

BIOCHE 01407

A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis

Nobuhiro Gō

Department of Chemistry, Faculty of Science, Kyoto University, Kyoto 606, Japan

Received 30 August 1989

Accepted 16 October 1989

Conformational fluctuation; Normal mode analysis; Mean square displacement; Low-frequency normal mode; Eckart condition

An exact theorem is proved and its implication is discussed. The theorem states that, if a large molecule, typically biological macromolecules such as proteins, undergoes small-amplitude conformational fluctuations around its native conformation in such a way that within the range of conformational fluctuations at thermal equilibrium the conformational energy surface can be approximated by a multidimensional parabola, then the mass-weighted mean-square displacement of constituent atoms is given by the sum of the contributions from each normal mode of conformational vibration, which in turn is proportional to the inverse of the square of its frequency. This theorem provides a firm theoretical basis for the fact hitherto empirically recognized in the conformational dynamics of, for instance, native proteins that very-low-frequency normal modes make dominant contributions to the conformational fluctuations at thermal equilibrium. Discussion is given on the implication of this theorem, especially on the importance of the concept of the low-frequency normal modes, even in the case where the basic assumption of the harmonicity of the energy surface does not hold.

1. Introduction

Biopolymers such as protein and DNA assume specific conformations under physiological conditions. However, they undergo small-amplitude conformational fluctuations around the mean specific conformation. Such conformational dynamics is essential for their biological functions. Three types of computer simulation methods have been used as powerful tools to study the dynamics: Firstly, the method of molecular dynamics, i.e., numerical integration of the classical Newtonian equation for cartesian coordinates of constituent atoms [1]. Secondly, the normal mode analysis (NMA) of small-amplitude conformational vibrations. This analysis has been carried

out in both the cartesian coordinate space [2] and the dihedral angle space [3,4]. Finally, the Monte Carlo simulation carried out in the dihedral angle space [5]. The second method is based on the assumption that the conformational energy surface can be approximated by a multidimensional parabola within the range of thermal fluctuations (assumption of harmonicity). The first and third methods are free from this assumption. Complex but intriguing dynamic pictures of protein conformations are emerging from simulations of these methods. A usual first step in the analysis of results of simulations is to compile amplitudes of fluctuations of atomic positions in thermal equilibrium. These amplitudes can be compared with X-ray crystallographic temperature factors. Even for the assumption of harmonicity, the second method of simulation can produce atomic equilibrium fluctuations that qualitatively agree with the results from the other two methods and also

Correspondence address: N. Gō, Department of Chemistry, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan.

with those deduced from X-ray temperature factors.

The results of the normal mode analysis have indicated that, in globular proteins, very-low-frequency modes make dominant contributions to mean square deviations of atomic positions in thermal equilibrium. In such very-low-frequency modes, an entire molecule is involved in a concerted motion. These results are significant in indicating that a relatively small number of degrees of freedom corresponding to these very-low-frequency modes have a dominant importance in protein conformational dynamics. Prompted by this indication, we are now developing a new method of describing protein dynamics by the coupled motion of normal mode variables (T. Horiuchi and N. Gō, manuscript in preparation) and a new method of X-ray crystallographic refinement based on the idea of the normal mode analysis (A. Kidera and N. Gō, manuscript in preparation).

In this paper, a theorem is presented which relates the mass-weighted mean-square atomic displacement in thermal equilibrium calculated by the normal mode analysis to frequencies of the normal modes. This theorem gives a clear theoretical justification for the above-mentioned empirical observation of the dominant importance of the very-low-frequency modes. We describe the theorem in section 2 and provide the proof thereof in section 3. Consequences arising from the theorem are discussed in section 4.

2. Description of the theorem

In the normal mode analysis, a macromolecule is treated as a dynamical system which obeys the classical Newtonian law. An empirical conformational energy function is assumed as the potential function. Normal mode analysis is based on the assumptions that dynamical conformational changes of a macromolecule are confined within a well of a single minimum of the conformational energy function, and that the energy function can be approximated by a multidimensional parabola within the range of the conformational dynamics. In the situation where these assumptions are

satisfied, the conformational dynamics of the macromolecule is described by a linear combination of normal modes. Each normal mode, e.g. the j -th normal mode, is characterized by its frequency, ν_j , and by the pattern of displacements of constituent atoms from their mean positions. Now the theorem is described as follows:

$$\sum_a m_a \langle \Delta r_a^2 \rangle = kT \sum_j (2\pi\nu_j)^{-2} \quad (1)$$

where m_a and Δr_a denote the mass and displacement vector of the a -th atom in the molecule, respectively, k Boltzmann's constant and T the absolute temperature. The right-hand-side of this equation represents the mass-weighted mean-square atomic displacement in thermal equilibrium. The theorem states that this mean atomic displacement is given as the sum of contributions from each normal mode, which in turn is proportional to the inverse square of the frequency. Note that the actual pattern of displacements of atoms in each normal mode does not appear explicitly in this relation.

3. Proof of the theorem

Proofs will be given for each of the two cases where the normal mode analysis is performed in cartesian coordinate space, and dihedral angle space, respectively. Justification for this apparent duplication exists in the complexity of the problem caused by the translational and rotational degrees of freedom of the molecule as a whole. The theorem of eq. 1 holds only when such external motions of the molecule are eliminated from the atomic fluctuations. Elimination of the translational degrees of freedom is trivial. One should simply choose a coordinate system in which the center of gravity of the molecule stays fixed in space. Elimination of the rotational degrees of freedom involves a complex problem. It should be carried out as prescribed by Eckart [6] many years ago. What is complex is that the condition of Eckart cannot be satisfied strictly by solutions of the Newtonian equation of motion. In other words, we have to impose an artificial condition, which is

not compatible with the equation of motion in the strict sense, in order to eliminate the rotational degree of freedom. Because this elimination is to be done differently in the normal mode analyses carried out in cartesian coordinate space and dihedral angle space, a proof will be given in each case in order to stress this point. To describe a proof, the usual mathematical procedure of the normal mode analysis must be reviewed to some extent.

3.1. Proof in cartesian coordinate space

When normal mode analysis is performed in cartesian coordinate space, external motions are not eliminated from the atomic motions, but appear as normal modes with vanishing frequencies. This situation must be considered carefully in order to prove eq. 1.

We consider an isolated macromolecule which is undergoing small-amplitude vibrational motions around its energy minimum conformation. Let the position vector and the cartesian coordinates of the a -th constituent atom in the minimum energy conformation be \mathbf{r}_a and r_{ak} ($k = x, y, \text{ and } z$), respectively. An instantaneous conformation of a vibrating molecule is characterized by the displacement vectors of its constituent atoms. Let the displacement vector and the cartesian components of the a -th atom be \mathbf{s}_a and s_{ak} ($k = x, y, \text{ and } z$), respectively. (The displacement vector of the a -th atom is designated by the usual notation of $\Delta\mathbf{r}_a$ in eq. 1, but is henceforth referred to as \mathbf{s}_a to simplify the notation). Then, the Lagrangian of the system is given by

$$L = (1/2) \sum_{ak} m_a \dot{s}_{ak}^2 - (1/2) \sum_{ak,bl} f_{ak,bl} s_{ak} s_{bl} \quad (2)$$

and

$$f_{ak,bl} = \partial^2 E / \partial s_{ak} \partial s_{bl} \quad (3)$$

where E is the conformational energy as a function of the displacement vectors of constituent atoms. Because the conformational energy does not depend on the spatial location and orientation of the molecule as a whole, the value of E should be invariant for displacement vectors of the fol-

lowing type, which express external (i.e., purely translational and rotational, and therefore involving no internal) motions of atoms:

$$\mathbf{s}_a^e = \boldsymbol{\delta} + \boldsymbol{\omega} \times \mathbf{r}_a \quad (4)$$

Here superscript e represents 'external', and $\boldsymbol{\delta}$ and $\boldsymbol{\omega}$ are arbitrary vectors of translation and rotation, respectively. The potential energy term in the Lagrangian of eq. 2 vanishes for a displacement vector of the form of eq. 4. If the displacement vector is not given exactly by the form of eq. 4, then the potential energy should be positive, because it is an energy around a minimum point. Therefore, the matrix F , whose elements are the second derivatives of eq. 3, is non-negative. In this case, the fact that the potential energy vanishes for vectors of the form of eq. 4 means that such vectors are eigenvectors belonging to eigenvalues which are equal to zero.

By writing down the equation of motion for the Lagrangian of eq. 2, we have:

$$m_a \ddot{s}_{ak} + \sum_{bl} f_{ak,bl} s_{bl} = 0 \quad (5)$$

A general solution of this equation is given as a linear combination of the motion of normal mode variables σ_i .

$$s_{ak} = \sum_i u_{ak,i} \sigma_i \quad (6)$$

Here the coefficients satisfy the following generalized eigenvalue equation:

$$(2\pi\nu_i)^2 m_a u_{ak,i} = \sum_{bl} f_{ak,bl} u_{bl,i} \quad (7)$$

They are orthonormalized to satisfy the following relation.

$$\sum_{ak} m_a u_{ak,i} u_{ak,j} = \delta_{i,j} \quad (8)$$

where $\delta_{i,j}$ is unity, if $i = j$, and vanishes otherwise. The normal mode variables satisfy the following harmonic equation of motion:

$$\ddot{\sigma}_i + (2\pi\nu_i)^2 \sigma_i = 0 \quad (9)$$

As shown earlier, the displacement vectors of the form of eq. 4 are eigenvectors of matrix F

belonging to vanishing eigenvalues. This fact together with eq. 7 means that they are also eigenvectors of the generalized eigenvalue equation of eq. 7 with $\nu_i = 0$. Because eq. 4 involves two arbitrary three-dimensional vectors, there are six vanishing eigenvalues and corresponding eigenvectors. All other eigenvalues $(2\pi\nu_i)^2$ in eq. 7 are positive. Therefore, eq. 9 expresses a harmonic oscillator, and ν_i is its frequency. The mean-square thermal-equilibrium amplitude of a harmonic oscillator is given by

$$\langle \sigma_i^2 \rangle = kT(2\pi\nu_i)^{-2} \quad (10)$$

These harmonic oscillators describe internal motions of the molecule, as opposed to the eigenvectors of eq. 4, which belong to vanishing eigenvalues, correspond to external motions. A general solution of internal motions is therefore given by

$$s_{ak} = \sum' u_{ak,i} \sigma_i \quad (11)$$

where summation extends over normal modes with non-vanishing frequencies. The theorem to be proved can be derived easily by using eqs. 8, 10 and 11. The summation in eq. 1, therefore, should be understood as extending over normal modes with non-vanishing frequencies. Note that elimination of the external motions of atoms is carried out only after the normal mode analysis has been performed.

Because of the orthogonality of eq. 8, internal motions of eq. 11 are orthogonal to external motions in the following sense.

$$\sum m_a s_a \cdot (\delta + \omega \times r_a) = 0 \quad (12)$$

Because the two vectors δ and ω are arbitrary, this equation means that

$$\sum m_a s_a = 0 \quad (13)$$

and

$$\sum m_a r_a \times s_a = 0 \quad (14)$$

The former equation means that the center of gravity does not change its position for small-amplitude internal motions of the molecule. The latter equation was first derived by Eckart [6] as a condition that atomic displacements s_a include no contributions from overall rotation of the mole-

cule. He derived this equation by requiring that the kinetic energy of the molecule be best separated into contributions from internal and external motions.

The above two equations mean that, once the quantities on the left-hand side of these equations have vanishing values, they stay vanishing during the course of motion governed by eq. 5. In other words, they are 'integrals of motion' of eq. 5. However, eq. 5 is an approximation to the original Newtonian equation of motion derived by expanding the potential energy surface up to the quadratic terms. The quantities of eq. 13 are also integrals of the original equation of motion. However, the quantities of eq. 14 are not. This means that solutions of the original Newtonian equation of motion cannot satisfy the Eckart condition strictly. In other words, the Eckart condition is an artificial one to be imposed in order to separate rotational motions from internal motions of complex polyatomic molecules. However, once the quadratic approximation of the potential energy function has been introduced, the Eckart condition can be satisfied strictly. When one carries out the normal mode analysis by solving the generalized eigenvalue problem of eq. 7, it is advisable to check whether the eigenvectors belonging to six vanishing eigenvalues are given by the form of eqs. 13 and 14.

3.2. Atomic fluctuations in the molecular dynamics calculation

We can give another interpretation of eq. 12 (and, consequently, of eqs. 13 and 14), which would provide an insight into what is performed in the normal mode analysis. Consider two conformations of the molecule; one is the minimum energy conformation characterized by position vectors r_a of the atoms, the other being an instantaneous vibrating conformation characterized by displacement vectors s_a of the atoms. The position vectors of the latter conformation are therefore $r_a + s_a$. Now we wish to bring the second conformation into the position which fits best with the first one by properly translating and rotating the latter. Let us define the best-fit position as that which gives the minimum value of the mass-weighted mean-square displacement. When this

definition of the best-fit position is adopted, we can prove that the above two conformations are in the best-fit position with no further translation and rotation, when eqs. 13 and 14 are satisfied. The proof is as follows: When we translate and rotate the second molecule by infinitesimal amounts represented by infinitesimal translation and rotation vectors δ and ω , respectively, the position vector $r_a + s_a$ is brought to $\delta + (I + R(\omega))(r_a + s_a)$, where I is the identity matrix and $R(\omega)$ is a matrix of rotation ω . Because the displacement vectors s_a are small vectors and the translation and rotation vectors are infinitesimal, we expand the above vector in terms of these small quantities and retain terms only up to the first order to obtain $r_a + s_a + \delta + \omega \times r_a$. Therefore, the mass-weighted mean-square displacement is given by

$$\sum m_a (s_a + \delta + \omega \times r_a)^2 \quad (15)$$

By equating a term of the above expression which is linear with respect to δ and ω to zero, we obtain the condition that $\delta = 0$ and $\omega = 0$ give the best-fit position. An equation thus obtained is none other than eq. 12. Therefore, eq. 14 signifies that atomic displacements s_a include no contribution from overall rotation of the molecule in the sense that the molecule is always oriented to keep the mass-weighted mean-square displacement minimum. The above proof of the theorem indicates that this elimination of the overall rotation is automatically performed in the normal mode analysis, when the summation in eq. 1 is extended over normal modes with non-vanishing frequencies. However, if the trajectory of the molecular dynamics calculation is to be compiled to discuss amplitudes of atomic fluctuations, the above proof indicates that each instantaneous conformation must be reoriented so as to satisfy the Eckart condition of eq. 14. If the molecular dynamics calculation is carried out under no external force for the molecule and no initial translational momentum of the molecule, eq. 13 is automatically satisfied.

3.3. Proof in dihedral angle space

We now provide a proof for the case where the normal mode analysis is carried out in dihedral

angle space. In this case, the conformational dynamics of a macromolecule is described as a motion of a state point in the dihedral angle space around a point corresponding to a minimum energy conformation. Changes in the dihedral angles from those corresponding to the minimum energy conformation, θ_k , are translated into changes in atomic positions, s_a , by

$$s_a = \sum_k (\partial s_a / \partial \theta_k) \theta_k \quad (16)$$

The derivatives on the right-hand side of this equation are assumed to have constant values calculated at the minimum energy point. They can be calculated analytically so as to satisfy the Eckart conditions of eqs. 13 and 14, i.e., by requiring that the molecule changes its conformation for changes of the dihedral angles in such a way that no overall translation and rotation are incurred [7]. This means that in the normal mode analysis in dihedral angle space the external motions are eliminated at the stage of formulation.

The Lagrangian is given by

$$L = (1/2) \sum h_{kl} \dot{\theta}_k \dot{\theta}_l - (1/2) \sum f_{kl} \theta_k \theta_l \quad (17)$$

where

$$h_{kl} = \sum_a m_a (\partial s_a / \partial \theta_k) \cdot (\partial s_a / \partial \theta_l) \quad (18)$$

$$f_{kl} = \partial^2 E / \partial \theta_k \partial \theta_l \quad (19)$$

Because the derivatives in eq. 18 are calculated so as to satisfy the Eckart conditions, the molecular motion described by the Lagrangian of eq. 17 involves no external motion provided the deviation is small.

By writing down the equation of motion for the Lagrangian of eq. 17, we have

$$\sum_l h_{kl} \ddot{\theta}_l + \sum_l f_{kl} \theta_l = 0 \quad (20)$$

A general solution of this equation is given as a linear combination of normal mode variables τ_i :

$$\theta_k = \sum_i u_{ki} \tau_i \quad (21)$$

Here the coefficients satisfy the following generalized eigenvalue equation:

$$(2\pi\nu_i)^2 \sum_l h_{kl} \theta_l = \sum_l f_{kl} \theta_l \quad (22)$$

They are orthonormalized to satisfy the following relation.

$$\sum_{kl} h_{kl} u_{ki} u_{lj} = \delta_{i,j} \quad (23)$$

The normal mode variables satisfy the following harmonic equation of motion.

$$\ddot{\tau}_i + (2\pi\nu_i)^2 \tau_i = 0 \quad (24)$$

The mean-square thermal-equilibrium amplitude of the harmonic oscillator is again given by

$$\langle \tau_i^2 \rangle = kT(2\pi\nu_i)^{-2} \quad (25)$$

The theorem to be proved can be derived easily by using eqs. 16, 18, 21, 23 and 25.

4. Discussion

As an illustration of the theorem the result obtained via the normal mode analysis of a globular protein, lysozyme, is presented. The X-ray crystallographic coordinates of hen egg white lysozyme [8] are taken from entry 2LYZ in the Protein Data Bank [9]. As an empirical conformational energy, we employ the one used in the UNICEPP program [10]. In this program, dihedral angles are treated as independent variables and some of the atoms are treated as united atoms. Thus, there are 385 backbone and 301 side-chain dihedral angles. A minimum energy conformation is at first obtained by starting from the X-ray conformation. Subsequently, the normal mode analysis is carried out at the minimum energy conformation.

The distribution of frequencies of the normal modes is given in fig. 1. About 64% of the modes have frequencies of less than 120 cm^{-1} . A sequential number is assigned to each of the normal modes in ascending order of frequency. Thus, the first mode is that with the lowest frequency, 3.9 cm^{-1} . The cumulative contribution to the mass-weighted mean-square fluctuation from the first n modes is plotted vs. n in fig. 2. More than 80% of the contribution to the mass-weighted mean-square deviation stems from a relatively small number of modes with frequencies below 30 cm^{-1} . This fact

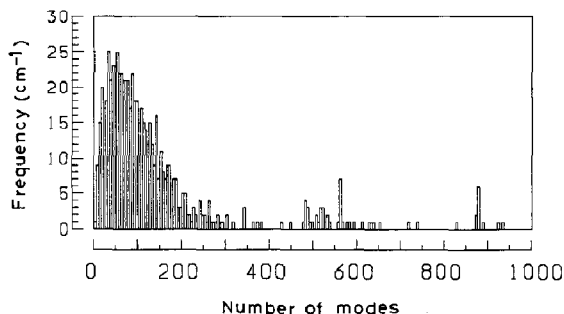


Fig. 1. Histogram of calculated frequencies of normal modes of vibration in lysozyme. Frequencies are shown in corresponding light wave numbers. Number of normal modes in each interval of 5 cm^{-1} is shown.

indicates that this small number of very-low-frequency modes play dominant roles in protein conformational dynamics.

Each normal mode is characterized by its frequency and eigenvector. The latter determines the pattern of atomic displacements in the mode. However, only the frequencies appear in the theorem of eq. 1. This fact indicates that very-low-frequency modes always play dominant roles in the conformational dynamics irrespective of the details of its pattern of atomic displacements.

In the normal mode analysis, the conformational dynamics of a macromolecule is expressed as a superposition of mutually independent mo-

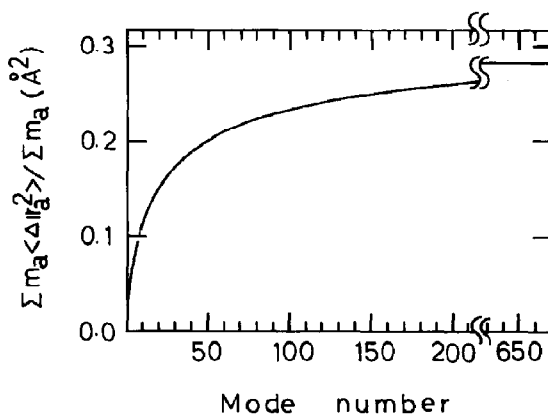


Fig. 2. Cumulative contribution to the mass-weighted mean-square fluctuation from the first n modes plotted vs. n .

tions, i.e., normal modes. Underlying in this picture of conformational dynamics, there is a basic assumption of the harmonicity of the energy surface, i.e., the energy surface is assumed to be given by a multidimensional parabola within the range of thermal conformational fluctuations. This assumption is already known to be invalid, in the case of proteins, but only in very-low-frequency modes. In the majority of not-very-low-frequency modes the assumption of harmonicity is valid [11]. The normal mode analysis is directly useful for understanding the dynamics of protein conformation in these modes. In this paper, we have given a theoretical justification to an observation, which hitherto has been made only empirically, i.e., that very-low-frequency modes have dominant roles in conformational dynamics of proteins. However, this justification must be reexamined, because the basic assumption of the harmonicity of the energy surface is known to be invalid for the very-low-frequency modes. Nonetheless, we believe in the significance of the normal mode analysis, because it provides a natural way of describing collective motions in proteins.

Collective motions are known to be important in protein functions [12]. Functionally important collective motions, which involve necessarily only small amounts of conformational energy change, are expected to be describable as nonlinearly coupled motions of these very-low-frequency normal modes. These normal modes, each of which is a collective motion with a characteristic frequency, can be calculated exactly via an algebraic procedure of the normal mode analysis. In order to realize this expectation, we are now developing a new method of describing protein dynamics by the coupled motion of normal mode variables (T. Horiuchi and N. Gō, manuscript in preparation). The above idea of using the normal mode analysis as a method to define a set of good collective motions for studying highly nonlinear conformational dynamics was also realized as the Monte Carlo simulation method with anisotropic step sizes [5]. In this method, variables corresponding to the normal modes are used as independent variables and the simulation based on this method was estimated to be more efficient than the molecular dynamics calculation by a factor of 5–50 in

sampling various conformations from an ensemble at thermal equilibrium.

Based on the indication that a relatively small number of normal modes make a dominant contribution to the conformational dynamics of proteins, we are also developing a new method of X-ray crystallographic refinement which uses the normal mode variables as independent variables to be refined (A. Kidera and N. Gō, manuscript in preparation).

The theorem proved in this paper provides a justification for the use of variables of very-low-frequency normal modes as good independent variables to describe significant conformational dynamics of proteins.

Acknowledgements

I would like to dedicate this article to Professor Akiyoshi Wada on the occasion of his 60th birthday, under whose guidance I began research investigation in the field of biophysics. I thank Dr. J. Higo for carrying out the normal mode analysis of lysozyme and for preparing the figures reported here. I also thank Dr. A. Kidera for reading the manuscript. Computations have been done at Computer Centers of Kyoto University and of Institute for Molecular Science. This work has been supported by grants from Ministry of Education, Science and Culture, Japan, and Science and Technology Agency, Japan.

References

- 1 M. Karplus and J.A. McCammon, *Annu. Rev. Biochem.* 53 (1983) 263.
- 2 B. Brooks and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* 80 (1983) 6571.
- 3 N. Gō, T. Noguti and T. Nishikawa, *Proc. Natl. Acad. Sci. U.S.A.* 80 (1983) 3696.
- 4 M. Levitt, C. Sander and P.S. Stern, *J. M. Biol.* 181 (1985) 423.
- 5 T. Noguti and N. Gō, *Biopolymers* 24 (1985) 527.
- 6 C. Eckart, *Phys. Rev.* 47 (1935) 552.
- 7 T. Noguti and N. Gō, *J. Phys. Soc. Jap.* 52 (1983) 3283.

- 8 C.C.F. Blake, D.F. Koenig, G.A. Mair, A.C.T. North, D.C. Phillips and V.R. Sarma, *Nature* 206 (1965) 757.
- 9 F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.* 112 (1977) 535.
- 10 L.G. Dunfield, A.W. Burgess and H.A. Scheraga, *J. Phys. Chem.* 82 (1987) 2609.
- 11 T. Noguti and N. Gö, *Nature* 296 (1982) 776.
- 12 R. Huber and W.S. Bennett, *Biopolymers* 22 (1983) 261.