RANDOM MATRIX THEORY IN BIOLOGICAL NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

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

ABSTRACT The statistical theory of energy levels or random matrix theory is presented in the context of the analysis of chemical shifts of nuclear magnetic resonance (NMR) spectra of large biological systems. Distribution functions for the spacing between nearest-neighbor energy levels are discussed for uncorrelated, correlated, and random superposition of correlated energy levels. Application of this approach to the NMR spectra of a vitamin, an antibiotic, and a protein demonstrates the state of correlation of an ensemble of energy levels that characterizes each system. The detection of coherent and dissipative structures in proteins becomes feasible with this statistical spectroscopic technique.

INTRODUCTION

Over the last few years, the increased quantity and greater quality of nuclear magnetic resonance (NMR) data obtained on biological systems, ranging from small metabolites to large macromolecular structures, reflect advances in technique and technology. Most of these studies were done in solution, whereas the NMR biological experiments performed in the solid state have had limited success up to now. Most solution NMR investigations deal with time-independent interactions such as magnetic shielding (chemical shift) and the indirect spin-spin coupling (J- or scalar coupling). Although time-dependent interactions monitored through relaxation measurements provide dynamic information, they will not be treated explicitly in this work. Here we apply random matrix theory (RMT) (1–3), or the statistical theory of energy levels, to the interpretation of chemical shift data in large biological systems. This theory has been successful in the understanding of spectroscopic data of excited nuclei, atoms, and certain molecular systems (1, 4, 5) but has never been applied to biological systems. Our purpose is to introduce to the biological community the assumptions, physical content, and results of RMT and to discuss the potential information that can be extracted from such analysis, as well as the inherent limitations of this approach. We examine these points in the Theory and Method section. In Results and Discussion we present an analysis and discussion of NMR data obtained from current literature with examples that include a vitamin, an antibiotic, and a protein. In the Conclusions and Summary, we suggest the application of this technique to some aspects of the problems of irreversibility in biological systems.

THEORY AND METHOD

The motivation for a statistical analysis of the energy level distribution is threefold. First, in large systems where the present methods of NMR are insufficient for a complete assignment of individual resonances, the energy level spacing distribution can characterize the system. Second, the method is general so that the interactions determining the energies or even the nature of the system need not be known. Finally, even if the assignment of the chemical shift is complete, this signature only describes the average local environment of each nuclear site. This microscopic approach neglects the presence of any correlation between sites. By examining an ensemble of local sites (or energy levels) we obtain a global picture of this many-body system which may indicate the occurrence of mutual or reciprocal relations between its constituents.

We are concerned with the statistical distribution of spacing between nearest-neighbor energy levels in NMR spectra of biomolecules. Let us recall a few basic ideas of probability and statistics. When dealing with an ensemble of values A_i of a property A in a system, one can characterize this sample space with the statistical ensemble average $\langle A \rangle$ (or the first moment), and the second moment describing the fluctuations of A_i about $\langle A \rangle$. The probability of finding a value in the interval $A_i + \mathrm{d}A_i$ in the ensemble is given by $P(A_i)\mathrm{d}A_i$, where $P(A_i)$ or P is a probability density function satisfying $\Sigma_i P(A_i)\mathrm{d}A_i = 1$. Furthermore, we can express the average of a discrete variable, $\langle A \rangle = \Sigma_i A_i P(A_i)$ with the probability P, which can easily be generalized through integration for the case of a continuous variable. The probability density function is loosely called a distribution function in statistical mechanics and characterizes the ensemble as a whole. The constraints employed in defining a representative ensemble are chosen to correspond to the maximal knowledge that we have of the system of interest. RMT permits the interpretation of the functional form of P in terms of qualitative correlation of the investigated property A. In statistics (6), correlation is a measure of how several variables vary together. Linear covariability of A and B is described by the covariance σ_{AB} given by

$$\sigma_{AB} = \langle AB \rangle - \langle A \rangle \langle B \rangle. \quad (1)$$

A correlation coefficient proportional to the covariance and standard deviation of the random variables can also be defined, but such a quantitative approach is of no concern here. Rather, correlation will imply the familiar and intuitive concept of relation between energy levels; the environment of a local site i is also determined by factors influencing another (or many) site(s) j. Cooperativity effects in allosteric enzymes and the concerted charge relay mechanism of action of chymotrypsin are biochemical examples of correlation or coherent couplings.

Theoretical distribution functions are available from the results of RMT (2) (Fig. 1). The Poisson distribution

$$P(x) = 1/D \exp(-x/D) \quad (2)$$

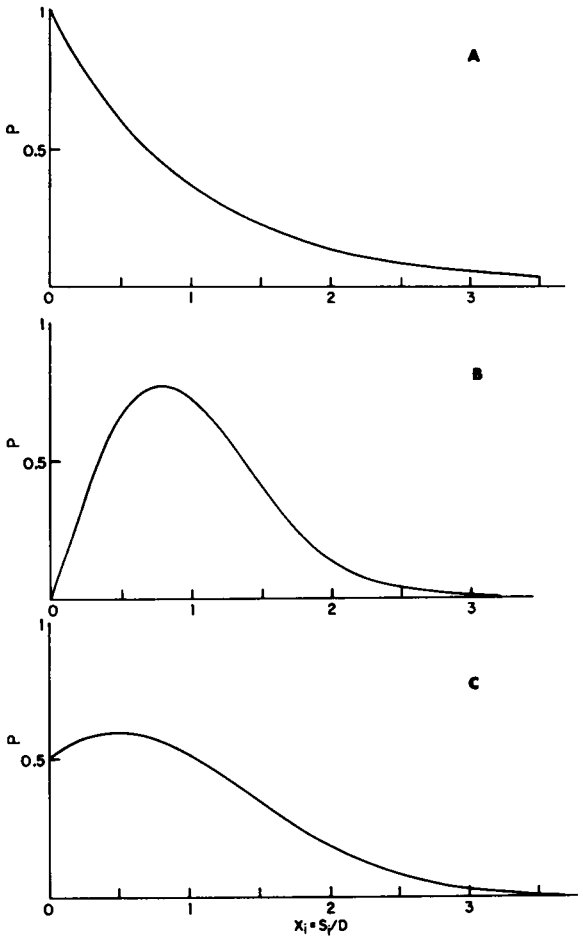

FIGURE 1 Probability density functions, P, for the nearest-neighbor energy level relative spacing x_i (A) Poisson distribution, (B) Wigner distribution, and (C) random superposition of two Wigner distributions.

corresponds to uncorrelated values of the variable x, the energy level spacing (see below), with normalization constant D, while in the correlated case, the Wigner distribution (7) applies and is given by

$$P(x) = \frac{\pi x}{2D^2} \exp\left(-\frac{\pi x^2}{4D^2}\right). \tag{3}$$

An intermediate condition consists of the random superposition of a few Wigner distributions. This decomposition can indicate a certain number, n, of correlated states in a system. A large superposition of Wigner functions, $n \rightarrow \infty$, leads to the Poisson distribution. In the case of a random superposition of two Wigner distributions, an analytic expression has been derived (8) for P given below

$$P(x) = \frac{1}{2D} \exp\left(-\frac{\pi x^2}{8D^2}\right) + \left(\frac{\pi x}{8D^2}\right) \exp\left(-\frac{\pi x^2}{16D^2}\right) \phi\left(\frac{\sqrt{\pi} x}{4D}\right), \tag{4}$$

where ϕ is the complementary error function,

$$\phi(y) = 1 - \frac{2}{\sqrt{\pi}} \int_0^y \exp(-t^2)\, dt. \tag{5}$$

Empirical distributions (9) such as the Porter-Thomas

$$P(x) = 4x/D \exp(-2x/D) \tag{6}$$

and the normal (Gaussian) distribution

$$P(x) = 1/\sigma(2\pi)^{1/2} \exp[-\tfrac{1}{2}(x - \mu/\sigma)^2], \tag{7}$$

where μ and σ^2 are the first moment and variance, respectively, have been used in the statistical analysis of energy levels. Because of their empirical origin, the interpretation and/or information content of such distributions have been discussed with respect to specific individual cases.

In spectroscopy, time-independent interactions lead to the experimental determination of characteristic values of the energy of the system, or eigenvalues E_i. The eigenfunctions $|i>$ represent the stationary states of the system. The nature of this kind of problem is embodied in the Schrödinger time-independent equation

$$H|i> = E_i|i>, \tag{8}$$

where H is the Hamiltonian operator corresponding to the energy. In the event where the system possesses a finite number of stationary states, as in NMR, Schrödinger's equation can be "solved" with matrix algebra. The Hamiltonian is represented by a matrix with elements H_{ij}, where the diagonal elements ($i = j$) are related to the eigenvalues, and the off-diagonal elements ($i \neq j$) carry information concerning the transition probability between stationary states i and j. Physical symmetry of the system permits a so-called block diagonalization transformation of the Hamiltonian which yields irreducible blocks of elements along the diagonal. Instead of working with the original $n \times n$ matrix, we can study the eigenvalue distribution function of each block of dimension smaller than n separately, since the elements of different blocks do not mix.

In the statistical theory of energy levels, the elements of the Hamiltonian are random variables distributed according to certain assumptions. We suppose the off-diagonal elements are distributed symmetrically about zero mean while the distribution of diagonal elements is taken to be symmetrical, but not necessarily about zero mean. The plausibility of these assumptions and others (2) outside the scope of this paper have been favorably tested with Monte Carlo computer calculations in the derivation of the spacing distribution functions (10, 11).

We wish to treat the chemical shift interaction Hamiltonian with RMT. For this interaction the most general Hamiltonian (12) is given by $H = -\gamma\hbar\mathbf{I} \cdot \overleftrightarrow{\sigma} \cdot \mathbf{B}$, where γ is the magnetogyric ratio; $\hbar$, Planck's constant $h/2\pi$; $\mathbf{I}$, the spin angular momentum operator; and $\mathbf{B}$, the magnetic field. The chemical shift $\overleftrightarrow{\sigma}$ is represented by a second-rank cartesian tensor with nine components σ_{ij}. These can be resolved with symmetry arguments into an isotropic component $\sigma_{iso} = \frac{1}{3}\Sigma_i \sigma_{ii}$, an antisymmetric tensor $\sigma_{ij} = \frac{1}{2}(\sigma_{ij} - \sigma_{ji})$, and a traceless symmetric tensor $\bar{\sigma}_{ij} = \frac{1}{2}(\sigma_{ij} + \sigma_{ji}) - \sigma_{iso}$, with three and five independent components, respectively. In solution, only the isotropic value survives random molecular motions as detected in the NMR spectrum. Under such a circumstance, H^i becomes $-\sigma^i_{iso}\,\omega_0\hbar\mathbf{I}^i_z$, where i denotes a particular site in the system; the energy is $E_i = -\sigma^i_{iso}\,\omega_0$ in frequency units $\omega_0 = \gamma\mathbf{B}$. Theories (12) have decomposed σ_{iso} into a paramagnetic and a diamagnetic part, σ_p and σ_d, which depend upon the spatial distribution and orbital angular momentum of the electrons. The interpretation of chemical shifts usually deals with the shielding of a nucleus by surrounding electrons. Any correlation between sites is not obvious from this approach. Chemical shift correlation spectroscopy (13) through scalar coupling is a promising two-dimensional NMR technique that treats correlation of different spins. From the standard NMR spectrum we obtain qualitative correlation information by analyzing the relative spacing between nearest-neighbor energy levels.

The method consists of ordering the set of energy levels E_i of a system due only to chemical shifts. The nearest-neighbor spacing $S_i = E_{i+1} - E_i$ is obtained for all i sites, from which sample space one establishes the average spacing D. One then computes the relative spacing $x_i = S_i/D$. A histogram of the number of spacings, NS, with value x_i vs. x_i permits the comparison of the experimental data with $P(x_i)$. All the distribution functions given here are dependent on a well-behaved spacing average, D. This requirement may be examined by plotting the number of levels N

having energy less than or equal to E vs. the energy E. A linear plot (constant slope) indicates a uniform energy level density that assures a well-behaved D. In the event of several slopes, the distribution of spacings is obtained separately for the levels belonging to the appropriate energy level density and are plotted on the same histogram (10). The possibility of high-level density or small spacings and low intensity of absorption lines could result in missing certain lines. In NMR, overlapping resonances will usually contribute most to this problem. If the missing levels form a small fraction of the total number of levels, e.g., <5%, we neglect such corrections and assume the overall distribution function remains essentially invariant (10). Quantitative determination of the statistical agreement between the experimentally measured and theoretical distributions is beyond the level of this paper (11). Suffice it to say that nonlinearity must then be introduced in the distribution function, $p\alpha x^q$, which we avoid for the simple cases treated here, the Poisson distribution where $q = 0$, and the Wigner function with $q = 1$. For a Poisson distribution the probability of spacing S between an energy level, E, and the adjacent one in $S + dS$ is independent of the spacing and is given by dS/D. The Wigner distribution is linearly dependent on S and the analogous probability is $S\, dS/D^2$. The different behavior of the Poisson and Wigner distributions for small spacings has been explained (2) with group theoretical arguments involving symmetry and level repulsion.

The method can easily be extended in principle for the treatment of time-dependent interactions, $H(t)$. Normally, relaxation measurements are interpreted in terms of a microscopic model of motion. The Wiener-Khintchine theorem shows that the Fourier transform of the autocorrelation function, $\langle H(t) \cdot H(t + \tau)\rangle$ (where the brackets denote an ensemble equilibrium average) yields the spectral density function, $J(\omega)$, thereby giving the frequencies of motions responsible for relaxation. Relaxation rates are directly proportional to $J(\omega)$. An expansion of $J(\omega)$ in terms of eigenvalues and eigenfunctions of a transition operator permits a frequency analysis of relaxation data (14). In the event of a discrete spectrum, one could probe directly the correlation of the different motions in the same way as we have presented it above for static chemical shift interactions.

For large biological systems such as proteins with domains or quaternary structures, the approach has inherent limitations for both static and time-dependent interactions. Continuous distributions and overlapping of resonance lines cannot be treated with RMT in its present form; one needs a discrete and well-resolved spectrum. Similarly for the power spectrum, $J(\omega)$, continuous distributions of correlation times are often found in large polymeric structures. This problem becomes less important with two-dimensional NMR experiments.

RESULTS AND DISCUSSION

The NMR spectrum that can be analyzed by RMT may not display J-coupling since the nearest-neighbor energy level spacing distribution will become a function of chemical shift and the indirect spin-spin interactions. In the detection of rare nuclei, e.g., ^{13}C, ^{15}N, broadband proton decoupling is normally employed, eliminating J-coupling to protons. As for detection of protons, several relatively new two-dimensional NMR techniques (13) permit the separation of different interactions in a spectrum, providing solely chemical shift information.

Recently, Keller et al. (15) assigned most of the resonances of ^1H NMR of the trypsin inhibitor homologue K, a protein from snake venom with 57 amino acids. In Fig. 2 *A*, a plot of the number of energy levels N with energy less than or equal to E, vs. E, is linear, indicating a well-behaved energy level density. The results of the statistical analysis of the energy level spacing distributions are pre-

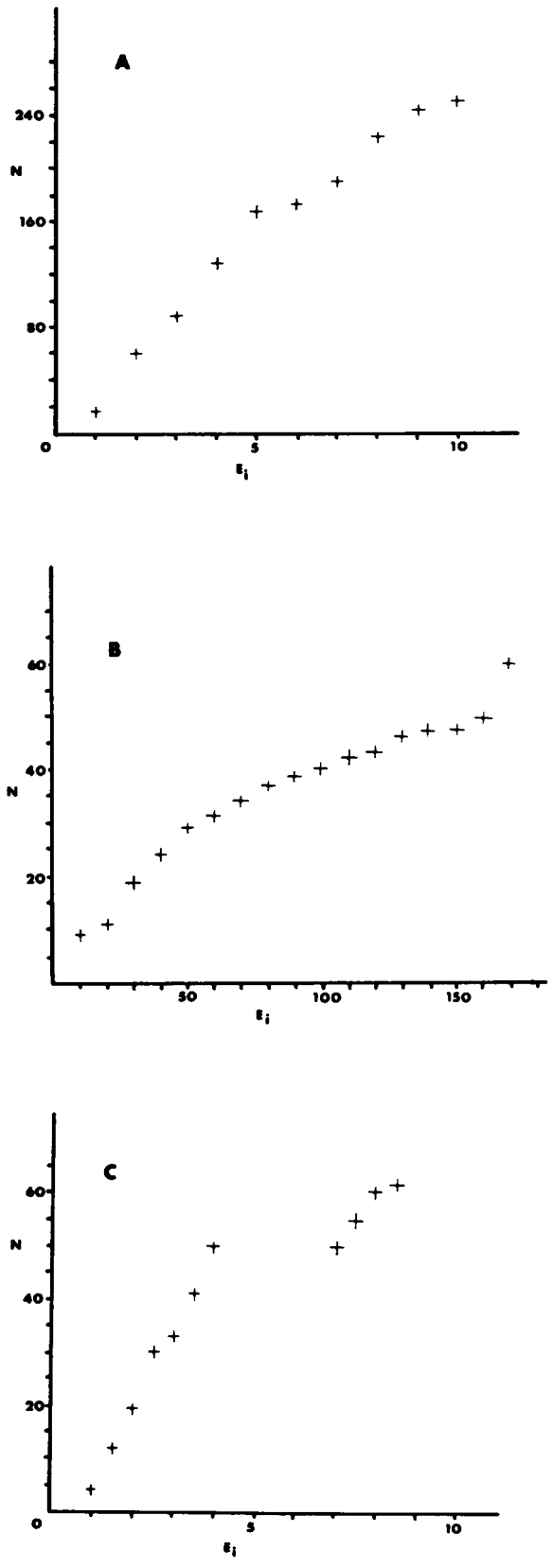

FIGURE 2 Number of energy levels, N, with energy less than or equal to E. All energies are expressed in ppm. (*A*) Trypsin inhibitor homologue K (15), (*B*) cyanocobalamin (16), (*C*) alamethicin (17).

sented in Fig. 3. Here we clearly see the presence of a Poisson distribution, indicating uncorrelated energy levels. In such a large system, one would not expect a correlation of all the energy levels. Tight coupling in the tertiary structure usually involves only certain moieties of the protein, but not necessarily all of the nuclei. Stronger correlations are to be expected in smaller regions of the molecule, such as domains or the secondary structures. Nevertheless, the Poisson distribution characterizes well the ensemble of nuclear energy levels of the trypsin inhibitor. A change in the state of correlation might be expected on binding to trypsin.

A smaller system without distinct domains or secondary structure features might reflect correlation of the tertiary structure involving most nuclei. Cyanocobalamin, or vitamin B_{12}, is well suited to test this proposal. Anton et al. (16) studied this system with ^{13}C NMR and assigned all resonances. Changes in the NMR of the carbonyl and imine carbon regions were monitored by varying the pH, and then compared to analogues. This approach of correlating resonances with several analogues helped in the interpretation of the spectrum. Our analysis of the data is presented in Figs. 2 *B* and 4 for cyanocobalamin only. We assumed the presence of a single energy level density though the curvature in the plot *N* vs. *E* might indicate the presence of two densities. Surprisingly, the Poisson distribution is in qualitative agreement with the experimental distribution. Therefore, the ^{13}C energy levels are uncorrelated on the NMR time scale. However, this does not exclude correlation of electronic, vibrational, or rotational energy levels which might be important in the functional aspect of the vitamin in vivo.

The structure of alamethicin, an antibiotic that displays interesting interactions with membranes, has been studied in solution with ^1H-two-dimensional NMR techniques by Banerjee et al. (17). Spectroscopic investigations have proposed the existence of secondary structures in the 20 amino acid peptide. The NH_2-terminal region would consist of an α-helix while the COOH-terminus would display an extended β-sheet structure. The presence of two structures implies possible correlation of energy levels within and among secondary structures. The NMR energy level density shows two distinct regions in the *N* vs. *E* plot (Fig. 2 *C*). To obtain the spacing distribution, a single average spacing *D* was used in Fig. 5 *A*. When two average spacings D_1 and D_2 were used to treat the data, and their respective relative spacings were plotted on the same histogram (Fig. 5 *B*), the results were very similar to those determined by a single average spacing *D*. A spacing with a value of 48 standard deviations from *D* (Fig. 2 *C*) as determined from the other 62 spacings was renormalized to one ($s_i = D$) in order to prevent the whole distribution from being determined solely by this single aberrant spacing. In both cases, the Poisson and Wigner distributions do not agree with the experimental distributions. The intermediate case of random superposition of two Wigner distributions agrees qualitatively with the data. Eq. 4 was developed assuming a single mean spacing *D*, though Harvey and Hughes (9) expect slight changes in the distribution in the case of two approximately similar energy level densities. From this, one might speculate that a Wigner distribution could be associated with the α-helix region, and another Wigner distribution could be associated with the β-pleated sheet. The presence of hydrogen bonds that stabilize each structure lends support to the correlation of energy levels in the ^1H NMR. From their studies, Banerjee et al. propose a conformation with the first nine amino acids involved in the α-helix, followed by an open spacer of four residues, with the last five residues forming the extended β-sheet. Here, we assumed for simplicity that the first 10 amino acids formed the α-helix and that the last 10 were involved in the extended β-sheet and analyzed the energy level spacing distribution separately, but plotted the results on the same histogram (Fig. 5 *C*). The Wigner distribution is in qualitative agreement with the data that indicates the presence of correlation of spacings. In any case, the ensemble of energy levels of alamethicin can be represented by the random superposition of two Wigner distributions (Eq. 4) that correspond to two correlated states.

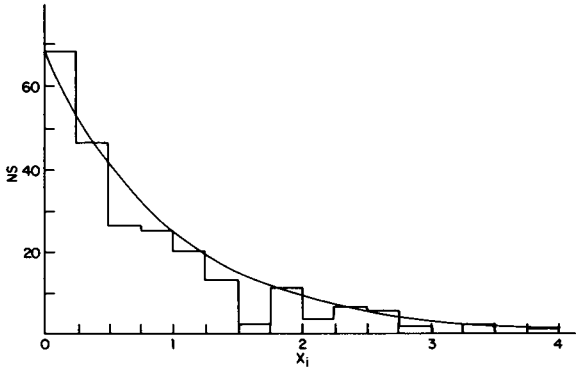

FIGURE 3 Histogram of the number of spacings, *NS*, vs. the relative spacing x_i for the trypsin inhibitor homologue K (15) superimposed with the Poisson distribution.

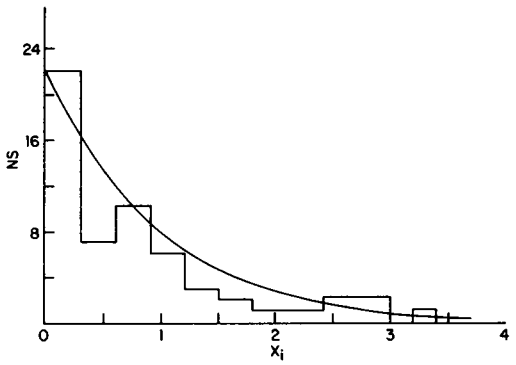

FIGURE 4 Histogram of the number of spacings, *NS*, vs. the relative spacing x_i for cyanocobalamin (16) superimposed with the Poisson distribution.

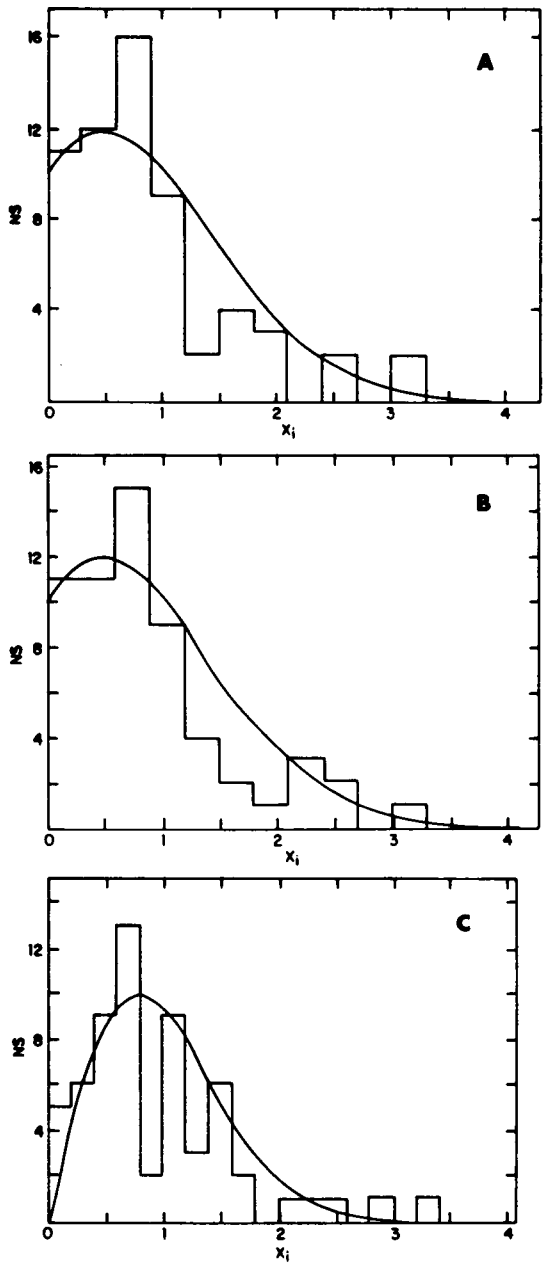

FIGURE 5 Histograms of the number of spacings, *NS*, vs. the relative spacing x_i for alamethicin (17), (*A*) assuming a single average spacing *D*, (*B*) assuming two average spacings D_1 and D_2 (both *A* and *B* are superimposed with the distribution function given by Eq. 4), (*C*) separation of α-helix from β-pleated sheet data as explained in the text, superimposed with the Wigner function. In all cases one level is found at $x_i = 7$ (not shown).

CONCLUSIONS AND SUMMARY

The characterization of an ensemble of energy levels by spacing distributions demonstrates the degree of correlation between energy levels. The ^{13}C NMR of vitamin B_{12} and ^1H NMR of a trypsin inhibitor both displayed uncorrelated energy levels while the ^1H NMR of the antibiotic alamethicin was in qualitative agreement with the random superposition of two Wigner distributions. Presumably, other types of distributions would also characterize biological systems.

In biological NMR, RMT has several advantages for describing a system, or part of it, by a spacing distribution. The existence of channels for either energy transfer or relaxation (18), or hydrogen bond network extending over long range in proteins (19), are dependent on structural or functional correlations (20) of several groups of atoms or amino acids with time-independent and -dependent properties. Correlations of this type of many-body collective modes can be conveniently "detected" with the analysis presented here.

Irreversibility in biological systems is ubiquitous. The state of equilibrium is an idealization that is never realized. Rather, stationary or steady states pervade a host of processes that lead to the presence of coherent and dissipative structures. These structures are the building blocks of the biological architecture; biological engineering and design is related to the evolution of these structures. Coherent and dissipative structures and their stability properties play an important role in modern theories of systems that are far from equilibrium (21). The correlated character of these structures and processes can be monitored by RMT analysis of an ensemble of energy levels obtained by NMR.

The support and encouragement from Professor B. C. Gerstein, discussions about random matrix theory with Dr. T. T. P. Cheung, and the help of Dr. L. A. Saucke and S. A. Standley in the preparation of the manuscript are gratefully acknowledged.

In the preliminary stages of this work, the author was supported by a Postgraduate Scholarship 1979–82 from the Natural Sciences and Engineering Research Council of Canada. Ames Laboratory-DOE is operated for the U. S. Department of Energy by Iowa State University under contract W-7405-Eng-82. This research was supported by the Office of Basic Energy Sciences, Chemical Sciences Division.

Received for publication 14 November 1983 and in final form 21 March 1984.

REFERENCES

1. Porter, C. E., editor. 1965. Statistical Theories of Spectra: Fluctuations. Academic Press, Inc., New York.
2. Mehta, M. L. 1967. Random Matrices and the Statistical Theory of Energy Levels. Academic Press, Inc., New York.
3. Carmeli, M. 1983. Statistical Theory and Random Matrices. Marcel Dekker, Inc., New York.
4. Schaefer, J. and R. Yaris. 1969. Random matrix theory and nuclear magnetic resonance spectral distributions. *J. Chem. Phys.* 51:4469–4474.
5. Haller, E., H. Koppel, and L. S. Cederbaum. 1983. On the statistical behaviour of molecular vibronic energy levels. *Chem. Phys. Lett.* 101:215–220.
6. Larson, H. J. 1974. Introduction to Probability Theory and Statistical Inference. Second ed. John Wiley & Sons, Inc., New York.
7. Wigner, E. P. 1957. Results and theory of resonance absorption. Gatlinburg Conference on Neutron Physics by Time of Flight. *Oak Ridge Natl. Lab. Rep. ORNL-2309.* 59–70.
8. Lane, A. M. 1957. Widths and spacing of nuclear resonance levels. Gatlinburg Conference on Neutron Physics by Time of Flight. *Oak Ridge Natl. Lab. Rep. ORNL-2309.* 113–123.

9. Harvey, J. A., and D. J. Hughes. 1957. Spacings of nuclear energy levels. *Phys. Rev.* 109:471–479.
10. Porter, C. E., and N. Rosenzweig. 1960. Statistical properties of atomic and nuclear spectra. *Ann. Acad. Sci. Fenn. Ser. A VI.* 44:1–60.
11. Brody, T. A., J. Flores, J. B. French, P. A. Miller, A. Pandey, and S. S. M. Wong. 1981. Random matrix physics: spectrum and strength fluctuations. *Rev. Mod. Phys.* 53:385–479.
12. Abragam, A. 1961. The Principles of Nuclear Magnetism. Clarendon Press, Oxford. 170–199.
13. Bax, A. 1982. Two-dimensional Nuclear Magnetic Resonance in Liquids. Delft University Pres, Boston. 50–98.
14. King, R., and O. Jardetzky. 1978. A general formalism for the analysis of NMR relaxation measurements on systems with multiple degrees of freedom. *Chem. Phys. Lett.* 55:15–18.
15. 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.
16. 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.
17. Banerjee, U., F. P. Tsui, T. N. Balasubramanicon, G. R. Marshall, and S. I. Chan. 1983. Structure of alamethicin in solution. *J. Mol. Biol.* 165:757–775.
18. Frölich, H. 1968. Long range coherence and energy storage in biological systems. *Int. J. Quant. Chem.* 2:641–649.
19. Metzler, D. E. 1979. Tautomerism in pyridoxal phosphate and in enzymatic catalysis. *Adv. Enzymol.* 50:1–40.
20. Wagner, G., A. Pardi, and K. Wüthrich. 1983. Hydrogen bond length and ^{1}H NMR chemical shift in proteins. *J. Am. Chem. Soc.* 105:5948–5949.
21. Prigogine, I. 1980. From Being to Becoming. Freeman Publications, San Francisco.