

Improved Hot-Spot Location Technique for 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—An improved technique for the location of hot spots in proteins based on the use of a bandpass notch (BPN) filter is described. The BPN filter is designed by specifying a stability margin and then minimizing the area under the amplitude response so as to achieve maximum selectivity for the chosen stability margin. Preliminary results indicate that the use of a BPN filter leads to improved hot-spot prediction compared to results obtained with classical filters investigated earlier by the authors such as inverse-Chebyshev filters. The results also show that the improved technique yields predictions that are consistent with results based on biological methodologies. In addition, the technique reveals certain new ‘potential’ hot-spot locations which could be investigated further by the biological community.

Index Terms—Proteins, hot spots, characteristic frequency, resonant recognition model (RRM), electron-ion interaction potential (EIIP), bandpass notch digital filters.

I. INTRODUCTION

Proteins, the building blocks of all life, consist of linear chains made of subunits known as *amino acids* [1]. There are 20 different types of amino acids, and all proteins are made from various combinations of these amino acids. A typical protein is represented by an alphabet sequence where each alphabet stands for a particular amino acid. The protein linear chains fold in certain unique ways to form complex 3-D structures. By virtue of these structures, proteins perform their biological functions through selective interactions with other molecules known as *targets*.

A general property of protein-target interactions identified by various studies over the past two decades (e.g., [2], [3]) is that most of the binding energy in an interaction is contributed by a small portion of the total number of amino acids comprising the interacting proteins. These few amino acids are termed *hot spots* and are responsible for the stability of the interacting sites as well as the protein-target complex as a whole. Due to the crucial role played by hot spots, thorough knowledge about their locations is essential in understanding protein function. Therefore, reliable and efficient techniques for locating hot spots are required. In previous work, we introduced a filter-based technique for identifying protein hot-spot locations [4], [5]. An IIR inverse-Chebyshev filter was employed for the filtering process. Here, we investigate the application of a bandpass notch (BPN) filter and present the preliminary results obtained using a set of protein examples.

The paper is organized as follows. Section II briefly reviews the technique of using digital filters for the location of hot

spots in proteins. Section III presents the details of designing the second-order BPN filter. Section IV presents the results obtained by applying the BPN filter to a set of example protein sequences.

II. LOCATION OF HOT SPOTS IN PROTEINS USING DIGITAL FILTERS

Hot spots in proteins can be located by employing a model of protein-target interactions known as the *resonant recognition model* (RRM) [6]. In this model, the protein alphabet sequences are first converted into numerical sequences using physical parameter values known as the *electron-ion interaction potentials* (EIIPs). Each amino acid corresponds to an EIIP value, which denotes the average energy of the valence electrons in that amino acid. A sufficient number of such protein numerical sequences, all belonging to the functional group of interest, are selected and their discrete Fourier transforms (DFTs) are computed. The pointwise product of the DFTs is then obtained, which is known as the *consensus spectrum*. According to the RRM, each protein functional group corresponds to a unique frequency component known as the *characteristic frequency* of the group. For the functional group of interest, this frequency is associated with a distinct peak in its consensus spectrum.

According to the RRM, the hot spots of a particular 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 the hot spots. This can be achieved by filtering the numerical sequence of the protein of interest using a narrowband bandpass digital filter in order to select the characteristic frequency and then plotting the energy of the filtered output. The locations of the peaks in this plot correspond to the hot-spot locations.

Filtering will introduce a delay in the processed numerical sequence. Calculation of this delay is essential in order to select the appropriate output samples. Unfortunately, this calculation is not very straightforward for an IIR filter whereas for an FIR filter it tends to be quite large. A simpler solution is to eliminate the filter delay by using *zero-phase filtering* (see Sec. 12.5 in [7]). In this approach, the numerical sequence of the protein is first fed to the filter and the resulting output is then reversed and fed to the filter again. The output of the second filtering operation is then reversed to obtain the final output. The delay introduced by the first filtering operation

is canceled by the second filtering operation since the signal is fed backwards the second time. Thus, upon zero-phase filtering, the processed numerical sequence is not delayed at the output, and the need to compute the phase response of the IIR filter is eliminated. For further details regarding the hot-spot location procedure, the reader is referred to [4].

The inverse-Chebyshev bandpass filter of order 8 employed in [4] provided good selectivity and yielded hot-spot locations of acceptable accuracy but there is scope for further improvement in accuracy by employing a filter of lower order and higher selectivity due to the following reasons. The higher the order of a digital filter, the longer would be its transient response. A long transient response would make the filtering of a protein sequence inefficient because a significant portion of the sequence would have already passed through the filter by the time steady state is achieved. Hence, for an accurate identification of hot-spot locations, it is critical to have as low a filter order as possible. Classical filters, including inverse-Chebyshev filters, can be designed to achieve any desired selectivity but for narrowband highly selective filters a high filter order is required. As mentioned above, an 8th-order filter was found to be necessary for the application at hand to achieve acceptable performance. However, high selectivity along with low filter order can be achieved by using a second-order BPN filter, and this possibility is explored in the rest of the paper.

III. DESIGN OF SECOND-ORDER BANDPASS NOTCH DIGITAL FILTER

A rudimentary strategy of designing a BPN filter is to use the relation

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

where $G_{BP}(z)$ and $G_{BS}(z)$ are the transfer functions of a bandpass and a bandstop notch filters, respectively. Figure 1 shows schematic representations of the bandstop and bandpass notch amplitude responses. $G_{BS}(z)$ takes the form

$$G_{BS}(z) = \frac{1}{2}[1 + A(z)] \quad (2)$$

where $A(z)$ is a second-order allpass transfer function given by

$$A(z) = \frac{d_0 + d_1 z^{-1} + z^{-2}}{1 + d_1 z^{-1} + d_0 z^{-2}} \quad (3)$$

From (1) and (2), we obtain

$$G_{BP}(z) = \frac{1}{2}[1 - A(z)] \quad (4)$$

Substituting (3) in (4) yields

$$G_{BP}(z) = \frac{N(z)}{D(z)} = \frac{1}{2} \left[\frac{(1 - d_0)(1 - z^{-2})}{1 + d_1 z^{-1} + d_0 z^{-2}} \right] \quad (5)$$

which is the transfer function of an allpass-based second-order BPN filter. The corresponding zero-pole plot is shown in Figure 2. For this transfer function, the zeros 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, the poles move along semicircles of radius $1 - \tau$ as the notch

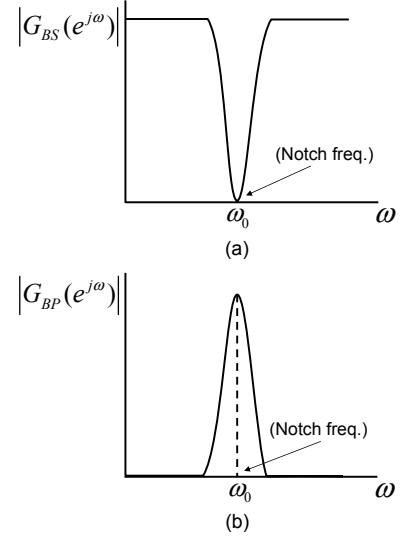


Fig. 1. Amplitude responses: (a) Bandstop notch filter, (b) bandpass notch filter.

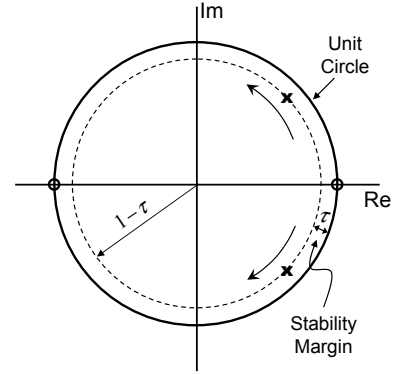


Fig. 2. A typical zero-pole plot of an allpass-based BPN transfer function. Zeros are denoted by 'o' and poles by 'x'.

frequency varies from 0 to π in a manner shown in Figure 2. From (5), it can be seen that the transfer function has two variables, namely, d_0 and d_1 which are the filter coefficients. It is well known that the feasible region in the (d_0, d_1) space is governed by the stability conditions and is a triangle as shown in Figure 3 [7]. For our problem, coefficient d_0 is equal to the square of the pole radius and hence can take only nonnegative values. This reduces the feasible region to the outer trapezoid at the right side of the d_1 axis. Allowing a stability margin η with respect to the parameter space, the feasible region is reduced to the area inside the trapezoid shown in Figure 3.

The selectivity of the BPN filter is determined by the proximity of the poles to the unit circle. The closer the locations of the poles are to the unit circle, the higher would be the selectivity. The best design for the hot-spot prediction technique turns out to be a compromise between high selectivity, on the one hand, and high sensitivity, on the other. As d_0 , which is equal to the square of the pole radius, approaches unity, a very high selectivity is achieved but at the same

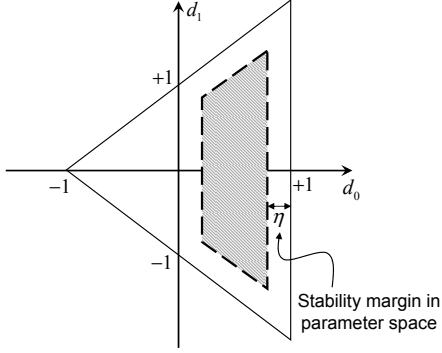


Fig. 3. Stability triangle in the (d_0, d_1) space.

time the sensitivity of the filter to roundoff errors increases whereas the stability margin decreases. The best design for the application under consideration would, therefore, be achieved by setting $d_0 = 1 - \eta$ and then determining d_1 such that the area under the amplitude response is minimized. The solution of the problem can be readily achieved by using a 1-dimensional optimization technique.

The design problem can be formulated as

$$\text{Minimize}_{d_1} J = \int_R |G_{BP}(e^{j\omega})|^2 d\omega \quad (6)$$

$$\text{for } R = [0, \omega_0 - \epsilon] \cup [\omega_0 + \epsilon, \pi]$$

$$\text{subject to } -2d_0 \leq d_1 \leq 2d_0 \quad (7)$$

where ϵ is a small positive scalar and ω_0 is the notch frequency of interest. The above interval for d_1 corresponds to the right edge of the shaded trapezoid in Figure 3. From (5) and (6), we have

$$J = \frac{1}{2} \int_R \frac{N(\omega)}{D(\omega)} d\omega \quad (8)$$

where

$$N(\omega) = (1 - d_0)^2(1 - \cos 2\omega) \quad (9)$$

and

$$D(\omega) = 1 + d_0^2 + d_1^2 + 2d_1(1 + d_0)\cos \omega + 2d_0 \cos 2\omega \quad (10)$$

The objective function is unimodal with respect to the constraint in (7) and takes the form shown in Figure 4. Many 1-dimensional optimization techniques can be used for the design of the BPN filter such as, for example, the golden-section search [8].

An alternative design approach for second-order BPN filters based on the Steiglitz-McBride procedure [9] has been described in [10].

A second-order BPN filter with a normalized notch frequency $\omega_0 = 0.3$ rad/s, stability margin $\eta = 0.03$, and $\epsilon = 10^{-3}$ was designed in 32 iterations using the golden-section search with a termination tolerance of 10^{-6} . The amplitude response achieved is shown in Figure 5.

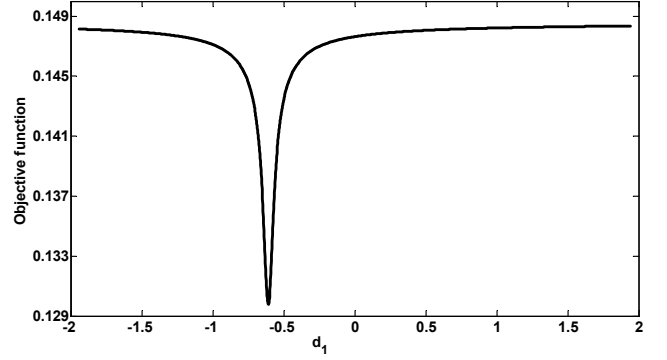


Fig. 4. Plot showing the unimodal nature of the objective function.

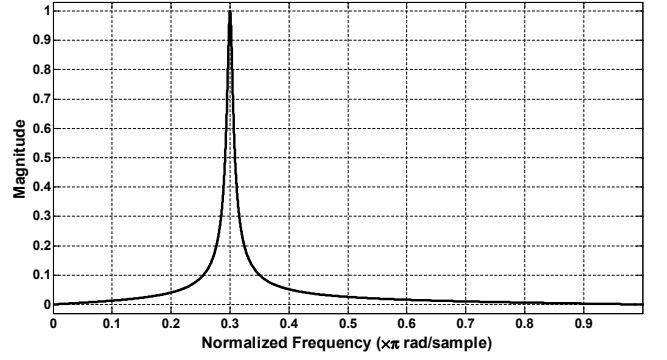


Fig. 5. Amplitude response of the filter example with $\omega_0 = 0.3$.

IV. APPLICATION OF THE BPN FILTER FOR THE LOCATION OF HOT SPOTS IN PROTEINS

In order to investigate the performance of the BPN filter, we applied it for the analysis of three example protein sequences. Figure 6 shows a sample hot-spot location plot and Table I lists the results corresponding to the three protein examples. Our results are compared with results obtained using biological methodologies reported in the alanine scanning energetics database (ASEdb) [11], [12] as well as with results we obtained earlier using an inverse-Chebyshev filter [4], [5].

ASEdb is a standard repository for hot-spot location data used by the biological community and is updated as and when new data become available. The procedure to obtain hot-spot locations for a protein from the ASEdb is as follows. We first note that in thermodynamic terms a hot spot is defined as an amino acid whose mutation to alanine leads to a *change in the binding free energy* (denoted as $\Delta\Delta G$) of at least 2.0 kcal/mol [13], [14]. ASEdb features a convenient search tool wherein the unique *protein data bank* (PDB) identifier of a given protein can be entered and a search can be executed for all documented mutations performed on the protein that correspond to a minimum $\Delta\Delta G$ of 2.0 kcal/mol. The resulting amino-acid locations would represent the hot spots that have been discovered for the protein so far.

The complete names of the example proteins are *Cellulomonas fimi* Endoglucanase C, bacteria tryptophan RNA-binding attenuator protein (TRAP), and *Escherichia coli*

TABLE I
HOT-SPOT LOCATIONS IDENTIFIED BY THE BPN AND THE INVERSE-Chebyshev FILTERS,
ALONG WITH THOSE REPORTED IN ASEdb

Example number	Protein name	Characteristic frequency	Hot-spot locations		
			Data reported in ASEdb	Inverse-Chebyshev filter	BPN filter
1	<i>C. fimi</i> endoglucanase C	0.093	19, 50, 75, 84	50	50, 75, 84
2	bacteria TRAP	0.247	37, 40, 56, 58	37, 40, 56	37, 40, 56, 58
3	<i>E. coli</i> IM9	0.190	33, 34, 41, 50, 51, 55	34, 41, 50, 51, 55	33, 34, 41, 50, 51, 55

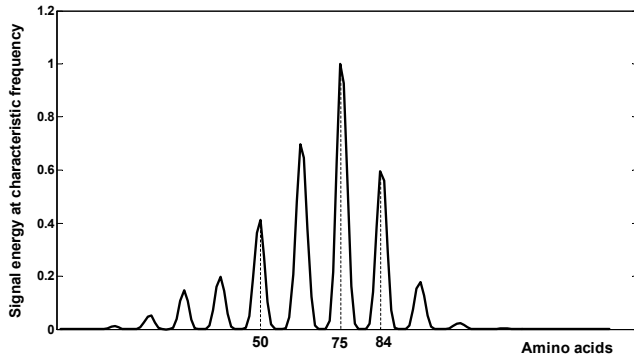


Fig. 6. Hot spots of *Cellulomonas fimi* Endoglucanase C Protein.

TABLE II
POTENTIAL HOT-SPOT LOCATIONS IDENTIFIED BY THE FILTER-BASED
TECHNIQUE

Example number	Protein name	Potential hot-spot locations
1	<i>C. fimi</i> endoglucanase C	14, 26, 36, 68, 90
2	bacteria TRAP	7, 15, 23, 35, 48, 64, 68, 72
3	<i>E. coli</i> IM9	14, 19, 25, 30, 46, 62

Colicin-E9 immunity protein (IM9). It can be observed from Table I that the BPN filter identified some hot-spot locations that were missed by the inverse-Chebyshev filter. Although these observations are preliminary, they indicate that the BPN filter can yield better accuracy in hot-spot locations. Comprehensive tests using a larger set of protein examples are currently being carried out in order to verify the validity of these results.

Finally, as can be seen in Table II, the filter-based technique identifies several potential hot-spot locations that have not been reported in the ASEdb independently of the type of filter used. This is evident from Figure 6. These new locations may prove to be crucial as and when more biological experiments reveal new insights in the future.

V. CONCLUSION

The application of a second-order BPN digital filter in a technique for the location of hot spots in proteins was investigated. The technique was applied to a set of protein examples and preliminary results indicate that the use of

the BPN filter can produce better results than the inverse-Chebyshev filter in terms of the accuracy of hot-spot locations. Computational hot-spot location techniques such as the one described are of immense help to biologists as they can provide good estimates of hot-spot locations for newly-discovered proteins. Furthermore, they can be carried out at virtually no cost. By using these estimates as guidelines, biologists can selectively perform expensive laboratory experiments to verify the predictions.

ACKNOWLEDGMENT

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

REFERENCES

- [1] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential Cell Biology*. New York: Garland Publishing, 1998.
- [2] T. Clackson and J. A. Wells, "A hot spot of binding energy in a hormone-receptor interface," *Science*, vol. 267, no. 5196, pp. 383–386, Jan. 1995.
- [3] T. Kortemme and D. Baker, "A simple physical model for binding energy hot spots in protein-protein complexes," *Proc. of the National Academy of Sciences (PNAS)*, vol. 99, no. 22, pp. 14 116–14 121, Oct. 2002.
- [4] P. Ramachandran and A. Antoniou, "Localization of hot spots in proteins using digital filters," in *IEEE Intl. Symp. on Signal Processing and Inf. Tech. (ISSPIT)*, Vancouver, Canada, Aug. 2006, pp. 926–931.
- [5] —, "A new technique for the identification of hot-spot locations in proteins using digital filters," *IEEE J. Sel. Topics Sig. Process.*, submitted.
- [6] I. Cosic, "Macromolecular bioactivity: is it resonant interaction between macromolecules?—theory and applications," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 12, pp. 1101–1114, Dec. 1994.
- [7] A. Antoniou, *Digital Signal Processing: Signals, Systems, and Filters*. New York: McGraw-Hill, 2005.
- [8] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. New York: Springer, 2007.
- [9] W.-S. Lu, S.-C. Pei, and C.-C. Tseng, "A weighted least-squares method for the design of stable 1-D and 2-D IIR digital filters," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 1–10, Jan. 1998.
- [10] C.-C. Tseng and S.-C. Pei, "Stable IIR notch filter design with optimal pole placement," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2673–2681, Nov. 2001.
- [11] K. S. Thorn and A. A. Bogan, "ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions," *Bioinformatics*, vol. 17, no. 3, pp. 284–285, 2001.
- [12] Alanine Scanning Energetics database (ASEdb). [Online]. Available: <http://nic.ucsf.edu/asedb/index.php>
- [13] A. A. Bogan and K. S. Thorn, "Anatomy of hot spots in protein interfaces," *Journal of Molecular Biology*, vol. 280, pp. 1–9, 1998.
- [14] X. Li, O. Keskin, B. Ma, R. Nussinov, and J. Liang, "Protein-protein interactions: Hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: Implications for docking," *Journal of Molecular Biology*, vol. 344, no. 3, pp. 781–795, 2004.