Biochem. J. (1983) **215**, 465–469
Printed in Great Britain

Identification of low-frequency modes in protein molecules

Kuo-Chen CHOU
Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853, U.S.A.

(*Received 27 April 1983/Accepted 3 August 1983*)

It is demonstrated that the observed low-frequency motions with wave numbers of $22\,cm^{-1}$ and $25\,cm^{-1}$ for insulin and lysozyme respectively originate from the accordion-like motions of the principal helices therein. The calculated results based on such a model are in good agreement with the observed values. During calculations the role of the internal microenvironment upon the low-frequency motion is naturally revealed, so as to elucidate as well why this kind of low-frequency motion is so sensitive to the conformations of proteins observed.

In the last decade, more and more evidence has emerged showing that low-frequency motions apparently exist in protein molecules. (Brown *et al.*, 1972; Genzel *et al.*, 1976; Painter *et al.*, 1982; Evans *et al.*, 1982). Meanwhile, many speculations have been made in an attempt to explore their biological functions (Green, 1974; Ji, 1974; Fröhlick, 1975; Careri *et al.*, 1975; Chou & Chen, 1977, 1978; Sobell *et al.*, 1979, 1982; Englander, 1980; Chou *et al.*, 1981; Zhou, 1981), which are no doubt very significant for promoting the investigation of the dynamic principle of biomacromolecule action. However, to approach such a subject seriously, a key prerequisite is to unequivocally identify the low-frequency mode of a protein molecule.

In principle the normal-mode calculation method (Wilson, 1939; Itoh & Shimanouchi, 1970; Fanconi *et al.*, 1971) can be used to calculate and analyse vibrational movements in any molecule, but in practice this is unfortunately computationally impossible owing to lack of symmetry in biomacromolecules and limitations on computer size and speed. When discussing the high-frequency vibrations of a molecule, which refer to very small relative displacements and very strong molecular forces (such as covalent bond) between neighbouring individual atoms, one has no alternative but resort to the discrete model. However, for the low-frequency motions in a biomacromolecule, which involve much bigger effective masses and much weaker force constants (Chou & Chen, 1977, 1978) and whose modes can be compared with heart pulsation, accordion-like motion, or something like this type of vibration that involves many atoms and spans a much bigger dimension than the length of a covalent bond, it is not only more convenient but also physically rational to adopt the continuity model as illustrated below.

Heart pulsation and accordion-like motion

Suezaki & Gō (1975) imagined that the observed low frequency (with wave number $\sim 30\,cm^{-1}$) in the native globular proteins originated from the mode of heart pulsation. They further compared a protein molecule with a continuous elastic sphere imitating heartbeat breathing motion, and used the formula in earth dynamics:

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

to calculate the low-frequency motions for globular proteins. In eqn. (1), ν is the fundamental frequency of the system concerned, $\tilde{\nu}$ the corresponding wave number, c the speed of light in vacuum, and E, ρ and r are the Young's modulus, mass density and radius of the globular protein respectively. The model of eqn. (1) does not touch on any internal conformation, and hence the change in microenvironment inside a protein molecule cannot be reflected through such a formulation. However, observations on α-chymotrypsin and pepsin indicated that the corresponding low-frequency peaks would immediately disappear after they were denatured (Brown *et al.*, 1972), which apparently tells us that this kind of low-frequency motion is very sensitive to the internal conformation of a protein molecule.

Rather than taking the whole molecule as a continuous elastic sphere, Chou (1983) paid more attention to the internal structure of a protein molecule. Special consideration has been made for the α-helix, the most fundamental structure element (Richardson, 1981) in protein molecules. It should

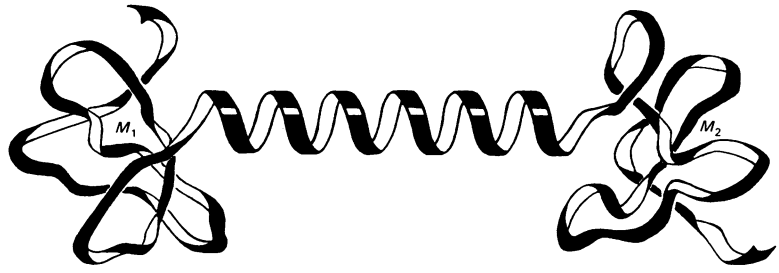

Fig. 1. *Accordion-like vibration system of an α-helix whose two ends are linked to two fragments with mass M_1 and M_2 respectively*

be pointed out that in the paper by Chou (1983) the formula was derived on such a model that the two ends of an α-helix are actually linked to two 'rigid walls' by two mass-negligible springs. However, for some α-helices in proteins a different terminal condition would be more realistic, depending on their actual microenvironment in protein molecules (see below). For instance, the illustration in Fig. 1 represents another type of model encountered quite often for accordion-like vibrations in protein molecules. As is seen there, the α-helix is compared with a mass-distributed spring, but its two ends are linked via a covalent bond to two fragments with masses M_1 and M_2 respectively.

For the vibration system illustrated in Fig. 1, using an almost parallel method to the previous one (Chou, 1983), the formula for calculating the fundamental frequency (lowest frequency) can be derived as follows:

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

where k is the stretching force constant of the helix, ρ the mass per unit length along the axis of the helix, L the length of the helix axis, and:

$$\beta_1 = M_2/(M_1+M_2), \qquad \beta_2 = M_1/(M_1+M_2) \quad (3)$$

$$M^* = M_1M_2/(M_1+M_2) \qquad (4)$$

Note that M_1 and M_2 in eqns. (3)–(4) can also represent the masses of any two segments of a sequence which attach to the two ends of the helix and have some free space to vibrate along the helix axis, undergoing very weak constraints from the other joint parts of the sequence so that they can be approximately treated as two mass fragments as illustrated in Fig. 1. Generally speaking, if the two ends of an α-helix are linked to those portions of a protein molecule which possess relatively much larger masses and are solidly packed as well, such as β-barrel, the previous model (Chou, 1983) is more suitable; however, when the two ends of an α-helix are linked to some short fragments (see the example of insulin below) or some rather flexible segments (see the example of lysozyme below), the model given in Fig. 1 will be more appropriate.

The stretching force constant, k, in eqn. (2) is essentially related to the constituent hydrogen bonds of an α-helix. 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. Consequently, an α-helix consisting of n amino acid residues generally has $n-4$ hydrogen bonds. Recently, a novel method was developed by which one can find the stretching force constant of an α-helix according to its constituent hydrogen bonds (Chou, 1983). By using induction as well as such a method, the stretching force constant of an α-helix with n amino-acid residues can be generally expressed as:

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

where:

$$\left.\begin{aligned} i &= \left[\!\left[\frac{n-4}{11} \right]\!\right] \\ j &= (n-4) - 11i \end{aligned}\right\} \text{(Note } n\text{, the number of amino acid residues, must be } \ge 5) \qquad (6)$$

are the integral part of the quotient $(n-4)/11$ and the corresponding remainder respectively, and:

$$k^* = \sqrt{(k_H^S \cos\theta)^2 + (k_H^B \sin\theta)^2} = 12\,\text{N/m}\ (0.12\,\text{mdyn/Å}) \qquad (7)$$

where $\theta \simeq 26°$ is the angle between the helix axis and the constituent hydrogen bonds, and $K_H^S = 13\,N/m$ (0.13 mdyn/Å) (Itoh & Shimanouchi, 1970) and $K_H^B = 3\,N/m$ (0.03 mdyn/Å) (Itoh & Shimanouchi, 1970) are the stretching and binding force constants of the hydrogen bond respectively. By means of eqns. (5)–(7), the stretching force constant of an α-helix with any number ($n \geqslant 5$) of constituent amino acid residues can be very easily worked out.

Both the heart-pulsation model (Suezaki & Gō, 1975) and the accordion-like model (Chou, 1983) have been employed to calculate the low-frequency wave number of α-chymotrypsin, and the results seem both to be in good agreement with the observation. We now examine the application to other protein molecules.

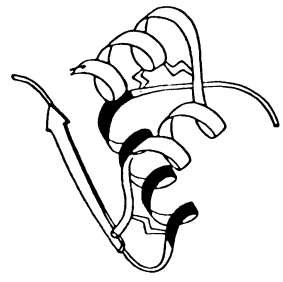

Fig. 2. *Schematic drawing of insulin molecule [after Richardson (1981) with permission], where, however, the principal helix (residues B9–B19) is shaded in*

Insulin

According to the study of Raman spectra of proteins, the low-frequency wave number of 22 cm^{-1} (Painter *et al.*, 1982) has been observed for both insulin monomer (M_r 5800) and dimer (M_r 11 600). Such a phenomenon can not be explained in terms of the model of eqn. (1), according to which it is obvious that the bigger the protein molecule is, the smaller the corresponding low-frequency wave number should be. Therefore, it is most likely that the observed low-frequency vibration originates from some special internal structure that exists in both monomer and dimer insulin. As is well known, an insulin molecule consists of A-chain and B-chain, each having 21 and 30 residues respectively (Blundell *et al.*, 1972). The A-chain consists of two helical regions, A2–A8 and A13–A20, that run very nearly antiparallel and are joined by a stretch of extended polypeptide chain from A9 to A12 (Fig. 2). There is a central piece of α-helix in the B-chain stretching from B9 to B19 from which both *N*- and *C*-terminal residues extend. Among these three helices, the longest helix (B9–B19) is most likely associated with the observed lowest-frequency peak of 22 cm^{-1}, since the fundamental frequency will generally be lower if generated by an α-helix with more residues (see eqns. 2, 5 and 6). Besides, the other two short helices (A2–A8 and A13–A20) are also rather distorted, and actually belong to irregular helix structure. Consequently, attention should be focused on the helix (B9–B19), which is termed the '*principal helix*' and is shown in the shaded portion of Fig. 2. The amino acid sequence of this principal helix is (Blundell *et al.*, 1972):

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

which contains eleven amino acid residues, i.e. $n = 11$, and hence, according to eqns. (5)–(7), we obtain the stretching force constant:

$$k = 12k^*/7 = 20\,N/m\ (0.20\,mdyn/Å) \qquad (8)$$

Besides, to a fair approximation, ρ, the mass per unit length of the helix, can be assumed to be uniform (Chou, 1983), thus ρL is actually the total mass of the helix that can be easily calculated by simply adding the masses of all the constituent residues together; i.e.

$$\rho L = 1228\,a.m.u. = 1228\,g/N \qquad (9)$$

where a.m.u. denotes the atomic mass units, and N the Avogadro constant.

The residues B1–B8 and residues B20–B30 that attach to the two ends of the above helix and can be approximately regarded 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 (Blundell *et al.*, 1972). Therefore, according to eqns. (3) and (4), we have:

$$M^* = 540\,a.m.u. = 540\,g/N, \quad \beta_1 = 0.60, \quad \beta_2 = 0.40 \qquad (10)$$

Substituting eqns. (8)–(10) into eqn. (2), we obtain:

$$\tilde{\nu} = \frac{1}{2\pi \times 3 \times 10^{10}} \sqrt{\frac{0.20 \times 10^5 \times 6.023 \times 10^{23}}{540 + \{[(0.60)^3 + (0.40)^3] \times 1228/3\}}} = 22.7\,cm^{-1} \qquad (11)$$

which is exactly the same as the result observed for insulin monomer and dimer (Painter *et al.*, 1982).

Lysozyme

The M_r of lysozyme is 14000, more than twice that of insulin monomer. But the observed low-

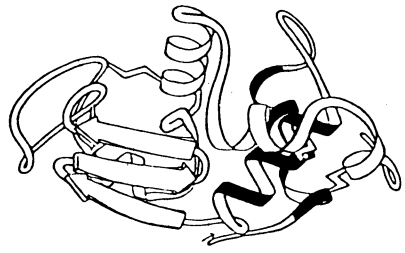

Fig. 3. *Schematic drawing of lysozyme molecule [after Richardson (1981) with permission], where, however, the two principal helices (residues 5–15 and 25–35) are shaded in*

frequency wave number is 25 cm^{-1} (Genzel *et al.*, 1976), 3 cm^{-1} higher than the observed value for insulin. This is once again an example in which it is difficult to explain if the model of eqn. (1) is adopted. However, this phenomenon can also be elucidated from the internal structure of the biomacromolecule. Lysozyme contains 129 residues. Those structures of polypeptide chain which adopt a helical conformation are residues 5–15, 25–35, 80–84, and 89–96 (Protein Data Bank, 1981). Among these four helices, the last two are too short, hence only helix 1 (residues 5–15) and helix 2 (residues 25–35) may generate low frequencies and can be deemed as the principal helices (Fig. 3).

The amino acid sequence of the helix is (Protein Data Bank, 1981):

Arg-Cys-Glu-Leu-Ala-Ala-Ala-Met-Lys-Arg-His
(5) (15)

and the two ends of the helix are linked by covalent bonds to the sequential segments

Lys-Val-Phe-Gly
(1) (4)

and

Gly-Leu-Asp-Asn-Tyr-Arg-Gly-Tyr-Ser
(16) (24)

respectively. Both of these two segments have some free space for vibrating along the helix axis (Protein Data Bank, 1981; Blake *et al.*, 1965; Imoto *et al.*, 1972), and therefore can be approximately treated as two fragments, attached to the two ends of the helix, as illustrated in Fig. 1. However, for helix 2, whose amino acid sequence is (Protein Data Bank, 1981):

Leu-Gly-Asn-Trp-Val-Cys-Ala-Ala-Lys-Phe-Glu
(25) (35)

the corresponding sequential segments that can be approximately treated as two attached fragments will become

Gly-Leu-Asp-Asn-Tyr-Arg-Gly-Tyr-Ser
(16) (24)

and

Ser-Asn-Phe-Asn-Thr
(36) (40)

respectively (Protein Data Bank, 1981; Blake *et al.*, 1965; Imoto *et al.*, 1972). Like the principal helix in insulin, here both helix 1 (residues 5–15) and helix 2 (residues 25–35) contain eleven amino acid residues. Consequently, the result given in eqn. (8) can also represent their stretching force constant.

Following the calculation method used for insulin, we immediately obtain from eqn. (2):

$$\tilde{\nu} = \begin{cases} 27.0\ cm^{-1} & \text{for helix 1} \\ 26.2\ cm^{-1} & \text{for helix 2} \end{cases} \quad (12)$$

results which are very close to the observed value (Genzel *et al.*, 1979) of 25 cm^{-1}. Consequently, the source of the observed low-frequency vibration is most likely the principal helices 1 and 2 of lysozyme plus their microenvironments. The derivation of about 1–2 cm^{-1} from the observed value is probably due to the approximate treatment related to the mass fragments at the two ends of the principal helices.

Low-frequency amplitude

Following almost the same derivation as given in the previous paper (Chou, 1983), the low-frequency amplitude for the model of Fig. 1 can be expressed as:

$$\sigma = \sqrt{\frac{2h\nu/(e^{h\nu/k_BT}-1)}{k}}\Bigg|_{h\nu \ll k_BT} \simeq \sqrt{\frac{2k_BT}{k}} \quad (13)$$

where k_B is the Boltzmann constant, h the Planck constant, and T the absolute temperature. Note that for low-frequency phonons whose wave numbers are less than 50 cm^{-1}, we have $h\nu \ll k_BT$ at room temperature. Substitution of eqn. (8) and $T = 300$ K into eqn. (13) will give $\sigma = 0.02$ nm (0.2 Å), the amplitude of the accordion-like vibration of the principal helices considered above in insulin and lysozyme molecules at room temperature. Such an estimation concerning the amplitude is in good agreement with the relevant result given by Peticolas (1979).

A discussion on the 'frictional' effects was already given in the previous paper (Chou, 1983), and therefore is not repeated here.

Conclusion

The low-frequency vibrations observed for insulin and lysozyme do not originate in the pulsation of the whole protein molecule, but rather more likely from the accordion-like motion of the internal principal

helices that generally contain more than ten amino acid residues. The calculations further indicate that the internal microenvironment also exerts an important influence upon the low-frequency motion, as reflected by the terminal conditions of the principal helices. For instance, both the principal helix in α-chymotrypsin (Chou, 1983) and the principal helix in insulin contain eleven amino acid residues, and their masses are almost also the same (1242 and 1228 a.m.u. respectively). But for α-chymotrypsin, one end of its principal helix is fixed by a salt bridge to the β-barrel 1 and the other end is linked to a segment which is buried inside the molecule (Chou, 1983). And for insulin, the two ends of its principal helix are linked to two short movable fragments. Because of such a difference in terminal condition, a different model and formula should be adopted here, thereby resulting in a quite different value for the wave number. This once again demonstrates that this kind of low-frequency motion is closely related to the conformation of a biomacromolecule (Brown *et al.*, 1972), and hence to its biological function as well (Chou, 1983).

Besides, it is also very important to realize that a protein molecule containing many helices does not necessarily generate an outstanding low-frequency peak if they interfere with each other in an unfavourable way or are located in some micro-environment where the accordion-like motions are severely blocked. Conversely, protein molecules containing only a few helices do sometimes generate outstanding low-frequency peaks if their principal helices are situated in a favourable micro-environment, as illustrated for insulin molecule here as well as for α-chymotrypsin previously (Chou, 1983).

References

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Philips, D. C. & Sarma, V. R. (1965) *Nature (London)* **206**, 757–761

Blundell, T., Dodson, G., Hodgkin, D. & Mercola, D. (1972) *Adv. Protein Chem.* **26**, 279–402

Brown, K. C., Erfurth, S. C., Small, E. W. & Peticolas, W. L. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1467–1469

Careri, G., Fasella, P. & Gratton, E. (1975) *CRC Crit. Rev. Biochem.* **3**, 141–164

Chou, K.-C. (1983) *Biochem. J.* **209**, 573–580

Chou, K.-C. & Chen, N. Y. (1977) *Sci. Sin. (Engl. Ed.)* **20**, 447–457

Chou, K.-C. & Chen, N. Y. (1978) *Huaxue Tongbao (Chin. Ed.)* no. 5, 29–34

Chou, K.-C., Chen, N. Y. & Forsèn, S. (1981) *Chem. Scr.* **18**, 126–132

Englander, S. W. (1980) *Comments Mol. Cell. Biophys.* **1**, 15–28

Evans, G. J., Evans, M. W. & Pething, R. (1982) *Spectrochim. Acta* **38A**, 421–422

Fanconi, B., Small, E. W. & Peticolas, W. L. (1971) *Biopolymers* **10**, 1277–1298

Fröhlick, H. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4211–4215

Genzel, L., Keilman, F., Martin, T. P., Winterling, G. Yacoby, Y., Fröhlich, H. & Makinen, M. W. (1976) *Bipolymers* **15**, 219–225

Green, D. E. (1974) *Ann. N.Y. Acad. Sci.* **227**, 6–45

Imoto, T., Johnson, L. N., North, A. C. T., Philips, D. C. & Rupley, J. A. (1972) *Enzyme 3rd Ed.* **7**, 665–868

Itoh, K. & Shimanouchi, T. (1970) *Biopolymers* **9**, 383–399

Ji, S. (1974) *Ann. N.Y. Acad. Sci.* **227**, 419–437

Painter, P. C., Mosher, L. E. & Rhoads, C. (1982) *Biopolymers* **21**, 1469–1472

Peticolas, W. L. (1979) *Biopolymers* **18**, 747–755

Protein Data Bank (1981) Brookhaven National Laboratory, Upton, NY

Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167–339

Sobell, H. M., Lozansky, E. D. & Lessen, M. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 11–19

Sobell, H. M., Banerjee, A., Lozansky, E. D., Zhou, G. P. & Chou, K. C. (1982) *J. Mol. Sci. (Engl. Ed.)* **2**, 113–128

Suezaki, Y. & Gō, N. (1975) *Int. J. Peptide Protein Res.* **7**, 333–334

Wilson, E. B. (1939) *J. Chem. Phys.* **7**, 1047–1052

Zhou, G. P. (1981) *Shengwuhuaxue Yu Shengwuwulin Jinzhan (Chin. Ed.)* no. 5, 19–22