Biophysical Chemistry 25 (1986) 105–116
Elsevier



BPC 01091

*Review*

# ORIGIN OF LOW-FREQUENCY MOTIONS IN BIOLOGICAL MACROMOLECULES *

## A VIEW OF RECENT PROGRESS IN THE QUASI-CONTINUITY MODEL

Kuo-Chen CHOU

*Protein Engineering Group, Corporate Research Laboratories, Eastman Kodak Company, Rochester, NY 14650, U.S.A.*

Received 10th April 1986
Revised manuscript received 1st October 1986
Accepted 1st October 1986

*Key words: Internal collective motion; α-Helix; β-Sheet; β-Barrel; DNA double helix; Microenvironment; Mass distribution; Weak-bond distribution*

The recent progress in the quasi-continuity model and its applications in studying the low-frequency internal motions of biological macromolecules have been surveyed. Emphasis is placed on revealing the origin of this kind of internal collective motion, which involves many atoms and has significant biological functions. In light of such a line, the low-frequency motions in α-helix structure, β-structure (including β-sheet and β-barrel), and DNA double-helix structure, the three most fundamental component elements in biological macromolecules, are discussed, and the corresponding physical pictures described. It turns out that the low-frequency motion in biological macromolecules originates from their two common intrinsic characteristics, i.e., they possess (1) a series of weak bonds, such as hydrogen bonds and salt bridges, and (2) a substantial mass distributed over the region containing those weak bonds.

## 1. Introduction

During the last decade more and more experimental results have emerged [1–7] indicating that low-frequency motions in protein and DNA molecules obviously exist. Meanwhile, various models have been proposed, attempting to reveal the possible mechanisms. They can be classified as follows:

### *1.1. The elastic global model [8]*

In this model a protein molecule was likened to a continuous elastic sphere imitating heart breathing or pulsation movement. And the following formula taken from earth dynamics was used to calculate the low-frequency mode for a globular protein:

$$\tilde{\nu} = \frac{1}{2\pi c}\sqrt{\frac{\pi E}{\rho r^2}} \tag{1}$$

where $\tilde{\nu}$ is the wave number, $c$ the speed of light in vacuum, and $E$, $\rho$, and $r$ are the Young's modulus, mass density, and radius of a protein molecule, respectively. This is a model full of imagination, which once played a role in stimulating development in this area, both theoretical and experimental. However, this model could not reflect the sensitivity of the observed low-frequency mode of a protein molecule to its conformational change, as observed by Brown et al. [1]. In particular, contradictory to eq. 1, it was found at times that the observed low-frequency wave number for a bigger protein molecule was higher than that for a smaller one [6]. All these problems are actually caused by the fact that the model itself treats the whole internal structure of a protein molecule as a 'black box' without considering its conformation.

* Part of this paper was presented at seminars held in Kyoto University, Japan, and in Los Alamos National Laboratory, U.S.A.

*1.2. The harmonic model [9–11]*

This model is actually a combination of the classic normal mode theory with the modern computer technique. In principle, it can reach to the level of the constituent atoms of a protein molecule, and hence much more detailed information can be provided. But also because of this, the approach is much more computationally demanding. The results thus obtained depend very sensitively on the choice or assumption of the empirical potential functions. Besides, using this model, one usually finds a frequency spectrum rather than the *dominant low-frequency mode*, the one whose phonons excited in number are in an overwhelming majority owing to various favorable factors such as the microenvironment [12,13], compared with the other possible low-frequency modes in the frequency spectrum thus found. Therefore, the so-called dominant low-frequency mode is actually the one that possesses the greatest amplitude and corresponds to the outstanding low-frequency peak as observed. In other words, the current harmonic model does not directly present an intuitive picture of the dominant low-frequency motion, which is a kind of collective internal motion involving substantial atoms in a biomacromolecule and hence is most likely to be of primary importance in biological functions [13–19].

*1.3. The quasi-continuity model [20–23]*

This model was established in an attempt to simplify the calculation of the dominant low-frequency motion by grasping the most essential factors and neglecting those that are unessential. As far as the depth of description is concerned, this model lies between the elastic global model and the harmonic model. Unlike the elastic global model, the quasi-continuity model can reflect the internal conformation of a biomacromolecule, but on the other hand, justified by the physical character of low-frequency motion [13], it treats the constituent atoms as a continuous mass distribution, thus greatly simplifying the calculation in comparison with the discrete, normal mode model. As an additional merit of the quasi-continuity model, an intuitive physical picture is provided for various types of dominant low-frequency motions in biological macromolecules, respectively, which is very useful for studying biological functions [17–19].

In this review paper, the quasi-continuity model will be summarized systematically. Before going on, however, let us first see what factors are most essential to the low-frequency internal motion of a biomacromolecule.

## 2. The essential factors of low-frequency motion in a biomacromolecule

In comparison with ordinary small molecules, a biological macromolecule generally bears the following two unique features: (1) a series of weak bonds, such as hydrogen bonds and salt bridges, which actually play a role in stabilizing some certain types of higher-order structures; (2) a great number of atoms distributed around those weak bonds. These two characteristics not only give variety to the conformation of a biomacromolecule, but are also the origin of its low-frequency internal motion. According to the report by Brown et al. [1], the observed low-frequency peaks for $\alpha$-chymotrypsin and pepsin would immediately disappear once they were denatured. This apparently indicates that this kind of low-frequency motion is very sensitive to the molecular conformation, or more explicitly, to the distribution, formation, or disruption of the weak bonds in a protein molecule.

The above conception can serve as a criterion, helping us to discern which factors are essential and must be taken into account, and which are unessential and may be treated by a rough approximation or even neglected in order to establish a simple but successful calculation model. Later, we shall apply this criterion to approach this subject from three different aspects in structure, i.e., the $\alpha$-helix structure, $\beta$-structure (including $\beta$-sheet and $\beta$-barrel), and the double-helix structure. The first two structures are the most fundamental component elements in protein molecules, and the third one stands for the basic structural pattern of a DNA molecule. The discussion following such a line will naturally disclose the origin of low-frequency motions in biological macromolecules, as well as provide a series of convincing demonstrations from different aspects.

## 3. α-Helical structure

The quasi-continuity model has been used quite successfully to calculate the low-frequency motions of α-helices with various microenvironments in protein molecules [12,13,20,22]. As is well known, a normal α-helix has 3.6 residues per turn, with a hydrogen bond between the CO of the $i$-th residue and NH of the $(i+4)$-th residue; i.e., an α-helix composed of $n$ amino-acid residues generally has $n-4$ hydrogen bonds. Consequently, a collective fluctuation of these hydrogen bonds along the axis of the α-helix will naturally give rise to an accordion-like motion (see fig. 1). Besides, the analysis in section 2 indicates that the covalent (strong) bonds in a biomacromolecule are not essential to the low-frequency internal motion. In fact, the effect of their fluctuations can be neglected because they either merely generate high-frequency modes or reach a temporary equilibrium (corresponding to a local minimum in energy) in a much shorter time than the period of the low-frequency motion considered here. Accordingly, for simplicity, rather than the discrete lattice model, we prefer to use the continuity model in which all the constituent atoms are treated as a continuous mass distributed according to some certain geometric shape, depending on the conformation of the structure concerned. Based on such a physical picture, an α-helix can be compared to a spring whose mass, however, is not negligible, viz., a spring with distributed mass. Such a continuity model implies the following two aspects of the origin: (1) a series of hydrogen bonds that stabilize the frame of a helical structure, and (2) a continuous mass distribution that further substantiates such a spring-like framework.

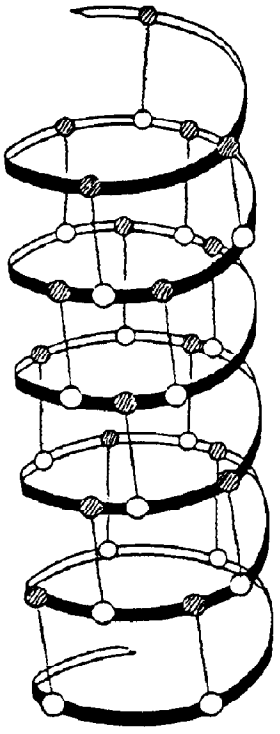

Fig. 1. Illustration of an α-helix. (○) Peptide oxygen, (⊘) peptide nitrogen, ( ~~~~ ) hydrogen bond.

However, an α-helix in a protein molecule may have various terminal conditions, depending on its microenvironment and location. For example, in proteins α-helices often possess the terminal conditions as illustrated in fig. 2a, b, and c, respectively. When in accordion-like motion, the formulas to calculate their low-frequency modes are as follows.

(1) If the two ends of a helix are linked to two hard walls * via two mass-negligible springs (e.g., chemical bonds) whose force constants are $K_1$ and $K_2$, respectively, as shown in fig. 2a, we have [20]

$$\tilde{\nu} = \frac{1}{2\pi c}\sqrt{\frac{k+K^*}{(\alpha_1^3+\alpha_2^3)\rho L/3}} \tag{2}$$

where $\rho$ is the mass per unit length along the axis of the helix, $L$ the total length of the helix axis, $k$ the stretching force constant of the helix, to be given later (see eq. 11), and

$$\alpha_1 = \frac{K_2}{K_1+K_2}, \quad \alpha_2 = \frac{K_1}{K_1+K_2} \tag{3}$$

$$K^* = \frac{K_1K_2}{K_1+K_2} \tag{4}$$

(2) If at the two ends of a helix there are two fragments attached whose masses are $M_1$ and $M_2$,

* Here the 'hard wall' represents the compact portion of a protein that has much greater mass than the helix.

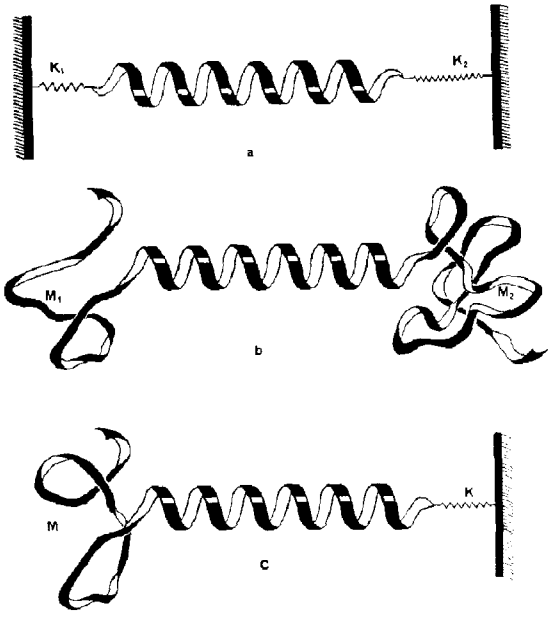


Fig. 2. The three different terminal conditions of an α-helix in protein molecules. See the text.

respectively (fig. 2b), then we have [12]

$$\tilde{\nu} = \frac{1}{2\pi c}\sqrt{\frac{k}{M^* + (\beta_1^3 + \beta_2^3)\rho L/3}} \tag{5}$$

where

$$\beta_1 = \frac{M_2}{M_1 + M_2}, \quad \beta_2 = \frac{M_1}{M_1 + M_2} \tag{6}$$

$$M^* = \frac{M_1 M_2}{M_1 + M_2} \tag{7}$$

(3) If one end of a helix is linked to a hard wall through a mass-negligible spring whose force constant is $K$, and at its other end is attached a peptide fragment whose mass is $M$ (fig. 2c), then we have instead [22]

$$\tilde{\nu} = \frac{1}{2\pi c}\sqrt{\frac{K^\dagger}{M + (1 + \gamma + \gamma^2)\rho L/3}} \tag{8}$$

where

$$\gamma = \frac{k}{k + K} \tag{9}$$

$$K^\dagger = \frac{kK}{k + K} \tag{10}$$

In the above equations the stretching force constant $k$ of the helix is associated with the constituent hydrogen bonds (fig. 1), and can be calculated by the following formula [12]

$$k = \begin{cases} 6k_H^\alpha/5i & \text{if } j = 0 \\ 12k_H^\alpha/(10i + 12/j) & \text{if } 1 \le j \le 4 \\ 12k_H^\alpha/[10i + 3 + 12/(j-4)] & \text{if } 5 \le j \le 8 \\ 12k_H^\alpha/[10i + 6 + 12/(j-8)] & \text{if } 9 \le j \le 10 \end{cases} \tag{11}$$

where

$$\begin{cases} i = \mathrm{INT}\left\{\dfrac{n_\alpha - 4}{11}\right\} \\ j = (n_\alpha - 4) - 11i \end{cases} \tag{12}$$

where $n_\alpha$ is the number of residues in the α-helix and hence must be greater than 4, the truncation operator INT represents taking the integral part for the number following it; e.g., INT(7/3) = 2, INT(1/4) = 0, and so forth, and

$$k_H^\alpha = \sqrt{(k_H^s \cos\theta)^2 + (k_H^b \sin\theta)^2} = 0.12 \text{ mdyn/Å} \tag{13}$$

where $\theta \simeq 26°$ is the angle between the helix axis and the constituent hydrogen bonds [20], and $k_H^s = 0.13$ mdyn/Å and $k_H^b = 0.03$ mdyn/Å [24] are the stretching and bending force constants of the hydrogen bond, respectively. By means of eqs. 11–13, the stretching force constant of any given α-helix can be easily determined.

Here let us take the insulin molecule as an example, through which the calculation model can be fully illustrated. It is well known that an insulin molecule consists of A- and B-chains. The A-chain contains two helical regions (residues A2–A8; A13–A20), and the B-chain contains one rather longer helix (residues B9–B19). Among these three helices, only the helix (B9–B19) is most likely responsible for the observed low-frequency peak of 22 $cm^{-1}$ [6] because the other two helices (A2–A8 and A13–A20) seem too short to generate

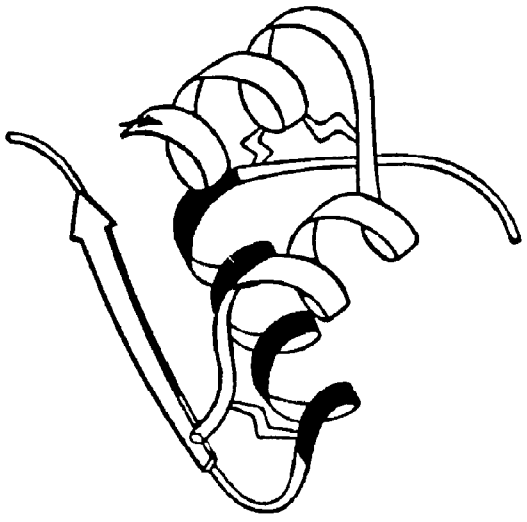

Fig. 3. Schematic drawing of an insulin molecule (with permission from J.S. Richardson), in which the helix (residues B9–B19) is shaded black.

any low-frequency modes within this range [12]. In other words, our attention should be first focused on the helix (B9–B19), which is specially marked with black shading (fig 3) and termed as the 'principal helix' [12]. The amino acid sequence of the helix (B9–B19) is [25]:

Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys
(B9) (B19)

which consists of 11 amino acid residues, i.e., $n_\alpha = 11$. Thus, according to eq. 12 we have $i = 0$ and $j = 7$. Substitution of these data, as well as of eq. 13 into eq. 11, will yield

$$k = 12k_H^\alpha/7 = 0.20 \times 10^5 \text{ dyn/cm} \tag{14}$$

On the other hand, we have [12]

$$\rho L = \text{the total mass of the helix (B9–B19)} = 1228 \text{ g}/N \tag{15}$$

where $N$ is Avogadro's constant. The residues B1–B8 and B20–B30 that attach to the two ends of the helix (B9–B19) and can be treated as two mass fragments are Phe-Val-Asn-Gln-His-Leu-Cys-Gly and Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Ala, respectively [25]. Accordingly, we have $M_1 = 901$ g/$N$ and $M_2 = 1346$ g/$N$, which via eqs. 6–7, further give

$$\beta_1 = 0.60, \quad \beta_2 = 0.40 \quad M^* = 540 \text{ g}/N \tag{16}$$

Substituting eqs. 14–16 into eq. 5, we obtain

$$\tilde{\nu} = \frac{1}{2\pi \times 3 \times 10^{10}} \times \sqrt{\frac{0.20 \times 10^5 \times 6.023 \times 10^{23}}{540 + \left[(0.60)^3 + (0.40)^3\right] \times 1228/3}} = 22.7 \text{ cm}^{-1} \tag{17}$$

which is in excellent agreement with the observed low-frequency peak of 22 $cm^{-1}$ [6]. This is also a good explanation for why the same low-frequency peak of 22 $cm^{-1}$ was observed for insulin dimer as well as insulin monomer, because both contain the principal helix (B9–B19). Moreover, the above illustration also discloses the essence of the dominant low-frequency motion in the insulin molecule, as reflected by the calculation process and the model itself.

For the $\alpha$-helix with different microenvironments in a protein molecule, the equations for calculating $\tilde{\nu}$ might be a little different in form (cf. eqs. 2, 5 and 8), but the essence of the accordion-like motion resulting from the collective fluctuation of their constituent hydrogen bonds remains the same. In table 1 more calculated results for other protein molecules, together with the corresponding observations, are listed, all indicating a good agreement between observation and calculation.

## 4. $\beta$-Structure

The model presented earlier for the accordion-like motion of an $\alpha$-helix is very successful in explaining the observed maximum low-frequency peaks for many protein molecules; however, remarkable low-frequency peaks have also been observed for some major-$\beta$ protein molecules, in which there is no $\alpha$-helix but extensive $\beta$-structures are observed. Therefore, it is necessary to extend the discussion on the origin of low-frequency motion to $\beta$-structures.

Table 1

A comparison between the observed low-frequency wave numbers for some biological macromolecules and their calculated results in terms of our models

| Biological macromolecule | Principal structure involved | Form of motion | Calculated wave number ($cm^{-1}$) | Observed wave number ($cm^{-1}$) |
|---|---|---|---|---|
| α-Chymotrypsin | helix (residues 235–245) | accordion-like motion | 30 [a] | 29 [b] |
| Pepsin | helix (residues 225–235) | accordion-like motion | 33 [a] | 32 [b] |
| Lysozyme | helix 1 (residues 5–15)<br>helix 2 (residues 25–35) | accordion-like motion | 27<br>26 } [c] | 25 [d] |
| Insulin | helix (residues B9–B19) | accordion-like motion | 23 * | 22 [e] |
| Immunoglobulin G | the β-barrel with 9 strands in $V_H$ domain | breathing motion | 28 [f] | 28 [e] |
| | the β-barrel with 7 strands in $V_L$ domain | breathing motion | 36 [g] | 36 [h] |
| | the β-barrel with 7 strands in $C_{HL}$ domain | breathing motion | 29 [g] | 28 [h] |
| Concanavalin A | the β-barrel with 14 strands | breathing motion | 19 * | 20 [e] |
| DNA | various individual intact DNA segments | standing wave | 30 * | 30 [i] |

* Examples given in this paper. [a] See ref. 13. [b] See ref. 1. [c] See ref. 12. [d] See ref. 3. [e] See ref. 6. [f] See ref. 23. [g] See ref. 19. [h] See ref. 4. [i] See ref. 5.

### *4.1. β-Sheet*

As is well known, in a β-sheet there are a series of hydrogen bonds between its two adjacent strands. The hydrogen bonds are formed between every other backbone C=O group of each strand and every other backbone NH group of its adjacent one (fig. 4a and b), and vice versa [26,27]. Thus, if an α-helix can be compared with a spring with distributed mass, as treated in section 3, a β-sheet can of course be compared to a 'rods and springs' mattress as shown in fig. 5. Each rod in this figure represents a strand, and each spring denotes a hydrogen bond whose force constant along the vibration direction is assumed to be $k_H^\beta$. When such a system is in accordion-like motion, the wave number is given by [23]

$$\tilde{\nu} = \frac{1}{\pi c}\sqrt{\frac{3\Lambda_S k_H^\beta}{(\mu^2-1)M_S}} \qquad (18)$$

where $M_S$ is the total mass of the β-sheet, and $\Lambda_S$ and $\mu$ are the numbers of its total hydrogen bonds and strands, respectively. Eq. 18 can also be used approximately to deal with a nonregular β-sheet if its nonregularity does not cause a large relative variance in the mass of each strand and in the number of hydrogen bonds between two adjacent strands. Actually, for many β-structures in proteins, the above condition can be satisfied quite well.

Now let us apply eq. 18 to a typical representative of β-sheets in proteins. A statistical analysis [28] for the known β-structures in proteins indicates that the average number of strands per sheet is 4.7, and most strands have 5–7 residues. Accordingly, as a representative, we can consider a β-sheet with five strands, each having six residues plus $CH_3CO$- and -$NHCH_3$ end groups. The number of total hydrogen bonds for such a β-sheet is

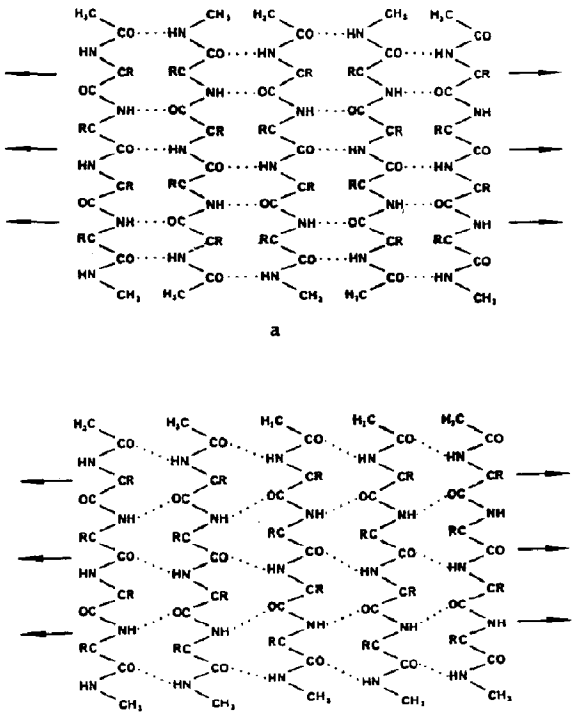

Fig. 4. Schematic drawing of (a) an antiparallel $\beta$-sheet, and (b) a parallel $\beta$-sheet. The dotted line represents a hydrogen bond between two adjacent strands. Note that in both cases the hydrogen atoms bound to $C^{\alpha}$ atoms are omitted.

$(6+1)\times(5-1)=4\times 7$ [29]; i.e., $\Lambda_S = 28$. But note that the hydrogen bonds in an antiparallel $\beta$-sheet are along the direction of its accordion-like motion (fig. 4a), whereas those in a parallel $\beta$-sheet tilt up or down with an angle of $\sim 20°$ [30] along that direction (fig. 4b). Therefore, we have

$$k_H^{\beta} = \begin{cases} k_H^s = 0.13\times 10^5 \text{ dyn/cm} \\ \quad \text{(for antiparallel sheet)} \\ \sqrt{\left(k_H^s \cos 20°\right)^2 + \left(k_H^b \sin 20°\right)^2} \\ \quad = 0.12\times 10^5 \text{ dyn/cm} \\ \quad \text{(for parallel sheet)} \end{cases} \qquad (19)$$

where $k_H^s$ and $k_H^b$ have the same definition as in eq. 13. The mass of each residue in such a representative $\beta$-sheet is 111 a.m.u., the average mass per residue for the known $\beta$-sheets [31]. Substituting

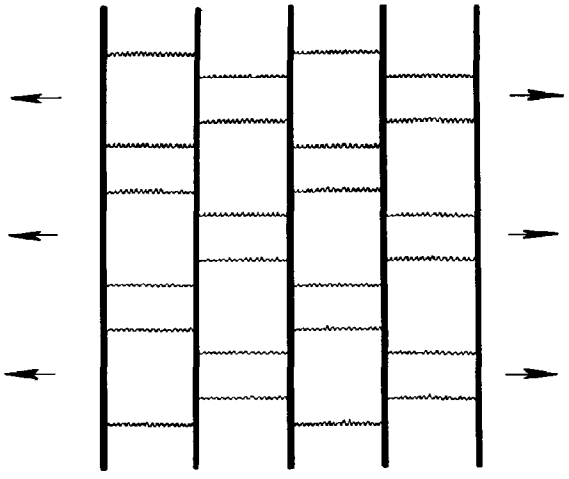

Fig. 5. A 'rods and springs' mattress used to imitate the accordion-like motion of a $\beta$-sheet in which there are five strands and six hydrogen bonds between any two adjacent strands.

$$\left.\begin{aligned} &\mu = 5, \quad \Lambda_S = 28 \\ &M_S \simeq 6\times 5\times 111 = 3330 \quad \text{a.m.u.} = 3330 \text{ g}/N \end{aligned}\right\} \qquad (20)$$

as well as eq. 19 into eq. 18, we obtain

$$\tilde{\nu} = \begin{cases} 30.4 \text{ cm}^{-1} & \text{for the antiparallel sheet} \\ 29.2 \text{ cm}^{-1} & \text{for the parallel sheet} \end{cases} \qquad (21)$$

The above calculation illustrates that a $\beta$-sheet with five strands, each having six residues, can also generate a low-frequency wave number of $\sim 30$ cm$^{-1}$. Besides the trivial difference between the two values in eq. 21 indicates that the effect due to a tilted angle of approx. 20° in hydrogen bonds is insignificant on the calculated low-frequency mode. Accordingly, it is not necessary in later calculations to make such a distinction between the two cases, as shown in eq. 19. This also justifies the approximate validity of the present model, even for a twisted $\beta$-sheet [27,32]. Further discussion and a comparison with observation will be given below.

### 4.2. β-Barrel

According to the same conception, a β-barrel can be compared to a 'rods and springs' cylinder (fig. 6a). When such a cylinder is carrying out a breathing motion (fig. 6b), its wave number can be derived as [23]

$$\tilde{\nu} = \frac{1}{\pi c}\left(\sin\frac{\pi}{\mu}\right)\sqrt{\frac{\Lambda_B k_H^s}{M_B}} \tag{22}$$

where $\Lambda_B$ is the number of the total hydrogen bonds in the β-barrel, and $M_B$ is its mass. Likewise, under the similar condition as mentioned above for a nonregular β-sheet, eq. 22 can also be applied approximately to a nonregular β-barrel.

Now let us take concanavalin A as an example. This molecule has been widely used as a molecular probe in studies of cell membrane dynamics and cell division. A striking feature of concanavalin A is that it contains a big β-barrel of 14 strands but no α-helix at all, as shown by the schematic drawing [33] in fig. 7. For such a β-barrel, we have $\mu = 14$, $M \simeq 13\,320$ g/$N$ [34], and $\Lambda_B = 110$ [35]. Substituting these data and $k_H^\beta \simeq k_H^s = 0.13 \times 10^5$ dyn/cm [24] into eq. 22, we obtain

$$\tilde{\nu} \simeq \frac{1}{\pi \times 3 \times 10^{10}}\left(\sin\frac{\pi}{14}\right) \times \sqrt{\frac{110 \times 0.13 \times 10^5 \times 6.023 \times 10^{23}}{13\,320}} = 19.0\ \text{cm}^{-1} \tag{23}$$

which is quite close to 20 $\text{cm}^{-1}$ [6], the maximum low-frequency peak observed for concanavalin A. Consequently, the dominant low-frequency motion observed in concanavalin A originates very likely from the breathing motion of the β-barrel therein.

## 5. DNA double-helix structure

If the conception regarding the origin of low-frequency internal motion in a protein molecule is correct, then this kind of motion must occur in a DNA molecule as well. This is because in a DNA molecule the following two factors that are essential to the low-frequency motion also exist: (1) a series of weak bonds as reflected by many consecutive hydrogen bonds between the complementary nitrogen bases along the double-helix chains; (2) a substantial number of atoms distributed over such a structure, which also justifies the continuity model applied here. We shall use the continuity model to approach the low-frequency internal motion of a DNA molecule, the most important genetic molecule in biology.

According to the continuity model as demonstrated earlier for the fundamental structures in protein molecules, the DNA double-helix struc-

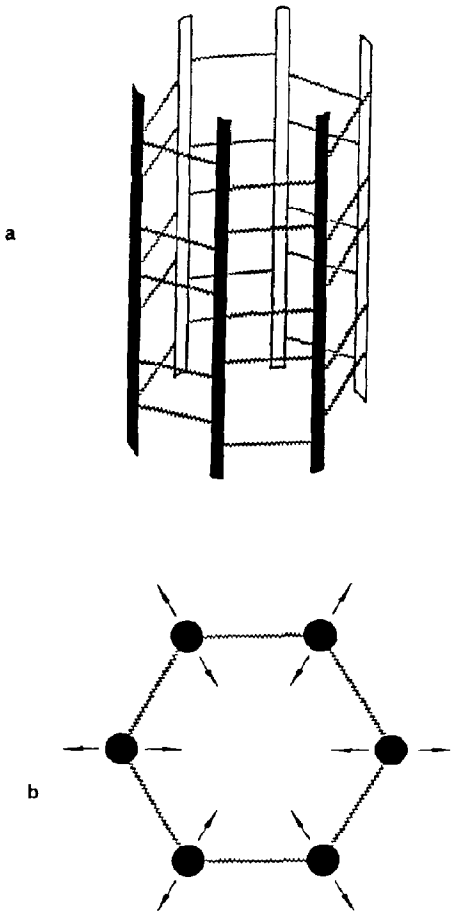


Fig. 6. Schematic drawing of (a) a 'rods and springs' cylinder used to imitate the breathing motion of a β-barrel, and (b) an illustration of such breathing motion when viewed from the top of the cylinder.

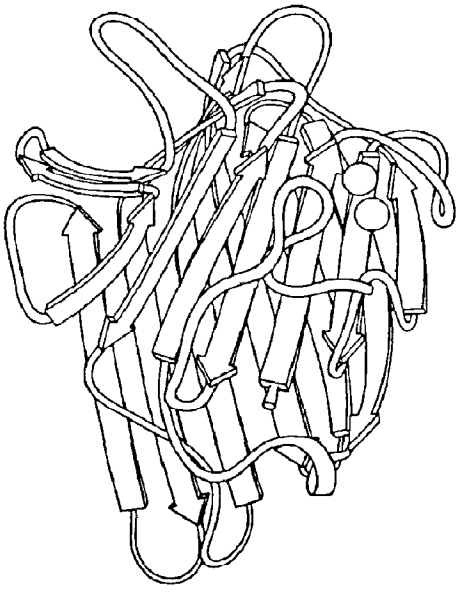

Fig. 7. Schematic drawing of concanavalin A (with permission from J.S. Richardson).

ture can be compared with a 'ribbons and springs' duplex twiner, as illustrated in fig. 8. In this schematic illustration the two DNA polynucleotide chains are compared to two right-handed helical ribbons intertwining around the same axis $Z$, while the hydrogen bonds between a pair of complementary bases are compared to a spring, holding the double ribbons together and consecutively rotating around the $Z$-axis by about 36°. However, under thermal equilibrium, not all these base-pairs are closed as held by the intact Watson-Crick hydrogen bonds; some will be open corresponding to a break in hydrogen bonds (or, according to the physical picture in fig. 8, corresponding to a snap of the springs). The closed-open motion is familiarly known as 'breathing', which is governed by the following thermodynamic relation [36]

$$\text{Closed base-pair} \overset{K_{\text{open}}}{\rightleftharpoons} \text{open base-pair} \tag{24}$$

The experiments gave [36]

$$\frac{1}{K_{\text{open}}} = \langle n_{\text{base}} \rangle = 38 \tag{25}$$

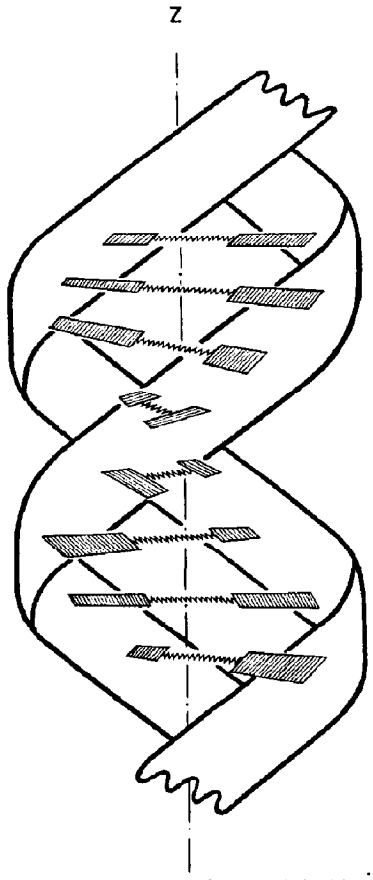


Fig. 8. A 'ribbons and springs' double twiner system used to imitate the collective fluctuations of the complementary hydrogen bonds in a DNA molecule, whose two polynucleotide chains are compared with two right-handed helical ribbons intertwining around the $Z$-axis, and whose hydrogen bonds between a pair of complementary bases (denoted by hatched rectangles) are compared with a spring, holding the double ribbons together and consecutively rotating around the $Z$-axis by about 36°.

which means that, on average, for every 38 closed base-pairs there is an open base-pair. Besides, NMR relaxation experiments have further demonstrated [37] that the open base-pair is usually sandwiched between closed base-pairs, i.e., base-pair opening is a single base-pair event. Therefore, along a DNA molecule there are actually many 'blocks,' which have a consecutive run of closed base-pairs and which are separated from each other by the open pairs (broken springs). Each of

these blocks is called an *intact DNA segment*, in which, however, the forces that hold the double helices are no longer uniform because of the effect of the open base-pairs close to its two ends (fig. 9a). Consequently, for any of these intact DNA segments, the vibration amplitude between a pair of bases along the hydrogen bond direction will depend on its location: the nearer the base-pair to an end of the segment, the larger its amplitude will be, and vice versa (fig. 9b). Furthermore, although eq. 24 represents a dynamic equilibrium, the conversion period between a closed base-pair and open base-pair is much longer [38] than that of the low-frequency vibration observed for DNA molecule. In fact the former is associated with a 'break of spring,' which is already beyond the elastic limit of a spring and hence no longer belongs to any kind of elastic movement. Therefore, within a certain interval of time that is much longer than the period of the low-frequency vibration but shorter than the conversion period of the closed-open motion, only the coupled effect of the low-frequency vibration with a particular closed-open state, but not with the closed-open motion itself, needs to be considered for the present calculation, at least as a first-order approximation. In other words, the low-frequency vibration of interest is presented as involving hydrogen-bond fluctuations within a particular conformation of a DNA molecule. Besides, for a given closed-open state, the number of closed base-pairs as a consecutive run in a block of intact DNA segment is generally quite large, as shown by eq. 25 (i.e., $\langle n_{\text{base}} \rangle \gg 1$). This means that there is no broken spring within quite long double intertwining ribbons. In this case, according to the above analysis on the forces, the amplitude at the middle of an intact DNA segment (corresponding to point C of fig. 9b) must be much smaller than those at its two ends (corresponding to points A and B of fig. 9b, respectively), and can be virtually neglected. Consequently, within each of these segments along a DNA molecule, there actually exists a fundamental vibration that has the feature of a standing wave, as illustrated in fig. 9c, where $Z = L/2$ (corresponding to C of fig. 9b) is the node, and the antinodes are at $Z = 0$ and $L$ (corresponding to A and B in fig. 9b, respectively). The continuous curve represents the picture of the standing wave at one instant, the broken curves depicting it at other instants. Based on such a physical picture of a standing wave, for any given intact segment in a homo-DNA double-helix molecule, the following formula has been derived [21] to calculate its fundamental frequency:

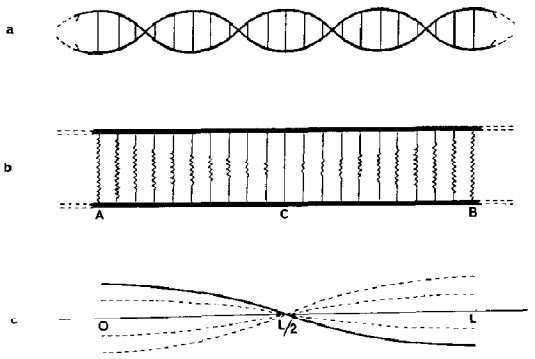


Fig. 9. Illustration to show (a) a piece of intact DNA segment, (b) how the amplitude of a complementary base-pair depends on its location in this segment, and (c) the feature of a standing wave of such a collective motion. In (a), right beyond the two ends of the segment are the open base-pairs which separate it from the other intact segments along a DNA molecule. In (b), the amplitude at A and B is the largest because they are, respectively, the closest to an open base-pair; and the amplitude at C, the middle point of the segment, is the smallest. In (c), the node is at $Z = L/2$, corresponding to C of (b), and the antinodes are at $Z = 0$ and $L$, corresponding to A and B of (b), respectively. The continuous curve represents the picture of the standing wave at one instant, the broken curves depicting it at other instants.

$$\tilde{\nu} = \frac{1}{2\pi c}\sqrt{\frac{(n_{\text{base}}+1)k_{\text{base}}}{n_{\text{base}}\langle m\rangle}} \tag{26}$$

where $n_{\text{base}}$ is the number of the total closed base-pairs in the intact DNA segment, $k_{\text{base}}$ is the resultant stretching force constant of the intact Watson-Crick hydrogen bonds holding a pair of complementary bases, and

$$\langle m\rangle = \frac{\text{total mass of the intact DNA segment}}{n_{\text{base}}} = \frac{\text{total mass of the homo-DNA molecule}}{\text{total number of its base-pairs}} \tag{27}$$

Now let us use eq. 26 to calculate the low-frequency modes in DNA molecules. As is well known, a novel feature of the complementary hydrogen bonds in a DNA molecule is that base A (adenine) is always bound to base T (thymine) with two hydrogen bonds, and base G (guanine) always bound to base C (cytosine) with three hydrogen bonds, as may be expressed by A ::: T and G ⋮⋮⋮ C. Therefore, we have [24]

$$k_{\text{base}} = \begin{cases} 2k_{\text{H}}^{\text{s}} = 0.26 \times 10^5 \text{ dyn/cm} \\ \quad \text{for poly(A ::: T) DNA} \\ 3k_{\text{H}}^{\text{s}} = 0.39 \times 10^5 \text{ dyn/cm} \\ \quad \text{for poly(G ⋮⋮⋮ C) DNA} \end{cases} \tag{28}$$

and

$$\langle m \rangle = \begin{cases} 615 \text{ g}/N & \text{for poly(A ::: T) DNA} \\ 616 \text{ g}/N & \text{for poly(G ⋮⋮⋮ C) DNA} \end{cases} \tag{29}$$

Substituting eqs. 28–29 as well as $n_{\text{base}} = \langle n_{\text{base}} \rangle = 38$ into eq. 26, we obtain

$$\tilde{\nu} = \begin{cases} 27.1 \text{ cm}^{-1} & \text{for poly(A ::: T) DNA} \\ 33.2 \text{ cm}^{-1} & \text{for poly(G ⋮⋮⋮ C) DNA} \end{cases} \tag{30}$$

However, for a general DNA molecule, the ratio of (A + T) : (G + C) is within the range 0.98–1.12 : 1 [39]; i.e., the ratio is very close to 1 : 1. Accordingly, for an intact segment in a general DNA molecule, we can instead approximately use the following data:

$$\left.\begin{aligned} &k_{\text{base}} \simeq 2.5k_{\text{H}}^{\text{s}} = 3.25 \times 10^5 \text{ dyn/cm} \\ &\langle m \rangle = 615.5 \text{ g}/N, \; n_{\text{base}} = 38 \end{aligned}\right\} \tag{31}$$

Substitution of eq. 31 into eq. 26 will yield

$$\tilde{\nu} = 30.3 \text{ cm}^{-1} \tag{32}$$

which is in very good agreement with the maximum low-frequency peak of 30 $\text{cm}^{-1}$ as observed by Painter et al. [5] for DNA molecules.

One question might be raised: The value of $\langle n_{\text{base}} \rangle$ just represents a statistical average under thermodynamic equilibrium, as implied in eqs. 24 and 25. Therefore, different intact segments along a DNA molecule will generally have different numbers of closed base-pairs. If so, will we also obtain very different low-frequency wave numbers if the calculations are based on different intact segments? The answer is no. This is because in eq. 26 the factor $n_{\text{base}}$ appears in both numerator and denominator, although in the numerator it has the form of $(n_{\text{base}} + 1)$. Simple calculations show that when $n_{\text{base}} \geq 10$ the relative change of $\tilde{\nu}$ for different $n_{\text{base}}$ will always be less than 0.05. For example, if $n_{\text{base}}$ in eq. 31 is taken between 20 and 50 [40], eq. 26 will give wave numbers within the range of 30.6 and 30.2 $\text{cm}^{-1}$, both of which are very close to the observed peak of 30 $\text{cm}^{-1}$ [5]. This is a very good elucidation of why the observed low-frequency peak is so sharp and stable at approx. 30 $\text{cm}^{-1}$ in spite of the difference in $n_{\text{base}}$, the number of consecutive closed base-pairs, for a different intact segment in a DNA molecule and its constant change with time as a consequence of the so-called closed-open breathing motion, as described by eq. 24. In other words, the low-frequency standing-wave motions occurring in different intact segments of a DNA molecule will generate almost the same wave number, 30 $\text{cm}^{-1}$.

The physical feature just described concerning the low-frequency internal motions in a DNA molecule is very useful for understanding the dynamic mechanism of DNA 'breathing' and drug intercalation, as described by Zhou [41]. This subject will be studied further in another report.

## 6. Conclusion

One of the frontiers today in molecular biology is the study of low-frequency internal motions in biological macromolecules and the relevant biological functions [13–19,41]. Consequently, it is crucially important to reveal the origin of this kind of interesting internal motion. It is reasoned that the origin is due to the existence of a series of weak bonds and a large number of atoms distributed around those weak bonds. The calculated results based on such a conception and the corresponding models are listed in table 1, where, in addition to the three concrete examples illustrated, some more examples are also incorporated. From table 1 we can see that almost all the calculated results are quite close to their corresponding ob-

served values. The reason why the results thus obtained are so good is that the adopted physical models, although quite simple, contain the essence of the matter, i.e., they reflect the real origin of low-frequency internal motion in a biological macromolecule. Consequently, although some very complicated but unessential factors were neglected during the current approximate treatment, the final results thus obtained are still very close to the observations. Apparently, this is because these kinds of unessential factors are either cancelled out by each other or reach an internal mutual equilibrium in a much shorter time in comparison with the period of the low-frequency motion considered here, as tacitly implied in the quasi-continuity model itself [13]. Especially, as an additional remarkable merit, the calculation model derived from such an origin provides us with a clear and intuitive physical picture of low-frequency internal motion for each of the three most fundamental structures in biological macromolecules, which is very useful for the study of biological functions [13–19,41].

## Acknowledgments

Valuable discussions with Professor Tatsuo Ooi, of Kyoto University, Japan, and Dr. I.J. Bigio, of Los Alamos National Laboratory, U.S.A, are gratefully acknowledged.

## References

1 K.G. Brown, S.C. Erfurth, E.W. Small and W.L. Peticolas, Proc. Natl. Acad. Sci. U.S.A. 69 (1972) 1467.
2 B. Fanconi and L. Finefold, Science 190 (1975) 458.
3 L. Genzel, F. Keilman, T.P. Martin, G. Winterling, Y. Yacoby, H. Frohlich and M.W. Makinen, Biopolymers 15 (1976) 219.
4 P.C. Painter and L.E. Mosher, Biopolymers 18 (1979) 3121.
5 P.C. Painter, L.E. Mosher and C. Rhoads, Biopolymers 20 (1981) 243.
6 P.C. Painter, L.E. Mosher and C. Rhoads, Biopolymers 21 (1982) 1469.
7 G.J. Evans, M.W. Evans and R. Pething, Spectrochim. Acta 38A (1982) 421.
8 Y. Suezaki and N. Go, Int. J. Peptide Protein Res. 7 (1975) 333.
9 T. Noguti and N. Go, Nature 296 (1982) 776.
10 B. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. 80 (1982) 6571.
11 R.M. Levy, A.R. Srinivasam, W.K. Olson and J.A. McCammon, Biopolymers 23 (1984) 1099.
12 K.C. Chou, Biochem. J. 215 (1983) 465.
13 K.C. Chou, Biophys. J. 45 (1984) 881.
14 K.C. Chou and N.Y. Chen, Sci. Sin. (Engl. Edn.) 20 (1977) 447.
15 K.C. Chou and N.Y. Chen. Huaxue Tongbao (Chinese Edn.) 5 (1978) 29.
16 K.C. Chou, N.Y. Chen and S. Forsen, Chem. Scr. 18 (1981) 126.
17 K.C. Chou, Biophys. Chem. 20 (1984) 61.
18 K.C. Chou and Y.S. Kiang, Biophys. Chem. 22 (1985) 219.
19 K.C. Chou, Biopolymers (1986) in the press.
20 K.C. Chou, Biochem. J. 209 (1983) 573.
21 K.C. Chou, Biochem. J. 221 (1984) 27.
22 K.C. Chou, Int. J. Biol. Macromol. 7 (1985) 77.
23 K.C. Chou, Biophys. J. 48 (1985) 289.
24 K. Itoh and T. Shimanouchi, Biopolymers 9 (1970) 383.
25 T. Blundell, G. Dodson, D. Hodgkin and D. Mercola, Adv. Protein Chem. 26 (1972) 279.
26 L. Pauling and R.B. Corney, Proc. Natl. Acad. Sci. U.S.A. 39 (1953) 253.
27 K.C. Chou and H.A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. 79 (1982) 7047.
28 M.J.E. Sternberg and J.M. Thornton, J. Mol. Biol. 110 (1977) 285.
29 K.C. Chou, M. Pottle, G. Nemethy, Y. Ueda and H.A. Scheraga, J. Mol. Biol. 162 (1982) 89.
30 G.J. Hol, L.W. Halie and C. Sander, Nature 294 (1981) 532.
31 S. Lifson and C. Sander, Nature 282 (1979) 109.
32 C. Chothia, J. Mol. Biol. 75 (1972) 295.
33 J.S. Richardson, Adv. Protein Chem. 34 (1981) 167.
34 F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, Jr, M.D. Brice, J.B. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, J. Mol. Biol. 112 (1977) 535.
35 G.N. Reeke, Jr, J.W. Becker and G.M. Edelman, J. Biol. Chem. 250 (1975) 1525.
36 G.S. Manning, Biopolymers 22 (1983) 689.
37 T.A. Early, D.R. Kearns, W. Hillen and R.D. Wells, Biochemistry 20 (1981) 3764.
38 G. Careri, P. Fasella and E. Gratton, CRC Crit. Rev. Biochem. 3 (1975) 141.
39 R.C. Bohinski, in: Modern concepts in biochemistry, 2nd edn., ch. 7 (Allyn and Bacon, Boston, 1976) p. 201.
40 C. Mandal, N.R. Kallenbach and S.W. Englander, J. Mol. Biol. 135 (1979) 391.
41 G.P. Zhou, Shengwu Huaxue Yu Shengwu Wuli Jinzhan (Chinese Edn.) 5 (1981) 19.