Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 13576–13581, December 1997
Biophysics

Mode matches and their locations in the hydrophobic free energy sequences of peptide ligands and their receptor eigenfunctions

ARNOLD J. MANDELL*†‡§, KAREN A. SELZ*†‡, AND MICHAEL F. SHLESINGER¶

*The Cielo Institute, 486 Sunset Drive, Asheville, NC 28804; †Department of Psychiatry and Behavioral Science, Emory University School of Medicine, Atlanta, GA 30322; ‡Department of Mathematics, Florida Atlantic University, Boca Raton, FL 33431; and ¶Physical Sciences Division, Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217

Communicated by Leo P. Kadanoff, University of Chicago, Chicago, IL, September 11, 1997 (received for review March 26, 1997)

ABSTRACT **Patterns in sequences of amino acid hydrophobic free energies predict secondary structures in proteins. In protein folding, matches in hydrophobic free energy statistical wavelengths appear to contribute to selective aggregation of secondary structures in "hydrophobic zippers." In a similar setting, the use of Fourier analysis to characterize the dominant statistical wavelengths of peptide ligands' and receptor proteins' hydrophobic modes to predict such matches has been limited by the aliasing and end effects of short peptide lengths, as well as the broad-band, mode multiplicity of many of their frequency (power) spectra. In addition, the sequence locations of the matching modes are lost in this transformation. We make new use of three techniques to address these difficulties: (*i*) eigenfunction construction from the linear decomposition of the lagged covariance matrices of the ligands and receptors as hydrophobic free energy sequences; (*ii*) maximum entropy, complex poles power spectra, which select the dominant modes of the hydrophobic free energy sequences or their eigenfunctions; and (*iii*) discrete, best bases, trigonometric wavelet transformations, which confirm the dominant spectral frequencies of the eigenfunctions and locate them as (absolute valued) moduli in the peptide or receptor sequence. The leading eigenfunction of the covariance matrix of a transmembrane receptor sequence locates the same transmembrane segments seen in *n*-block-averaged hydropathy plots while leaving the remaining hydrophobic modes unsmoothed and available for further analyses as secondary eigenfunctions. In these receptor eigenfunctions, we find a set of statistical wavelength matches between peptide ligands and their G-protein and tyrosine kinase coupled receptors, ranging across examples from 13.10 amino acids in acid fibroblast growth factor to 2.18 residues in corticotropin releasing factor. We find that the wavelet-located receptor modes in the extracellular loops are compatible with studies of receptor chimeric exchanges and point mutations. A nonbinding corticotropin-releasing factor receptor mutant is shown to have lost the signatory mode common to the normal receptor and its ligand. Hydrophobic free energy eigenfunctions and their transformations offer new quantitative physical homologies in database searches for peptide-receptor matches.**

The importance of the sequential arrangements of amino acid side chain hydrophobicities in the determination of peptide and protein secondary structures has been established knowledge in protein biology and physics for many decades (1, 2). Determinants of polypeptide interactions, such as those between peptide segments in protein folding, are coded in the one-dimensional sequence space of physical characteristics (3), most prominently the hydrophobic free energies of their individual amino acids (4). These chemical potentials have been approximated as relative solubility in organic vs. aqueous solvents in which glycine = 0 and quantified as hydrophobic free energies in kcal/mol (5, 6). The hydrophobic attractive forces in newtons between hydrophobic moieties as a function of surface area and radius of curvature are measured by using techniques such as atomic force microscopy (7) and are up to two orders of magnitude larger than those predicted by van der Waals theory and extend spatially in a slower-than-exponential decay to beyond 20 nanometers (8). Complete substitution of hydrophobically equivalent amino acids in peptides maintains and sometimes increments their peptide-receptor mediated physiological potency (9, 10). Helical secondary structures of differing turn lengths can be designed with sequences of amino acids of high and low hydrophobicities, independent of the specific amino acids chosen within each hydrophobicity class (11).

Secondary structures with matching hydrophobic amplitude variational frequencies, such as in the β-strands of interleukin 1β, have been shown to bind together and initiate protein folding (12) in a process called the "hydrophobic zipper" (13). Two long, helical secondary structures with congruent hydrophobic frequencies bind to create the central "hydrophobic knot" that stabilizes the structure of phospholipase A2 (14). Recent studies of the binding of extracellular domains of growth hormone receptor by polyclonal antibodies to ovine growth hormone was shown to be related to common helical, loop, and/or disordered secondary structure (15).

In earlier studies, we found that when the amino acid sequences of neuropeptides and peptide hormones were transformed into their individual hydrophobic free energies, functional families demonstrated similarities in hydrophobic free energy power spectral modes. We refer to the statistically dominant power spectral (inverse frequency) wavelengths in amino acid residues as $h(\omega)$. Family members shared the same $h(\omega)$, though differing in their ordered amino acid content by as much as 60% (16, 17). The range of $h(\omega)$ included the well known $h(\omega) = 3.6$ and $h(\omega) = 2.0$ of the α-helix and β-strand, respectively (18), but many others as well, ranging from the $h(\omega) = 13.10$ amino acid residue mode of acid fibroblast growth factor to the $h(\omega) = 2.18$, which dominates the hydrophobic free energy power spectrum of corticotropin releasing factor.

The AIDS coat protein manifests a waxing and waning $h(\omega) =$ 7–9 [observed by sliding a 50 residue windowed Fourier transform (19) along its sequence], which appears to be conserved across many of its mutations (17). Fibroblast growth factor (FGF; at the time, of unknown molecular mechanism) was predicted and confirmed to have a regulatory influence on ribonuclease A, with which it was found to share dominant hydrophobic modes (20). The specific amino acid sequences of the calcitonins, the peptide hormone family that regulates the rate of enzymatic bone catabolism, vary by ≈60% across species, but all are dominated by a $h(\omega) = 3.6$ (21). Our system of analysis demonstrates the expected

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: FGF, fibroblast growth factor; EGF, epidermal growth factor.

§To whom reprint requests should be sent at: The Cielo Institute, 486 Sunset Drive, Ashville, NC 28804.

hydrophobic mode invariance under translation and reflection that has been studied experimentally in new peptide design by using nucleotide complementation (22).

Among the difficulties encountered in our previous work is that the usual Fourier, power spectral transformation is at best a crude approximation when used on these short and irregular data series with aliasing, end effects and broad-band multimodality. This study reports the results of the new use of three consecutive linear decompositions and transformations, making hydrophobic mode isolation and identification less ambiguous. Their successful application to examples from a peptide and ligand receptor sequence database[‖] support the claim that these peptide ligand modes are functionally relevant.

The component of the peptide-receptor ligand binding process of interest here is the initial, long-range, faster-than-diffusive, selective aggregation ("hydrophobic zippering") of peptides with extramembranous transmembrane receptor segments. This process precedes their internalization and equilibrium charge or ionic binding within the membrane (23–25). Inside the low dielectric constant environment of the lipid membrane, amino acid side chains with charged and/or ionic groups are protected from charge-screening by the aqueous and ionic extramembranous fluid, and dominate the binding mechanism. In contrast, hydrophobic mode matches in peptide ligand-receptor pairs are assumed to play a significant role in the extramembranous, high dielectric constant, charge-screened, aqueous portions of the receptor sequences.

de Gennes' work (26) on the statistical mechanics of constrained polymers and polymeric reptation and models of the behavior of one-dimensional polymers with immobilized ends making "trains" and/or "loops" on a surface or in a membrane (27) predict that extramembranous ligand-receptor hydrophobic mode aggregation would increase the local hydrophobic free energy density above some critical threshold for internalization. In the vicinity of the metastable boundaries between the intramembranous and extramembranous portions of the receptor sequence, this would favor the internalization of that ligand-receptor segment into the membrane where charge and ion-dependent equilibrium binding can occur more easily.

Construction of Hydrophobic Free Energy Sequences and Their Eigenfunctions from Orthogonal Decompositions

Among the many amino acid hydrophobicity scales (28, 29), we have consistently used the results of partitions in binary, aqueous-organic solvents begun in the pioneering work of Edsall in the 1930s (5, 6). We chose this system because of its long history of successful use and our finding that it had a statistically significant correlation ($R^2 = 0.723, P = 0.00475$) with a measurable physical property, amino acid partial specific volume (in ml/gm) (30), in the physiological circumstance for which we wish to make predictive inferences. Tanford and coworkers (31) demonstrated that amino acid free energy relations were associated with aqueous cavity surface area derived from these measures made by using the aqueous partition. Our results are consonant with those achieved with the currently more popular, but less physiological, method of condensed vapor partition (32). This agreement in the patterns of the hydropathy plots of the seven transmembrane segments is seen, for example, in graphs of the very well studied rhodopsin I (see Fig. 1, upper left).

To assign hydrophobic free energy values, H_i, to each amino acid, A_i, of the peptide ligand and its transmembrane receptor, we used the results of studies in which each amino acid's H was determined from its relative concentration at equilibrium in the two phases of a binary solvent, and quantified as a free energy of transfer, in kcal/mol, from a nonpolar hydrocarbon solvent to water at 37°C, with glycine = 0 as a reference (5, 6). This results

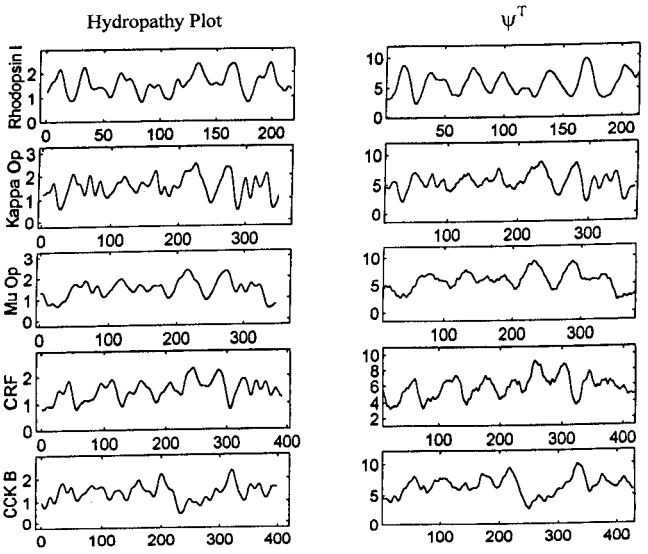

FIG. 1. (*Left*) Graphic results of asymptotically smoothing, 32 iterated nearest neighbor averaging computations, the "hydropathy plots" of the hydrophobic free energy series, H_i, of the amino acid sequences, A_i, of rhodopsin I, the kappa and mu opioid, the corticotropin releasing factor, and the cholecystokinin B seven-transmembrane receptors. (*Right*) Graphs of the leading (transmembrane) eigenfunction, Ψ^T, of the same H_i. Note that for the same H_i, the hydropathy plots and those of Ψ^T are quite similar.

in the following set of amino acid hydrophobic free energy equivalents, H_i, in kcal/mol: G, Q = 0.0; S, T = 0.07; N = 0.09; D = 0.66; E = 0.67; R = 0.85; A, H = 0.87; C = 1.52; K = 1.65; M = 1.67; V = 1.87; L = 2.17; Y = 2.76; P = 2.77; F = 2.87; I = 3.15; and W = 3.77.

Let each amino acid sequence of length N, $A_1, A_2, \ldots, A_N$, be represented by a sequence of hydrophobic free energy values, $H_1, H_2, \ldots, H_N$, where H_i represents the hydrophobic free energy of the amino acid A_i in the ith place in the amino acid sequence, using the H values listed above.

From the lagged vectors, $V_1 = (H_1, H_2, \ldots, H_M)$, $V_2 = (H_2, H_3, \ldots, H_{M+1}), \ldots, V_{N-M+1} = (H_{N-M+1}, H_{N-M+2}, \ldots, H_N)$, with $M \approx 16$, an $M \times M$ covariance matrix, C_M, is formed in which $K = N - M + 1$ and N ranges along the receptor sequence lengths from ≈400 to 1,200 residues. M is chosen to maximize the least square fits of the first eigenfunctions (Ψ^T, Fig. 1 *Right*) with the standard hydropathy plots (Fig. 1 *Left*) used to demonstrate the (seven) transmembrane segments (32).

$$C_M = \frac{1}{K}\begin{pmatrix} \sum_{i=1}^{K} H_iH_i & \sum_{i=1}^{K} H_iH_{i+1} & \cdots & \sum_{i=1}^{K} H_iH_{i+k-1} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{K} H_{i+K-1}H_i & \sum_{i=1}^{K} H_{i+K-1}H_{i+1} & \cdots & \sum_{i=1}^{K} H_{i+K-1}H_{i+K-1} \end{pmatrix}.$$

We compute the eigenvalues, $\{\nu_i\}_{i=1}^{M}$, and the associated eigenvectors, $X_i(j)$, of C_M, where $i = 1 \ldots M$ labels the eigenvector, and $j = 1$ to M refers to the jth component of the eigenvector $X_i(j)$. The $\{\nu_i\}_{i=1}^{M}$ are ordered largest to smallest as are the corresponding $X_i(j)$. The ordered $X_i(j)$ are then used as multiplicative "weights" to transform the $H_1, H_2, \ldots, H_N$ into M statistically weighted eigenfunctions, $\Psi_i(j)$, where $i = 1 \ldots M$ labels the eigenvector and $j = 1 \ldots N - M$ indexes its jth component. The $\Psi_i(j)$, for $j - k + 1 > 0$, are given by

$$\Psi_i(j) = \sum_{k=1}^{M} X_i(k) H_{j-k+1}.$$

In effect, C_M contains a scan for hydrophobic modes across a range of covariance/correlation lengths from 1 to M. Because

[‖]Swiss-Prot (1996) Protein Sequence Data Bank (Med. Biochem. Department, University of Geneva, Switzerland), Release 32.5.

C_M is by definition real, symmetric, and normal, its $\{\nu_i\}_{i=1}^{M}$ are real, nonnegative, and distinct, and its associated $X_i(j)$ constitute natural bases for orthonormal projections on $H_1, H_2, \ldots, H_N$. The set of $\Psi_i(j)$ are orthonormally decomposed sequences of moving average values (33–35). We designate the leading eigenfunction representing the transmembrane segments of G-coupled receptor proteins as Ψ^T, the secondary eigenfunction containing the peptide receptor modes as Ψ^R, and the leading peptide ligand eigenfunction as Ψ^L, when the peptide is long enough to permit its construction.

Best Bases, Discrete Trigonometric Wavelet Transformation of Hydrophobic Free Energy Eigenfunctions

Wavelet transformations of the receptor and ligand hydrophobic free energy eigenfunctions generate wavelet graphs, $W(\Psi^R) \equiv W^R$ and $W(\Psi^L) \equiv W^L$. Wavelet transformation, $W(a,b)$, consists of decomposing the Ψ^R values into translated $W(n) \rightarrow W(n-b)$ and scaled $W(n) \rightarrow [W(n)/a]$ (scale is analogous to the inverse radian frequency of a trigonometric function) versions of the mother wavelet, w, a waveform with an average value of zero ($\int_{-\infty}^{\infty} w(n)dn = 0$), of finite length, arbitrary regularity and symmetry, and which is composed with data series, $H_{n=1}, H_{n=2}, \ldots, H_{n=N}$, as

$$W(a,b) = \frac{1}{\sqrt{a}} \int_0^n H(n) w\left(\frac{n-b}{a}\right) dn.$$

For w we chose the family of windowed, Gaussian (the dilate-location product bounded from below by finite Gaussian variance), discrete sine transformations globally maximized with respect to the trade-off in resolution between sequence frequency (dilation, l) and location (translation, a)

$$w_{i,k}(n) = \frac{2}{\sqrt{2l_i}} p_i(n) \sin\left[(2k+1)\frac{\pi}{2l_i}(n-a_i)\right],$$

in which $p_i(n)$ is a "partition-like" window function supported in $[a_i - l_{i/2}, a_{i+1} + (l_{i+1})/2]$ such that $\Sigma p_i^2(n) = 1$. The wavelet plane portrays discrete steps of dilation of the local trigonometric wave graphed along the ordinate. The sequence location of these windowed dilates is indicated along the abscissa.

"Best basis" refers to optimization of the Heisenberg trade-off between specifying the locations in the sequence vs. those in dilate (wavelength) space, such that the "Shannon entropy of the expansion," a Pythagorian distance in a probability Hilbert-like space between the chosen basis and the function, is minimized (36). Discrete, rather than continuous wavelet techniques, allow cutting smooth windows of differing lengths while preserving orthogonality during pattern identification in W (37). The graphs of the wavelet transformations of the receptor and ligand eigenfunctions, W^R and W^L, are represented as isopotential plots of their interpolated moduli.

Maximum Entropy, Complex Poles Power Spectral Transformation of Hydrophobic Free Energy Sequences and Their Eigenfunctions

The Ψ^R and Ψ^L, as well as the undecomposed hydrophobicity sequences, H_i, of the shorter peptide ligands, were transformed into their dominant inverse hydrophobic frequencies, ω_i,

$$h(w) = \frac{\sigma^2}{2\pi} \frac{1}{|1 + a_1 \exp(-iw_i) + \ldots + a_k \exp(-ikw_{k+1})|^2},$$

by using maximum entropy, complex poles, power spectra: $h(\omega)$. The Fourier coefficients, a_i, match a very small set of k known autocorrelations, chosen so that the entropy of the spectral estimate, $\mathcal{H}(\omega) = \int \ln h(\omega) d\omega$, is maximal. Beyond the limited information of a small set of autocorrelation-matched Fourier coefficients (38), the process is extended into a Gaussian process such that $\mathcal{H}(\omega)$ is maximized (39). In these studies $k \leq 8$ for receptor eigenfunctions of sequence lengths of several hundred amino acids, to avoid "splitting" $h(\omega)$ into spurious modes. The minimum, $k = 2$, was used for the hydrophobic free energy sequences of the short peptide ligands.

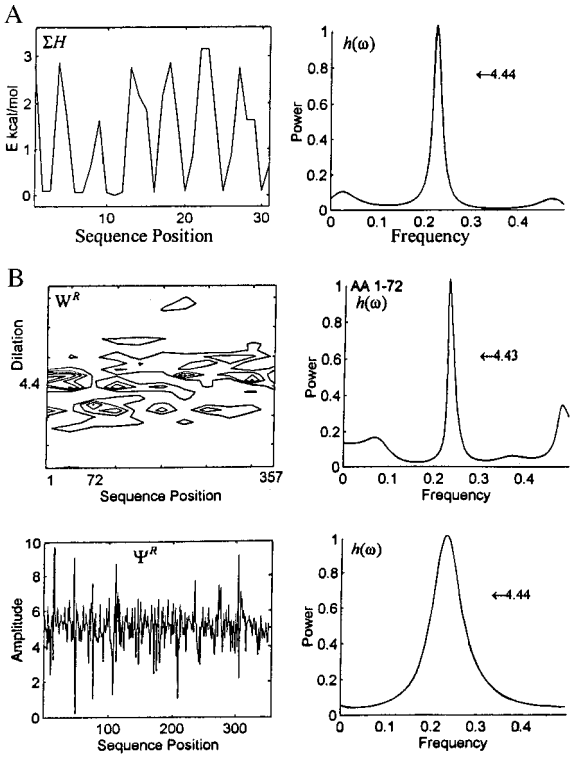

FIG. 2. (*A Left*) Graph of the H_i of β-endorphin, the definitional peptide ligand for the delta opioid receptor. (*Right*) Plot of its maximum entropy, complex poles power spectrum indicating an $h(\omega) = 4.44$ wavelength in amino acid residues. (*B Upper Left*) Graph of the wavelet transform, $W^R = 4.4$, of the Ψ^R (*Lower Left*) of the delta opioid receptor. A plot of the $h(\omega) = 4.43$ of its extracellular, N-terminal H_i (*Upper Right*) is identifiable in the isopotential plot of the W^R (*Upper Left*) and the global $h(\omega) = 4.44$ of its Ψ^R is portrayed in its graph (*Lower Right*).

$h(\omega)$ is much like an autoregressive, maximum-likelihood spectral estimate in that it is not model-dependent but is derived directly from the data (40) and behaves like a filter that yields the one or two leading complex poles of discrete hydrophobic variational frequency in undistorted form from hydrophobic free energy sequences and/or their eigenfunctions.

Locating the TMH_{1-7} in G Protein-Coupled Receptor Sequences by Using the Graphs of Their Leading Hydrophobic Free Energy Transmembrane Eigenfunctions, Ψ^T

The locations of the transmembrane segments, TMH_{1-7}, along the sequences of the G protein-coupled receptors for hormones, neurotransmitters, odorants, and light can be predicted graphically with reasonable accuracy by using hydropathy plots of the n-block averages of their amino acid sequences as hydrophobic free energies (32, 41–43). The hydropathy plots and the associated electron microscopic, electron diffraction studies, and projection maps of bacteriorhodopsin and rhodopsin I are generic for the seven-transmembrane receptor family and define four extracellular domains: the extracellular N-terminal chain, the first extracellular loop called *e1* between TMH_{2-3}, *e2* between TMH_{4-5}, and *e3* between TMH_{6-7}. The top row of Fig. 1 compares the graphs of the 32 times iterated, asymptotically smoothed, nearest neighbor averages (called an hydropathy plot) (*Left*) and the leading, transmembrane, hydrophobic free energy eigenfunction, Ψ^T (*Right*) derived from the H_i of rhodopsin I. The seven peaks characteristic of the G

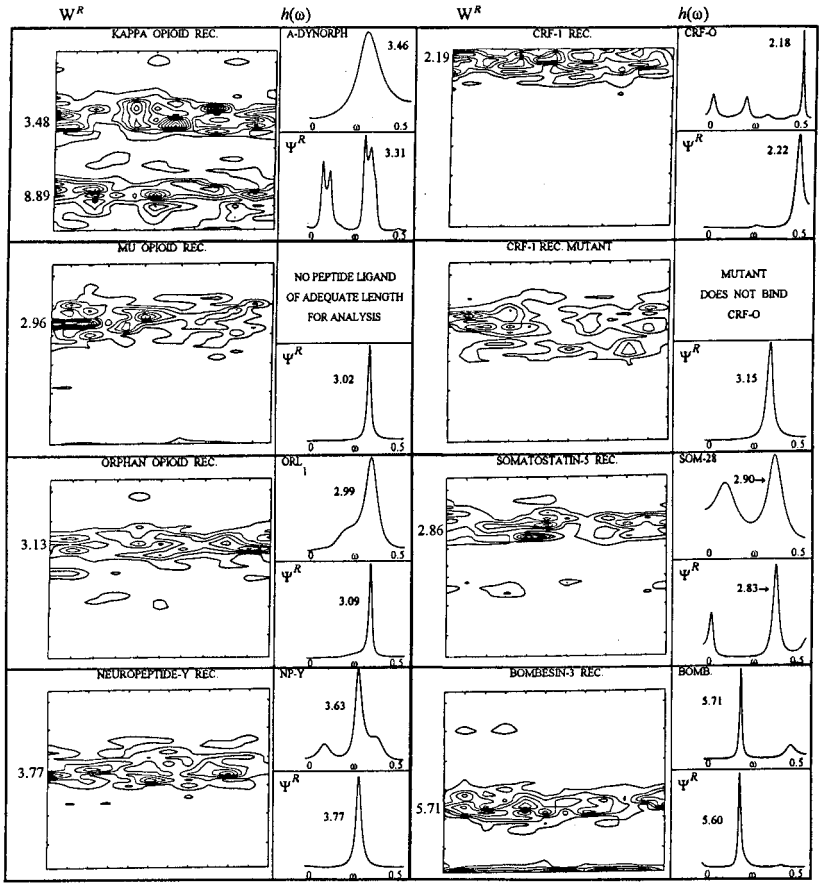

FIG. 3. Summaries of the graphs of the discrete wavelet transforms W^R of each Ψ^R (*Left* of each panel), the maximum entropy, complex poles power spectra $h(\omega)$ values of each peptide ligand (*Upper Right* of each panel), and the $h(\omega)$ of its receptor Ψ^R (*Lower Right* of each panel) for a representative set of peptide ligands and their G protein-coupled, transmembrane receptor sequences. The sequence lengths along the abscissa of the W^R plots (receptor length in amino acid residues minus the lag of the autocovariance matrix) are as follows: KAPPA OPIOID RECEPTOR (REC). = 362, MU OPIOID REC. = 402, ORPHAN OPIOID REC. = 352, NEUROPEPTIDE-Y REC. = 368; CRF-1 REC. = 307, CRF-1 REC. MUTANT = 307, SOMATOSTATIN-5 REC. = 373, and BOMBESIN-3 REC. = 486. See text for details of each peptide-receptor match and the relevant experiments in chimeric exchanges and point mutations. Note particularly the top two panels on the right that associate the loss of the peptide ligand-matched dominant hydrophobic free energy receptor mode in a nonbinding mutant receptor for ovine corticotropin releasing factor.

protein-coupled receptor family are easily recognizable. Structural analogies with rhodopsin proteins are made by comparing the results of site mutagenic studies, chimeric exchanges, and fluorescent emission spectroscopy of membrane loop binding domains (23–25).

The second to fifth rows of Fig. 1 demonstrate the close similarity between the paired graphs of hydropathy and the Ψ^T values for the kappa and mu opioid, corticotropin releasing factor, and cholecystokinin B seven transmembrane (TMH_{1-7}) receptors. In obtaining information about transmembrane structures, the hydropathy technique eliminates other mode structures of the receptor sequence, whereas the extraction of Ψ^T leaves Ψ^R intact, allowing further analyses for peptide ligand binding mode(s).

Characterizing and Localizing the Dominant Hydrophobic Modes of Peptide Ligand H_i Sequences and Their G Protein-Coupled Receptor Eigenfunctions, Ψ^R

Fig. 2*A Left* is a graph of the hydrophobic free energy sequence of H_i values of the delta opioid peptide ligand, β-endorphin. Fig. 2*A Right* shows its maximum entropy, complex poles power spectrum, in which the dominant hydrophobic free energy inverse frequency, $h(\omega) = 4.44$ amino acids. The shortness of the β-endorphin sequence, H_i, $i = 32$, broadened the band of its conventional power spectral transformation and included a larger wavelength peak, but the selection of the leading poles $k = 2$ in the maximum entropy method minimizes these effects.

Fig. 2*B Upper Left* portrays the W^R of the second leading, β-endorphin-activating, delta opiate receptor eigenfunction, Ψ^R (*Lower Left*). The $h(\omega) = 4.43$ residues of the extracellular N-terminal chain, H_i, $i = 1$–72 (*Upper Right*) is clearly located in the isopotential plot of the W^R (*Upper Left*). The more opaque isopotential densities in the modular graph of W^R signify the expression of the dominant dilate (wavelength) across larger variations in hydrophobic free energies. For example, the broad modular density on the left end of W^R representing the N-terminal extracellular portion of the delta opiate receptor's Ψ^R results from the dominant statistical wavelength expressed in both relatively low average hydrophobic values, M, R, T, L, N, T, S, A, M, D, G, T, and G = 1.67, 0.85, 0.07, 2.17, 0.09, 0.07, 0.07, 0.87, 1.67, 0.66, 0.1, 0.07, and 0.1, respectively, and relatively higher average hydrophobic values, I, L, T, A, C, F, L, S, L, L, I, and L = 3.15, 2.17, 0.07, 0.87, 1.52, 2.87, 2.17, 0.07, 2.17, 2.17, 3.15, and 2.17, respectively. To the right of the N-terminal, extracellular segment, W^R also portrays modular peaks in sequence positions roughly corresponding to the cytoplasmic loops, *e1*, *e2*, and *e3* of Ψ^R. The distributed character of this hydrophobic free energy mode is consistent with the finding that $h(\omega) = 4.44$ (Fig. 2*B Right Lower*) when computed on the entire sequence of Ψ^R (Fig. 2*B Left Lower*). The hydrophobic mode matches between β-endorphin and the N-terminal domain, and the extracellular loops of the delta opiate receptor are consistent with results of site-directed mutagenesis studies performed on opiate receptors, which suggest a "discriminatory role" for these receptor extracellular loop-peptide ligand interactions (44). In addition, chimeric exchange experiments involving the delta opiate receptor have suggested that *e2* and *e3* (and a portion of TMH_{5-7}) are required for opioid peptide ligand (45, 46) but not necessarily for small molecule binding (47, 48).

Fig. 3 summarizes comparisons of the dominant hydrophobic free energy modes of several representative peptide ligand-receptor pairs. Each panel contains three subsections: (*Left* in each) a graph of the discrete wavelet transformation, W^R, of each peptide receptor's Ψ^R; (*Upper Right* in each) a graph of the maximum entropy, complex poles power spectral transformation, $h(\omega)$, of the peptide ligand sequences as H_i' values; and (*Lower Right* in each) is the $h(\omega)$ of the peptide receptor's Ψ^R. The dominant hydrophobic mode in the amino acid sequence of dynorphin A, $h(\omega) = 3.46$ residues, is relatively close to the global $h(\omega) = 3.31$ of the Ψ^R of the kappa opioid receptor. The peptide

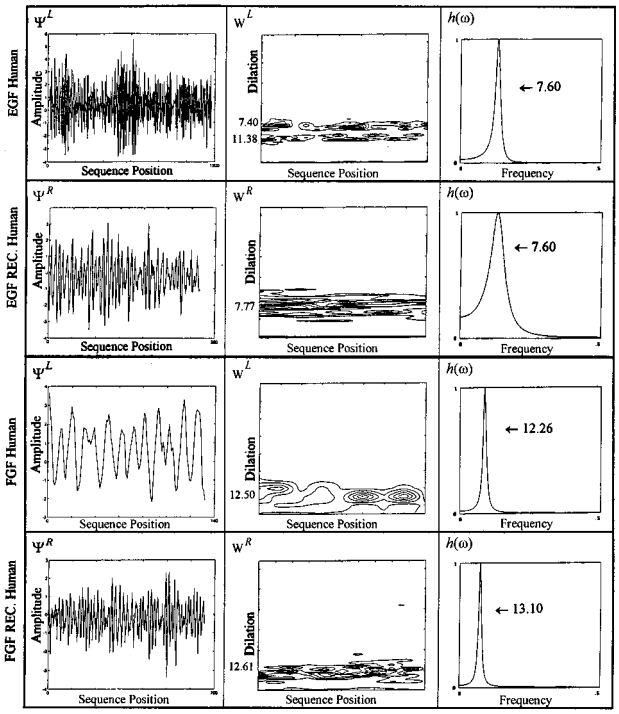

FIG. 4. (*Left*) Plots of the leading ligand and receptor eigenfunctions, Ψ^L, Ψ^R, of EGF and FGF. (*Center*) Corresponding graphs of their wavelet transformations, W^L, W^R. (*Right*) Maximum entropy, complex poles power spectra, $h(\omega)$ values of their Ψ^L and Ψ^R. The sequence positions along the abscissas of the W^L and W^R plots are $n - 18$ of the indicated amino acid sequence lengths where n of the EGF propeptide = 1,188; extracellular EGF REC. = 355; FGF = 154; extracellular FGF REC. = 700.

ligand does not manifest the $h(\omega) = 8.8$ of the Ψ^R, which is also seen as a set of distributed not necessarily receptor-related wavelets modes of $W \cong 8.88$. The $W^R \cong 3.48$, however, is shared with dynorphin A and is particularly well localized in the e_2 region between TMH_{4-5}, which chimeric exchange studies have shown to be required for kappa receptor binding of the dynorphins (45, 46).

The dominant hydrophobic free energy wavelength, $h(\omega) = 3.02$, of the Ψ^R of the mu opioid receptor evidences high absolute-valued wavelet coefficients at $W^R \cong 2.96$ in the N-terminal region and more diffusely distributed in e_2 and e_3. The results of these studies have varied from those suggesting the importance of segments in the N-terminal region (49, 50), to those pointing to e_3 (47). The generic pentapeptide agonists for the mu opioid receptor, the enkephalins, are too short, A_i, i = 1–5, to justify an analysis by even the high resolution, maximum entropy, complex poles power spectra, as are the recently discovered, very high affinity mu receptor ligands (the endomorphins) with A_i, i = 1–4 (51).

The Ψ^R of the orphan opioid receptor, OLR_1 (52) (which does not manifest high affinity, equilibrium binding of the delta, kappa, or mu peptide opioid ligands, though it has a high sequence homology with the opioid receptor family) and its recently sequenced endogenous ligand (54), share a $h(\omega) = 2.99–3.09$ dominant mode. Its $W^R \cong 3.13$ manifests a well localized density at the C-terminal end of the Ψ^R of the receptor sequence; but of greater potential physiological significance is the graphically diffuse mode, roughly corresponding to e_2, that OLR_1 shares with the more densely marked e_2 region in kappa opioid receptor's W^R, the extracellular loop required for binding with the dynorphins (54).

The wavelet graph of the Ψ^R of the human neuropeptide Y's cloned Y_1 receptor demonstrates several modular dilate densities along the sequence of $W^R \cong 3.774$ residue wavelength, which correspond to the $h(\omega) = 3.63$ of neuropeptide Y's sequence of H_i's and the $h(\omega) = 3.77$ of the peptide receptor's Ψ^R. Site mutagenic studies have suggested that extracellular loops e_1, e_2, and e_3 of the Y1 receptor are essential ligand binding domains (55) and that alanine substitution for the amino acids in the region of the latter significantly reduced binding of neuropeptide Y (56).

The top two right panels of Fig. 3 consist of graphic comparisons of the mode matches among the $h(\omega) = 2.18$ of the hydrophobic sequence of the ovine corticotropin releasing factor, the $h(\omega) = 2.22$ dominant mode of the Ψ^R, and the $W^R \cong 2.19$ of the normal corticotropin releasing factor 1 receptor. The $h(\omega) = 3.15$ residue wavelength of the Ψ^R of the nonbinding, mutant corticotropin releasing factor 1 receptor demonstrates a mismatch with the normal ligand (57). Poor localization in the putative e_2 and e_3 mode densities can be seen in the W^R of the mutant's Ψ^R in comparison with that of the normal receptor.

Site mutagenesis and chimeric exchanges involving the receptors of somatostatin 14 and somatostatin 28 (SOM-28) have suggested that their third (58, 59) and sixth (60) transmembrane segments as well as extracellular loops e_2 and e_3 (61) are important in peptide ligand binding and functionality. The $h(\omega) = 2.90$ of the H_i of the peptide ligand corresponds to the $h(\omega) = 2.83$ wavelength of Ψ^R of the somatostatin receptor 5, the receptor with the greatest affinity for SOM-28. Although some regions of the wavelet moduli are relatively indistinct across the sequence, there is a well defined peak at $W^R \cong 2.86$, in the general vicinity of TM_3 and e_2. The receptor's "hydrophobic pocket," posited to play an important role in somatostatin binding, has been located in e_3 (62), which is more diffusely localized in the graph of W^R.

The W^R graph of the Ψ^R of human bombesin-gastrin releasing peptide receptor sequence suggests multiple distributed isoenergetic modular peaks of average wavelength (dilate) $W^R \cong 5.714$, which corresponds to the $h(\omega) = 5.71$ residues of the bombesin peptide sequence, as well as the $h(\omega) = 5.60$ of the receptor's eigenfunction, Ψ^R. Chimeric exchanges involving e_3 or the carboxyl end of the bombesin receptor (note the variegated densities on the right side of its W^R graph) and the corresponding portion of the M_3 muscarinic cholinergic receptor eliminated ligand-receptor internalization and recycling normally initiated by the bombesin peptide ligand (63, 64). The modular densities in the W^R to the left of e_3 may correspond to the second and third intracellular loops composed of hydrophilic, charged (low hydrophobic valued) amino acid residues, which chimeric studies have shown to be required for receptor coupling to phospholipase C-dependent intracellular processes (65).

Characterizing and Localizing the Dominant Hydrophobic Modes of Peptide Ligand H_i Sequences and Their Tyrosine Kinase-Coupled Receptor Eigenfunctions, Ψ^R

The leading eigenfunction, not shown, of the 1,188 amino acid precursor of epidermal growth factor (EGF, urogastrone) in H_i was dominated by an average $h(\omega) \approx 53–54$ residues, consistent with its known composition of 53 amino acid EGF repeats. Fig. 4 (top row, *Left*) is a graph of the second (ligand) eigenfunction of the EGF precursor, Ψ^L, for which $h(\omega) = 7.60$ (*Right*, top row). As is generally the case with peptide precursors, this mode also dominated the maximum entropy, complex poles transformation of the H_i's of the 53 residue EGF peptide ligand itself. The wavelet transformation of Ψ^L of the EGF precursor (*Center*, top row) manifests two bands of dilates, $W^L \cong 7.40$ and $W^L \cong 11.38$, hidden in the nonconvergent tail to the left side of the peak of the EGF propeptide Ψ^L values $h(\omega) = 7.60$. The Ψ^R of the extracellular segment, i = 1–335 residues, of the EGF receptor sequence (Fig. 4 *Left*, second row) manifests a broad band $h(\omega) = 7.60$ (*Right*, second row) and a diffusely localized, average $W^R \cong 7.77$ residues (*Center*, second row). The dominant wavelet mode of transforming growth factor α (TGF-α), which binds to the EGF receptor, is $W^L \cong 7.44$ (data not shown), and demonstrates a well-marked density in the sequence location of about A_i, i = 83–120. EGF/TGF chimeric exchanges have shown this region to

be required for the high affinity binding of TGF-α to the EGF receptor (66).

Fig. 4 *Left* (third row) portrays the leading eigenfunction of the H_i, i = 1–154 sequence of acid fibroblast growth factor, aFGF, its leading hydrophobic mode, $h(\omega)$ = 12.26 residues (*Right*, third row) and its discrete, trigonometric wavelet transformation (*Center*, third row) with $W^L \cong 12.50$ residues for the N-terminal densities. Studies have suggested that the wavelet mode density at the N-terminal end of aFGF is essential for cell nuclear localization (67) and, when given intracerebrally, the induction of consummatory behavior in rodents (68). The density at the N-terminal end of the W^R of the aFGF receptor has also been found to be essential for the discriminative binding of basic FGF and aFGF (69). The Ψ^R of the predominantly extracellular domain of the aFGF receptor (*Left*, fourth row) manifests a globally dominant $h(\omega)$ = 13.10 (*Right*, fourth row) and a $W^R \cong 12.61$ residue band, on average (*Center*, fourth row). The multiplicity and distributed character of the isopotential densities in W^R is consistent with chimerical studies that indicate multiple Ig-like binding regions distributed in the central portions of the extracellular aFGF receptor sequence (70). In a pilot study of the eigenfunctions of representative lambda, kappa, and gamma chains of the Ig superfamily, we have found Ψ^R values with $h(\omega)$ = 12.5–13.7.

Technical Comments and Limitations

The eigenfunction construction technique used here (34, 35) is derived from the singular value decomposition of autocovariance matrices, which has been successfully applied to many discrete, relatively short sample length biological data series including neuronal interspike intervals, digitized electromyographic, and brain wave potentials (71). Consistency is achieved across results, as in our ligand-receptor families of mode matches, rather than through the use of physiologically unrealistic sample lengths. These approaches to physiological data often involve the partition and representation of apparently continuous quantitative data by a small discrete alphabet composing a finite set of data values (72, 73), quite like our use of the finite set of discrete values for amino acid hydrophobic free energies used in these analyses. It should also be noted that with respect to both the ligands and receptors, we have found that hydrophobic free energy receptor eigenfunction modes from physiologically similar families vary around only 10 or so commonly found wavelengths, for example, $h(\omega)$ of $\Psi^R \approx 2.1, 3.0, 3.6, 4.0, 4.4, 5.6, 7.5, 8, 10$, and 13.0 amino acid residues (16, 17).

In addition, we offer the caveat that we do not expect similarities in hydrophobic mode distributions among ligands and receptors to necessarily order relative binding affinities, discriminate between agonist and antagonist influences, or predict subsets of the range of biochemical and physiological actions differentially evoked by members of a family of peptides.

1. Rose, G. D. (1978) *Nature (London)* **272,** 586–590.
2. Lesk, A. M. & Chothia, C. (1980) *J. Mol. Biol.* **136,** 225–270.
3. Anfinsen, C. B. (1973) *Science* **181,** 223–230.
4. Chothia, C. (1984) *Annu. Rev. Biochem.* **53,** 537–572.
5. Nozaki, Y. & Tanford, C. (1971) *J. Biol. Chem.* **246,** 2211–2217.
6. Manavalen, P. & Ponnuswamy, P. K. (1978) *Nature (London)* **275,** 673–674.
7. Israelachvili, J. & Wennerstrom, H. (1996) *Nature (London)* **379,** 219–225.
8. Pashley, R. M., McGuiggan, P. M., Ninham, B. W. & Evans, D. F. (1985) *Science* **229,** 1088–1098.
9. Kaiser, E. T. & Kèzdy, F. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1137–1143.
10. Kaiser, E. T. & Kèzdy, F. J. (1984) *Science* **223,** 249–255.
11. Kamtekar, S. & Hecht, M. H. (1995) *FASEB J.* **9,** 1013–1017.
12. Gronenbron, A. M. & Clore, G. M. (1994) *Science* **263,** 536–538.
13. Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1942–1946.
14. Lumry, R. (1995) *Methods Enzymol.* **259,** 1–103.
15. Beattie, J., Shand, J. H. & Flint, D. J. (1996) *Eur. J. Biochem.* **239,** 479–485.
16. Mandell, A. J. (1984) *Annu. Rev. Pharmacol. Toxicol.* **24,** 237–274.
17. Mandell, A. J., Russo, P. V. & Blomgren, B. (1987) *Ann. N.Y. Acad. Sci.* **504,** 88–117.
18. Ramachandran, G. N. & Sassiekharan, V. (1968) *Adv. Protein Chem.* **28,** 283–437.
19. Gabor, D. (1946) *J. Inst. Electron. Eng.* **93,** 429–457.
20. Zetin, F. N., Rusk, S. F., Raymond, V., Mandell, A. J. (1988) *Endocrinology* **122,** 1121–1128.
21. Mandell, A. J. (1988) in *Dynamic Patterns in Complex Systems*, eds, Kelso, J. A. S., Mandell, A. J. & Shlesinger, M. F. (World Scientific, Singapore), pp. 219–235.
22. Araga, S., Leboef, R. D. & Blalock, J. E. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8747–8751.
23. O'Dowd, B. F., Lefkowitz, R. J. & Caron, M. G. (1989) *Annu. Rev. Neurosci.* **12,** 67–83.
24. Ostrowski, J., Kjelsberg, M. A., Caron, M. G. & Lefkowitz, R. J. (1992) *Annu. Rev. Pharmacol. Toxicol.* **32,** 167–183.
25. Wagner, R. L., Apriletti, J. W., McGrant, M. E., West, B. L., Baxter, J. D. & Gletterick, R. J. (1995) *Nature (London)* **378,** 690–697.
26. de Gennes, P. G. (1985) *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, NY).
27. Di Marzio, E. A. & Guttman, C. M. (1985) *J. Chem. Phys.* **95,** 1189–1197.
28. White, S. H. & Jacobs, R. E. (1990) *Biophys. J.* **57,** 911–921.
30. Degli Esposti, M., Crimi, M. & Venturoli, G. (1990) *Eur. J. Biochem.* **190,** 207–219.
30. Zamyatnin, A. A. (1972) *Prog. Biophys. Mol. Biol.* **24,** 107–123.
31. Reynolds, J. A., Gilbert, D. V., Tanford, C. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 2925–2927.
32. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
33. Golub, G. H. & Van Loan, C. F. (1993) *Matrix Computations* (Johns Hopkins Univ. Press, Baltimore), pp. 409–474.
34. Broomhead, D. S. & King, G. P. (1986) *Physica D* **20,** 217–236.
35. Broomhead, D. S., Jones, R. & King, G. P (1987) *J. Phys. A* **20,** L563–L569.
36. Wickerhauser, MV. (1994) *Adapted Wavelet Analysis from Theory to Software* (Peters, Wellesley, MA), pp. 103–150.
37. Daubechies, I., Jaffard, S. & Jean-Lin, J. (1991) *SIAM J. Math. Anal.* **22,** 554–573.
38. Madan, R. N. (1993) in *Maximum Entropy and Bayesian Methods*, eds. Mohammad-Djafari, A. & Demoments, G. (Kluwer, Dordrecht, The Netherlands), pp. 49–54.
39. Burg, J. P. (1972) *Geophysics* **37,** 375–376.
40. Priestly, M. D. (1981) *Spectral Analysis and Time Series* (Academic, San Diego), p. 600.
41. Henderson, R., Baldwin, J. M. & Ceska, T. A. (1990) *J. Mol. Biol.* **213,** 899–929.
42. Strader, C. D., Fong, T. M., Tota, M. R., Underwood, D. & Dixon, R. A. (1994) *Annu. Rev. Biochem.* **63,** 101–132.
43. Dohlman, H. G., Thoner, J., Caron, M. G. & Lefkowitz, R. J. (1991) *Annu. Rev. Biochem.* **60,** 653–667.
44. Metzger, T. G. & Ferguson, D. M. (1995) *FEBS Lett.* **375,** 1–4.
45. Meng, F., Hoversten, M., Thompson, R. C., Taylor, L., Watson, S. J. & Akil, H. (1995) *J. Biol. Chem.* **270,** 12730–12736.
46. Wang, W. W., Shahrestanifar, M., Jin, J. & Howells, R. D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 12436–12440.
47. Xue, J. C., Chen, C.-G., Zhu, J. M., Kunapuli, S., DeRiel, J. D., Lei, Y. & Liu-chen, L.-Y. (1994) *J. Biol. Chem.* **269,** 30195–30199.
48. Gether, U., Johansen, T. E., Snider, R. M., Lowe, J. A., Nakanishi, S. & Schwartz, T. W. (1993) *Nature (London)* **362,** 345–348.
49. Fukuda, K., Terasako, K., Kato, S. & Mori, K. (1995) *FEBS Lett.* **373,** 177–181.
50. Watson, B., Meng, F. & Akil, H. (1996) *Neurobiol. Dis.* **3,** 87–96.
51. Zadina, J. E., Hackler, L., Ge, L.-J. & Kastin, A. J. (1997) *Nature (London)* **386,** 499–502.
52. Chen, Y., Fau, Y. & Liu, J. (1994) *FEBS Lett.* **347,** 279–283.
53. Meunier, J. C., Mollereau, C., Toll, L & Suaudeau, C. (1995) *Nature (London)* **377,** 532–535.
54. Wang, J. B., Johnson, P. S., Wu, J. M., Wang, W. F. & Uhl, G. R. (1994) *J. Biol. Chem.* **269,** 25966–25969.
55. Walker, P., Munoz, M., Martinez, R. & Peitsch, M. C. (1994) *J. Biol. Chem.* **269,** 2863–2869.
56. Sautel, M., Rudolf, K., Wittneben, H. & Herzog, H. (1996) *Mol. Pharmacol.* **50,** 285–292.
57. Ross, P. C., Kostas, C. M. & Ramabhadran, T. V. (1994) *Biochem. Biophys. Res. Commun.* **205,** 1836–1842.
58. Strnad, J. & Hadcock, J. R. (1995) *Biochem. Biophys. Res. Commun.* **216,** 913–921.
59. Nehring, R. B., Meyerhoff, W. & Richter, D. (1995) *DNA Cell Biol.* **14,** 939–944.
60. Ozenberger, B. A. & Hadcock, J. R. (1995) *Mol. Pharmacol.* **47,** 82–87.
61. Liapakis, G., Fitzpatrick, D., Hoeger, C. & Rivier, J. (1996) *J. Biol. Chem.* **271,** 20331–20339.
62. Bell, G. I. & Reisine, T. (1993) *Trends Neurosci.* **16,** 34–38.
63. Tseng, M. J., Coon, S., Stuenkel, E., Struk, V. & Logsdon, C. D. (1995) *J. Biol. Chem.* **270,** 17884–17891.
64. Tseng, M. J., Detjen, K., Struk, V. & Logsdon, C. D. (1995) *J. Biol. Chem.* **270,** 18858–18864.
65. Benya, R. V., Akeson, M., Mrozinski, J., Jensen, R. T. & Battey, J. F. (1994) *Mol. Pharmacol.* **46,** 495–501.
66. Kramer, R. H., Lengerink, A. E., van Bueren-Koornneef, I. L & van der Meer, A. (1994) *J. Biol. Chem.* **269,** 8708–8711.
67. Fantal, W. J., Johnson, D. E. & Williams, L. T. (1993) *Annu. Rev. Biochem.* **62,** 453–481.
68. Lin, Y. Z., Yao, S. Y. & Hawiger, J. (1996) *J. Biol. Chem.* **271,** 5305–5308.
69. Reich-Slotkey, R., Shaoul, E., Berman, B., Graziani, G. & Ron, D. (1995) *J. Biol. Chem.* **270,** 29813–29818.
70. Zimmer, Y., Givol, D. & Yayon, A. (1995) *J. Biol. Chem.* **268,** 7899–7903.
71. Hamer, R. N. (1990) *Brain Topography* **3,** 43–47.
72. Mandell, A. J. & Selz, K. A. (1997) *Chaos* **7,** 67–81.
73. Walters, P. (1991) *Symbolic Dynamics and Its Applications* (Am. Math. Soc., Providence, RI)