

TUNING TECHNIQUE FOR THE LOCATION OF HOT SPOTS IN PROTEINS USING A BANDPASS NOTCH DIGITAL FILTER

Parameswaran Ramachandran, Wu-Sheng Lu, and Andreas Antoniou

Department of Electrical and Computer Engineering
University of Victoria, BC, Canada, V8W 3P6
Email: rpara26@ieee.org, wslu@ece.uvic.ca, aantoniou@ieee.org

ABSTRACT

A tuning technique is proposed for the location of hot spots in proteins using a bandpass notch (BPN) digital filter. The technique involves tuning the notch frequency of the filter over a small frequency range. To facilitate tuning, a user-friendly graphical interface has been implemented using MATLAB. Simulations have shown that the use of the tuning technique leads to improved hot-spot predictions compared to results obtained using untuned BPN filters.

1. INTRODUCTION

Accurate location of hot spots in proteins is an important aspect in the broader problem of understanding protein function. To locate hot spots, biologists perform *site-directed mutagenesis* [1] experiments which are complex and expensive in terms of time and resources. Several computational techniques have been developed that provide good estimates of hot-spot locations using computer-generated models of protein function. Using these estimates, biologists can selectively perform laboratory experiments to confirm the hot-spot locations thereby saving a considerable amount of time and resources.

A simple and effective strategy to computationally predict hot spots based solely on a protein's amino acid sequence is to employ the *resonant recognition model* (RRM) [2]. According to the RRM, each protein functional group corresponds to a unique frequency component known as the *characteristic frequency* of the group. The hot spots of a protein are the amino acids corresponding to the regions in the protein numerical sequence where the characteristic frequency is dominant. Identifying these regions amounts to locating hot spots. In previous work, we introduced a filter-based technique for identifying hot-spot locations using a second-order bandpass notch (BPN) digital filter [3, 4]. The technique was shown to yield accurate results. Here, we propose an enhancement that can

be used to improve the accuracy further by tuning the BPN filter. The enhancement uses a least-squares polynomial model to characterize the variations of the filter coefficients in response to small changes in the notch frequency. This model can then be used for computing the filter coefficients for slight changes in the notch frequency without having to redesign the filter. The paper also describes a graphical interface that implements the technique in the MATLAB environment. Simulations show that the use of the tuning technique yields improved hot-spot predictions compared to results obtained using untuned BPN filters.

The paper is organized as follows. Section 2 briefly describes the RRM and the filter-based hot-spot location technique. Section 3 presents the proposed tuning technique. Section 4 describes the implementation details in terms of the graphical interface developed using MATLAB. Section 5 demonstrates the usefulness of the technique by applying it to a set of example protein sequences.

2. THE RRM AND FILTER-BASED HOT-SPOT LOCATION

Predicting the locations of hot spots using the resonant recognition model involves first mapping a set of protein character sequences belonging to the functional group of interest into numerical sequences. For this mapping, a physical measure known as the *electron-ion interaction potential* (EIIP) is employed [2]. An EIIP value exists for each of the 20 amino acids. The discrete Fourier transforms (DFTs) of the EIIP sequences are then computed and are multiplied pointwise to obtain what is known as a *consensus spectrum*. A distinct peak in the consensus spectrum yields the characteristic frequency of the functional group. Identifying the regions in a protein EIIP sequence where the characteristic frequency is dominant would locate its hot spots.

Once the characteristic frequency is identified, different techniques can be adopted for locating hot spots.

In [3,4], we introduced a filter-based technique in which a narrowband bandpass digital filter is designed to select the characteristic frequency component and the EIIP sequence is filtered using a zero-phase filter [5]. A plot of the output signal power versus location reveals the hot spots as distinct peaks. An inverse-Chebyshev IIR filter was initially used and, subsequently, in [4], a second-order BPN filter was proposed which is designed as follows. The transfer function of the BPN filter is given by

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

where d_0 and d_1 denote the coefficients of the filter. Coefficient d_0 is equal to the square of the pole radius and controls the selectivity of the filter. As d_0 approaches unity, the selectivity is increased but at the same time the sensitivity of the filter to roundoff errors is also increased. The best design for a specified stability margin can be achieved by fixing d_0 and determining d_1 such that the area under the amplitude response curve is minimized. This is a unimodal minimization problem, and hence its solution can be readily obtained by using a 1-dimensional optimization technique such as the *golden-section search* [6].

3. PROPOSED TUNING TECHNIQUE

Although the BPN filter-based technique is effective in locating hot spots, the accuracy of the results obtained depends on how accurately the characteristic frequency has been determined. This, in turn, depends on the lengths of the EIIP sequences used to compute the consensus spectrum. Typically, protein sequences are short of the order of a couple of hundred amino acids. This limits the accuracy of the characteristic frequency as determined by the consensus spectrum, which may be offset by a small amount relative to the true characteristic frequency. This problem can be circumvented by tuning the notch frequency, ω_0 , of the BPN filter in order to achieve an accurate estimate of the true characteristic frequency. Such a tuning must be computationally efficient and must not require redesigning the filter so that the effects of changing the notch frequency can be instantaneously observed in terms of the hot-spot results. In what follows, we model the variations of the filter coefficients in response to small changes in the notch frequency using a least-squares polynomial of order two. The model is then used to tune the notch frequency of the BPN filter.

The zero-pole plot of the BPN filter is shown in Figure 1. The zeros of the BPN filter are always located on the real axis at $z = \pm 1$ and, for a fixed stability margin, τ , with respect to the unit circle of the z plane,

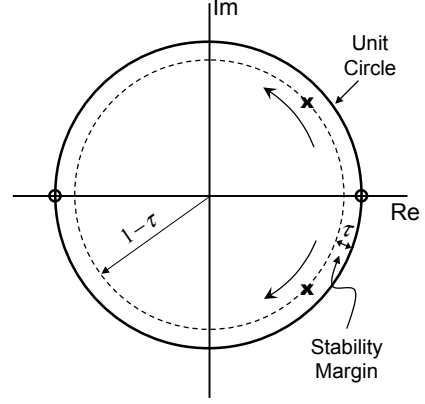


Figure 1. Zero-pole plot corresponding to the BPN filter. Zeros are denoted by ‘o’ and poles by ‘x’.

$d_0 = (1 - \tau)^2$ and hence the poles move along semicircles of radius $1 - \tau$ as the notch frequency varies from 0 to π in a manner shown in Figure 1. With d_0 fixed, coefficient d_1 is a function of the notch frequency, ω_0 , as illustrated in Figure 2. For small variations in ω_0 , coefficient d_1 can be expressed as a quadratic function of ω_0 as

$$d_1(\omega_0) = d_{10} + \alpha_1(\omega_0 - \omega_{0U}) + \alpha_2(\omega_0 - \omega_{0U})^2 \quad (2)$$

where d_{10} and ω_{0U} are the values of d_1 and the notch frequency in the untuned BPN filter, and α_1 and α_2 are coefficients to be determined. The notch frequency is bounded by the interval $\omega_l \leq \omega_0 \leq \omega_u$ centered at ω_0 , where $\omega_l = \omega_0 - \epsilon$ and $\omega_u = \omega_0 + \epsilon$ with ϵ being a small positive constant. To determine α_1 and α_2 , a set of N BPN filters, say 10, can be designed with notch frequencies uniformly distributed in the interval $[\omega_l, \omega_u]$. The values of d_1 for these filters can then be used to form the overdetermined system

$$\mathbf{\Omega} \mathbf{\alpha} \approx \mathbf{d} \quad (3)$$

where

$$\mathbf{\Omega} = \begin{bmatrix} \omega_l - \omega_0 & (\omega_l - \omega_0)^2 \\ \vdots & \vdots \\ \omega_u - \omega_0 & (\omega_u - \omega_0)^2 \end{bmatrix} \quad \mathbf{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

and

$$\mathbf{d} = \begin{bmatrix} d_{1\omega_l} - d_{1\omega_0} \\ \vdots \\ d_{1\omega_u} - d_{1\omega_0} \end{bmatrix}$$

The shortest-length least-squares solution for (3) can be obtained by

$$\mathbf{\alpha} \approx \mathbf{\Omega}^+ \mathbf{d} \quad (4)$$

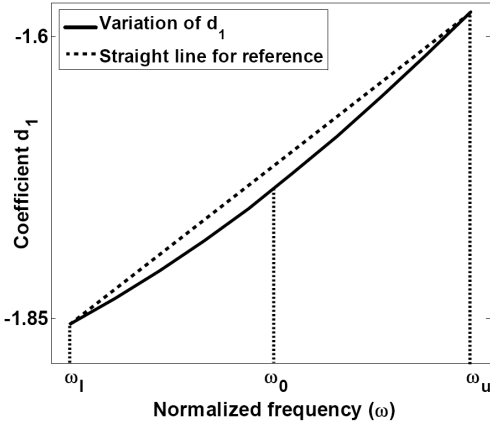


Figure 2. Variation of d_1 with small changes in the notch frequency. The values of ω_l , ω_0 , and ω_u are 0.1, 0.15, and 0.2, respectively.

where Ω^+ denotes the Moore-Penrose pseudo-inverse of Ω [7]. The values of α_1 and α_2 thus obtained can be used in (2) to determine coefficient d_1 for any BPN filter with notch frequency in the interval $[\omega_l, \omega_u]$. In effect, with the coefficients of the untuned BPN filter and the values of α_1 and α_2 known, the coefficients of any BPN filter with notch frequency in the interval $[\omega_l, \omega_u]$ can be obtained with good accuracy using the model in (2) without going through the filter design process. Tuning can, therefore, be easily achieved by simply changing the notch frequency and instantaneously observing the effects of the change in terms of the hot-spot results.

4. IMPLEMENTATION

To facilitate the application of the tuning technique, a graphical user interface (GUI) has been implemented in MATLAB.¹ A screen shot of the GUI is shown in Figure 3. The basic operation of the software is as follows. The user first selects a set of protein sequences using the listbox and clicks on “Compute Consensus Spectrum.” This displays the consensus spectrum identifying the characteristic frequency as a distinct peak. The software also assigns default values for the filter design parameters. The user can keep these values or alter them as desired. Then the user clicks on “Design Filter” to design the untuned BPN filter. Clicking on “Predict” will predict hot-spot locations for the selected protein sequences using the untuned BPN filter and display corresponding plots in separate windows. Finally, clicking on “FreqVarGUI” will deduce the least-squares model and invoke the slider shown

¹The MATLAB source code can be obtained by contacting the authors.

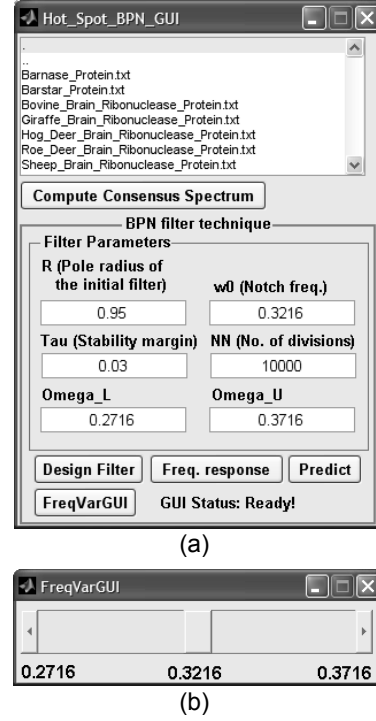


Figure 3. Screen shot of the tuning GUI: (a) the hot-spot GUI to design the untuned BPN filter and predict default hot-spot locations; (b) the slider for the notch frequency.

in Figure 3(b) which can be used to tune the notch frequency of the filter. As the notch frequency changes with the movement of the slider, the software computes the new filter coefficients and updates the plots to reflect the new hot-spot locations. For more details regarding the filter design parameters in Figure 3(a), the reader is referred to [4].

To predict the hot-spot locations of an unknown protein sequence, a training set of known protein sequences belonging to the same functional group as the unknown sequence can first be processed to calibrate the GUI. The software can then be applied to the unknown sequence for accurate prediction of its hot spots. In a typical tuning session, the user varies the notch frequency from ω_l to ω_u and records the frequency that yields the maximum number of hot-spot locations. This frequency can be taken to be an accurate estimate of the true characteristic frequency.

5. RESULTS

To demonstrate the usefulness of the proposed tuning technique, we applied it to a set of four example protein sequences. The values used for N and ϵ were 10 and 0.05, and ω_l , ω_0 , and ω_u were set to 0.2716, 0.3216,

Table 1. Hot-spot Locations Obtained by Tuning the BPN Filter

Protein name	Char. frequency		Hot-spot locations		
	Untuned BPN filter	Tuned BPN filter	Data reported in ASEdb	Untuned BPN filter	Tuned BPN filter
<i>C. fimi</i> endoglucanase C	0.093	0.098	19, 50, 75, 84	50, 75, 84	19, 50, 75, 84
human basic FGF	0.904	0.900	24, 96, 103, 140	24, 96, 103	24, 96, 103, 140
barnase	0.321	0.323	27, 54, 59, 60, 73, 82, 83, 87, 102, 103	27, 54, 59, 60, 73, 83, 87, 102, 103	27, 54, 59, 60, 73, 82, 83, 87, 102, 103
barstar	0.321	0.323	33, 35, 38, 39, 42, 73, 76, 80	33, 35, 38, 42, 73, 76, 80	33, 35, 38, 39, 42, 73, 76, 80

and 0.3716, respectively. The results are summarized in Table 1. For comparison, the table also lists results obtained using biological methodologies reported in the alanine scanning energetics database (ASEdb) [8]. From the table, it is evident that by using the tuning technique to adjust the notch frequency of the BPN filter, we were able to predict certain hot-spot locations that were not predicted using the untuned BPN filter. Locations 19 in endoglucanase C, 140 in FGF, 82 in barnase, and 39 in barstar correspond to these locations. It should be mentioned that proteins barnase and barstar belong to the same functional group and hence share the same characteristic frequency.

6. CONCLUSION

A tuning technique has been proposed for improving the accuracy of hot-spot locations predicted using a second-order BPN filter. The technique has been implemented in the form of a graphical interface in the MATLAB environment. The usefulness of the technique has been demonstrated by applying it to a set of example protein sequences. Simulation results have shown that by using the proposed tuning technique to adjust the notch frequency of the BPN filter, certain hot-spot locations that were missed by the untuned BPN filter can now be identified.

7. ACKNOWLEDGMENT

The authors are grateful to the Natural Sciences and Engineering Research Council of Canada for supporting this research.

8. REFERENCES

- [1] J. A. Wells, "Systematic mutational analyses of protein-protein interfaces," *Methods in Enzymology*, vol. 202, pp. 390–411, 1991.
- [2] I. Cosic, "Macromolecular bioactivity: Is it resonant interaction between macromolecules?—Theory and applications," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 12, pp. 1101–1114, Dec. 1994.
- [3] P. Ramachandran and A. Antoniou, "Identification of hot-spot locations in proteins using digital filters," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 378–389, Jun. 2008.
- [4] P. Ramachandran, W.-S. Lu, and A. Antoniou, "Improved hot-spot location technique for proteins using a bandpass notch digital filter," in *IEEE International Symposium on Circuits and Systems*, Seattle, May 2008, pp. 2673–2676.
- [5] A. Antoniou, *Digital Signal Processing: Signals, Systems, and Filters*. New York: McGraw-Hill, 2005.
- [6] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. New York: Springer, 2007.
- [7] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore: The John Hopkins University Press, 1996.
- [8] Alanine Scanning Energetics database (ASEdb). [Online]. Available: <http://nic.ucsf.edu/asedb>