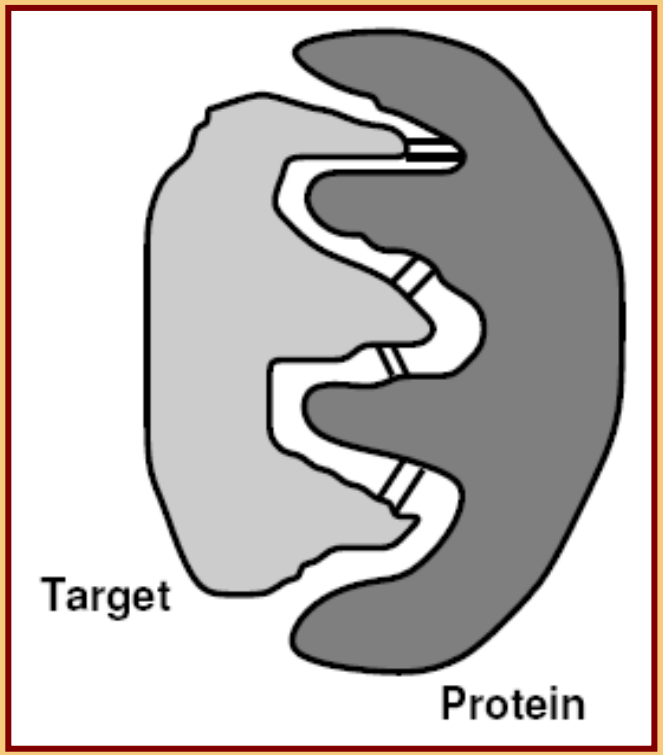
Ribosomal subunit

Credit: Argonne Photo Gallery

Hot Spots in Proteins

- Interactions among different regions of a protein molecule enable the molecule to fold into complex 3-dimensional (3-D) structures.

- Proteins perform their functions by selectively binding to other molecules (referred to as targets) by virtue of their 3-D structures.



Binding of a protein and its target.

Hot Spots (cont'd)

- The protein-target interactions occur at specific regions in the 3-D structure of a protein, known as active sites.
- Hot spots are small groups of amino acids surrounding the active sites; they stabilize the active sites and carry out the energy transfers in a protein-target interaction.

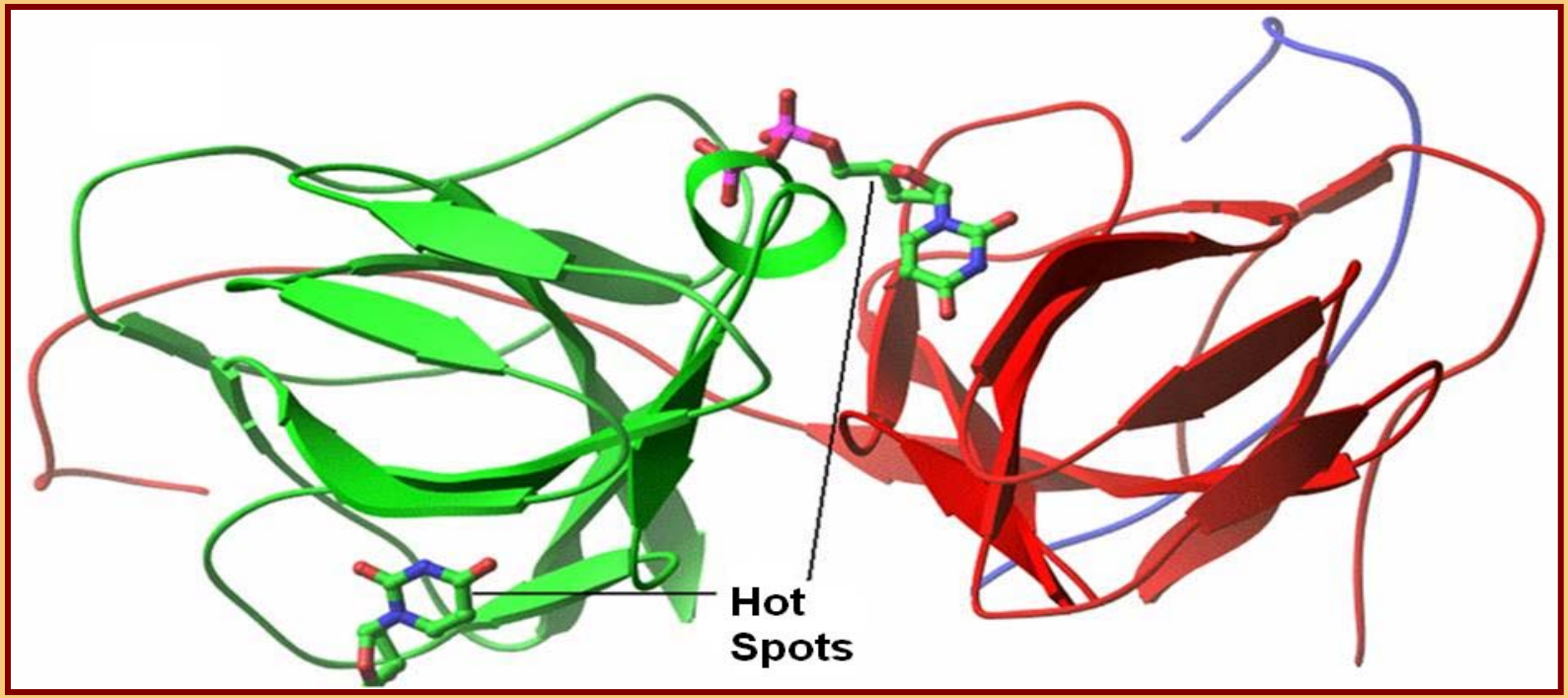


Illustration of hot spots.

The Resonant Recognition Model (RRM)

- The energies of free electrons in the amino acids can be represented by numbers called electron-ion interaction potential (EIIP) values.

  By assigning these numbers to the amino acids, a protein can be represented by a numerical sequence.

- The amplitude spectrums of the numerical sequences (i.e., magnitudes of their DFTs) of proteins belonging to a functional group exhibit a peak at a unique frequency referred to as the characteristic frequency of the group (Cosic, 1994).

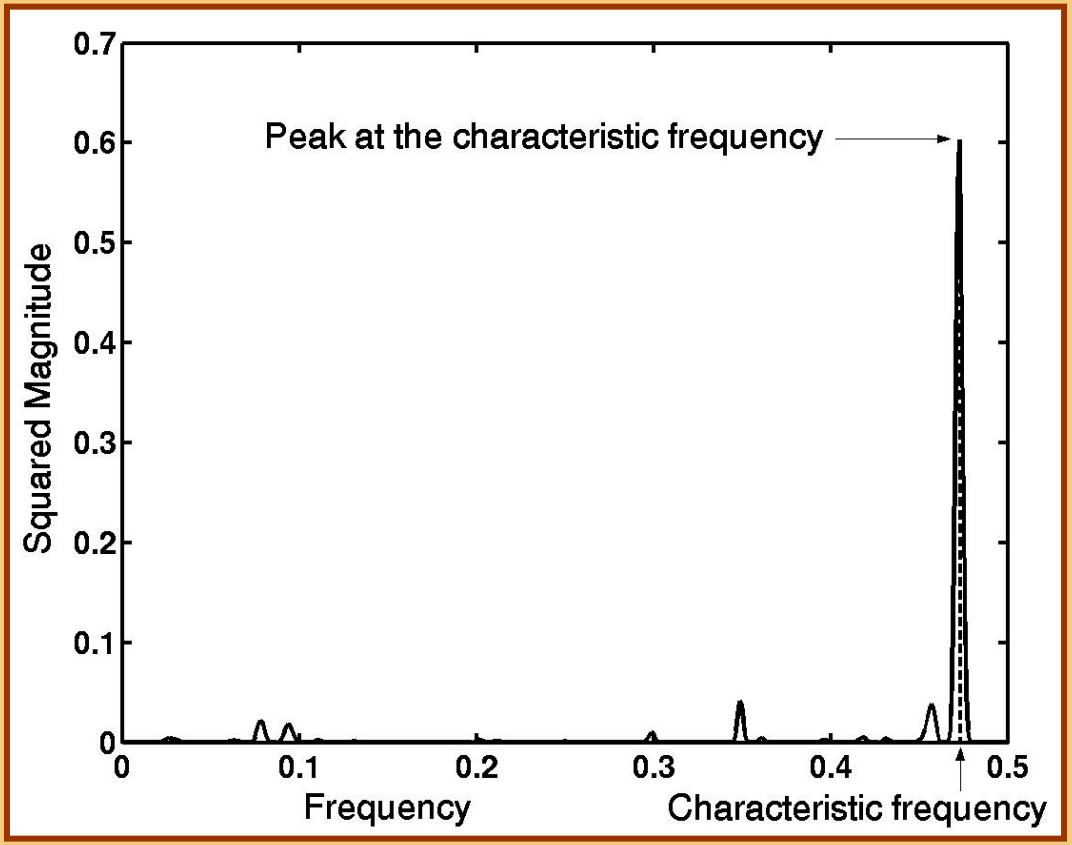
RRM (cont'd)

- The characteristic frequency can be determined from

$$S(e^{j\omega}) = \left| X_1(e^{j\omega}) X_2(e^{j\omega}) \cdots X_M(e^{j\omega}) \right|$$

$S(e^{j\omega})$ is known as the consensus spectrum.

- Hot-spot locations correspond to the regions in the numerical sequence where the characteristic freq. is dominant.



Consensus spectrum of cytochrome C functional group.

Hot-Spot Location Using Digital Filters

- The Algorithm:

1) Convert several protein sequences of the functional group of interest into numerical sequences.
2) Compute their DFTs and consensus spectrum.
3) Design a narrowband bandpass digital filter that would select the characteristic frequency.

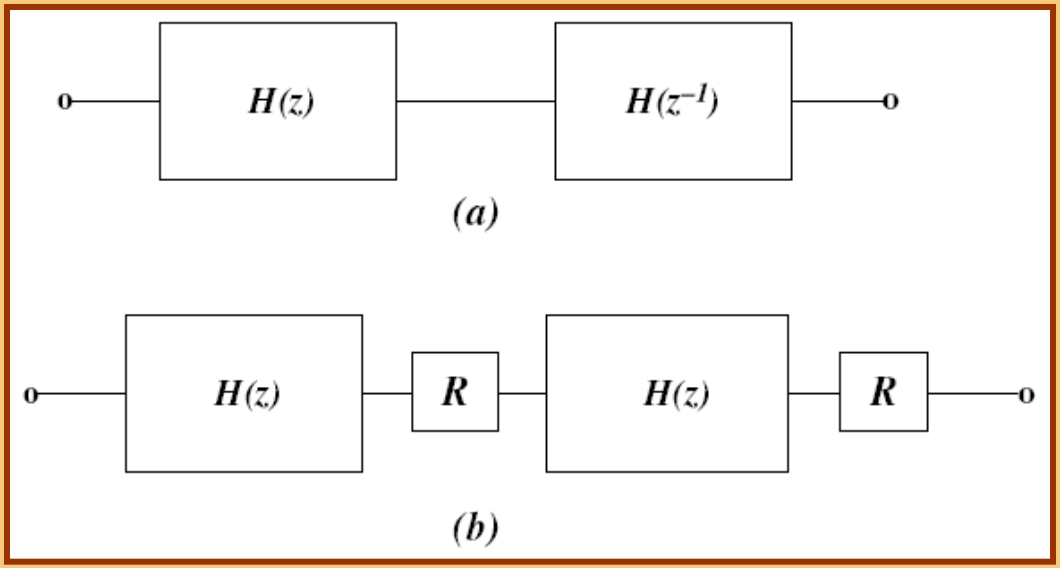
Hot-Spot Location Using Digital Filters

- Algorithm cont'd:

4) Filter the protein numerical sequence of interest by using the digital filter.
5) Compute the energy of the filter output.
6) Find the hot spots by locating the energy peaks.

Zero-Phase Filtering

- Filter delay can be eliminated altogether by using zero-phase filtering.

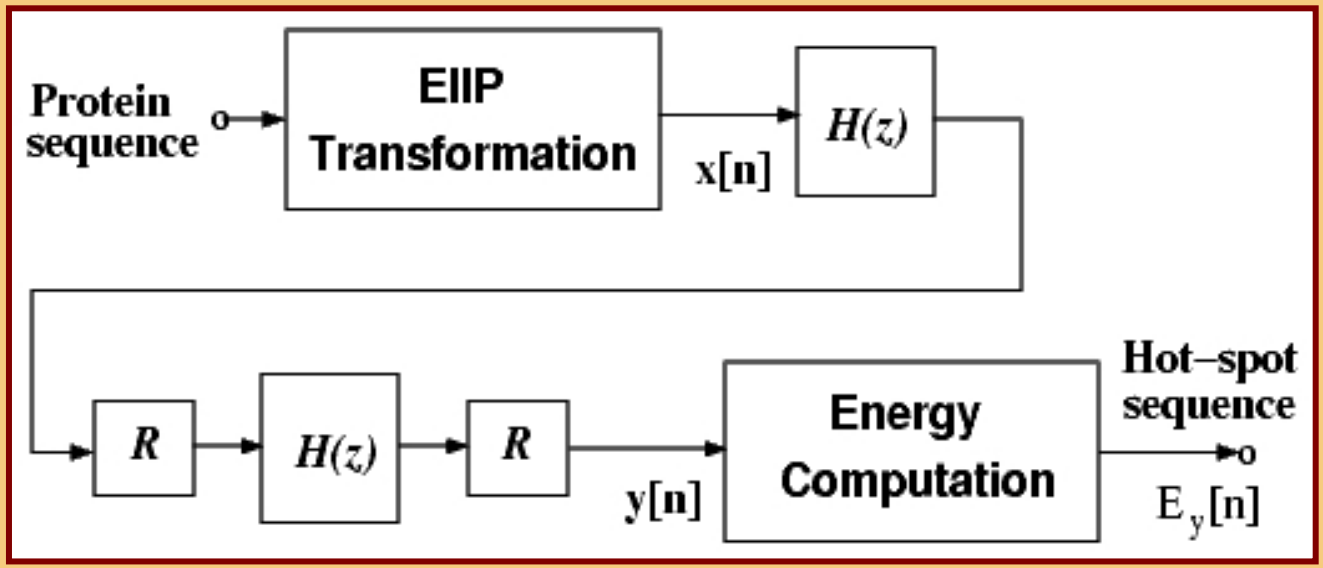


(a) Zero-phase filter, (b) Implementation.

- The frequency response of the arrangement is

$$H_0(e^{j\omega T}) = H(e^{j\omega T})H(e^{-j\omega T}) = \left|H(e^{j\omega T})\right|^2$$

Block Diagram



The digital-filter based hot-spot location system.

- Energy of the output sequence is given by

$$E_y[n] = \left| y[n] \right|^2$$

Our Previous Work

- We used an inverse-Chebyshev digital filter (IIR) since it gives good selectivity and its amplitude response does not exhibit passband oscillations.

  Filter order required for satisfactory performance: 8.

Proposed Technique

- We propose a technique to design an optimal second-order bandpass notch (BPN) filter.
- Advantages: Very low order, i.e., 2^{nd} order but high selectivity. Hence, more accurate hot-spot location.

Optimization Problem Formulation

A rudimentary strategy of designing a bandpass notch (BPN) filter would be

$$G_{BP}(z) = 1 - G_{BS}(z) \rightarrow (1)$$

$G_{BP}(z)$ → transfer function of bandpass filter

$G_{BS}(z)$ → transfer function of bandstop filter

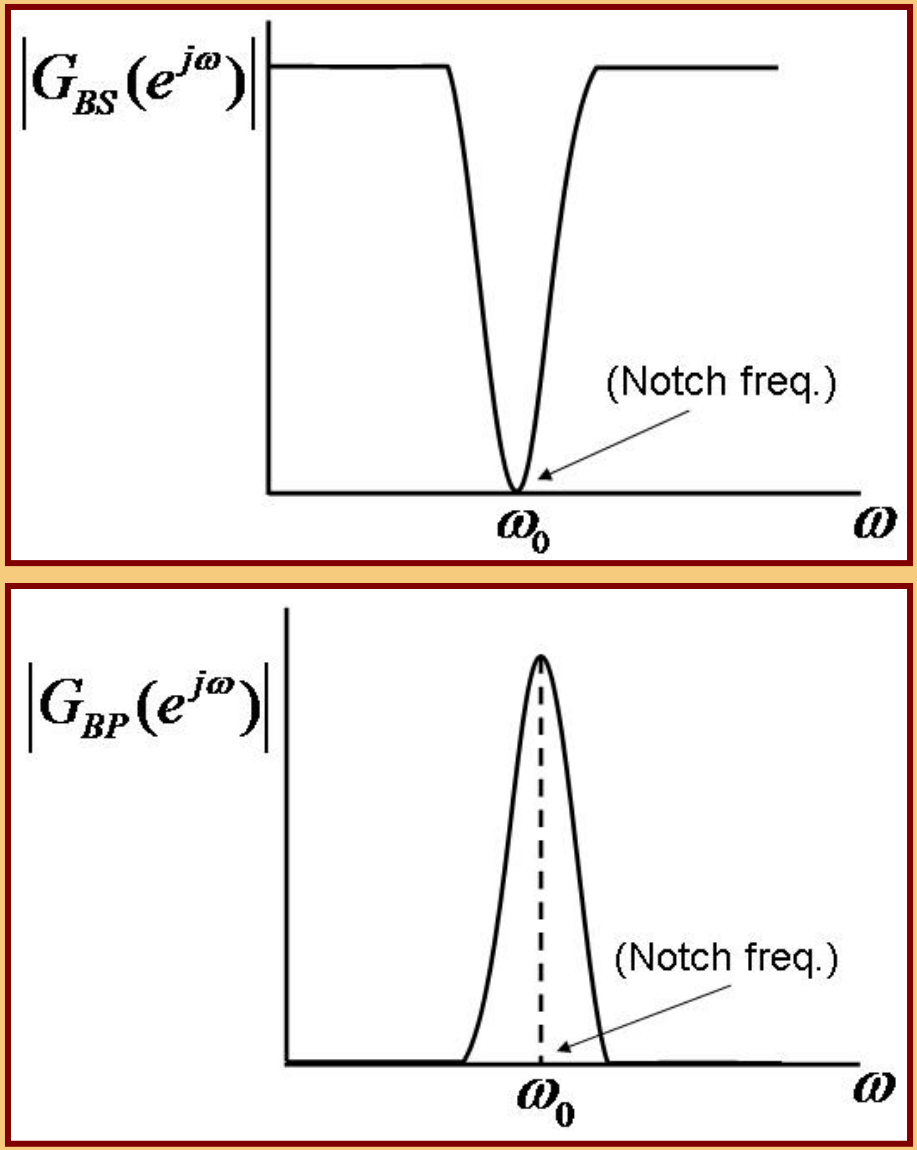
$G_{BS}(z)$ takes the form

$$G_{BS}(z) = \tfrac{1}{2}\left[1 + A(z)\right] \rightarrow (2)$$

where

$$A(z) = \frac{d_0 + d_1 z^{-1} + z^{-2}}{1 + d_1 z^{-1} + d_0 z^{-2}}$$

corr. to a 2nd order allpass filter.



Amplitude responses.

Problem Formulation (cont'd)

From (1) and (2), we obtain

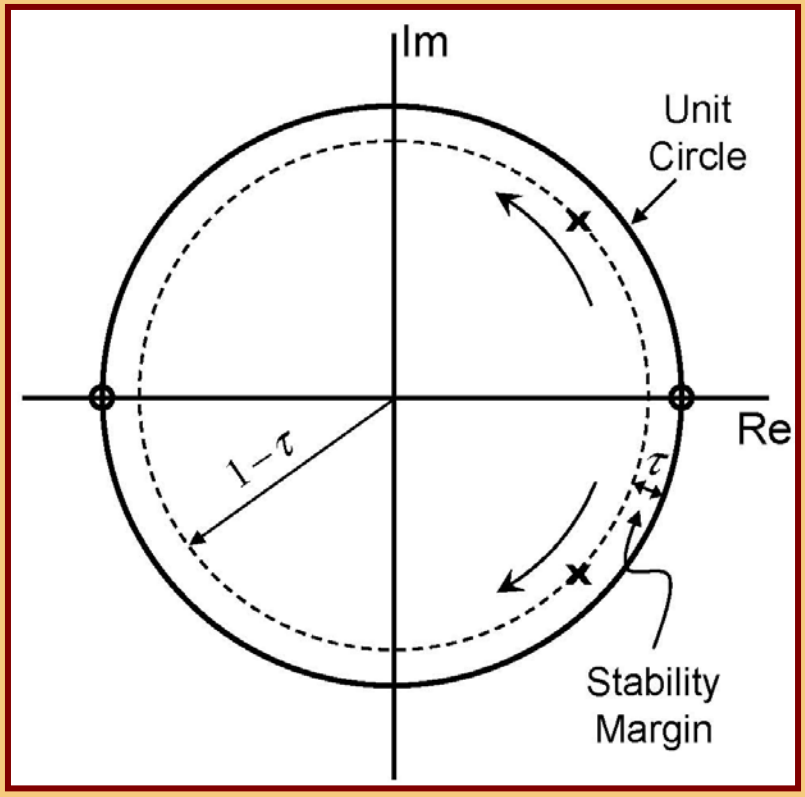
$$G_{BP}(z) = \tfrac{1}{2}\left[1 - A(z)\right] = \frac{1}{2}\left[\frac{(1-d_0)(1-z^{-2})}{1+d_1 z^{-1}+d_0 z^{-2}}\right]$$

which is the transfer function of an allpass-based second-order BPN filter.

Zeros → always on REAL axis at $z = \pm 1$

Poles → move along semicircles of radius $1-\tau$ as freq. varies from 0 to π.

The above transfer function has two variables, d_0 and d_1, which are the filter coefficients.



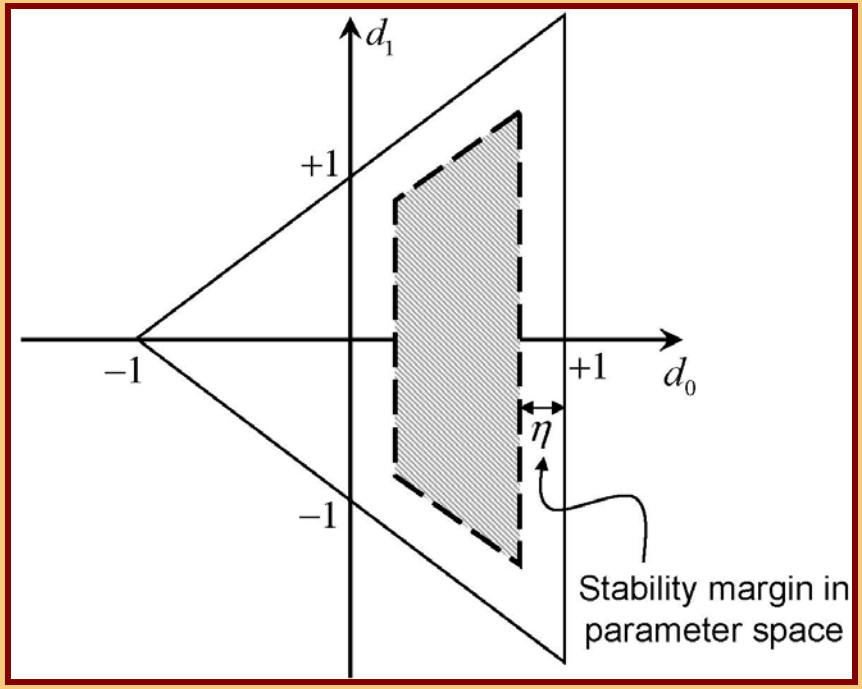
Zero-pole plot.

Problem Formulation (cont'd)

Feasible region in (d_0, d_1) space is governed by the stability conditions and is a triangle.

Since $d_0 = r^2$, d_0 takes only nonnegative values.

Allowing a stability margin, η, the feasible region is reduced to the inner trapezoid.



Tradeoff exists between selectivity and sensitivity.

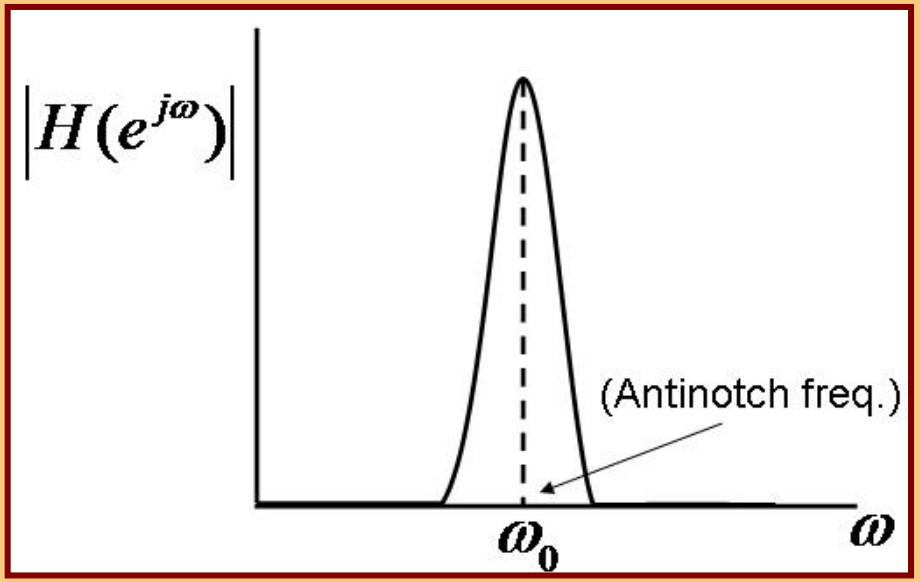
Best design → Fix η, set $d_0 = 1 - \eta$, and then determine d_1 such that the area under the amplitude response is minimized.

Problem Formulation (cont'd)

Minimize $\underset{d_1}{J} = \int_R \left| G_{BP}(e^{j\omega}) \right|^2 d\omega$

for $R = [0, \omega_0 - \varepsilon] \cup [\omega_0 + \varepsilon, \pi]$

subject to $-2d_0 \le d_1 \le 2d_0$



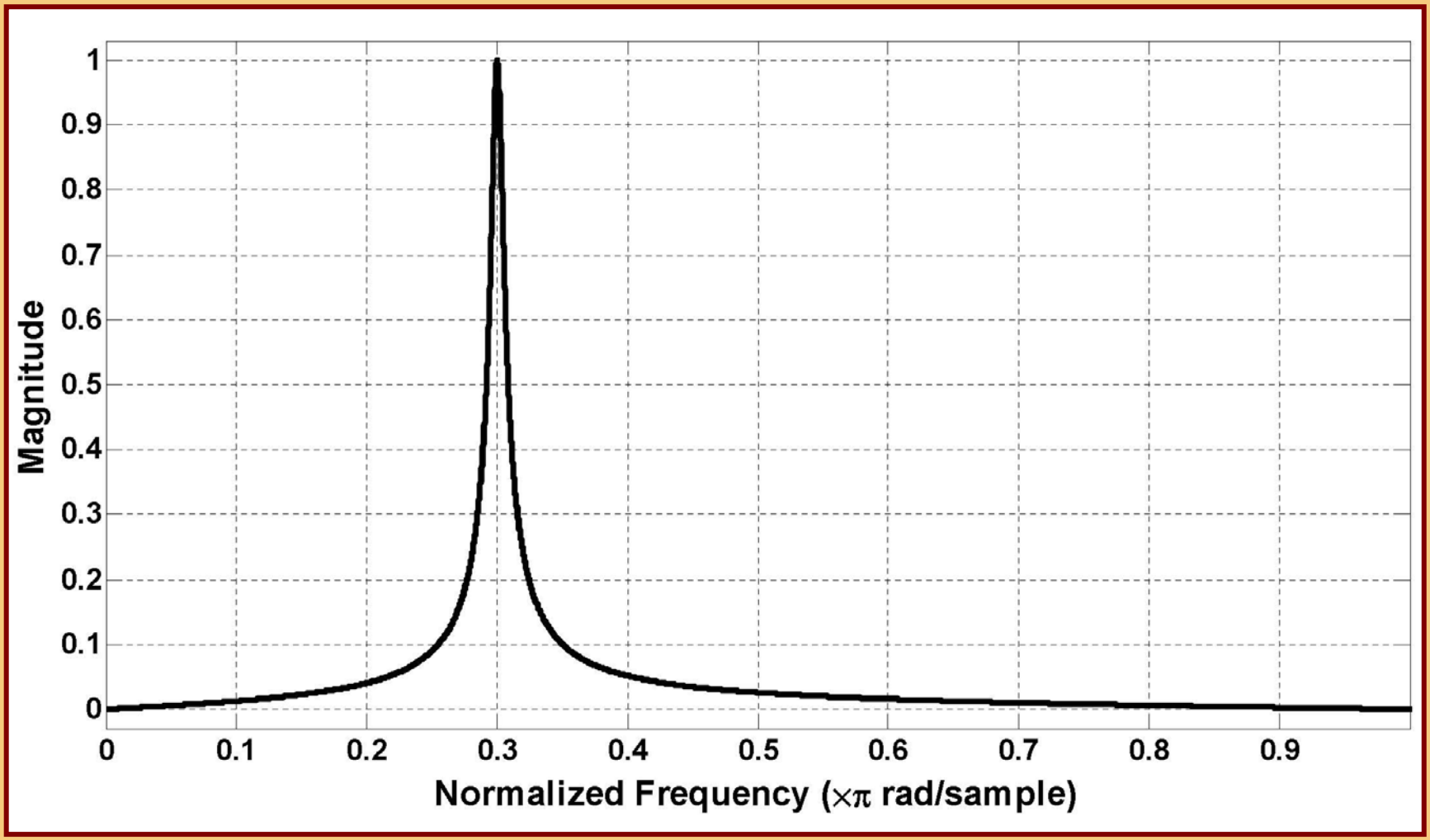
ε is a small positive scalar.

On simplification, J becomes

$$J = \tfrac{1}{2} \int_R \frac{(1-d_0)^2(1-\cos 2\omega)}{1+d_0^2+d_1^2+2d_1(1+d_0)\cos\omega+2d_0\cos 2\omega} d\omega$$

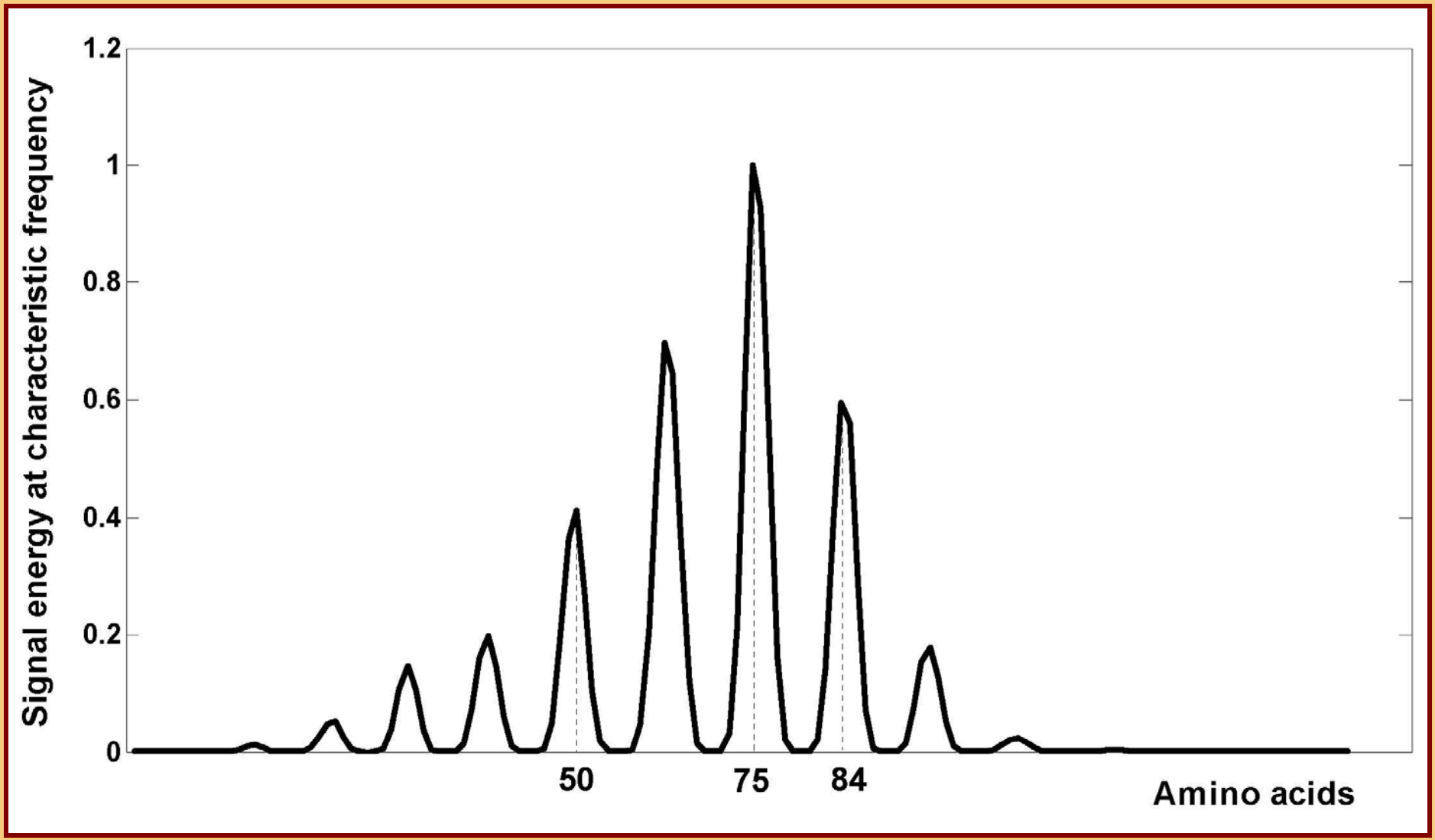
Implementation: Using one-dimensional search methods, such as golden section.

Examples and Results



Amplitude response of filter example with $\omega_0 = 0.3$ rad/s, $\eta = 0.03$, $\varepsilon = 10^{-3}$. The filter was designed using golden section search in 32 iterations with a termination tolerance of 10^{-6}.

Results (cont'd)



Hot Spots of *Cellulomonas fimi* Endoglucanase C Protein.

(Identified using the proposed BPN filter)

Performance Comparison

Protein Name	Hot Spots		
	Inv.-Cheby.	BPN	ASEdb
C. *fimi* endoglucanase C	50	50, 75, 84	19, 50, 75, 84
bacteria tryptophan RNA-binding attenuator protein (TRAP)	37, 40, 56	37, 40, 56, 58	37, 40, 56, 58
E. Coli colicin-E9 immunity protein (IM9)	34, 41, 50, 51, 55	33, 34, 41, 50, 51, 55	33, 34, 41, 50, 51, 55

Potential Hot Spots

Protein Name	Potential Hot Spots Identified by the Filter-Based Technique
C. *fimi* endoglucanase C	14, 26, 36, 68, 90
bacteria tryptophan RNA-binding attenuator protein (TRAP)	7, 15, 23, 35, 48, 64, 68, 72
E. Coli colicin-E9 immunity protein (IM9)	14, 19, 25, 30, 46, 62

Conclusions

- A high-selectivity second-order BPN digital filter was designed and was then applied for finding the locations of hot spots in proteins.
- Preliminary results have shown that the BPN filter can identify the locations of hot spots with better accuracy than the inverse-Chebyshev filter used earlier.
- The technique can be used to build inexpensive hot-spot location systems that could be used by biologists as a first step in analyzing newly discovered proteins.
- The results obtained can be used to narrow down the search for hot-spot locations, thus significantly reducing the number of expensive wet lab experiments.