Predicting Peptide−Receptor, Peptide−Protein, and Chaperone−Protein Binding Using Patterns in Amino Acid Hydrophobic Free Energy Sequences†

Arnold J. Mandell* and Karen A. Selz

Cielo Institute, 486 Sunset Dr., Asheville, North Carolina 28804, and Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia 30322

Michael F. Shlesinger

Physical Sciences Division, Office of Naval Research, 800 N. Quincy St., Arlington, Virginia 22217

Received: October 27, 1999; In Final Form: February 1, 2000

Much of the current study of protein organization is aimed at understanding emergent polymeric structure in three rather than one dimensions. This is largely because the weak bonds between amino acid side chains in one-dimensional peptide heteropolymers that determine their equilibrium tertiary structures involve large loop interactions between sequentially distant sites. For this reason, searches for sequential patterns seem intuitively irrelevant even though the coding for 3D protein structure is present in the 1D peptide chain, On the other hand, we have found that there are particular circumstances in which matches in sequential patterns in amino acid side chain thermodynamic properties postdict signatory modularity in the tertiary structure of protein families and polypeptide−protein interactions, such as peptide−receptor, peptide−membrane transporter, nuclear factor−protein docking, and chaperone−protein binding. Here we describe and justify our computational approach to matching sequential organization, with examples from experimentally demonstrable, polypeptide−protein interactions. Using established values for each amino acid's hydrophobic free energy, "hydrophobicity", in kcal/mol, derived from their normalized free energy of transfer from nonpolar to polar bulk phases of binary solutions,[34] we study the variations in hydrophobicity along primary sequences and seek matching patterns among them. We analyze this discrete data series using the all-poles power spectrum for shorter polypeptides. For longer polypeptides and proteins, including membrane receptors, relevant protein domains, nuclear factors, and chaperones, we first decompose the data using autocovariance matrices into orthogonal eigenvectors, compose these eigenvectors with the original series to generate eigenfunctions, and then compute their power spectra. We show examples of sequential pattern-matched peptides that postdict binding of a non-peptide (estrogen) receptor, protein (calcineurin) binding to its T-cell nuclear factor (NFAT) docking site, and chaperone (GroEL) binding to one of its target enzyme proteins (β-lactamase). Polypeptide sequences and the associated proteins that bind them share distinguishing spectral features in their all poles power spectrum. This approach achieves practical significance because inversion of these analytic methods are now being used to successfully design de novo peptides which demonstrated their predicted physiological activity.

Introduction

Proteins are one-dimensional chains composed of the 20 essential (nutrition dependent) amino acids which are found in mammalian systems. Each amino acid is chained to the next by strong, covalent peptide bonds of approximately 100 kcal/mol. The three-dimensional shape assumed by the polymeric chain in its aqueous, physiological environment, however, is determined by four types of "weak" intermonomeric bonds of 2−5 kcal/mol. The weak bonds include van der Waals contacts, ≈0.5−1 kcal/mol, hydrogen bonds, ≈3−6 kcal/mol (though significantly distracted from full expression of polymeric self-relations by the surrounding water), ionic bonds (reduced functionally from as high as 80 kcal/mol to 1 kcal/mol due to shielding by high dielectric constant water) and the energetically most significant, hydrophobic bonds of up to 5 kcal/mol resulting from the attractive interaction of nonpolar substances which aggregate for surface minimization in polar water.[39] These weak bonds are functionally expressed against a background of ambient molecular kinetic energy at physiological (≈25 °C) temperatures of about 0.5 kcal/mol. Quantitative estimates of the hydrophobic, lipophilic seeking property of amino acids, free energies of solvation,[22,8] can be derived from their relative concentrations in organic versus water bulk phases of a binary solution,[34] normalized for glycine = 0 and parametrized by the Boltzmann constant. In the relative hydrophobicity scheme used here,[24] the amino acids fall generally into four categories of five members each.

The free energies of transfer, the hydrophobic free energies in kcal/mol of the amino acids are respectively glycine/0, glutamine/0, serine/0.07, threonine/0.07 asparagine/0.09; aspartate/0.66, glutamate/0.67, arginine/0.85, alanine/0.87, histidine/0.87; cysteine/1.52, lysine/1.64, methionine/1.67, valine/1.87, leucine2.17; and tyrosine/2.76, proline/2.77, phenylaline/2.87, isoleucine/3,15, tryptophan/3.77.

There is much evidence that the protein primary sequence codes information essential for the secondary and tertiary structures of a protein in the form of amino acid hydrophobicity sequences.[18] Guidance by global hydrophobic free energy

† Part of the special issue "Harvey Scher Festschrift".

10.1021/jp993810+ CCC: $19.00
Published on Web 03/18/2000

minimizing arrangements among sequential patterns of hydrophobic groups take the form of long-range (>20 Å versus 2−3 Å for van der Waals interactions), water hydrogen bond strain minimizing attractive force between hydrophobic surfaces.[35] This force has been measured directly (in newtons) by atomic force spectroscopy.[17] For examples, experiments have shown that chains that fold into helices or β-sheets can tolerate positional exchanges in specific amino acids of equal hydrophobicity without a change in thir secondary structures.[19,20,40] With respect to peptide−peptide hydrophobic attraction and aggregation, the folding of the protein interlukin-1β into sheets starts with the closing of a "hydrophobic zipper" that matches and draws together mode matched hydrophobic sequences to minimize their contact with water.[6,7,14]

For an N amino acid long protein, we treat the primary protein hydrophobicity sequence as a discrete N point data series. Early on, it was discovered that hydrophobic rotation numbers,[37] graphs representing the sequence of amino acids on "helical wheels",[19] hydrophobic moments,[9] and direct Fourier transformation of the data[25,27] revealed distinguishing hydrophobic cycles and/or spectral peaks for polypeptides and proteins, including membrane receptors. For example, generic α-helices have wavelengths of 3.6 aa and ≈1.5 Å linear physical distance per residue; generic β-strands have 2.1 aa per hydrophobic rotation and ≈3.3 Å distance per residue. Hydrophobic wavelengths ranging from 2.0 to 13.6 aa and beyond have been found in polypeptides and proteins.[30]

Polypeptides with as much as 60% disparity in the specific amino acid primary sequence manifest similar spectral modes and physiological behavior. For example, across species, power spectra of hydrophobic series for growth hormone-releasing factor and calcitonin in human, pig, rat, salmon, and cow possess the same dominant Fourier modes. Snake toxins and the Torpedo shark cholinergic receptors which bind them share Fourier modes.[25,27]

The direct Fourier transform has the disadvantage of requiring long data sets for precise resolution of frequencies and freedom from end effects, while polypeptide sequences, even in generic proteins are, relative to the usual length of data used in processing of signals, relatively short. This has encouraged us to seek more suitable methods of spectral analyses of hydrophobic sequences for the prediction of polypeptide binding to membrane receptor proteins,[30] membrane protein transporters,[26] and chaperones and their associated proteins.[28] We begin with a type of power spectrum analysis geared toward capturing peaked spectra from short data sets, such as that representing small peptides that bind to larger protein receptors. For the longer sequences of receptor and other proteins in the form of their autocovariance matrices, we first decompose the sequence into orthogonal eigenvectors, which, when composed with the original sequence, generate eigenfunctions. We study the spectra of these eigenfunctions and match their spectral peaks, "hydrophobic modes," to those found in the spectra of the short peptides.

Characterizing these hydrophobic modes and their matches has practical implications because our new (proprietary) work involves the inversion of some of these analytic procedures for the purpose of generating new protein-targeted peptides, which pilot studies in receptor transfected cell systems demonstrate the predicted physiological activitiy. The approach presented here is thus part of the analytic portion of a new and potentially orders-of-magnitude more efficient program of alternatives to searches for functionally relevant "lead compounds" in very large, randomly generated, peptide libraries.

Here we apply these techniques to (1) find signatory peptide hydrophobic binding wavelengths in the non-peptide estrogen receptor, ERα, thought to be involved in escape from tamoxifen chemotherapy for breast cancer for which active, potentially therapeutic peptides have been sought by using 18 phage-displayed peptide libraries, each with a complexity of 1.5×10^9 phage;[32] (2) find signatory peptide sequence binding modes of the calcineurin docking site to the nuclear factor of activated T-cells, NFAT; peptides for which have been sought as a less cytotoxic, immunological drug using serial optimization of affinity screens of randomly generated peptide libraries[1]; (3) find the signatory peptide sequence modes that postdict the binding of a heat shock protein, GroEL, which hydrophobically binds to and facilitates the proper folding of the enzyme β-lactamase as its chaperone.[41] These and similar settings involving unfolded, partially unfolded and/or solvent exposed amino acid strings are circumstances in which efforts to characterize polypeptide sequence patterns and using one-dimensional techniques appear rewarding.

All-Poles Power Spectrum

One of the most common methods used to seek latent, statistical, wavelike sequential organization in apparently irregular data is the power spectrum, $S(\omega)$, which is used to estimate the relative weights and the frequency content of a sequence of observables, the signal.

For the ideally infinite and stationary set of data x_i, the power spectrum is a real positive function, given by (where $z = \exp(i\omega)$)

$$S(\omega) \equiv |\sum_k x(k) \exp(ik\omega)|^2 = \sum_{k=-\infty}^{\infty} c_k z^k \qquad (1)$$

where the positive coefficents c_k represent an average over points separated by k spaces. For a finite set of N data points

$$c_k = \langle x(n)x(n+k)\rangle = \frac{1}{N-k}\sum_{n=1}^{N-k} x(n)x(n+k)$$

i.e., the autocovariance coefficients, c_k, are known quantities calculated directly from the data. In eq 1, when the number of data points is N, k will only range from $-N/2$ to $(N/2) - 1$. Since this restricted sum is a finite polynomial in z, it has only zeros (and no poles). This is called the all-zero power spectrum. If the power spectrum has peaks, then many coefficients are needed in the all-zero method to inscribe sharp peaks. If, as in the case of polypeptides and proteins, the number of data points is relatively small, then a combination of end effects and broad band modes can cause the all-zero power spectrum to poorly represent the frequency content of the data.

As an alternative to eq 1 let us choose to write $S(\omega)$ as[36]

$$S(\omega) = \frac{1}{|1 + \sum_{k=1}^{N} a_k z^k|^2} \approx \sum_{k=-N}^{N} c_k z^k \qquad (2)$$

Expanding the denominator in eq 2, we equate powers of z with the power of the sum on the rhs, for exponents from $-N$ to N, and in this fashion find the coefficients a_k in terms of the known c_k coefficients. The expansion of the denominator, of course, will contain all powers of z, and, when expanded, gives an infinite polynomial. Think of $(1 + z + z^2)$ vs $1/(1 - z)$. The expression for $S(\omega)$ can have poles because the denominator in

eq 2 can have zeroes. That is why eq 2 is called the all-poles power spectrum. For values of $z(\omega)$ where the denominator is a minimum, $S(\omega)$ will have relatively sharp peaks, even when derived from relatively short data sequences. It is in this way that the all-poles method is superior to the all-zeros method for representing the intrinsically short sample lengths of polypeptides and proteins. We have used the all-poles method directly to obtain reliable and related spectra of amino acid series as short as 15−20 points that represent the sequence of hydrophobic values of some of the peptide sequences the mediate their binding to receptors and other proteins.[30,26]

As an intuitive heuristic, we will sketch a proof that eq 2 is the form of the power spectrum that maximizes the entropy of the signal thus explaining the exchangeable use of its two names: all poles and maximum entropy (MEM) power spectra. From the Wiener−Khinchine theorem, we can write the power spectrum as the Fourier transform of $\langle x(n)x(n+k)\rangle$. Inverting this transform, in the $k = 0$ case gives

$$\langle x^2(t)\rangle = \int_0^\infty S(\omega)\, d\omega \tag{3}$$

Let us now assume that $x(t)$ is a Gaussian random process that we decompose into its Fourier modes, $x(t) = \sum_{k=1}^{\infty}(p_k \cos(\omega kt) + q_k \sin(\omega kt))$ and thus

$$\langle x^2(t)\rangle = \sum_{k=1}^{\infty}(\langle p^2k\rangle \cos^2(\omega kt) + \langle q^2k\rangle \sin^2(\omega kt))$$

$$= 2\sum_{k=1}^{\infty}\langle p^2k\rangle = \sum_{k=1}^{\infty} S(\omega k) \tag{4}$$

where a discrete version for of eq 3 has been used and $\langle p_k{}^2\rangle = \langle q_k{}^2\rangle = 2\sigma^2$ gives the variance in the kth spectral mode. The probability distribution for the kth spectral mode of $x(t)$ is $p(x_k) = 1/\sqrt{2\pi\sigma_k{}^2}\exp(-x_k{}^2/2\sigma_k{}^2)$ and the entropy H_k associated with the kth mode is $H_k = -\int p(x_k)\log p(x_k)\, dx_k$. For this Gaussian probability

$$H_k = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_k}{2\pi\sigma^2}\right)\left(-\frac{1}{2}\log(2\pi\sigma^2) - x_k{}^2/2\sigma^2\right) dx_k$$

$$\approx \log(\sigma^2) \approx \log\langle p_k{}^2\rangle \approx \log S(\omega)$$

Integrating over all frequencies gives the entropy of the signal as

$$H = -\int \log S(\omega)\, d\omega \tag{5}$$

We maximize the entropy of $S(\omega)$ subject to the known contraints calculated from the data, ie., the c_k's for $k = -N$ to N, by using the method of Lagrange multipliers, λ_k's

$$\delta(-\int \log S(\omega)\, d\omega + \sum_{k=-N}^{N} \lambda_k[\int S(\omega)z^k(\omega)\, d\omega - c_k]) = 0$$

$$\Rightarrow \left(\frac{1}{S(\omega)} - \sum_{k=-N}^{N} \lambda_k z^k\right)\delta S = 0 \tag{6}$$

This gives $S(\omega)$ in for all-poles form of eq 1 when pairing + and − exponents to give real terms. As noted above, other names for this all-poles form is the maximum entropy method (MEM) and an autoregressive model (ARM).[36]

Karhuenen−Loeve Orthogonal Decomposition

For longer data series we expand the series in terms of the orthogonal eigenfunctions of its autocovariance operator, which we call KL modes since we use a method derivative of the expansion of Karhuenen and Loéve.[15,21] This is followed by computation of the all-poles spectra of these KL modes. The generalization of the KL expansion we use is due to Broomhead and King (BK),[3,4] who used this approach to estimate the intrinsic dimensionality of a sequence of observables.[11] The BK modification of this method will be discussed more specifically in the following section.

Generally, the sum of the KL modes converge rapidly to the function representing the series being decomposed and KL modes are not arbitrarily predetermined as sines, cosines, Bessel, or other known orthogonal functions, but emerge uniquely from each specific hydrophobic sequence. Given that there are N data points x_i ($i = 1, ..., N$), we will assume that we can decompose the data series into N orthogonal functions ψ_n, i.e., $x_i = \sum_{n=1}^{N}\lambda_n\psi_n(i)$, where $\sum_{i=1}^{N}\psi_n(i)\psi_m(i) = \delta_{nm}$, $n = 1, ..., N$, $m = 1, ..., N$, and then we will show that these KL modes satisfy a certain eigenvalue problem. We could also provide the opposite proof that if the ψ_n satisfy the eigenvalue problem, then they constitute a proper orthogonal mode decomposition of the data. We want the ψ_n to appear as eigenfunctions (in the BK procedure, we find them by eigenvector composition with the original sequence) and the λ_n to be related to the associated eigenvalues. To accomplish this we form A_{ij} the autocovariance matrix of the data

$$A_{ij} - \langle x_i x_j\rangle = \sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m\langle\psi_n(i)\psi_m(j)\rangle \tag{7}$$

where the average is over an ensemble of data, and we assume that the average $\langle\psi_n(i)\psi_m(j)\rangle = \delta_{nm}\psi_n(i)\psi_m(j)$, giving

$$A_{ij} = \sum_{n=1}^{N}\lambda^2 n\psi_n(i)\psi_m(j) \tag{8}$$

Treating **A** as a matrix operator, we will show that the ψ_n's are its eigenvectors and the $\lambda^2 n$'s its eigenvalues, i.e.

$$\sum_{j=1}^{N} A_{ij}\psi_m(j) = \sum_{j=1}^{N}\sum_{n=1}^{N}\lambda^2 n\psi_n(i)\psi_n(j)\psi_m(j)$$

$$= \sum_{j=1}^{N}\lambda^2 n\psi_n(i)\delta_{nm} = \lambda^2 m\psi_m(i) \tag{9}$$

This means if one forms the autocovariance matrix of the data, its eigenfunctions realize the orthogonal decomposition of the data.

Broomhead−King Compositional Eigenfunctions

The $x_{i,i=1,...,n}$ sequences were used to generate an M-lagged data matrix from which $M \times M$ covariance matrices, C_M, were computed. These C_M were decomposed into a particular form of l orthogonal eigenfunctions, $\psi_l(j)_{l=1,...,M,j=1,...,n-M+1}$.[4,3,13,30,38] From the data column vectors (T ≡ transpose) $V_1^T = (x_1, x_2 ..., x_{n-M})$, $V_2^T = (x_2, x_3, ..., x_{n-M+1})$, ..., $V_M^T = (x_M, x_{M+1}, ..., x_n)$ and where $K = n - M + 1$, the sequence averaged dyadic product,

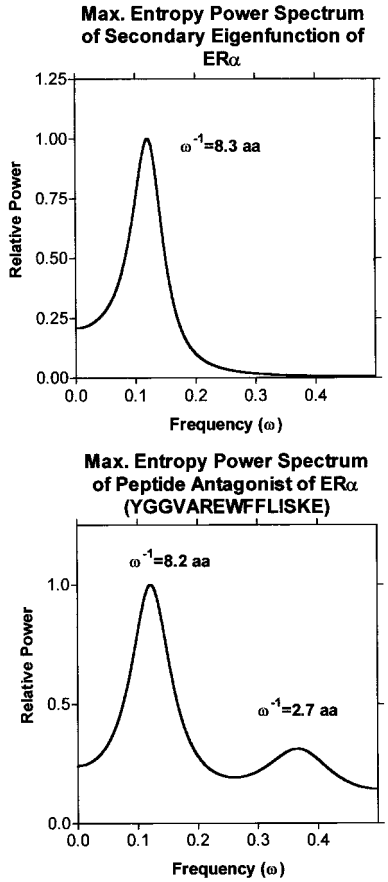

Figure 1. A dominant hydrophobic mode match is demonstrated by the $S(\omega)$ of the ψ_2 of ERα (top) and the $S(\omega)$ of the 15-residue peptide (bottom) that belongs to a small set of peptides that bind it that were found by surveys of 18 phage-displayed peptide libraries.[32] (See text.)

$x_i x_i^T$, is used to obtain the autocovariance matrix, a Hermitean $\mathbf{M} \times \mathbf{M}$ matrix, $C_M = 1/K\{x_i x_i^T\}$. M was chosen to minimize the least-squares error of many of the protein's ψ_l from their hydropathy plots. We compute the ordered eigenvalues, $\{\lambda_i\}_{i=1,...,M}$, and the associated ordered eigenvectors, $X_i(j)$, of C_M, where $i = 1, ..., M$ and labels the eigenvector, and $j = 1,...,M$ and refers to the jth components of the eigenvector X_i. The $\{\lambda_i\}_{i=1,...,M}$ are ordered from largest to smallest and constitute the eigenvalue spectrum of C_M. The similarly ordered associated $X_i(j)$ are then convolved with $x_1, x_2, ..., x_n$ into this particular form of $\psi_l(j)$, where $l = 1, ..., M$ labels the eigenvector and $j = 1, ..., n - M$ indexes the eigenfunction's jth component.

In words, the convolution of each of the leading eigenvectors with the hydrophobic free energy series is carried out by computing the sums of the scalar products of the M length eigenvector with an M length of the hydrophobic series to produce a point in the eigenfunction; this process is translated down the data series by one step and repeated to generate each of the sequential points of the eigenfunction that corresponds to its ordered eigenvalue associated eigenvector in the computation. In these studies, the ψ_l, $l = 1$, 2, or 3, of M are plotted as a function of the M lag-reduced sequence position.

Intuitively, C_M scans for hydrophobic modes across a range of autocovariance lengths from 1 to M, the range of the lags in the autocovariance matrices. Because C_M is real, symmetric ($x_{ij} = x_{ij}$) and normal ($C_M C_M^T = C_M^T C_M$), its $\{\lambda_i\}_{i=1,...,M}$ are real, nonnegative and distinct, and its associated $X_i(j)$ constitute a natural bases for orthonormal projections on $x_1, x_2, ..., x_n$.[13] The set of ψ_l can be regarded as orthonormally decomposed sequences of eigenvector-weighted, moving average values.[4]

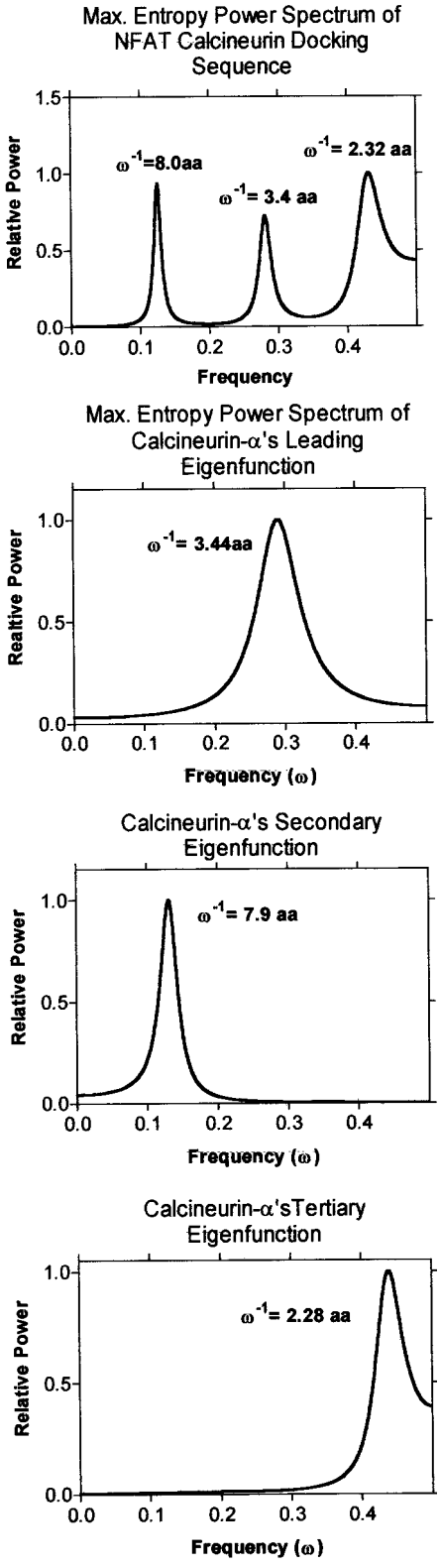

Figure 2. Multifrequency $S(\omega)$ of the calcineurin docking site on NFAT,[5] the nuclear factor of activated T-cells, (top row) and the $S(\omega)$'s of the $\psi_{1,2,3}$ of calcineurin (second, third and fourth rows) demonstrate correspondence among the modes. (See text.)

Whereas Broomhead and King were attempting to estimate the intrinsic dimensionality of a chaotic signal through a precise computation of M[3], we choose the value of M empirically such that the fit of the leading eigenfunction to the function resulting from the asymptotic, iterative, nearest-neighbor smoothing of the original hydrophobicity series minimizes least-squares

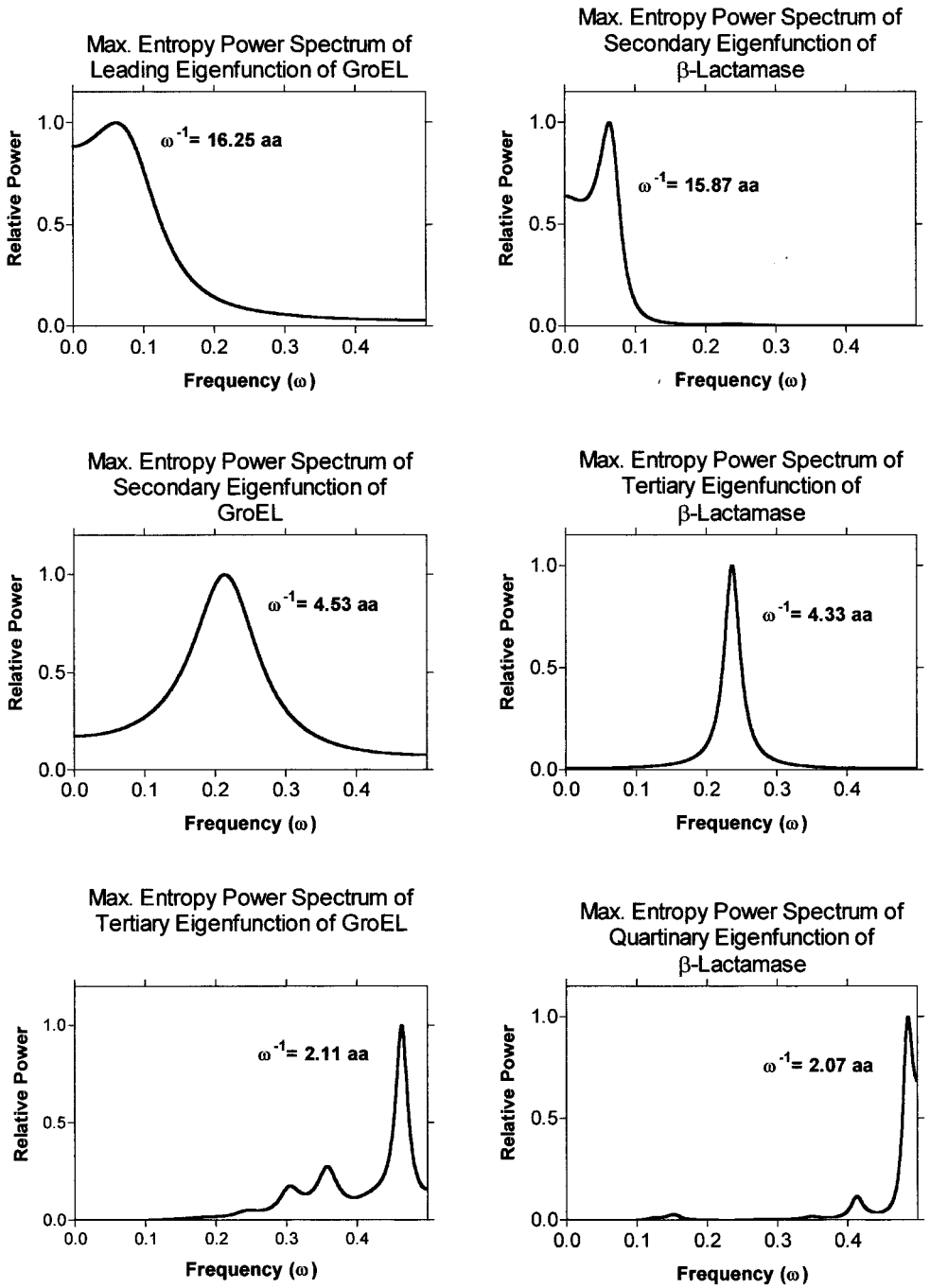

Figure 3. $S(\omega)$'*s* of the $\psi_{1,2,3}$ of the chaperone, GroEL (left column), evidence mode correspondences with the $S(\omega)$'*s* of the $\psi_{2,3,4}$ of the enzyme, β-lactamase for which it serves as a chaperone.[41]

error.[30] This normalization procedure turned out to be particularly helpful in our studies of G-protein coupled receptors with high-amplitude, long-wavelength modes representing the seven quasi-regularly spaced, highly hydrophobic transmembrane segments. This component of the hydrophobic sequence's signal dominates both the smoothed function and the leading eigenfunction of the $\mathbf{M} \times \mathbf{M}$ matrix. It was also useful in discriminating types of membrane channel proteins on the basis of their primary sequences.[38] Having "removed" much of the highly hydrophobic transmembrane components as the leading eigenfunction, the secondary eigenfunctions can then be studied for otherwise hidden, potential hydrophobic peptide binding modes. This and related problems can be approached by investigating the all-poles power spectra of some or all of the *M* eigenfunctions.

Whereas the choice of forming $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ would have produced a large $(N - M + 1) \times (N - M + 1)$ matrix, from studies of many ligand−receptor systems and the successful inverse algorithmic peptide design experiments noted above, we have found empirically (see below) that hydrophobically transformed polypeptide and protein sequences may contain almost all the required spectral information in the first few distinct eigenfunctions of the smaller $\mathbf{M} \times \mathbf{M}$ matrix. Viral proteins such as the HIV's viral coat may require as few as two.[27]

Three Examples of Functionally Relevant Peptide−Protein Matches in Spectral Signature

1. About half of all breast cancers express the estrogen receptor, ERα. Some anti-estrogens, such as tamoxifen, can inhibit cancer cell growth by binding to ERα, changing its

conformation (three-dimensional shape) and its associated tumor-promoting sensitivity to estrogen. Escape from this anticancer effect has been attributed to a spontaneous reversion of the conformation of ERα. Peptides are being sought which might bind ERα at some "away-from-the-active site" loop and either prevent or reverse this putative conformational transition thus preventing the emergence of resistance to tamoxifen. Using randomly generated, very large phage-displayed peptide libraries, peptides were found which bound ERα with high affinity.[32] We use the methods described above as the analytic part of an alternative technique to more efficiently find lead peptides that might bind ERα, and apply them here in a comparison of the hydrophobic modes of the peptides that were found using the phage display technique with those of ERα. The discovery of matching modes would contribute to the application of our peptide design strategy for finding lead peptide compounds. Though the peptides sought are short and receptors are several hundred amino acids long, we have discovered that the all-poles power spectra of the BK compositional eigenfunctions of the receptors and spectra of the hydrophobic sequences of the peptides that bind to them have dominant peaks located at similar frequencies.

We calculated the power spectrum of these candidate 15 residue, ERα-targeted peptides[32] directly via the all-poles method. Our method for elucidating the power spectrum of the 519 amino acid long ERα receptor hydrophobicity sequence was first to compute their compositional eigenfunctions, $\psi_i(j)$, as described above. We then calculated the all-poles power specta on the first few of the resulting $\psi_i(j)$.

Figure 1 (top) displays the all poles power spectrum of the secondary eigenfunction of the 519 residue ERα, which indicates a dominant hydrophobic free energy wavelength of 8.3 amino acids. The primary wavelength manifested by the power spectrum of the indicated ERα peptide antagonist is a close match at 8.2 amino acids, Figure 1 (bottom). Similar hydrophobic mode matches have been found in a number of neurotransmitter and hormonal peptide-receptor systems.[30,26]

2. Nuclear transcription involving the nuclear factors of activated T-cells, the NFAT family, are involved in immunological processes and are triggered by calcium mobilized by cell surface membrane receptors. The message is mediated by the calcium-activated protein, calcineurin with a catalytic domain, calcineurin-α, which binds NFAT at its conserved, "docking site".[5] Globally active and therefore toxic drugs, such as cyclosporin, have been used to interfere with the calcineurin-NFAT interactions. More specific and therefore less toxic peptides are sought to replace these and similar agents using affinity screening of randomly generated, large peptide libraries.[1] Figure 2 represents the results of using our methods to elucidate and compare the dominant hydrophobic modes in the calcineurinin docking site on NFAT and those of calcineurin-α. These results would lead naturally to the design of de novo peptides with the potential for binding to this site and interfering with the calcinureinin−NFAT interaction.

Figure 2 (top) is the power spectrum of the calcineurin docking site on NFAT,[1] which demonstrates leading hydrophobic spectral modes of 8.0, 3.4, and 2.32 aa wavelengths. Figure 2 (second, third, and fourth rows) display the power spectra of calcineurin-α's three leading eigenfunctions, with dominant spectral modes similar to those of NFAT: 3.44, 7.9, and 2.28 aa.

3. GroEL, a heat shock protein,[10] also known as chaperonin 60,[16] binds many newly synthesized proteins and helps facilitate, correct and/or maintain their optimal folded conformations by their weak binding and release in an interative process that prevents self-aggregation and/or semistable states in unproductive, conformational false minima.[12] Binding of proteins by chaperones usually involves the former in a partially unfolded, "molten gobular" state, such that one-dimensional subsequences are exposed for mutual recognition and binding. We have previously reported hydrophobic mode matches in a system involving chaperone-aided oligomeric protein assembly.[28] Figure 3 (left) represents the power spectra of the leading three hydrophobic free energy eigenfunctions of the chaperone, GroEL, manifesting modes of 16.25, 4.53, and 2.11 aa wavelengths. Figure 3 (right) contains the power spectra of the dominant modes of the second, third, and fourth leading eigenfunction of an enzyme, β-lactamase, known to be bound by and aided in finding its proper folded conformation by GroEL.[41] The wavelength modes of β-lactamase resemble those of GroEL are found with 15.87, 4.33, and 2.07 aa. This suggests that in this instance, the inversion of our analytic methods may lead to the de novo design of relatively short peptide "chaperones" for modulating physiological and/or correcting pathological self-aggregation in the protein folding process.

References and Notes

(1) Aramburu, J.; Yaffe, M. B.; Lopez-Rodriguez, C.; Cantley, L. C.; Hogan, P. G.; Rao, A. *Science* **1999**, *285*, 2129−2132.

(2) Brooks, C. L., 3rd; Gruebele, M.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11037−11038.

(3) Broomhead, D. S.; Jones, R.; King, G. P. *J. Phys. A* **1987**, *20*, L563−L569.

(4) Broomhead, D. S.; King, G. P. *Physica D* **1986**, *20*, 217−236.

(5) Crabtree, G. R. *Cell* **1999**, *96*, 611−614.

(6) Dill, K. A. *Biochemistry* **1990**, *29*, 7133−7155.

(7) Dill, K. A.; Fersht, A. R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 1−2.

(8) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199−203.

(9) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 140−144.

(10) Fayet, O.; Ziegelhoffer, T.; Georgopoulos, C. *J. Bacteriol.* **1989**, *171*, 1379−1385.

(11) Fukunaga, K.; Olsen, D. R. *IEEE Trans. Comput.* **1971**, *20*, 176−183.

(12) Gething, M.-J. *Guidebook to Molecular Chaperones and Protein-Folding Catalysts*; Sambrook-Tooze-Oxford: Oxford, UK, 1997.

(13) Golub, G. H.; Van Loan, C. F. *Matrix Computations*; Johns Hopkins Press: Baltimore, MD, 1993.

(14) Gronenborn, A. M.; Clore, G. M. *Science* **1994**, *263*, 536.

(15) Helstrom, C. W. *Statistical Theory of Signal Detection*; Pergamon Press: Oxford, UK, 1968.

(16) Hemmingsen, S. M.; Woolford, C.; van der Vies, S. M.; Tilly, K.; Dennis, D. T.; Georgopoulos, C. P.; Hendrix, R. W.; Ellis, R. J. *Nature* **1988**, *333*, 330−334.

(17) Israelachvili, J.; Wennerstrom, H. *Nature* **1996**, *379*, 219−225.

(18) Janin, J. *Prog. Biophys. Mol. Biol.* **1995**, *64*, 145−66.

(19) Kaiser, E. T.; Kezdy, F. J. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 1137−1143.

(20) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680−1685.

(21) Kittler, J.; Young, P. C. *Pattern Recognit.* **1971**, *5*, 335−352.

(22) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105−132.

(23) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8721−8725.

(24) Manavalen, P.; Ponnuswamy, P. K. *Nature* **1978**, *275*, 673−674.

(25) Mandell, A. J. *Annu. Rev. Pharmacol. Toxicol.* **1984**, *24*, 237−274.

(26) Mandell, A. J.; Owens, M. J.; Selz, K. A.; Morgan, W. N.; Shlesinger, M. F.; Nemeroff, C. B. *Biopolymers* **1998**, *46*, 89−101.

(27) Mandell, A. J.; Russo, P. V.; Blomgren, B. W. *Ann. N.Y. Acad. Sci.* **1987**, *504*, 88−117.

(28) Mandell, A. J.; Selz, K. A.; Shlesinger, M. *Physica A* **1997**, *244*, 254−262.

(29) Mandell, A. J.; Selz, K. A.; Shlesinger, M. F. *Proc. Int. Sch. Phys. "Enrico Fermi"* **1997**, *134*.

(30) Mandell, A. J.; Selz, K. A.; Shlesinger, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13576−13581.

(31) Mandell, A. J.; Selz, K. A.; Shlesinger, M. F. *J. Stat. Phys.* **1998**, *93*, 673−697.

(32) Norris, J. D.; Paige, L. A.; Christensen, D. J.; Chang, C.-Y.; Huacani, M. R.; Fan, D.; Hamilton, P. T.; Fowlkes, D. M.; McDonnell, D. P. *Science* **1999**, *285*, 744−746.

(33) Norris, J. D.; Paige, L. A.; Christensen, D. J.; Chang, C.-Y.; Huacani, M. R.; Fan, D.; Hamilton, P. T.; Fowlkes, D. M.; McDonnell, D. P. *Science* **1999**, *285*, 744−746.

(34) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1971**, *246*, 2211−2217.

(35) Pashley, R. M.; McGuiggan, P. M.; Ninham, B. W.; Evans, D. F. *Science* **1985**, *229*, 1088−1089.

(36) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C.; The Art of Scientific Computing*; Cambridge University Press: Cambridge, UK, 1988.

(37) Rose, G. D. *Nature* **1978**, *272*, 586−590.

(38) Selz, K. A.; Mandell, A. J.; Shlesinger, M. F. *Biophys. J.* **1998**, *75*, 2332−2342.

(39) Stryer, L. *Biochemistry*; Freeman: New York, 1988.

(40) Taylor, J. W.; Miller, R. J.; Kaiser, E. T. *J. Biol. Chem.* **1983**, *258*, 4464−4471.

(41) Zahn, R.; Pluckthun, A. *J. Mol. Biol.* **1994**, *242*, 165−174.