Haworth, R. A., Hunter, D. R., & Berkoff, H. A. (1980) *FEBS Lett. 110*, 216-218.
Hughes, B. P., & Barritt, G. J. (1978) *Biochem. J. 176*, 295-303.
Irvine, R. F., Letcher, A. J., Lander, D. J., & Downes, C. P. (1984) *Biochem. J. 225*, 237-243.
Jansen, C. B., Sistare, F. D., Hamman, H. C., & Haynes, R. C. (1983) *Biochem. J. 210*, 819-827.
Joseph, S. K., Thomas, A. P., Williams, R. J., Irvine, R. F., & Williamson, J. R. (1984) *J. Biol. Chem. 259*, 3077-3082.
Marshall, S. E., McCormack, J. G., & Denton, R. M. (1984) *Biochem. J. 218*, 249-260.
McCormack, J. G. (1985) *Biochem. J. 231*, 597-608.
McCormack, J. G., & England, P. J. (1984) *Biochem. J. 214*, 581-585.
Nedergaard, J. (1984) *Eur. J. Biochem. 144*, 159-168.
Nicholls, D. G. (1974) *Eur. J. Biochem. 50*, 305-315.
Nicholls, D. G. (1978) *Biochem. J. 176*, 463-474.
Oviasu, O. A., & Whitton, P. P. (1984) *Biochem. J. 224*, 181-186.
Puskin, J. S., Gunter, T. E., Gunter, K. K., & Russell, P. R. (1976) *Biochemistry 15*, 3834-3842.
Reinhart, P. F., Taylor, W. M., & Bygrave, F. L. (1982) *J. Biol. Chem. 257*, 1906-1912.
Reinhart, P. F., Taylor, W. M., & Bygrave, F. L. (1984) *Biochem. J. 223*, 1-13.
Rottenberg, H. (1984) *J. Membr. Biol. 81*, 127-138.
Scholz, H. (1980) *Handb. Exp. Pharmacol. 54*, 651-733.
Sies, H., Graf, P., & Crane, D. (1983) *Biochem. J. 212*, 271-278.
Siess, E. A., Fahimi, F. M., & Weiland, O. H. (1981) *Hoppe-Seyler's Z. Physiol. Chem. 362*, 1643-1651.
Studer, R. K., & Borle, A. B. (1982) *J. Biol. Chem. 257*, 7987-7993.
Sugano, R., Shiota, M., Khono, H., Shimada, M., & Oshino, N. (1980) *J. Biochem. (Tokyo) 87*, 465-472.
Taylor, W. M., Prpic, V., Exton, J. H., & Bygrave, F. L. (1980) *Biochem. J. 188*, 443-450.
Taylor, W. M., Van de Pol, E., & Bygrave, F. L. (1986) *Biochem. J. 233*, 321-324.
Thomas, A. P., Alexander, J., & Williamson, J. R. (1984) *J. Biol. Chem. 259*, 5574-5584.
Titheridge, M. A., & Coore, H. G. (1976) *FEBS Lett. 71*, 73-78.
Titheridge, M. A., Slinger, J. L., & Haynes, R. C. (1979) *Eur. J. Biochem. 102*, 117-124.
Vaghy, P. L., Johnson, D. J., Matlib, M., Wang, T., & Schwarz, A. (1982) *J. Biol. Chem. 257*, 6000-6002.
Wingrove, D. E., Amatruda, J. M., & Gunter, T. E. (1984) *J. Biol. Chem. 259*, 9390-9344.
Yamazaki, R. K., Mickey, D. L., & Storey, M. (1980) *Biochim. Biophys. Acta 592*, 1-12.

# Thermal Expansion of a Protein

Hans Frauenfelder,[‡] Hermann Hartmann,[§] Martin Karplus,[∥] I. D. Kuntz, Jr.,[⊥] John Kuriyan,[∥,#] Fritz Parak,[§] Gregory A. Petsko,[*,#] Dagmar Ringe,[#] Robert F. Tilton, Jr.,[⊥,#] Michael L. Connolly,[△] and Nelson Max[×]

*Department of Physics, University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois 61801, Department of Physics, Technische Universität, Munich, West Germany, Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, Department of Pharmaceutical Chemistry, University of California at San Francisco, School of Pharmacy, San Francisco, California 94143, Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037, and Lawrence Livermore National Laboratory, University of California, Livermore, Livermore, California 94550*

*Received May 28, 1986; Revised Manuscript Received September 17, 1986*

ABSTRACT: The thermal expansion of a protein, metmyoglobin, was investigated by analysis of the refined X-ray crystal structures at 80 and 255-300 K. On heating from 80 to 300 K, the volume occupied by myoglobin increases by approximately 3%. The linear thermal expansion coefficient is estimated to be 115 $\times 10^{-6}$ $K^{-1}$. This value is more than twice as large as that of liquid water but less than that of benzene. As the temperature is raised, the internal volume change does not come from the large, atom-sized internal cavities in the structure but from an increase in the small, subatomic free volumes between atoms. The largest expansion occurs in the region of the CD and GH corners; both these regions move away from the center of the protein. The remainder of the expansion results from the lengthening of contacts between segments of secondary structure.

In any condensed medium, there is a dynamic balance between the cohesive forces holding atoms or molecules together and the thermally driven random motion of these atoms. At higher temperatures, the availability of more thermal energy implies greater motion, which usually increases the average distance between atoms. In a protein, unlike a simple atomic solid or liquid, both structure and motion are highly irregular.[1]

* Correspondence should be addressed to this author.
‡ University of Illinois at Urbana-Champaign.
§ Technische Universität.
∥ Harvard University.
⊥ University of California at San Francisco.
# Massachusetts Institute of Technology.
△ Research Institute of Scripps Clinic.
× University of California, Livermore.

[1] Proteins are dynamic systems, and their function depends not only on their average structures but also on fluctuations about that average structure. General reviews of protein dynamics are given by Careri et al. (1975), Gurd and Rothgeb (1979), Debrunner and Frauenfelder (1982), Karplus and McCammon (1983, 1986), and Petsko and Ringe (1984).

0006-2960/87/0426-0254$01.50/0 

Thus, the thermal expansion coefficient of a protein molecule is expected to be anisotropic, and an analysis of the local variations in this quantity should reflect the distribution of interactions that stabilize the average structure and govern the fluctuations. One approach to such an analysis is to examine the structure and dynamics of a protein at two different temperatures.

Recently, X-ray diffraction measurements on the crystalline protein sperm whale metmyoglobin were described at 80 and 300 K, and the changes in the mean-square displacements that occur on cooling were reported (Hartmann et al., 1982). Both the unit cell volume of the crystal and the volume of the myoglobin molecule itself were observed to be larger at 300 K than at 80 K. The availability of refined crystal structures for a protein at temperatures 220 K apart provided the opportunity to analyze the volume and structural changes that result from a large change in temperature. Walter et al. (1982) in a report comparing the structures of trypsinogen at 103 and 173 K have remarked that the protein seems to expand anisotropically, but they do not analyze the thermal expansion in any detail.

In this paper, the analysis is performed in three stages. First, the average thermal expansion coefficient is estimated. This value is compared with that of simple substances such as water and benzene. Second, the anisotropy of the expansion is examined by calculating distances between atoms in different segments of secondary structure and analyzing the changes in the internal regions of the molecule. Third, the expansion of the unit cell is described in the context of the changes in the intermolecular crystal contacts. Correlations between these changes in the molecular packing and changes in the protein structure are discussed.

## Experimental Procedures

Sperm whale myoglobin, a small (molecular weight 17 500; 153 amino acid residues), mostly $\alpha$-helical, oxygen-storage protein[2] containing a heme prosthetic group, crystallizes readily in the met form from 75% saturated ammonium sulfate solutions in the form of dark red plates or rhombs. The crystals have the symmetry of space group $P2_1$ with one myoglobin molecule per asymmetric unit. Four complete, independent, X-ray data sets to 2.0-Å resolution were collected in the temperature range of 255–300 K (one at 255 K, two at 290 K, and one at 300 K) on a Nicolet $P2_1$ diffractometer in the conventional manner (Frauenfelder et al., 1979). One data set to 2.0-Å resolution was collected photographically at 80 K (Hartmann et al., 1982). Measurement at 80 K without employment of a cryosolvent required cooling the crystal below the freezing point of its mother liquor by a "flash freezing" technique that has been described earlier (Hartmann et al., 1982).

The metmyoglobin model of Takano (1977) was used as a starting point for refinement against both the 300 K and the 80 K data sets, by a restrained-parameter least-squares method (Konnert & Hendrickson, 1980; Hendrickson, 1985). To estimate the errors in these refined coordinates, the newly refined 300 K model was used as the starting point for refinements against the three independent data sets at 290 and 255 K. Because only one data set was available at 80 K, a different myoglobin structure, that of (carbon monoxy)myoglobin at 260 K (Kuriyan et al., 1986), was used as the starting point for a second refinement against the 80 K data set. Thus, four structures in the temperature range of 255–300 K were obtained from the refinements against four independent data sets, and two structures at 80 K were obtained by refinements with different starting models, but against the same data set. The structures in the 255–300 K temperature range will be referred to as RT ("room temperature") structures, and the two 80 K structures will be referred to as LT ("low-temperature") structures. The two LT structures were refined to $R$ factors of 20% and the four RT structures to $R$ factors of 18.5%. The root mean square deviation from ideality of bond lengths was less than 0.03 Å in all the structures.

Comparison of the various LT and RT structures allows a rough estimate of the errors in the atomic coordinates to be made (Kuriyan, 1986). The standard deviations in the positions of backbone and side-chain atoms are estimated to be 0.1 and 0.3 Å, respectively, in the RT structures and 0.2 and 0.6 Å, respectively, in the LT structures. Examination of electron density maps indicated that about 15 surface side chains are disordered in the 80 K structure, which is one reason for the large positional uncertainty. The LT structures are also less precise because the data were measured photographically and fewer reflections were included in the refinement (Hartmann et al., 1983). These errors suggest that differences in backbone distances (between an LT and an RT structure) that are less than 0.20 Å are not likely to be significant. In the calculation of expansion coefficients and other quantities, the results are averaged over the eight possible pairs of RT and LT structures to further reduce errors.

## Results and Discussion

As noted previously (Hartmann et al., 1982), the unit cell volume of metmyoglobin increases by 5% on going from 80 to 255–300 K. The average unit cell parameters are $a = 64.46 \pm 0.09$ Å, $b = 30.95 \pm 0.06$ Å, $c = 34.83 \pm 0.02$ Å, $\beta = 105.82° \pm 0.27°$, $V = 66\,860 \pm 230$ Å$^3$, for the 255–300 K crystals. The measured unit cell parameters at 80 K are $a = 63.44$ Å, $b = 30.45$ Å, $c = 34.05$ Å, $\beta = 105.60°$, and $V = 63\,350$ Å$^3$. The parameters for the 80 K structure have been derived from photographic measurements on a single crystal, and their standard deviations have not been estimated.

Overall, the structure of the protein is similar at the two temperatures; no gross conformational changes have occurred. The heme orientation is unaltered over the temperature range studied, and there is also no change in the position of the iron atom relative to the heme plane. Considerable bound solvent is visible at 80 K and, to a lesser extent, at 255–300 K. Inclusion of these bound molecules in the refinement is necessary for accurate determination of atomic positions for surface amino acids.

In addition to the expansion of the unit cell, there is an increase in the size of the protein itself. Analysis of this increase was carried out, first, by determining the overall thermal expansion coefficient, second, by analyzing the changes in the relative packing of the secondary structural elements and the volume of the internal cavities, and, finally, by examining the effects of changes in the unit cell on crystal packing.

*Coefficient of Expansion.* The X-ray structure of a protein gives average coordinates, $\bar{r}_i(T)$, of each non-hydrogen atom. If the coordinates are measured at two different temperatures, $T_1$ and $T_2$, expansion coefficients, $\alpha_{ij}(T_1,T_2)$, which express the fractional change in distance between two atoms per degree kelvin, can be defined through the equation:

$$\alpha_{ij}(T_1,T_2) = \frac{r_{ij}(T_2) - r_{ij}(T_1)}{(T_2 - T_1)\langle r_{ij}\rangle} \quad (1)$$

[2] The molecule may be approximated by a prolate ellipsoid of radii 19 × 10 × 10 Å.

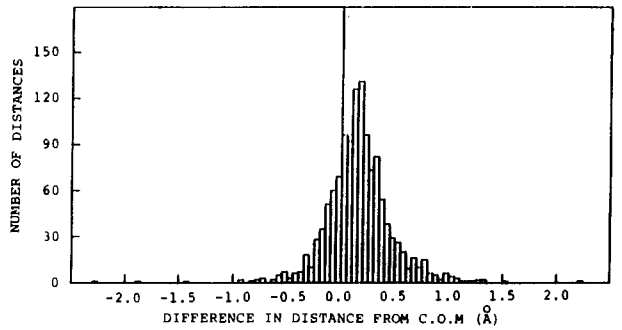


FIGURE 1: Distance of every atom from the center of mass in the structure of myoglobin at 80 K subtracted from the corresponding distance at 300 K plotted as a histogram with 0.3-Å bins.

Here, $r_{ij}(T)$ is the distance between the positions of atoms $i$ and $j$ at temperature $T$, and $\langle r_{ij} \rangle$ is the average distance. The coefficients $\alpha_{ij}(T_1,T_2)$ can be expected to vary with $i$ and $j$. For sperm whale metmyoglobin, which has 1261 non-hydrogen atoms, the X-ray structure at two temperatures yields about $8 \times 10^5$ values of $\alpha_{ij}$. The analysis of the $\alpha_{ij}$ is correspondingly complex but can be simplified in various ways.

An overall linear thermal expansion coefficient was calculated by computing $\alpha_{ij}$(LT,RT) according to eq 1, taking $i$ to be the center of mass of the protein and allowing $j$ to vary over all the atoms and then averaging to give $\langle \alpha(\text{LT,RT}) \rangle$. Figure 1 shows a histogram of the average difference in distance at RT and LT of every atom in the myoglobin molecule from the center of mass. If there were no thermal expansion, the envelope would be symmetric about zero. Calculation of the same graph for the independently refined structures of myoglobin at RT shows precisely such a symmetric distribution for this control case. The small, systematic, positive deviation observed on going from LT to RT indicates that the calculated value for the thermal expansion coefficient of myoglobin reflects an overall increase in the volume of the protein. The average difference in distance from the center of mass is +0.16 Å, while the average of the absolute values of the differences is 0.26 Å. Calculating the linear expansion coefficient, $\alpha_{ij}$ (eq 1), for all atoms $j$ where $i$ is the center of mass, and averaging, yields $\langle \alpha(\text{LT,RT}) \rangle = 50 \times 10^{-6}$ K$^{-1}$.

The $\langle \alpha \rangle$ obtained from averaging the values from eq 1 cannot be compared directly to standard values for simple substances because these values usually refer to a particular temperature, while the calculated $\langle \alpha \rangle$ has been obtained by comparing LT and RT distances. To compensate for this, the solid-state approximation is made (Kittel, 1971; Ashcroft & Mermin, 1976), namely, that $\langle \alpha \rangle$ in proteins is proportional to $T^3$ from absolute zero up to the Debye temperature ($T_D$) and increases linearly with $T$ above this value. Taking into account only the contribution from the Debye temperature to 300 K gives the approximate relation

$$\langle \alpha(T_2) \rangle \approx \frac{2(T_2 - T_1)}{T_2 - T_D} \frac{T_2}{T_2 + T_D} \langle \alpha(T_1,T_2) \rangle \qquad (2)$$

The appropriate value of the Debye temperature $T_D$ for a protein is not known. Mossbauer experiments imply a value between 160 and 200 K (Parak et al., 1981). Arbitrarily taking $T_D = 180$ K, and with $T_1 = 80$ K and $T_2 = 300$ K, the correction factor is equal to 2.3 so that $\langle \alpha(300\text{ K}) \rangle = 115 \times 10^{-6}$ K$^{-1}$. Alternatively, if $\langle \alpha(300\text{ K}) \rangle$ is calculated using *all* interatomic distances, a value of $160 \times 10^{-6}$ K$^{-1}$ is obtained. These values may be compared with the linear expansion coefficient of $140 \times 10^{-6}$ K$^{-1}$ calculated from the volume expansion coefficient published for bovine metmyoglobin.[3]

Table I: Linear Thermal Expansion Coefficients for Various Substances

| substance | $\alpha$ ($\times 10^{-6}$ K$^{-1}$) | substance | $\alpha$ ($\times 10^{-6}$ K$^{-1}$) |
|---|---|---|---|
| water (ice) | 5 | water (liquid) | 70 |
| glass | 8 | myoglobin | 115 |
| concrete | 12 | glycerol | 167 |
| iron | 12 | gasoline | 317 |
| copper | 18 | ethanol | 373 |
| aluminum | 24 | benzene | 410 |
| mercury | 60 | | |

Another way of calculating an average linear thermal expansion coefficient is to consider distances between the backbone $C_\alpha$ atoms, which are generally the best determined atoms in the structure. The value of $\langle \alpha(\text{LT,RT}) \rangle$ obtained in this way ($49 \times 10^{-6}$ K$^{-1}$) is the same as that obtained from all atoms. An interesting feature of the $C_\alpha$–$C_\alpha$ distances is that the thermal expansion coefficient is uniform over various length scales. Considering $C_\alpha$–$C_\alpha$ distances in groups that lie between 5–10, 10–15, 15–20, 20–25, and 25–30 Å, the values of $\langle \alpha(\text{LT,RT}) \rangle$ obtained are $46 \pm 3$, $49 \pm 2$, $43 \pm 3$, $45 \pm 5$, $49 \pm 4$ ($\times 10^{-6}$ K$^{-1}$), respectively. (The values are averages over the eight pairs of LT and RT structures, $\pm 1$ standard deviation.) Scatter plots for $C_\alpha$–$C_\alpha$ distances are given in Figure 2 and, it can be seen that $\alpha$ is approximately constant; i.e., the increase in interatomic distances is largest for large distances.

In Table I, values are given for the linear expansion coefficient for a number of simple substances. The calculated expansion coefficient of the protein myoglobin is between that of liquid water and hydrocarbons. This result is interesting since the forces that stabilize the protein and determine the protein expansion behavior include both hydrogen-bonding and hydrophobic interactions.

*Local Variations in* $\alpha$. To explore the atomic movements that occur as a function of temperature, the local behavior of the expansion coefficient $\alpha$ is examined. Only a small correlation is observed between the change in distance from the center of mass for any $C_\alpha$ atom and its distance from the center of mass (correlation coefficient = 0.23, 153 points, one per residue). Examination of the locally averaged expansion coefficient of a 10-Å sphere of neighboring atoms centered around each $C_\alpha$ atom (Figure 3) reveals that the turns between the helices generally show a larger degree of thermal expansion than observed within the individual helices. An exception is the C–D corner region, which has a small local expansion coefficient but which moves as a unit away from the rest of the protein (see below). Figure 3 also shows the residue-averaged backbone mean-square displacements at 300 K, as determined from the experimental isotropic Debye–Waller factors (Hartmann et al., 1982). The correlation coefficient between the mean-square displacements and the local expansion coefficient is low (0.14).

A more direct analysis of this anisotropic expansion involves comparison of all possible $C_\alpha$ interatomic distances in both structures. Conventional distance matrices (153 × 153, half of each being redundant) for the LT and the RT myoglobin structures contain absolute distances between all the $C_\alpha$ atoms (Phillips, 1970; Kuntz, 1975). To represent relative movement between structures, a "difference distance" matrix is generated

[3] The volume expansion coefficient, $305 \times 10^{-6}$ cm$^3$ g$^{-1}$ K$^{-1}$, may be converted to a mass-independent linear expansion coefficient by dividing by 3 and multiplying by 1.35 g cm$^{-3}$, the density of myoglobin. The volume expansion coefficient is given in the *CRC Handbook of Biochemistry* (1976) and also in Jolicoeur (1981) and Bull and Breese (1973).

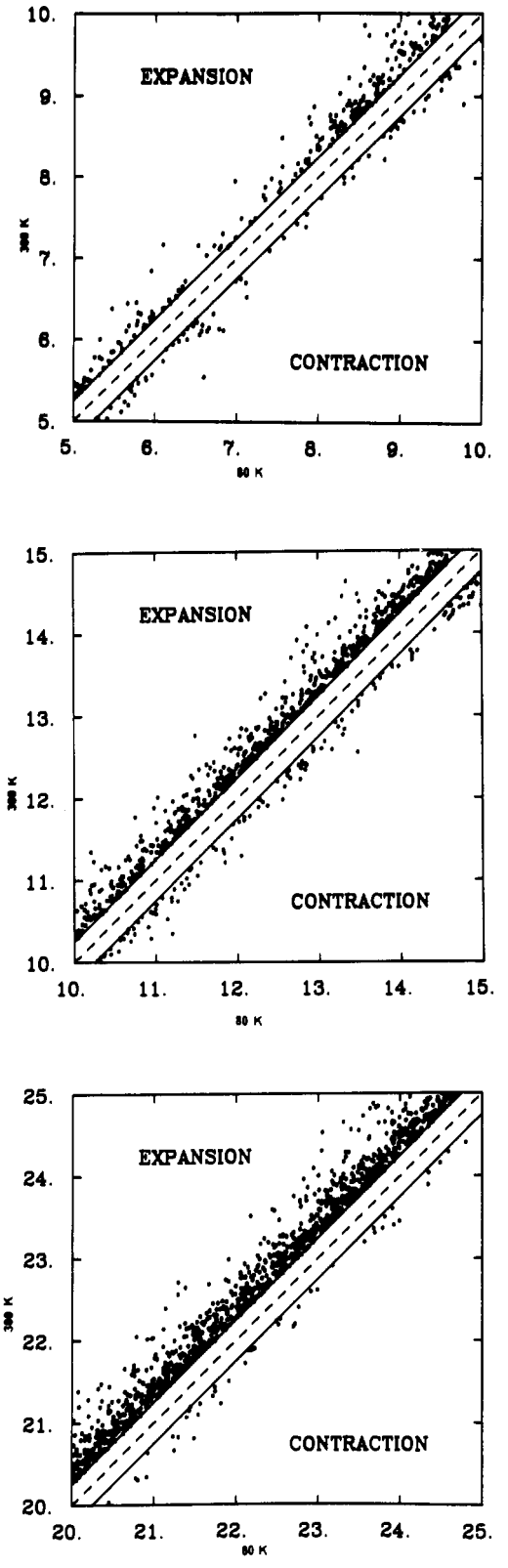


FIGURE 2: Three figures comparing $C_\alpha$–$C_\alpha$ distances in an LT structure with the corresponding distances in an RT structure. All distances which deviate by less than 0.25 Å between RT and LT are excluded in the plots. The comparison is done in three ranges of distances: (top) 5–10 Å; (middle) 10–15 Å; (bottom) 20–25 Å.

as the difference between an LT and an RT distance matrix. Eight such matrices are calculated for each pair of LT and RT structures, and the individual entries are averaged and plotted in Figure 4.

In Figure 4, the open circles below the diagonal of the matrix represent a thermal *contraction* of between 0.25 and 0.5 Å between the two $C_\alpha$ atoms on heating from LT to RT, while those above the diagonal correspond to a net thermal expansion of between 0.25 and 0.5 Å. Crosses correspond to net thermal contraction (below the diagonal) or expansion (above the diagonal) of greater than 0.5 Å (blank spaces denote differences at or below the "noise level" of the analysis, corresponding to a net movement between −0.25 and +0.25 Å). Particular regions of the structure contribute more than others to the overall thermal expansion. The most striking region is the CD corner (residues 37–57), a protruding segment containing a number of positively charged amino acids, which moves as a unit away from the protein center.

A summation of the interatomic distance changes over the secondary structure elements is given in Table II. In a preliminary report (Hartmann et al., 1982), some of the helices were described as undergoing changes in length. The present, more comprehensive, analysis suggests that this effect is not a significant contribution to the total expansion. As summarized in Table II, the smallest changes are within the individual helices, indicating that the helices themselves do not undergo large changes in volume (this is confirmed by radius of gyration calculations for the individual helices, which show that the helix lengths do not change significantly). The bulk of the overall effect is due to the separation of secondary structure units as the temperature increases.

*Radius of Gyration.* The radius of gyration and its three orthogonal components are measures of the size and shape of the molecule and can be used to characterize the expansion. The radii of gyration for the four RT structures and the two LT structures are given in Table IIIA, including all atoms in the calculation. The radius of gyration and its components increase by about 1% on going from LT to RT (this corresponds to a 3% increase in the volume of the corresponding ellipsoid). If the loop regions and the entire CD region of the molecule are excluded from the calculation (Table IIIB), the linear expansion is reduced only slightly, to 0.8%. This indicates that even though the motions of the loops and the CD region are the most noticeable features of the expansion (see Figure 4), the rest of the protein does expand.

A structural picture of the overall expansion of the molecule emerges on calculating the radial distribution of atoms around a central atom in the protein, at both LT and RT. The normalized number density is calculated in shells around the atom closest to the centroid of the protein (the CBB atom of the heme) and averaged separately over the four RT structures and the two LT structures (Figure 5). Both distributions show about four to five peaks at distances from 5 to 15 Å from the central atom, and the distribution is shifted outward by about 0.5 Å in the RT distribution relative to the LT distribution. The differences between these spherically averaged distributions are enough to account for the increase in the radius of gyration, $R_g$, of the protein. $R_g$ for two spherical bodies with mass distributions given by the number densities in Figure 5 is 21.25 Å at LT and 21.46 Å at RT. Because of spherical averaging, these values are not identical with those obtained by using the protein structure; however, the increase of 0.2 Å in $R_g$ is the same as that obtained by using the actual protein coordinates (Table IIIA).

*Surface/Volume Analysis.* Another way to investigate the expansion of the myoglobin molecule is to compute the volume of the protein at low and high temperatures by means of accessibility to a small spherical probe (Lee & Richards, 1978, Connolly; 1983). Accessible surface calculations indicate a 2–3% increase in total molecular surface area and volume of myoglobin on going from 80 to 300 K. The volume change

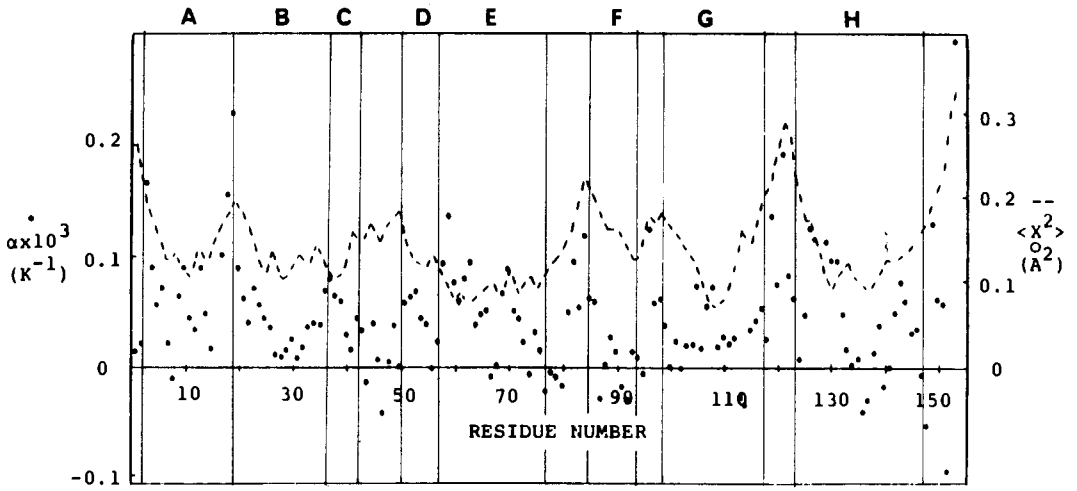


FIGURE 3: Locally averaged thermal expansion coefficients, $\alpha \times 10^3$, plotted as a function of residue number. Each point represents the average of all atoms within 10 Å of the α-carbon for that residue. The dashed curve is a plot of residue-averaged mean-square displacements for the backbone atom of each residue at 300 K (Hartmann et al., 1983).

Table II: Difference Distance Matrix Averages for Secondary Structural Features[a]

| | | A 3–18 | B 20–35 | CD 37–57 | E 59–77 | EF 78–85 | F 86–94 | FG 95–99 | G 100–118 | GH 119–124 | H 125–148 | terminal 149–153 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 3–18 | –0.03 | 0.35 | 0.55 | 0.28 | 0.11 | 0.21 | 0.30 | 0.43 | 0.19 | 0.20 | 0.49 |
| B | 20–35 | | –0.005 | 0.13 | 0.16 | 0.27 | 0.05 | 0.12 | 0.12 | 0.36 | 0.10 | 0.36 |
| CD | 37–57 | | | 0.08 | 0.24 | 0.46 | 0.23 | 0.33 | 0.29 | 0.60 | 0.35 | 0.58 |
| E | 59–77 | | | | 0.10 | 0.21 | 0.08 | 0.05 | 0.26 | 0.36 | 0.14 | 0.44 |
| EF | 78–85 | | | | | 0.11 | 0.18 | 0.10 | 0.29 | 0.32 | 0.13 | 0.39 |
| F | 86–94 | | | | | | –0.04 | –0.15 | 0.01 | 0.34 | 0.01 | 0.27 |
| FG | 95–99 | | | | | | | –0.02 | 0.13 | 0.51 | 0.13 | 0.40 |
| G | 110–118 | | | | | | | | 0.13 | 0.51 | 0.11 | 0.22 |
| GH | 119–124 | | | | | | | | | 0.05 | 0.31 | 0.65 |
| H | 125–148 | | | | | | | | | | 0.03 | 0.26 |
| terminal | 149–153 | | | | | | | | | | | 0.18 |
| | av[b] | 0.28 | 0.18 | 0.35 | 0.21 | 0.23 | 0.11 | 0.17 | 0.23 | 0.35 | 0.15 | 0.33 |

[a] Sums of difference distances over all pair interactions in the regions indicated (in angstroms). [b] Average for each secondary structure element.

Table III

| | $R_g$ (Å) | $R_x$ (Å) | $R_y$ (Å) | $R_z$ (Å) | volume (ellipsoid) (Å$^3$) |
|---|---|---|---|---|---|
| (A) Radius of Gyration for All Atoms | | | | | |
| LT(1), 80 K | 21.15 | 10.87 | 11.79 | 13.79 | |
| LT(2), 80 K | 21.11 | 10.88 | 11.73 | 13.77 | |
| RT(1), 300 K | 21.35 | 10.98 | 11.89 | 13.92 | |
| RT(2), 290 K | 21.35 | 10.97 | 11.90 | 13.92 | |
| RT(3), 290 K | 21.34 | 10.95 | 11.91 | 13.91 | |
| RT(4), 255 K | 21.31 | 10.93 | 11.89 | 13.89 | |
| av(LT) | 21.13 | 10.87 | 11.76 | 13.78 | 7378.6 |
| av(RT) | 21.34 | 10.96 | 11.90 | 13.91 | 7599.3 |
| Δ | 0.21 | 0.09 | 0.14 | 0.13 | |
| (B) Radius of Gyration for All Atoms Except Loops and C–D Corner | | | | | |
| LT(1), 80 K | 19.15 | 10.28 | 10.67 | 12.14 | |
| LT(2), 80 K | 19.10 | 10.21 | 10.66 | 12.11 | |
| RT(1), 300 K | 19.28 | 10.32 | 10.75 | 12.23 | |
| RT(2), 290 K | 19.28 | 10.32 | 10.74 | 12.23 | |
| RT(3), 290 K | 19.27 | 10.32 | 10.74 | 12.23 | |
| RT(4), 255 K | 19.24 | 10.33 | 10.70 | 12.20 | |
| av(LT) | 19.12 | 10.25 | 10.67 | 12.12 | 5552.4 |
| av(RT) | 19.27 | 10.32 | 10.73 | 12.22 | 5668.1 |
| Δ | 0.15 | 0.07 | 0.06 | 0.10 | |

amounts to 400–600 Å$^3$ (Table IV). These results are relatively independent of probe size or assumed van der Waals parameters.

A natural question is the following: Where does the additional space arise? One possibility is an increase in the volume of the local packing defects, i.e., atomic (or larger) sized cavities distributed throughout the protein matrix (Connolly, 1981; Tilton & Kuntz, 1982; Tilton et al., 1984). However, the larger defects, defined with a probe radius of 1.2–1.4 Å, are essentially conserved in both structures, with a combined volume increase of only 30 Å$^3$. While this number represents approximately 10–15% of the volume associated with the large internal cavities, it only accounts for about 5% of the total molecular volume change and hence is a negligible contribution to the expansion. By default, the new volume created as the atoms move apart must be due an increase in the volume of the large number of subatomic spaces remaining when spheres are close-packed, i.e., the "free" volume.

As part of the examination of the large internal cavities, it was found that there is a rather sharp threshold in probe radius below which the large cavities connect to the external surface. This radius depends on the specific atomic van der Waals radii chosen but is on the order of 0.7 Å (Table IV). This probe radius cutoff occurs at a smaller value for the LT structure than for the RT structure, indicating a higher density for the protein at low temperature.

*Intermolecular Contacts.* To examine whether the changes in the unit cell parameters are reflected in the protein structure, the intermolecular crystal contacts between one protein molecule and all its neighbors in the $P2_1$ lattice were calculated by using the low-temperature and high-temperature unit cells. A "contact" is defined as an intermolecular distance that is less than 4.5 Å. In the top panel of Figure 6, the number of such contacts per residue is plotted for one of the RT structures, using the RT unit cell. The effect of the change in unit cell is examined by using the LT unit cell to calculate the intermolecular contacts for the same RT structure; i.e., the

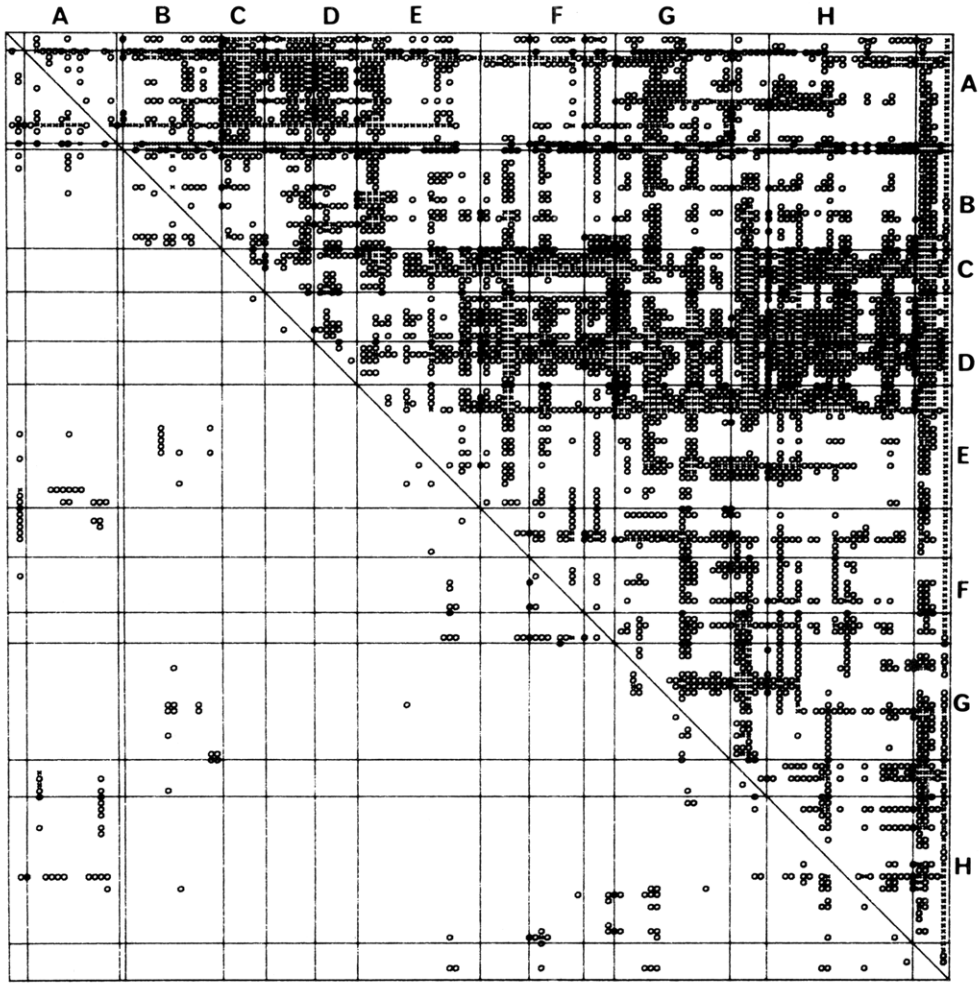


FIGURE 4: $C_\alpha$ difference distance matrix averaged over the eight pairs of LT and RT structures. The entries below the diagonal represent contraction on going from LT to RT; the entries above the diagonal represent expansion on going from LT to RT. The open circles represent changes in $C_\alpha$–$C_\alpha$ distances that are between 0.25 and 0.50 Å, and the crosses represent changes that are greater than 0.50 Å. The helices, A–G, are marked on the figure.

Table IV

(A) Total Surface Area and Volumes Calculated from Molecular Surface of 300 and 80 K Myoglobin

| probe radius (Å) | 300 K surface area (Å²) | 300 K volume (Å³) | 80 K surface area (Å²) | 80 K volume (Å³) | volume(300 K)/ volume(80 K) |
|---|---|---|---|---|---|
| 1.2 | 7068 | 22162 | 6840 | 21592 | 1.026 |
| 1.4 | 6742 | 22423 | 6539 | 21784 | 1.023 |
| 1.6 | 6460 | 22533 | 6323 | 21874 | 1.030 |

(B) Total Surface Area and Volume of Internal Cavities of 300 and 80 K Myoglobin[a]

| probe radius (Å) | 300 K surface area (Å²) | 300 K volume (Å³) | 80 K surface area (Å²) | 80 K volume (Å³) |
|---|---|---|---|---|
| 0.6[b] | 89 | 29 | 101 | 37 |
| 0.7[b] | 228 | 78 | 833 | 429 |
| 0.8[b] | 1025 | 551 | 782 | 384 |
| 1.0[b] | 814 | 456 | 632 | 329 |
| 1.2 | 409 | 235 | 369 | 215 |
| 1.4 | 181 | 109 | 138 | 82 |
| 1.6 | 86 | 52 | *c* | *c* |

[a] The cavity volumes include the distal $O_2$ binding cavity. [b] Individual internal cavities become connected to the exterior surface, making the volume comparisons meaningless with a probe of this radius. [c] No cavities found with the probe radius.

protein is kept rigid while the unit cell is contracted. The resulting increase in the number of contacts is plotted per residue in the middle panel of Figure 6. As expected, most of the residues involved in intermolecular contacts show an increase in the number of contacts, the largest increase being in the A helix and the CD region. However, changes in the structure of the protein between RT and LT affect the intermolecular packing, as shown in the bottom panel of Figure 6. The increase in the number of contacts relative to the RT structure is now plotted by using an LT structure in the LT unit cell. The increase is much less than that predicted by using a rigid protein (middle panel of Figure 6), especially in the A helix and the CD region.

Changes in the protein structure are clearly correlated with changes in the unit cell, in such a way that the reduction in unit cell volume at LT does not result in a large increase in

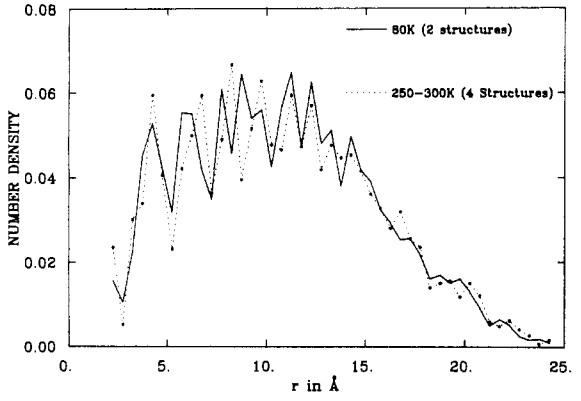


FIGURE 5: Spherically averaged radial distribution of atoms around the CBB atom of the heme calculated and averaged over the two LT structures (solid line) and the four RT structure (dotted line). The normalized number density, defined as the number of protein atoms within a 0.5-Å shell divided by the volume of the shell, is plotted as a function of the distance of the shell from the origin.

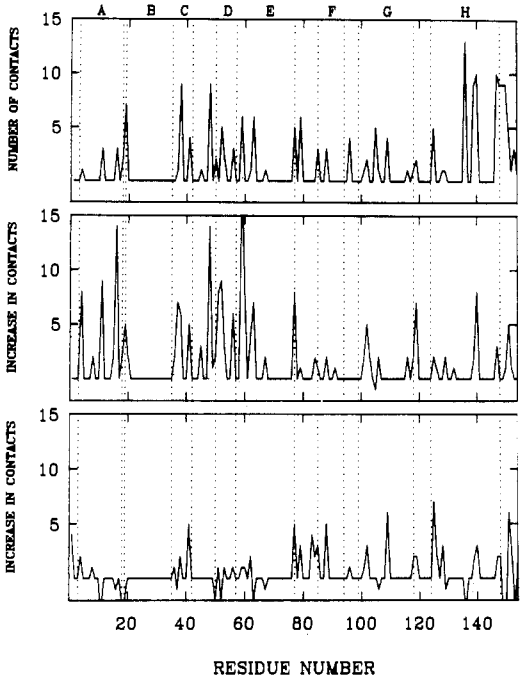


FIGURE 6: (Top) Number of contacts vs. residue number for the RT structure in the RT unit cell. This plot is typical for all the RT structures. (Middle) *Increase* in the number of contacts when the same RT structure is placed in the 80 K unit cell. (Bottom) *Increase* in the number of contacts (with reference to the figure at top) when the 80 K structure is placed in the 80 K unit cell.

crystal contacts. The left side of Figure 7 shows the regions of the protein that are connected through intermolecular contacts at 255 K, and the right side of Figure 7 shows which contact pairs involve correlated changes in protein structure and crystal contacts when the temperature is reduced. Some of the large shifts seen in the difference distance matrix (Figure 4) are in regions where there are changes in the intermolecular packing (the N-terminal region, the A–B corner, the whole C–D region, and the C-terminal region).

## Conclusions

A nonuniform thermal expansion of crystalline sperm whale metmyoglobin has been observed on going from 80 K to room temperature. The average positions of many of the atoms in myoglobin are different at the two temperatures, leading to an overall volume expansion of ≈3% on heating. The expansion leads to an increase in the volume of the numerous tiny packing defects in the protein. The larger atomic-sized

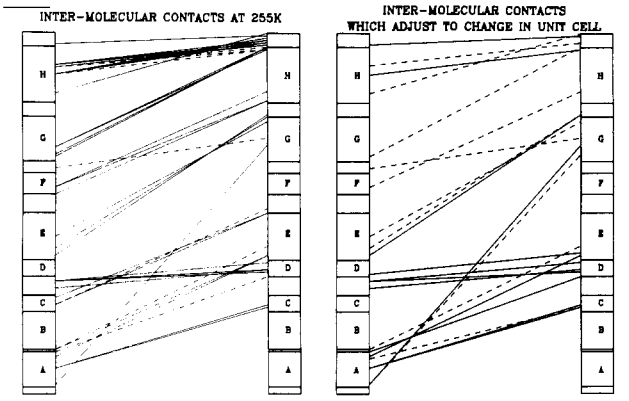


FIGURE 7: (Left) Intermolecular contacts made between residues in the met-Mb 255 K structure. The secondary structure elements are shown in the two vertical bars. Intermolecular contacts are represented as lines connecting residues in one bar to residues in the other. Solid lines represent contacts between 3.1 and 4.0 Å. Dashed lines represent residue pairs where at least one contact is less than 3.1 Å. (Right) Intermolecular distances which are correlated with changes in the unit cell. The shortest intermolecular distance for every residue pair is considered for this figure. Contacts in the 80 K structure which are between 0.25 and 0.75 Å *greater* than that predicted by the change in the unit cell parameters are shown by a dashed line. Contacts which are more than 0.75 Å *greater* are shown by a solid line.

cavities also expand, but the proportional increase is much smaller. The volume change is relatively small within the helices, and a significant component of the overall effect arises from the movement of the C–D region away from the rest of the protein.

Some of the changes in the protein structure are correlated with the expansion of the unit cell of the crystal in such a way that the intermolecular contacts are approximately preserved. One component of the overall thermal expansion therefore involves local structural changes associated with the crystal lattice; however, there is also a systematic expansion of the entire myoglobin molecule which results in the radius of gyration increasing by about 1%.

After this work was completed, a metmyoglobin data set at 102 K was collected on a CAD4 diffractometer equipped with a low-temperature device. The unit cell at this temperature and the model refined against the 2.0-Å data set confirm the results reported in this paper. The details regarding the 102 K structure will be published separately (R. F. Tilton and J. Dewan, unpublished results).

Myoglobin is an all-helical protein in which helix–helix contacts are predominantly hydrophobic. The core of the protein is thus stabilized by a large number of very weak interactions. Many other proteins have β-pleated structures in their interior; i.e., discontinuous segments of the linear amino acid sequence are held together by hydrogen bonds. It is possible that such proteins will display quite different thermal expansion behavior than myoglobin, and it will be interesting to examine proteins in other structural classes by the same methods used in this paper; such experiments are in progress. In addition to providing the first quantitative and spatial description of the thermal expansion of a protein, the present observations supply information about the deformable regions in myoglobin that is important for a theoretical understanding of the nature and distribution of the forces that stabilize the structure of myoglobin.

## References

Ashcroft, N. W., & Mermin, N. D. (1976) *Solid State Physics*, Holt, Reinhart and Winston, New York.

Bull, H. B., & Breese, K. (1973) *Biopolymers 12*, 2351–2358.

Careri, G., Fasella, P., & Gratton, E. (1975) *CRC Crit. Rev. Biochem. 3*, 141–146.
Connolly, M. L. (1981) Ph.D. Thesis, University of California, Berkeley, CA.
Connolly, M. L. (1983) *Science (Washington, D.C.) 221*, 709–713.
*CRC Handbook of Biochemistry* (1976) 3rd ed., p 595, CRC Press, Boca Raton, FL.
Debrunner, P., & Frauenfelder, H. (1982) *Annu. Rev. Phys. Chem. 33*, 283–299.
Frauenfelder, H., Petsko, G. A., & Tsernoglou, D. (1979) *Nature (London) 280*, 558–563.
Gurd, F. R. N., & Rothgeb, T. M. (1979) *Adv. Protein Chem. 33*, 73–165.
Hartmann, H., Parak, F., Steigemann, W., Petsko, G. A., Ponzi, D. R., & Frauenfelder, H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4967–4971.
Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252–270.
Jolicoeur, C. (1981) *Methods Biochem. Anal. 27*, 171–287.
Karplus, M., & McCammon, J. A. (1983) *Annu. Rev. Biochem. 53*, 263–300.
Karplus, M., & McCammon, J. A. (1986) *Sci. Am. 254*, 42–51.
Kittel, C. (1971) *Introduction to Solid State Physics*, Wiley, New York.
Konnert, J. H., & Hendrickson, W. A. (1980) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A36*, 344–349.
Kuntz, I. D., Jr. (1975) *J. Am. Chem. Soc. 97*, 4362–4366.
Kuriyan, J. (1986) Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
Kuriyan, J., Wilz, S., Karplus, M., & Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133–154.
Lee, B., & Richards, F. M. (1978) *J. Mol. Biol. 55*, 379.
Parak, F., Frolov, E. N., Mossbauer, R. L., & Goldanskii, V. I. (1981) *J. Mol. Biol. 145*, 825–833.
Petsko, G. A., & Ringe, D. (1984) *Annu. Rev. Biophys. Bioeng. 13*, 331–371.
Phillips, D. C. (1970) in *British Biochemistry, Past and Present* (Goodwin, T. W., Ed.) pp 11–28, Academic Press, London.
Takano, T. (1977) *J. Mol. Biol. 110*, 569–584.
Tilton, R. F., Jr., & Kuntz, I. D., Jr. (1982) *Biochemistry 21*, 6850–6857.
Tilton, R. F., Jr., Kuntz, I. D., Jr., & Petsko, G. A. (1984) *Biochemistry 23*, 2849–2857.
Walter, J., Steigemann, W., Singh, T. P., Bartunik, H., Bode, W., & Huber, R. (1982) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B38*, 1462–1472.

# Crystal and Molecular Structure of the Serine Proteinase Inhibitor CI-2 from Barley Seeds†

C. A. McPhalen and M. N. G. James*
*Medical Research Council of Canada Group in Protein Structure and Function, Department of Biochemistry, The University of Alberta, Edmonton, Alberta, Canada T6G 2H7*

*Received June 18, 1986; Revised Manuscript Received September 15, 1986*

ABSTRACT: Chymotrypsin inhibitor 2 (CI-2), a serine proteinase inhibitor from barley seeds, has been crystallized and its three-dimensional structure determined at 2.0-Å resolution by the molecular replacement method. The structure has been refined by restrained-parameter least-squares methods to a crystallographic $R$ factor ($=\sum||F_o| - |F_c||/\sum|F_o|$) of 0.198. CI-2 is a member of the potato inhibitor 1 family. It lacks the characteristic stabilizing disulfide bonds of most other members of serine proteinase inhibitor families. The body of CI-2 shows few conformational changes between the free inhibitor and the previously reported structure of CI-2 in complex with subtilisin Novo [McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242–7246]. However, the reactive site loop has some significant conformational differences between the free inhibitor and its complexed form. The residues in this segment of polypeptide exhibit relatively large thermal motion parameters and some disorder in the uncomplexed form of the inhibitor. The reactive site bond is between Met-59I and Glu-60I in the consecutive sequential numbering of CI-2 (Met-60-Glu-61 according to the alignment of Svendsen et al. [Svendsen, I., Hejgaard, J., & Chavan, J. K. (1984) *Carlsberg Res. Commun. 49*, 493–502]). The network of hydrogen bonds and electrostatic interactions stabilizing the conformation of the reactive site loop is much less extensive in the free than in the complexed inhibitor.

Serine proteinases are hydrolytic enzymes that perform a wide variety of biological functions, from activation of blood clotting to digestion of food, protein processing during viral replication, and processing of peptide hormones (Neurath, 1984). Protein inhibitors of these enzymes play important roles in regulating some of these serine proteinase activities. A knowledge of the structures of such inhibitors provides a basis for the rational design of drugs that could limit proteolysis selectively. Such small, well-characterized proteins can also provide a good system for studying protein–protein interactions and structure–function relationships through site-directed mutagenesis.

† This work was funded by grants from the Medical Research Council of Canada to the MRC Group in Protein Structure and Function at The University of Alberta. C.A.M. was the recipient of an Alberta Heritage Foundation for Medical Research Studentship. The work described in this paper formed part of the Ph.D. thesis of C.A.M. (McPhalen, 1986).

0006-2960/87/0426-0261$01.50/0 