SCALING IN BIOLOGICAL NUCLEAR MAGNETIC RESONANCE SPECTRAL DISTRIBUTIONS

SERGE LACELLE
Ames Laboratory, Department of Energy, and Department of Chemistry, Iowa State University, Ames, Iowa 50011

ABSTRACT A statistical analysis of the distribution of the eigenvalues of the chemical shift interaction as detected by nuclear magnetic resonance (NMR) spectroscopy in large biological systems is presented in the light of random matrix theory. A power law dependence is experimentally observed for the distribution of the number of eigenvalues, N, of the shielding hamiltonian with $\epsilon_i \leq E$ as a function of the energy E. From this cumulative distribution of energy levels, $N(E)$, we also obtain a density of states $\rho(E)$. The exponent of the energy variation of $N(E)$ and $\rho(E)$ are correlated with the dimensionality of the molecular system. A crossover in the values of the exponents is found in passing from low to higher energy in the spectra. Our method classifies and reduces the chemical shift data base of proteins and also demonstrates a degree of regularity in seemingly irregular spectral patterns.

INTRODUCTION

The visual complexity of an NMR chemical shift stick spectrum of a protein, constructed from the data of Keller et al. (1), is displayed in Fig. 1. The purpose of the approach presented here and elsewhere (2) is to extract from such one-dimensional spectral patterns qualitative information which physically characterizes the system. With the advent of technological and technical advances, it becomes possible to separate and study the different interactions contributing to an NMR spectrum. In this respect, we note that state of the art two-dimensional NMR techniques applied to proteins with known primary structures permit the assignment of up to 250 resonance lines for ^{1}H chemical shift in small proteins (mol wt $\sim$ 6,000) (1). Clearly, these upper limits should increase with progress. The questions then become, how can we handle this huge data base, what does it tell us about these systems, and how can we compare data from one protein to another? Our answer to these questions has been to analyze the collection of chemical shifts from a molecule rather than to treat shifts of individual atoms. In doing so, we have essentially separated the problem of understanding the chemical shift of complicated systems into a nonstatistical part (i.e., the shift of individual atoms) and a statistical part (i.e., the ensemble of shifts of a molecule).

From a microscopic point of view a description of interactions in many-body systems such as magnetic shieldings in a protein can be difficult to attain because of the large number of particles involved. The microscopic approach is replaced by statistical methods to describe the collective behavior of a representative ensemble of systems in statistical mechanics, and by a phenomenological approach in thermodynamics. In the event of discrete energy levels in a many-body system, assignment of quantum numbers to individual levels, based on various models, becomes prohibitive as the number of particles increases. In random matrix theory (3, 4), it is assumed that the statistical behavior of energy levels of a complicated system with either an unknown or unmanageable hamiltonian is represented by the eigenvalues obtained from an ensemble average of random matrices. This method permits a qualitative description of the general appearance and the degree of regularity/irregularity of the level structure observed experimentally in a spectrum. Statistical properties such as spacings of characteristic values of the magnetic shielding hamiltonian can be studied successfully with NMR spectroscopy (2, 5). The extension of this approach to the density or cumulative distribution of eigenvalues of this interaction in large biological systems is the subject of this paper.

Nearest-neighbor spacing distribution functions representing uncorrelated, correlated, and random superposition of correlated energy levels have been observed for the chemical shift interactions in the spectra of a vitamin, an antibiotic, and a protein (2). Such functions represent global properties of an energy level ensemble which serve to characterize the molecule, in addition to the usual characterization of the magnetic shielding of individual levels from a local point of view. The statistical analysis of the cumulative eigenvalue distribution will be correlated with the dimensionality of the molecular geometry. This analysis will also serve to distinguish systems with the same energy spacing distribution functions.

Dr. Lacelle's present address is Département de chimie, Faculté des sciences, Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1

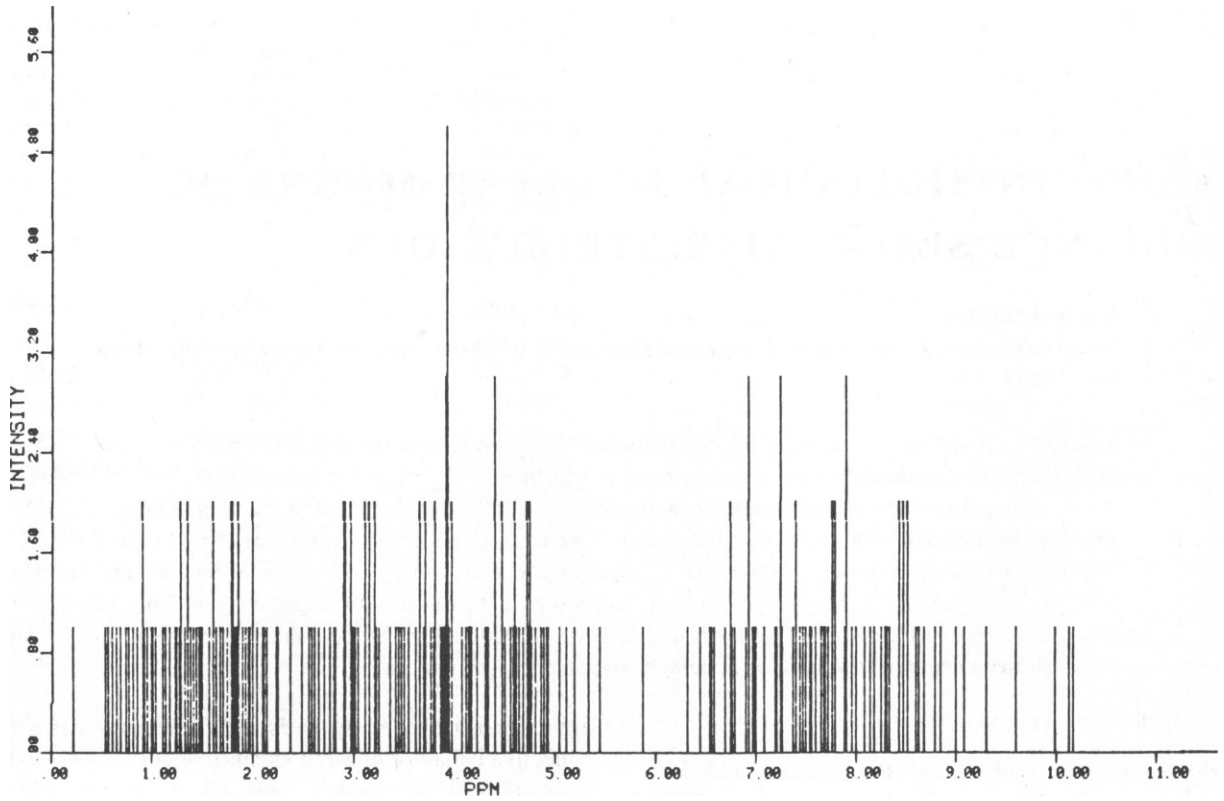

FIGURE 1 ^{1}H NMR chemical shift stick spectrum of BTI, constructed from the data of two-dimensional NMR experiments (1). Energy in ppm increases from *left* (high field) to *right* (low field).

In the next section, we discuss the chemical shift, the cumulative distribution, and the density of a sequence of energy levels in an energy interval in the context of random matrix theory. An empirical scaling relationship is proposed to treat the cumulative distribution of discrete chemical shift obtained from spectral patterns of large biomolecules. This method is then applied to data from current literature in the Results and Discussion section. Limitations, a possible implication for continuous spectra, and a passing mention to randomness in a sequence of eigenvalues are presented in the conclusions.

BACKGROUND

In the most general case, the chemical shift of a site in a molecule is a property represented by a second rank cartesian tensor. Classically speaking, the physical origin of this interaction in a system with no macroscopic electron orbital or spin angular momentum is due to a"small" magnetic field at the nuclei induced by the motions of electrons in a "large" magnetic field (6). A microscopic approach to the understanding of the chemical shift tensor gives two contributions, a diamagnetic and a paramagnetic part

$$\sigma_{\alpha\beta} = \sigma^{d}_{\alpha\beta} + \sigma^{P}_{\alpha\beta}, \tag{1}$$

where

$$\sigma^{d}_{\alpha\beta} = \frac{e^2}{2mc^2}\left\langle 0 \left| \sum_i \frac{r_i^2\delta_{\alpha\beta} - r_{i\alpha}r_{i\beta}}{r_i^3} \right| 0 \right\rangle \tag{2}$$

$$\sigma^{P}_{\alpha\beta} = \frac{e^2}{2mc^2}\sum_k (E_0 - E_k)^{-1} \cdot \left[\left\langle 0 \left| \sum_i L_{i\alpha} \right| k \right\rangle \left\langle k \left| \sum_i \frac{L_{i\beta}}{r_i^3} \right| 0 \right\rangle + \left\langle 0 \left| \sum_i \frac{L_{i\alpha}}{r_i^3} \right| k \right\rangle \left\langle k \left| \sum_i L_{i\beta} \right| 0 \right\rangle\right]. \tag{3}$$

α, β represent the components of the tensor, and r_i is the distance from the nucleus of interest to the ith electron. The α component of the orbital angular momentum operator of the ith electron is $L_{i\alpha}$. E_k is the electronic (vibrational) energy of the state $|k>$ with $|0>$ representing the ground state. The other symbols are standard constants. In solution, only the rotationally invariant part (scalar) of the tensor survives rapid random molecular motions. The observed magnetic shielding σ_{iso} in the laboratory (or rotating) frame is related to the components σ_{jj} of the shielding tensor in its principal axis system by

$$\sigma_{iso} = \tfrac{1}{3}[\sigma_{11} + \sigma_{22} + \sigma_{33}]. \tag{4}$$

From the sums in Eqs. 2 and 3, one notes that a particular nucleus is coupled in principle to all the electrons in a molecule. Numerous approximation schemes have been devised to simplify these calculations to account for effects such as neighbor anisotropy, ring currents, electric fields, and solvents (7). It is obvious from the nature of the physical problem that this complicated many-body interaction cannot be"solved" explicitly for cases of interest in biological NMR. Hence, the methods of random matrix theory (the statistical theory of energy levels) provide a means of studying the properties of an ensemble of complicated levels rather than attempting to understand individual levels.

To describe the regularity/irregularity in a spectrum such as level repulsion and clustering, one resorts to statistical properties like the rth nearest neighbor spacing distribution functions, and density of levels (or mean spacings). The local density $\rho(E)$ for a discrete spectrum (delta

functions) is given by

$$\rho(E) = \frac{\Delta N}{\Delta E}, \tag{5}$$

with ΔN levels in an energy interval ΔE. A related quantity is the cumulative distribution $N(E)$ of energy levels as a function of the energy

$$N(E) = \int_{-\infty}^{E} \rho(y)\mathrm{d}y. \tag{6}$$

In other words, how many levels have energy $\epsilon_i < E$? In mathematics an analogous problem, the prime number theorem, asks how many primes $N(n)$ can be found less than the integer n (4).

The slope of the staircase function $N(E)$ gives the level density from which the average spacing D is obtained, $D = \rho^{-1}$. The secular energy variation in the level density ρ (the curvature) is removed by local averaging procedures which map ρ to a constant (3). This unfolding step is technically difficult since one wants to preserve the fluctuation measures or statistics about the smoothed or uniform spectrum. The translational invariance of the spectral averaged density over the unfolded spectrum is a stationary property of stochastic processes relevant to the question of ergodicity, i.e., the spectral average being equal to the matrix ensemble average. In a previous paper (2, Fig. 2 therein) a constant level density was assumed though a secular energy variation in ρ was observed, leading to nonstationarity of ρ. Instead of unfolding the spectrum, a locally averaged density in the different parts of the spectrum were used for the nearest neighbor spacing distributions. Similarly here, the difficulties associated with the nonstationarity of ρ and its unfolding are circumvented by considering the scaling behavior of the cumulative distribution $N(E)$.

We assume empirical scaling relationships for $N(E)$ and $\rho(E)$ of the form

$$N(E) \sim E^{X} \tag{7}$$

$$\rho(E) \sim E^{S}. \tag{8}$$

Because Eqs. 7 and 8 are not exact (note the absence of equality), these relationships imply statistical scaling. From Eqs. 6, 7, and 8, we obtain

$$X = S + 1. \tag{9}$$

Therefore, a logarithmic transformation of Eq. 7 gives

$$\log N = X \log E \tag{10}$$

or when plotting $N(E)$ on doubly logarithmic plots

$$\log\left(\frac{N_2}{N_1}\right) = X \log\left(\frac{E_2}{E_1}\right), \tag{11}$$

where the variables N_2 and E_2 measured in one part of the cumulative distribution are referenced to another part N_1 and E_1 of the distribution. From Eqs. 8, 9, and 11, we can determine the exponent S of the density of levels. The exponents of the energy variation of the density of states is an interesting quantity since it is model dependent. For example (8), elastic vibrations and deformations of solids have the following density of states

$$\rho(E) \sim E^{d-1}, \tag{12}$$

whereas for a free particle subject to periodic or boundary conditions (e.g., electron gas in a metal, particle in a box)

$$\rho(E) \sim E^{d/2-1}, \tag{13}$$

where d is the Euclidean and/or fractal dimension characterizing the system. In statistical mechanics, the Laplace transform of $\rho(E)$ with respect to E yields the partition function, while the log $\rho(E)$ is proportional to the entropy of the system (9). For the exponents of the empirical relationships 7 and 8, no a priori model will "explain" their dependence in the next section. However, a classification of levels according to distinct exponents in the different parts of the spectrum will (*a*) correlate with the dimensionality of the molecular geometry, (*b*) reduce the molecules chemical shift data base, and (*c*) demonstrate a degree of regularity in a seemingly irregular spectrum.

RESULTS

We examine experimental NMR chemical shift data available from current literature. The investigated systems comprise two small proteins, Bovine trypsin inhibitor, BTI (1); Bull seminal inhibitor, BSI (11); a vitamin, cyanocobalamine — B_{12} (12); and an antibiotic, Alamethicin (13). For these molecules, with the exception of BSI, the primary, secondary, and tertiary structures are well characterized. We attempt to interpret the slopes obtained from the doubly logarithmic plots, log N vs. log E, in terms of the dimensionality of the tertiary structures.

These plots are presented in Fig. 2 for BTI, in Fig. 3 for BSI, in Fig. 4 for B_{12}, and in Fig. 5 for Alamethicin. For each molecule, instead of fitting all the available chemical shift data, a"coarse-grained" distribution of N vs. E was used so that the analysis could be easily carried out on a pocket calculator. The number of slopes was kept to a minimum by visually examining the fit, and by determining the standard deviations and correlation coefficients of the slopes. We must keep in mind that we are interested in the scaling of the chemical shift over a distribution that should reflect an intermediate scale rather than a local (few levels) or global (all levels) scale property of the sequence of levels. Table I summarizes the slopes, standard

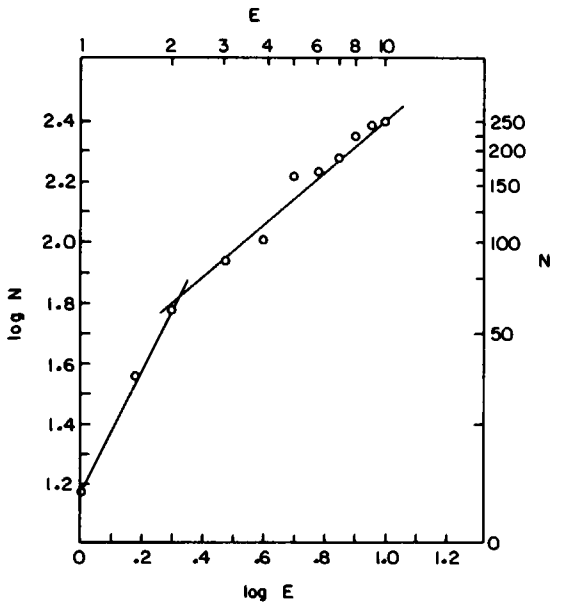

FIGURE 2 Distribution of the log N vs. log E for ^{1}H NMR chemical shift of BTI (1). N represents the number of energy levels (or transitions) with energy $\leq E$.

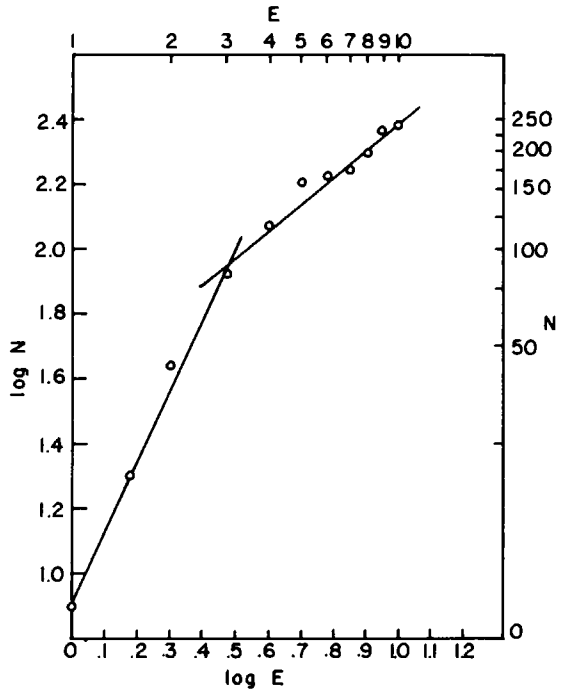

FIGURE 3 Distribution of the log N vs. log E for ^{1}H NMR chemical shifts of BSI (13).

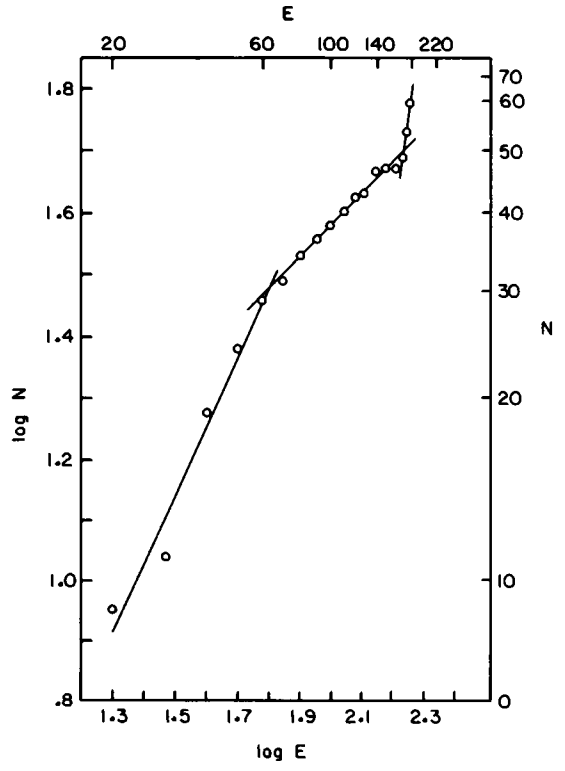

FIGURE 4 Distribution of the log N vs. log E for ^{13}C NMR chemical shifts of B_{12} (14).

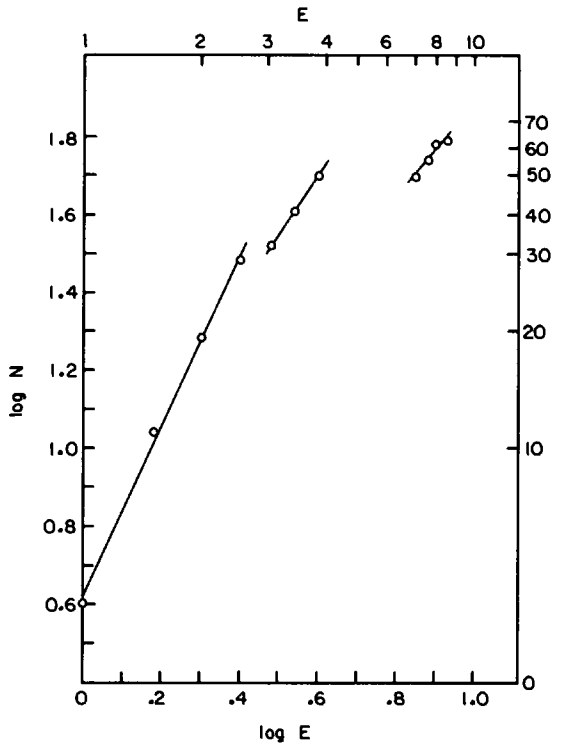

FIGURE 5 Distribution of the log N vs. log E for ^{1}H NMR chemical shifts of Alamethicin (15).

deviations, and correlation coefficients calculated from a linear least squares fit, the number of levels or transitions contributing to a particular slope, and the exponent, S, characterizing the density of states dependence on the energy. When the number of levels is small (~10), we find error limits of up to 25%.

TABLE I
DATA OBTAINED FROM DOUBLY LOGARITHMIC PLOTS

System	Slope* $S+1$	Corr†	S‡	Levels	Number of levels
Cyano-	1.1 ± 0.1	0.979	0.10	0–29	29
cobalamine	0.51 ± 0.02	0.996	−0.49	29–49	20
B12	3.4 ± 0.9	0.982	2.4	49–60	11
Alamethicin	2.2 ± 0.1	0.998	1.2	0–30	30
	1.5 ± 0.0	1.00	0.5	30–50	20
	1.2 ± 0.3	0.960	0.2	50–62	12
Bovine	2.0 ± 0.2	0.990	1.0	0–60	60
trypsin	0.90 ± 0.05	0.918	−0.10	60–251	191
inhibitor BTI					
Bull seminal	2.2 ± 0.2	0.995	1.2	0–84	84
inhibitor	0.83 ± 0.10	0.977	−0.17	84–239	155
BSI at 18°C					
Bull seminal	2.1 ± 0.2	0.993	1.1	0–79	79
inhibitor	0.84 ± 0.09	0.978	−0.16	79–239	160
BSI at 45°C					

*Error limit on slope is one standard deviation.
†Correlation coefficient from linear least-squares fit.
‡Error limit on s is the same as that of the slope.

DISCUSSION

For both proteins, we may group the distribution of levels according to their energy dependence into two groups. In the case of B_{12} and Alamethicin, three groups of levels are necessary to account for the distribution. For $N(E)$, all systems display roughly an E^2 dependence at low energies (or high field), except B_{12}, which has an E^1 dependence. At higher energies (or low field) we find a crossover to $E^{.9}$ for BTI and BSI, whereas the distributions for B_{12} and Alamethicin cross over twice to different power laws. The exponent of the density of states' energy dependence, $\rho(E)$, is obtained by subtracting one from the value of the slope of the doubly logarithmic plot (see Eq. 9 and Table I).

Crossover is a general feature observed for phonon dispersion in the percolation network model (13). For these extended states, at lower energies, one finds a normal or classical behavior for $N(E)$ and $\rho(E)$ (see Eq. 12), i.e., d is an integer. At higher energies, a crossover occurs in $\rho(E)$ to a fractal dimension $\bar{d}$ which characterizes nontopological aspects of the form of these systems. The shape of some hemoproteins at liquid helium temperatures has been shown (14) with ESR relaxation measurements that are proportional to the density of vibrational states to occupy a space of fractal (and spectral) dimensionality $1.19 \leq \bar{d} \leq 2.24$. A striking similarity is observed and discussed below in the crossover of the chemical shift spectral distributions.

If we assume (see Eqs. 7–9) that

$$x = d - 1 \tag{14}$$

and therefore

$$s = d - 2, \tag{15}$$

then these relationships can be used to classify the initial slope in $N(E)$ according to the dimensionality d of the molecular geometry as presented below. At low energy (see Table I), $N(E)$ of BTI, BSI, and Alamethicin all follow roughly E^2, whereas the distribution of the atoms contributing to that part of the spectrum are distributed over the whole molecule, i.e., $d = 3$; therefore $x = 2$ if the above assumptions are correct. In the case of B_{12}, the initial distribution of ^{13}C chemical shifts follows an $E^{1.1}$. Of the 30 atoms that contribute to this part of the spectrum, 26 can be found in the plane like environment defined by the corrin ring, i.e., $d = 2$ and $x = 1$. At higher energies, the distribution $N(E)$ of chemical shifts crosses over to some fractional power of E for BSI and BTI, $x \sim 0.8$ to 0.9, which would imply $d = 1.8$–1.9. Two such crossovers are observed for B_{12} and Alamethicin, which give $d \sim 1.2$–1.5. One of the crossover for B_{12} gives $d = 4.4 \pm 0.9$, but with a high uncertainty in view of the small number of levels.

At this point we caution the reader that we have not proposed a microscopic interpretation of the scaling $N(E)$ of the chemical shift, nor have we rationalized the scaling in terms of the percolation cluster model. Indeed, our observtions bear a striking resemblance to some of the predictions of this model. It is evident that further investigations are needed to elucidate the statistical scaling that we propose on empirical grounds to classify the energy levels.

CONCLUSIONS

The distribution of chemical shifts in large biological systems has been shown to follow a power law relationship in relative energies. For the investigated systems, we observed for the distribution of the log N vs. log E a behavior quite analogous to the predictions of a model for vibrational states on a percolation cluster. Indeed, the distribution of chemical shifts at low energies seemed to reflect the dimension of the structure in which the atoms were embedded. Crossover to different power law behavior with fractional energy exponent is reminiscent of fractal behavior of mode localization in solids. To understand the empirical observation for the exponent of $N(E)$, $x = d - 1$, further studies are needed before a claim to "universality" can be envisioned.

Our analysis of a collection of chemical shifts in large biological systems is presently limited to spectra with discrete resonances. For continuous spectra, the common slope $x = d - 1$ points to an intriguing possibility. If the energy E_c at which the crossover occurs is found to be a constant, while the total energy range E_T is approximately a constant, say $E_T \approx 12$ ppm for ^1H chemical shift in liquids, then Eq. 11 becomes a constant, i.e.,

$$\log\left(\frac{N_2}{N_1}\right) = X \log\left(\frac{E_T}{E_c}\right) = \text{constant}, \tag{16}$$

where N_2 is the total number of levels, while N_1 is the number of levels at which the crossover occurs. Therefore, the ratio of the area of the spectrum below E_c to the total area would be a constant. As two-dimensional NMR techniques evolve, the possibility to unravel the spectra of larger molecules will make the approach presented here and elsewhere (2) more attractive and statistically more reliable. The coarse grained distribution of N vs. E smoothed the data. This step kept the analysis simple enough that calculations could be done on a pocket calculator. At the same time, we bypassed the unfolding procedures mentioned in the background section. Instead of smoothing the data, a staircase distribution involving all the available data could be used with appropriate computational facilities.

The analysis of the distribution of discrete chemical shifts in large biological systems bears strong resemblance to the problem of randomness in other disciplines. Accordingly, we can benefit by briefly mentioning some general models that are available for the discussion of randomness in a sequence of eigenvalues. Ordered sets obtained from discrete spectra can be modeled as eigenvalues of random matrix ensemble, points on a line (disconnected sets, e.g.,

Cantor dust [15, 16]), and noninvertible one dimensional maps (17). The degree of randomness can be computed via information theoretical approach with Kolmogorov entropy (18), and algorithmic complexity (19). We mention these models since in favorable cases, the correlation of the chemical shift energy levels with global molecular structures has been successful with other approaches such as graph and group theoretical methods, and numerical studies (20–22).

The support and encouragement from Professor B. C. Gerstein, the help from Dr. C. G. Fry and Dr. L. A. Saucke, and the typing of the manuscript by G. Lessard are gratefully acknowledged.

Operated for the U. S. Department of Energy by Iowa State Unviersity under contract no. W-7405-Eng-82. This research was supported by the Office of Basic Energy Sciences, Chemical Sciences Division.

Received for publication 21 June 1985 and in final form 26 September 1985.

REFERENCES

1. Keller, R. M., R. Bauman, E. H. Hunziker-Kurk, F. J. Joubert, and K. Wüthrich. 1983. Assignment of the ^{1}H NMR spectrum of the trypsin inhibitor homologue K. *J. Mol. Biol.* 163:623–649.
2. Lacelle, S. 1984. Random matrix theory in biological nuclear magnetic resonance spectroscopy. *Biophys. J.* 46:181–186.
3. Brody, T. A., J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. M. Wong. 19 . Random matrix physics: spectrum and strength fluctuations. *Rev. Mod. Phys.* 53:385–479.
4. Porter, C. E., editor. 1965. Statistical Theories of Spectra: Fluctuations. Academic Press Inc., New York.
5. Schaeffer, J., and P. Yaris. 1969. Random matrix theory and nuclear magnetic resonance spectral distributions. *J. Chem. Phys.* 51:4469–4474.
6. Ramsey, N. F. 1950. Magnetic shielding of nuclei in molecules. *Phys. Rev.* 78:699–703.
7. Ando, L., and G. A. Webb, 1983. Theory of NMR Parameters. Academic Press Inc., New York. 47–82.
8. Kittel, C. 1976. Introduction to Solid State Physics. 5th edition. John Wiley & Sons, New York. 131–136.
9. Kubo, R. 1965. Statistical Mechanics. Elsevier/North Holland, New York. 8–11, 102.
10. Strop, P., G. Wilder, and K. Wüthrich. 1983. ^{1}H spectra of bull seminal inhibitor. *J. Mol. Biol.* 166:641–667.
11. Anton, D. L., H. P. C. Hagenkamp, T. E. Walker, and N. Maturyoff. 1982. C-13 NMR studies of cyanocobalamine and several of its analogues. *Biochemistry.* 21:2372–2378.
12. Banerjee, V., F. P. Tsui, T. N. Balasubramanicon, G. R. Marshall, and S. I. Chan. 1983. Structure of alamethicin on solution. *J. Mol. Biol.* 165:757–775.
13. Alexander, S., and R. Orbach. 1982. Density of states on fractals: "fractons." *J. de Physique Lettres.* 43:625–631.
14. Colvin, J. T., and H. J. Stapleton. 1985. Fractal and spectral dimensions of biopolymer chains. *J. Chem. Phys.* 82:4699.
15. Mandelbrot, B. B. 1983. The Fractal Geometry of Nature. Freeman, New York. 74–83.
16. Cederbaum, L. S., E. Haller, and P. Pfeifer. 1985. Fractal dimension function for energy levels. *Phys. Rev.* A31:1869.
17. Collet, P., and J. P. Eckman. 1980. Iterated Maps on the Interval as Dynamical Systems. Birkhauser, Boston. 20.
18. Kosloff, R., and S. A. Rice. 1981. The influence of quantization on the onset of chaos in Hamiltonian systems: the Kolmogorov entropy interpretation. *J. Chem. Phys.* 74:1340.
19. Casati, G., I. Guarneri, and F. Valz-Gris. 1984. Degree of randomness of a sequence of eigenvalues. *Phys. Rev.* A30:1586–1588.
20. Randic, M. 1984. Average carbon-13 chemical shifts in benzenoid dications. *J. Magn. Res.* 59:34–40.
21. Jones, R. G. 1969. The use of symmetry in nuclear magnetic resonance. *In* NMR Basic Principles and Progress. Vol. 1. P. Diehl, E. Fluck, and R. Kosfeld, editors. Springer-Verlag, New York. 97–174 and references therein.
22. Hoch, J. C., C. M. Dobson, and M. Karplus. 1982. Fluctuations and averaging of proton chemical shifts in bovine pancreatic trypsin inhibitor. *Biochemistry.* 21:1118–1125.