Biol Res 29: 77-100 (1996)

Active-site studies of enzymes by secondary structure prediction and by molecular modeling

HILDA CID and MARTA BUNSTER

Laboratorio de Biofísica Molecular, Facultad de Ciencias Biológicas,
Universidad de Concepción, Concepción, Chile

Since the determination of the tertiary structure by X-ray crystallography has been achieved only for a limited number of proteins, alternative approaches are being sought. In this article, the use of secondary structure prediction and of molecular modeling is discussed. Several examples are analyzed in detail.

Key terms: *beta-lactamase; fructose 1,6 bisphosphatase; molecular modeling; molecular dynamics; secondary structure prediction; snake venom phospholipases.*

STATEMENT OF THE PROBLEM

In order to understand the mechanism of action of an enzyme, a hormone or a macromolecular carrier, it is necessary to determine, as precisely as possible, the tertiary structure of the protein and, when relevant, its quaternary structure. Since the first tertiary structure of a protein was determined (Kendrew *et al*, 1960), many important developments in the biological sciences have been a direct consequence of the advances in the knowledge of the structure of biological macromolecules.

However, the determination of the tertiary structure of a protein (performed by X-ray diffraction) has not yet become a routine task. Several problems, such as growing protein crystals, or searching for heavy atom derivatives, sometimes take several years before they are adequately solved.

The secondary structure of a protein describes the spatial folding of the polypeptide chain which, in this specific way, acquires a more stable conformation. The process of folding allows the shielding of the hydrophobic amino acid residues in the interior of the molecule, whereas the hydrophilic ones are exposed to the surrounding solvent. The folding of the polypeptide chain also facilitates the proximity of residues capable of forming stabilizing bonds such as disulfide bridges or hydrogen bonds between the turns of a helical structure or between the strands of a b-pleated sheet.

For years, the secondary structure of a protein was obtained as a by-product of tertiary structure determination. It provides a simpler way to describe the complicated architecture of the macromolecule than the spatial location of the thousands of atoms, or the hundreds of amino acid residues. Figures 1a and 1b illustrate the presentation of the structure of human carbonic anhydrase B (today named HCA I) by a secondary structure model, and by a chain containing just the a carbon of each amino acid residue. It is only in the last 20 years that the determination of the secondary structure, without the previous knowledge of the tertiary structure, has been considered possible.

The discovery of the genetic code and the determination of DNA sequences showed that the primary structure of a protein is the main information genetically transcribed. This fact means that the amino acid sequence already contains all the necessary information to determine the folding of the polypeptide chain; that is, it determines both the secondary and the tertiary structure. This

Correspondence to: Dra Hilda Cid, Laboratorio de Biofísica Molecular, Facultad de Ciencias Biológicas, Universidad de Concepción, Casilla 4077, Correo 3, Concepción, Chile. Fax: (56-41) 245-975.

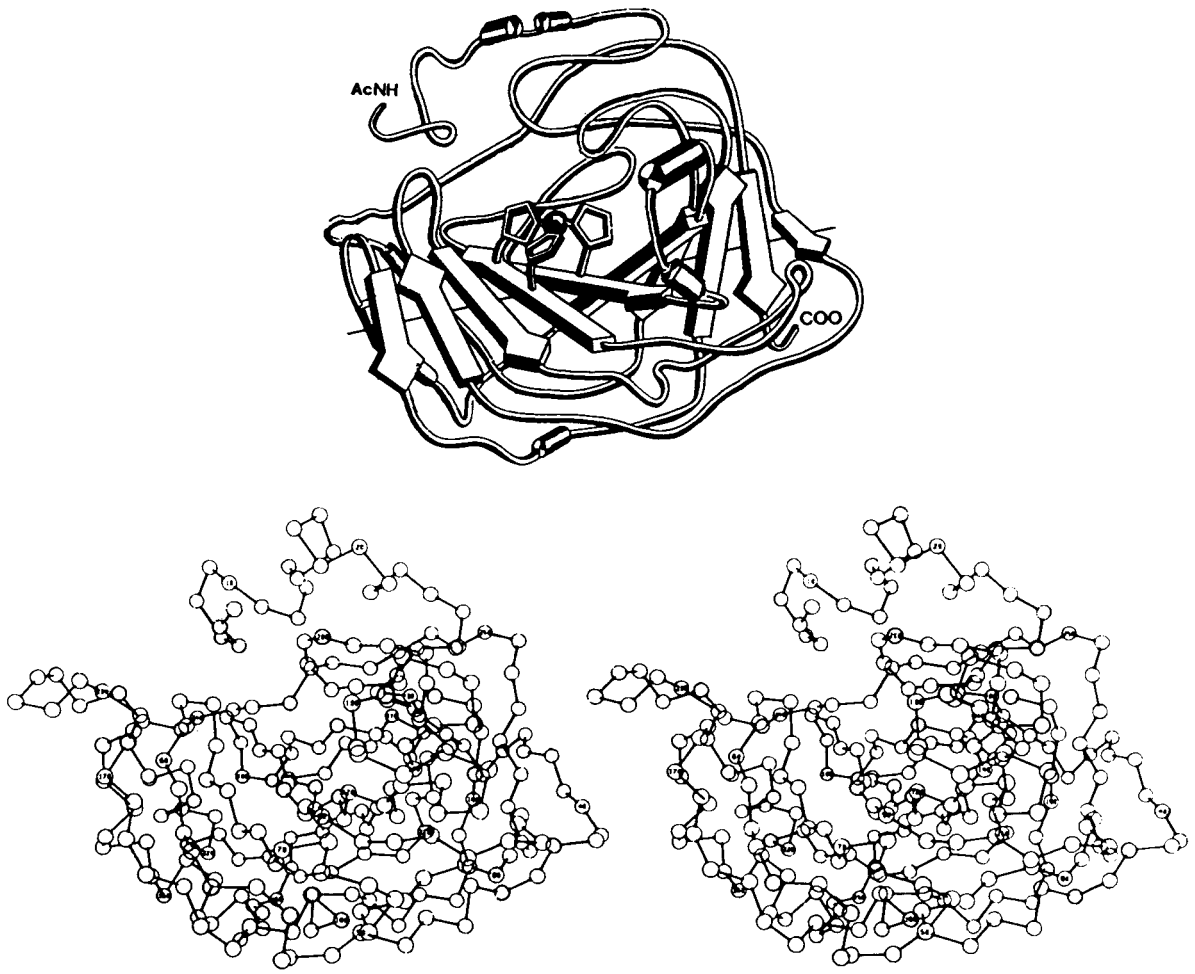

Fig 1. (a) Secondary structure of human carbonic anhydrase I (HCAII), in the cylinder and arrow representation. (b) Tertiary structure of HCAII. Stereoscopic computer drawing of the α-carbons only (From Nostrand, 1974). Reprinted with permission.

conclusion is supported by denaturation-renaturation experiments on several enzymes. However, today it is known that some proteins, called 'chaperones' play a role in facilitating protein-folding, by catalyzing the slow-folding reactions or by preventing 'wrong' protein aggregations (Ellis & van der Vies, 1991).

Following the principle that the primary structure is the main determinant factor in protein folding, several methods have been designed to predict the secondary structure of a protein from a knowledge of its amino acid sequence. Some of these methods are based on empirical probabilities (Chou & Fasman, 1974), numerical algorithms (Lim, 1974), a combination of physicochemical measurements from a data base of known protein structures (Cid *et al*, 1982, 1992; Parker & Song, 1990), or the use of 'neural networks' for recognition of secondary structure patterns (Holley & Karplus, 1991) .

The possibility that the secondary structure of a protein may be predicted in the absence of knowledge of the tertiary structure, opens new trends in enzyme active-site studies. Even if the complete enzyme structure cannot be known precisely, the location of the active-site and some information about its morphology, would be a valuable tool in the comprehension of its mechanism of action.

In the first version of this article (Cid, 1987), we presented three examples of enzyme families with known primary structures but unknown tertiary structures, where information about the active-site was searched by a combination of secondary

structure prediction and chemical studies. A spatial folding of the polypeptide chain was proposed, based on a combination of secondary structure prediction and simple model building, which fulfilled the accepted packing of secondary structure elements. The tertiary structure of some members of these three families have been determined in the meantime at good resolution. Both the predictions and the new results are included in this review; thus, the limitations of the prediction methods employed can now be analyzed.

The central idea in our secondary-structure prediction approach is that the active-site region in a family of enzymes should present a much higher degree of structural invariability in the secondary and tertiary structures than is usually found in the primary structures. The shape of the active 'pocket' must be conserved since all enzymes accommodate the same substrate or part of the substrate within it. This type of approach has been used in the study of the following protein families: snake-venom phospholipases A_2, ß-lactamases obtained from different microorganisms, and fructose-1,6-bisphosphatase from various animal tissues.

PREDICTION OF THE SECONDARY STRUCTURE OF A PROTEIN

Of the several methods designed to predict the secondary structure of a protein from the amino acid sequence, none has proved to be 100% reliable when applied to proteins whose tertiary structure (and therefore, secondary structure) have been fully determined by X-ray diffraction methods. Success is variable. The predictions shown below are based on two independent methods, one proposed by Chou and Fasman (1974) and the other by Cid and collaborators (1982), which have shown between 60-80% reliability when applied to globular proteins.

Chou and Fasman's method for secondary structure prediction

The Chou and Fasman method is based on empirical probabilities. It defines conformational parameters P_a, P_b and P_t for each of the 20 natural amino acids. These parameters represent the probability of each amino acid of participating in a helix, a ß-structure or a turn structure. They are, in fact, the normalized frequency of occurrence of each amino acid residue in that particular type of structure, as established from a data base of 29 fully determined protein structures. A probability average greater than 1.0, obtained for a group of amino acids taken in sequence (6 for a helix, 5 for a ß-strand, and 4 for a ß-turn) is an indication that a specific type of structure is likely to occur in that region of the sequence. To improve the sensitivity of the method in the vicinity of the limit value 1.0, the probability average can be replaced by a product of the conformational parameters (Dufton & Hider, 1977), This and other two modifications of the method, one that considers four conformational parameters for each amino acid residue in a turn structure (Chou & Fasman, 1978) and another that allows a differentiation between ß-strands participating in parallel or antiparallel ß-structure (Lifson & Sander, 1979), have proved to be very useful in secondary structure prediction.

It has been reported in recent years, that if the conformational parameters are separately determined for proteins belonging to the same structural class, as defined by Levitt and Chothia in 1976, the reliability of the prediction is improved (Chou, 1989).

Hydrophobicity profiles' method for secondary structure prediction

This method combines physico-chemical measurements of the solubilities of amino acids in polar and non-polar solvents with information obtained from a data base of 21 known protein structures (Cid *et al*, 1982). The method gives the relative position of portions of the polypeptide chain with respect to the protein surface since a linear correlation exists between the 'surrounding hydrophobicity' H_t, as defined by Ponnuswamy *et al* (1980), and the average distance of an amino acid residue to the protein surface, as measured on 21 known protein structures. The 'surrounding hydrophobicity' for a given residue is defined as

the sum of the Tanford hydrophobicities of all the amino acids included in an 8 Å-radius sphere, centered at the α-carbon of the residue in consideration. Since the calculation of the surrounding hydrophobicity implies the knowledge of the tertiary structure, a 'bulk hydrophobic character' <Ht>, has been calculated as an experimental average value from the tertiary structures available for 21 proteins. Recently, these coefficients have been recalculated for a data base of 64 protein structures, and also for proteins belonging to each one of the four structural classes (Cid *et al*, 1992).

The hydrophobicity profile is simply a plot of <Ht> versus the amino acid number in the sequence. Four basic profiles have been defined for 4 types of secondary structure: helix, ß-turn, and buried, and exposed ß-strands. These basic profiles are represented in Figure 2. The identification of these basic patterns in the hydrophobicity profile of the protein yields the predicted structure.

It should be observed that the helical profile only describes amphipathic helical structures; they represent about 50% of all helices found in proteins. The existence of mainly hydrophobic or mainly hydrophilic helical structures was a possibility not considered in the method. This fact explains why this method tends to underpredict helical structures and to overpredict ß-strands. Even though it does not seem possible to define a profile that differentiates the hydrophobic helical structure from the buried ß-strand, these two structures can be distinguished by the nature of the hydrophobic amino acid residues involved: alanine and leucine characterize helices, valine and isoleucine are mainly found in ß-strands. Hydrophilic helices are also possible, but they are short and usually present glutamic acid and lysine. Other stabilization factors of hydrophilic helices are: 1) opposite charged residues situated at positions i,i+3, or i,i+4 allow ion-pair interactions, and 2) a stabilization of the helical dipole moment is achieved by negative charges located at the N-cap and positives ones at the C-cap of the helix. (H Cid, F Gazitua & M Bunster, 1994, unpublished results).

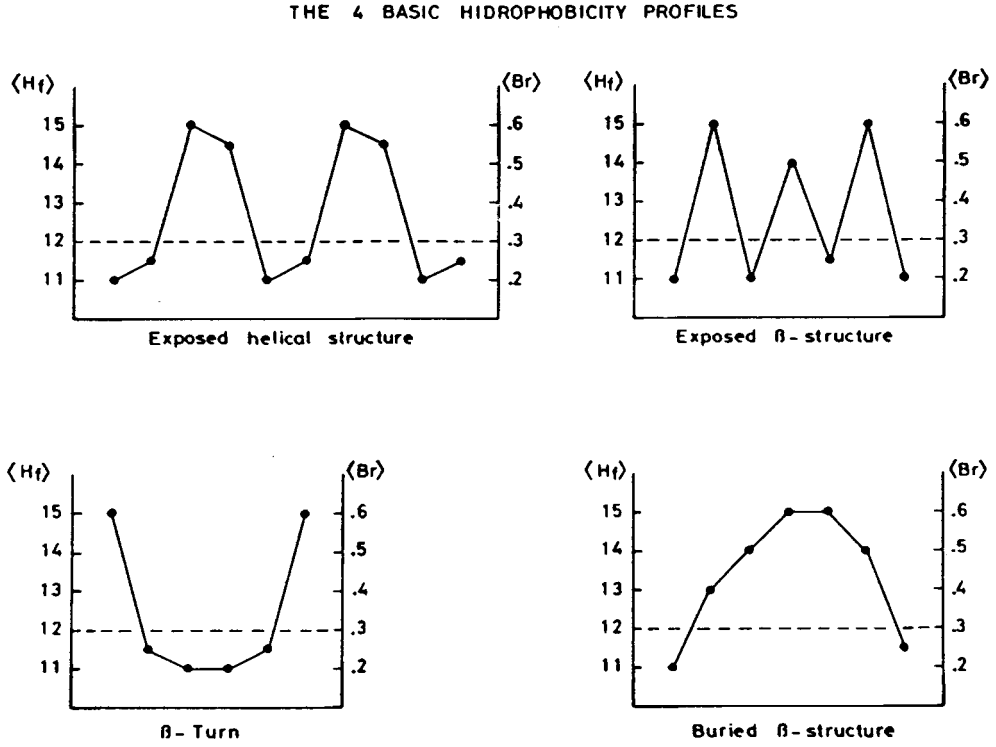

Fig 2. The four basic hydrophobicity profiles. The number indicating the position of the amino acids in the sequence has been plotted against the 'bulk hydrophobic character'. (From Cid *et al*, 1982)

Three-dimensional fold of predicted secondary structures

In order to obtain as much information as possible from a predicted secondary structure, a three-dimensional fold can be proposed by building simple models with rigid arrows, cylinders and 'hair pins' to represent ß-strands, α-helices and ß-turns, respectively. These elements are joined by mobile connections and by flexible wires that represent random coil regions. The lengths of the building elements are scaled to the number of amino acids involved, and to the distance between α-carbons in that particular type of structure. In building such models, the following complementary information must be considered:

1. Distinction between exposed and buried α-strands, as predicted by the hydrophobicity profiles.
2. Structural class of the protein, as predicted from amino acid composition (Chou, 1989).
3. Preferences for ß-strands to be part of a parallel or antiparallel ß-sheet, as given by the modified Chou and Fasman method.
4. Stabilization of helical and ß-structures in one of the three super secondary structures defined by Levitt and Chothia (1976): aa, bb, bab.
5. Proximity of some amino acid residues based on chemical evidence. Examples of these models are given in Figures 7 and 14.

Results obtained from application of criteria 1 to 4 should agree with the predicted structural class.

COMPARATIVE MODELING OF PROTEIN STRUCTURES

A step further in predicting the secondary structure of proteins, and specially their three-dimensional folding, is provided by molecular modeling methods. The use of these methods involves extrapolation from one or more known protein structures to construct a new structure when the sequences of various members of the family, as well as that of the problem structure, are known (Greer, 1990). These methods work better when the known tertiary structures have been experimentally determined at good resolution. It is then possible to differentiate regions with an invariant structure, or 'structurally conserved regions' from 'variable regions'. This procedure involves molecular graphics and molecular mechanics modeling techniques.

The successive steps of the method are the following:

Model Building

These steps are illustrated on Figure 3.

1. Superposition of the 3-dimensional structures in order to search for the structurally conserved regions (SCRs) and to locate the variable regions (VRs). This process is usually carried out manually on a graphics display system.
2. Sequence alignment based on the superposition of the SCRs, and not on amino acid identity or sequence homology. The alignment of the VRs is more or less arbitrary, unless there are sequence homologies, or a conformation is found in more than one structure.
3. Determination of patterns of sequence homology in the aligned sequences, and alignment of the 'new' sequence with these structurally conserved patterns. It should be possible to align a stretch of the 'new sequence' with each one of the SCRs. The remaining residues will form the VRs. This is a crucial step for building the model of the 'new' structure.
4. Model building can start with the backbone coordinates of the SCRs of any one of the solved proteins The side chains are mutated to those of the 'new' sequence, trying to achieve the best overlapping with the beginning of the 'old' side chain.
5. Search for sequences to model the VRs, either from the VRs of one of the known structures or from homologous sequences in Data Bases (Summers & Karplus, 1990). Energy minimization procedures are necessary at this step, as described below.

Model Refinement

1. Search for errors in the proposed structure, such as bond angles and distances too

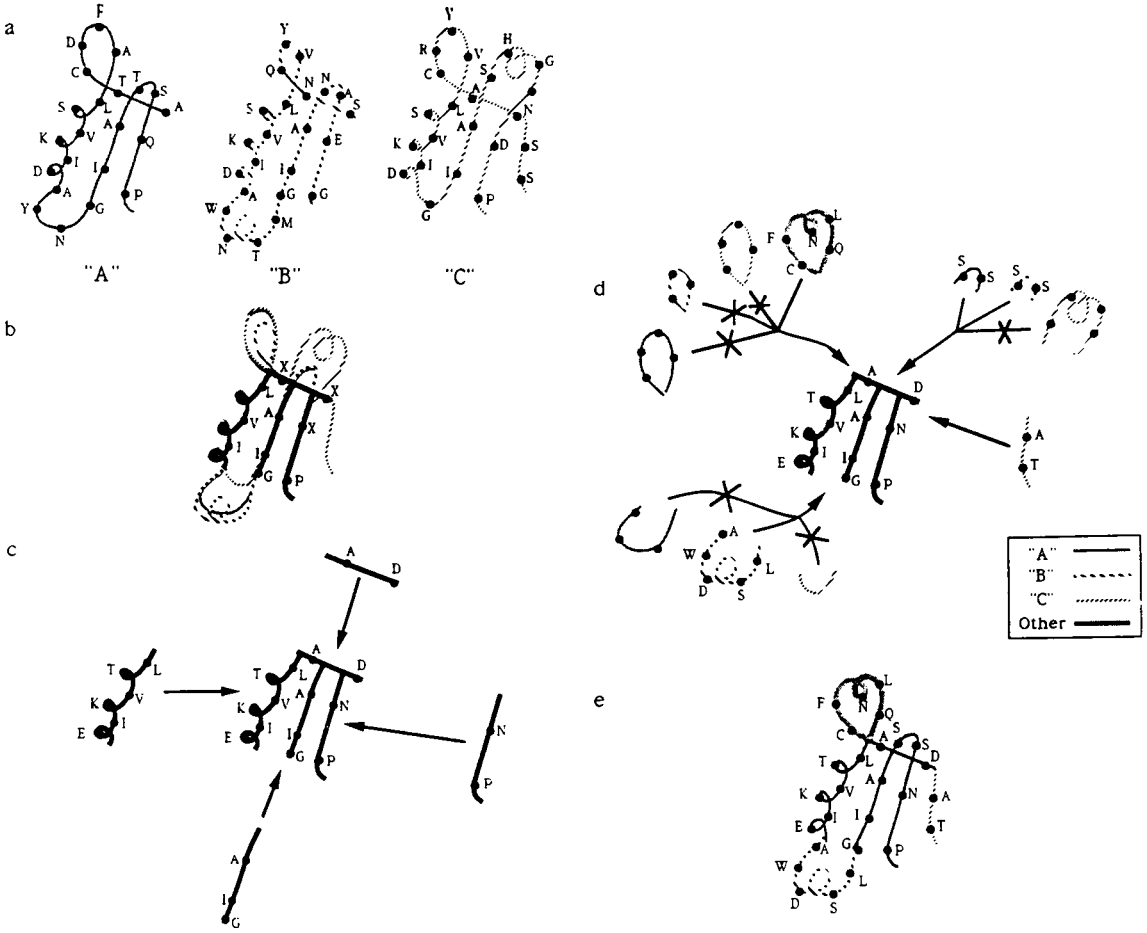

Fig 3. Schematic representation of the comparative modeling method. (a) A, B, and C are three proteins of known structure of an homologous family, each one represented by a characteristic dashed or dotted pattern, shown in d). (b) The three structures are superimposed to show the structurally conserved regions (SCRs) in bold lines, and the variable regions (VRs) in dashed or dotted lines. (c) First step in the construction of the 'new structure', the SCRs are built from the chain coordinates of any of the known structures. The side chains are mutated to the 'new sequence' when necessary. (d) The various VRs conformations found in the known structures are considered for the VRs of the 'new protein' (compare with the VRs shown in a and b). Those that do not fit are rejected, as shown by the crossed arrows, and the most suitable is selected. If none of them is adequate, other conformational search or energy minimization methods must be employed (see VR in upper left corner with the sequence C-F-N-L-Q as an example of a necessarily different new conformation). (e) The composite structure showing the source of the respective 'spare parts' selected for the model structure. The model is now ready for refinement. (From Greer, 1990; reprinted by permission from John Wiley & Sons, Inc.)

far from those accepted, or superposition of atoms, due to side chain mutations which can be corrected by changing the^c angles.

2. Introduction of a set of water molecules (using the coordinates of the water molecules of one of the solved structures), and introduction of the hydrogen atoms. This process is usually included in the minimization packages.

3. Energy minimization of the model to obtain a minimum energy conformation. These calculations, known as 'molecular mechanics' are based on a Potential Energy representation of the molecule as a function of the internal coordinates and the inter atomic distances, as shown on equation [1].

$$V= \Sigma\{D_b[1-e^{-a(b-bo)}]^2 -D_b\} + 1/2\ \Sigma H_o(\theta-\theta_o)^2 + 1/2\ \Sigma Hf\ (1+ s\cos n\phi) + 1/2\ Hc\chi^2 + \Sigma\Sigma\ \varepsilon[(r^*/r_{ij})^{12} -2(r^*/r_{ij})^6] + \Sigma\Sigma\ q_iq_j/r_{ij} \quad [1]$$

The first four terms represent the energy necessary to distort the bonds (b),valence angles (θ), torsion angles (ϕ), and angles out of the plane (χ); b_o and θ_o are ideal values. H_k represents the force constant necessary to

distort the respective internal coordinate 'k', whose values have been determined from small model compounds. The last two terms represent non-bonded interactions between atoms i and j, r^* being the most favorable interatomic distance and e the energy of interaction at this distance (Sessions *et al*, 1992).

With the potential energy of the system defined by [1], its minimization involves the solution of the equation:

$$\delta V/\delta x_i = 0 , \quad i = 1,3n \qquad [2]$$

where x_i represents the three components of the vector r_i defining the position of each one of the N atoms of the system we want to minimize.

In the first cycles of minimization of the complete molecule the movement of atoms should be restricted to the hydrogens. After these have found the best location, the minimization process proceeds, usually until the derivatives of the energy with respect to the coordinates x_i are less than 0.01.

Model Evaluation

1. Check that the overlapping of the backbones of the SCRs is maintained within a root mean square (rms) deviation of less than 2Å. If not, the building of the neighboring VRs should be revised.

2. Check that the model ϕ and φ angles fall within the Ramachandran's allowed regions.

3. Check that the ratio between the calculated solvent-accessible surface of the model to the protein molecular weight (SURFMW), lies in the range accepted for globular proteins:

$$0.320 < \text{SURFMW} < 0.520 \text{ Å}^2$$

4. Check that the structural information provided by the model fulfills the requirements for an adequate functioning of the protein. This condition can be proved by experimental tests. At this stage, molecular dynamics simulations (MD) are a useful tool that complement and help to explain the existing data by designing new experiments. These simulations are increasingly useful for refinement of active-site models generated by molecular modeling, NMR or X-ray diffraction methods. (McCammon & Harvey, 1987). The understanding of the mechanisms of action of proteins, a dynamic event, has been greatly improved by the use of MD.

Molecular Dynamics Simulations

This method calculates the future positions and velocities of atoms based on their current positions; for example the atomic positions obtained after the refinement in the molecular modeling, or in the X-ray determination of the structure. A simulation determines first the force (F_i) on each atom, as a function of time, equal to the negative gradient of the potential energy (V), with respect to the position of the atom i, determined by the vector r_i.

$$V = V(r_1, r_2, r_3 \ldots\ldots r_N) \qquad [3]$$

$$F_i = \frac{\delta V}{\delta r_i} \qquad [4]$$

Once the force F_i is obtained from equation [4], the acceleration a_i on each atom can be calculating dividing this force by the mass mi of the atom :

$$a_i = \delta^2 v_i/\delta t^2 = F_i/m_i \qquad [5]$$

The change in velocity Δv_i is obtained by integrating equation [5] over a time interval Δt, and the change in the position Δr_i is equal to the integral of v_i over the same time interval. The kinetic energy is defined in terms of Newton's equation [6]

$$K = (1/2) \sum_{i}^{N} m_i \, v_i^2 \qquad [6]$$

The Hamiltonian, or total energy of the system, is the sum of the potential energy given in equation [1], and the kinetic energy given by equation [6].

It should be noted that integration of the equations of motion is done using very small time steps Δt. At each step, the energy and forces of the system are evaluated; the Dt values used are of the order of femtoseconds

(10^{-15} s). Time steps of this order allow simulations that integrate adequately high frequency motions of the system, like bond stretching vibrations on the order of several picoseconds (10^{-12} s).

The second important parameter to consider in the molecular dynamics simulation process is Δt, the temperature increment, that will provide the necessary changes in the kinetic energy (at T near 0°K, atoms have zero velocity). When the simulation process starts, it needs to assign velocity values which are realistic for that molecular system at a certain temperature. It is convenient to reach the desired temperature in several 'heating 'steps. (McCammon & Harvey, 1987)

The first important application of MD is in the refinement of parts of a structure obtained by molecular modeling: In applying energy minimization algorithms, one searches for a minimum energy configuration of the system, by moving along the gradient of the potential energy in the energy hypersurface describing a macromolecular system. This surface may contain a very large number of local minima, and because energy minimization moves only downhill over the surface, it can reach only a local minimum energy configuration in the neighborhood of the initial one. The relaxation of the system strains by small local positional adjustments can be made by MD, since this method accounts for thermal motion, and molecules may contain enough thermal energy to cross potential barriers and reach a true minimum energy configuration.

Molecular Dynamics allows the study of concerted motions that can occur, for example, during ligand binding, enzymatic activity and molecular recognition, because it is possible to store 'snapshots' of a molecular dynamic trajectory (set of positions and velocities of atoms over time), and use them to detect possible conformational changes in the region under study. These facilities make the method very valuable to study docking of substrates to proteins, since it can give, not only the 'best' position of the substrate in the active site to start the catalytic process, but it can also describe the whole docking process in time, and the different conformational changes that the substrate and the active site can undergo. (See for example Sessions *et al*, 1992; Di Nola *et al*, 1994).

One drawback of the method is that, so far, it is unable to handle time intervals of the order of nanoseconds or longer, that characterize most chemical or physical processes; therefore, it is not useful to study long-term processes like protein folding.

EXAMPLE 1: SEARCH FOR A 'TOXIC SITE' IN SNAKE-VENOM PHOSPHOLIPASES A_2

Protein toxins from snake venom block transmission across the cholinergic neuromuscular junction, either by a post-synaptic, curare-like action on the nicotine acetylcholine receptors of the muscle end plate, or by a presynaptic interference with the release of acetylcholine from the motor nerve terminals.

The sequences of more than 50 post-synaptic snake-venom neurotoxins have been reported, and the typical structure of the two main groups of toxins involved have been determined by X-ray diffraction methods (Walinshaw *et al*, 1980; Tsernoglou & Petsko, 1976).

By contrast, until the mid-eighties, very little was known about the three-dimensional structure of the pre-synaptic neurotoxins. It is clear, however, that there is no common structure that can describe them all. In general, they present multiple polypeptide chains, and have different degrees of toxicity, although it has been found that all of them show a phospholipase A_2 (PLA_2) activity. Therefore, one can say that pre-synaptic neurotoxins are either phospholipases A_2, or at least one of their polypeptide chains is a PLA_2. Great similarities are found among toxic phospholipases and with the mammalian phospholipases A; at least 50% of the sequences are conserved, and all have about 120 amino acid residues and 7 disulfide bridges. The high homology in primary structure makes it difficult to explain the remarkable differences in specific action between the snake venom phospholipases and the homologous phospholipases from mammals. Alkaline venom phospholipases present a high degree of toxicity and can

show neurotoxic, mitotoxic, cardiotoxic, or anticoagulant activity; acidic or neutral phospholipases A are less toxic or non-toxic. It is reasonable, then, to look for differences in the secondary structures that could account for their toxic action.

The relationship between secondary structure and toxicity were the aim of studies on the secondary structure of snake-venom PLA_2 (Arriagada & Cid, 1980, 1989). The characteristics of the toxins included are listed in Table I. The secondary structures, predicted both by the Chou and Fasman method and by the hydrophobicity profiles method, were compared to the predicted and observed secondary structure of bovine phospholipase A_2, whose tertiary structure has been fully determined by high resolution X-ray diffraction (Dijkstra *et al*, 1981). The predicted secondary structure of the mammalian phospholipase, compared to that obtained from the tertiary structure, gave an estimation of the reliability of the method in this case (Arriagada & Cid, 1980). Figure 4 illustrates the relationships between predicted secondary structures of the toxins

Table I

Some characteristics of snake-venom PLA_2 as compared to a mammalian PLA_2

Phospholipase A_2	AA residues	Character	Toxicity
Bovine pancreas PLA_2	123	Basic	Non toxic
Notexin	119	Basic	Highly neurotoxic & myotoxic
Notechis II-5	119	Basic	Highly neurotoxic & myotoxic
Notechis II-1	119	Neutral	Non-toxic
Enhydrina schistosa PLA_2	119	Basic	Myotoxic & neurotoxic
$ß_1$-Bungarotoxin (chain A)	120	Basic	Highly neurotoxic
Naja nigricollis PLA_2	118	Basic	Moderately toxic
Naja naja atra PLA_2	119	Acidic	Weakly toxic
Taipoxin g- subunit	125	Acidic	Non-toxic
H. haemachatus PLA_2	119	Neutral	Weakly toxic

listed on Table I, among themselves and with the X-ray determined secondary structure of bovine phospholipase A_2. The helical zones have been represented by springs and the zones of extended structure (that form part of ß-structures) by stippled blocks. A remarkable similarity among the snake-venom and mammal enzymes can be seen

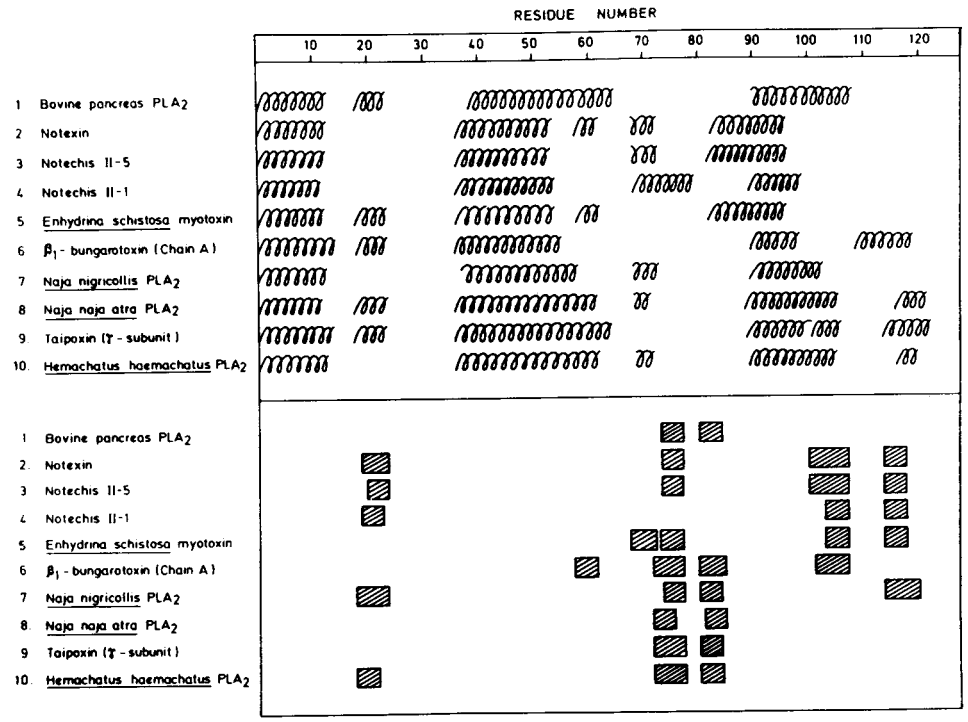

Fig 4. Structural relationships between phospholipases. The predicted secondary structures of nine snake-venom phospholipases A_2 have been compared with the secondary structure of bovine pancreas PLA_2, as obtained from its known tertiary structure. Helical zones are represented by springs and regions of ß-strands by stippled blocks. (From Arriagada & Cid, 1989).

in the first part of the sequence up to residue 60. Important differences are observed in the second half of the sequence; several extended structures predicted for the snake-venom phospholipases are not found in the experimentally determined structures of the mammalian enzyme.

The hydrophobicity profiles of notexin and bovine phospholipase A_2 are aligned to best fit in Figure 5 in order to emphasize the differences between a highly neurotoxic and miotoxic snake-venom phospholipase, and the mammalian enzyme. The profiles are almost identical up to amino acid residue 60. In the mammalian phospholipase the sequence is then followed by 5 amino acids not included in notexin, which, according to the X-ray data, form part of a small helix (helix D). This helix is absent in all venom phospholipases studied so far with the single exception of γ-taipoxin from *Oxyranus scutellatus* (Dufton *et al*, 1983).

In addition to the absence of helix D, another striking feature is the accumulation of positive charges in two regions of the highly neurotoxic basic phospholipases: one in the vicinity of amino acid 60 and the other between residues 88 and 90 (using the sequence numbering of the mammalian enzyme) as shown in Figures 5 and 8. In the latter region, notexin (as well as Notechis II-5 and *Enhydrina schistosa* PLA_2) has the sequence Lys 88, Lys 89, Lys 90; the corresponding sequence in the bovine phospholipase being Asn, Asn, Ala. Moreover, in the non-toxic phospholipase Notechis II-1, obtained from the same snake venom as notexin, the corresponding sequence is Lys, Tyr, Gly.

A secondary structure study on 32 phospholipases A_2, of which 29 are from snake venoms, has been done by Dufton *et al* (1983). The prediction method used is a modified version of the method of Chou and Fasman, which was complemented with studies of circular dichroism and relative interface hydrophobicity, as well as with a careful study of the sequences of such phospholipases (Dufton & Hider, 1983). These results, together with those reported above, give the following clues to explain the behavior of the snake venom phospholipases, as compared to the non-toxic mammalian phospholipases A_2:

1. The neurotoxins must be able to bind to the nerve terminal membrane by means of a special recognition site. This site is

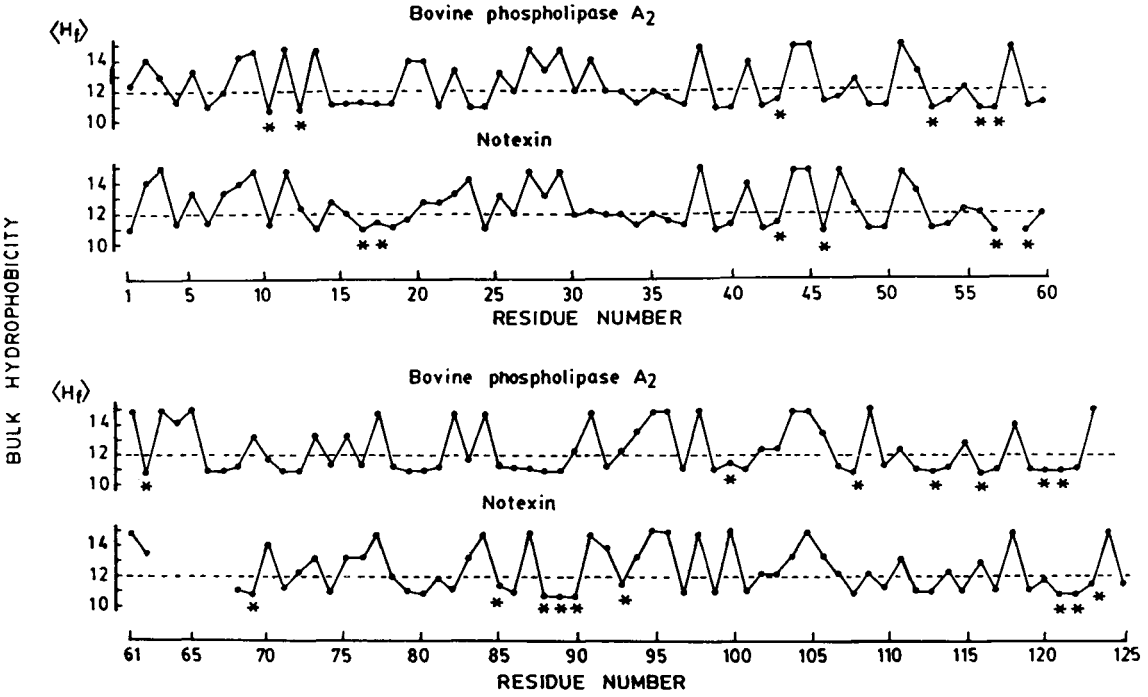

Fig 5. Hydrophobicity profiles of notexin and bovine PLA_2 aligned for best fit. Note the similarity of the profiles, that indicate an identity in secondary structure, from amino acid 1 to 60. The positions of the positively charged residues are indicated by*.

not accessible when helix D (Fig 6) is present, as it occurs in the mammalian PLA_2. Helix D is absent in most snake venom PLA_2s, the only known exception being the non-toxic Taipoxin γ-subunit.

2. The presence of the positively charged lysine residues in the vicinity of residue 60 plays a fundamental role in the neurotoxic properties of the snake venom phospholipases. This fact may account for the observed lack of toxicity of some venom phospholipases, such as Notechis II-1, which do not have Lys 58 and Lys 69, even though helix D is also absent.
3. A second 'toxic site' could be assigned to the region 88-90 where the 3 lysine residues occur in notexin and in other toxic phospholipases such as Notechis II-5 and Enhydrina schistosa myotoxin. This region is located just after the only ß-structure found in bovine PLA_2 (Fig 6).

New data on the structure of snake-venom phospholipases A_2

The X-ray structures of several snake-venom PLA_2 are now known; some of these are a dimeric PLA_2 from the venom of Crotalus atrox (Brunie *et al*, 1985), a cobra venom PLA_2 (White *et al*, 1990) and notexin (Westerlund *et al*, 1992). A comparison between the structures of the bovine and the *Crotalus atrox* PLA_2s indicate a close similarity between these two enzymes, with two exemptions: the antiparallel ß-strands of the dimeric enzyme (amino acids 74-84), present a more 'open' conformation, with differences as large as 7 Å in the C_α positions; the other region with large differences was the contact surface between both subunits in the snake venom-enzyme, where intersubunit hydrogen bonds appreciably deviate the polypeptide chain. The catalytic network of the conserved residues His 48, Tyr 52, Tyr 73 and Asp 99 are identical. Since the *C. atrox* PLA_2 did not contain calcium, differences found in the orientation of the side chains in the "calcium-binding loop" formed by Tyr 28, Gly 30 and Gly 32 were expected .

The availability of the tertiary structure of notexin, a monomeric presynaptic phospholipase at 2.0 Å resolution (Westerlund *et al*, 1992) has given other clues that may explain the toxic behavior of snake-venom PLA_2: the helical regions A, B, C and E are very well conserved, but large structural differences are found precisely in the region 59-71 where helix D is absent, and in the positively charged region where the 'ß-wing' ends, containing Arg 85, Lys 88, Lys 89, Lys 90 and Arg 93 . Close to this region is also Lys 69, which is a Tyr in the bovine PLA_2 and in the non-toxic Notechis II-1 (Arriagada & Cid, 1989) .

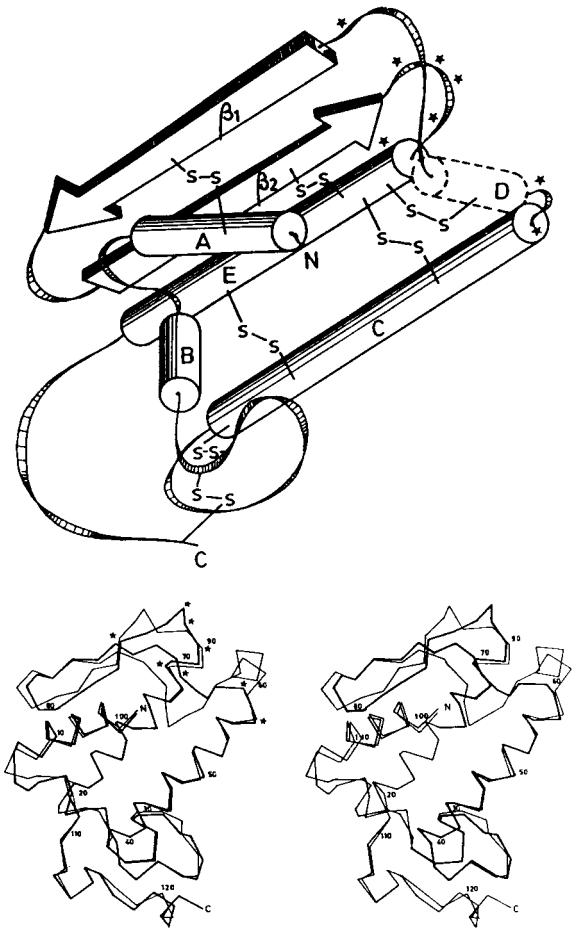

Fig 6. (a) Consensus secondary structure of most phospholipases A_2 in the cylinder and arrow representation. Helix D is absent in all toxic snake-venom PLA_2. The location of charged residues, marked with an asterisk, close to the sequence where helix D is absent, is the proposed toxic site for snake-venom PLA_2 (From Arriagada & Cid 1989). (b) Superposition of the α-carbon backbones of the X-ray structures of notexin (in heavy line) and bovine phospholipase A_2 (in thin line). As in a), the location of the positively charged residues in the proposed toxic sites are marked with*. (From Westerlund *et al*, 1992; reprinted with permission).

The catalytic site of phospholipases A_2 has been described as a cleft lined by the

following amino acid residues : Leu 2, Phe 5, Asn 6, Leu 19, Asn 23, Cys 29, Gly 30, Leu 31, Gly 32, Cys 45; in the bottom of it are located the catalytic residues His 48 and Asp 99 (Sessions *et al*, 1992); therefore the structural invariance of this zone on all venom phospholipases is well justified. Notexin, C. atrox and cobra venom PLA_2s maintain the invariance and the hydrogen-bond network of the catalytically involved amino acid residues. The site occupied by the calcium ion in the bovine enzyme is filled in by a water molecule in the notexin structure.

Coincidences and differences between the predicted and the X-ray determined structural data

Secondary structure prediction. The hydrophobic helix E was predicted as a buried ß-strand in the snake venom enzymes, due to the lack of definition of a "buried helix" hydrophobicity profile.

Structural invariance of the first 60 aa residues. Confirmed by the structure determination of the three snake venom PLA_2s, and by studies of the catalytic mechanism of all PLA_2s by molecular mechanics and inhibitor binding (Sessions *et al*, 1992).

Location of a toxic site. As secondary structure studies pointed out, the toxic behavior of snake venom PLA_2s must be related to the absence of helix D, and to the concentration of positively charged residues (Arriagada & Cid, 1989). The "neurotoxic" and the "myotoxic" sites should be very close together, and would include the positively charged residues, as indicated in Figure 6.

EXAMPLE 2: SECONDARY STRUCTURE STUDIES IN ß-LACTAMASES

ßlactamases are enzymes responsible for the bacterial resistance to ß-lactam antibiotics such as penicillins and cephalosporines. The loss of antibiotic action is produced by a disruption of the ß-lactam ring. ß-lactamases form a group of enzymes which show a great diversity in molecular weights, amino acid composition and activity against various antibiotics, although they perform a common function (Citri & Pollock, 1966). The primary structure of several ß-lactamases has been determined (Ambler, 1980; Dale *et al*, 1985), and they show that the number of invariant amino acid residues amounts to no more than 20% of the total.

It is obvious that a precise knowledge of the mechanism of action of these enzymes would be of the greatest importance for medical purposes. Therefore, much effort has been spent trying to solve the three-dimensional structure of ß-lactamases and of their active-sites. However, although some low resolution X-ray diffraction data had been reported as early as 1976, the tertiary structure of none of these enzymes was completely solved until 1987 (Herzberg & Moult, 1987).

In the absence of tertiary structure information, knowledge of secondary structure can provide information about the conformation of the active-site, and, at the same time, it can furnish a better pattern for comparison than does the primary structure. One would expect a structural homology far better than 20% in the vicinity of the active-sites of enzymes that act upon the same substrate.

A common model for the structure of four ß-lactamases

Bunster and Cid (1984) made a secondary structure prediction for the ß-lactamases obtained from *Escherichia coli, Bacillus cereus, Bacillus licheniformis, and Staphylococcus aureus*, using both prediction methods previously described. The primary structures were those reported by Ambler (1980). The first important result that came out of the predictions was that on the average, 56% of all amino acid residues were involved in zones with a conserved secondary structure. Models of the four secondary structures predicted were built, according to the technique described before. One of these models is represented in Figure 7.

The models suggested various interesting results:

First, the four ß-lactamases can be described as two-domain structures; this is consistent with the number of amino acid residues and with the report for the *E. coli*

enzyme from a low-resolution structure obtained by X-ray methods (Knox *et al*, 1976).

Second, a large part of the secondary structure was conserved in the four enzymes, as shown by the shadowed structural elements in Figure 7. The constant structures may differ in the lengths of the ß-strands, helices or random coiled regions, but they are still easily recognized (Bunster & Cid, 1984). The small number of invariant amino acid residues present in a zone with a conserved secondary structure is remarkable, as illustrated in Figure 8.

Third, the 'first domain' (residues 31 to 157) presented 84% of all its amino acids involved in invariant secondary structures, while only 22% of the sequence is conserved. This structural constancy was considered as an indication that the active-site, and presumably the catalytic function of these ß-lactamases, was restricted to the first

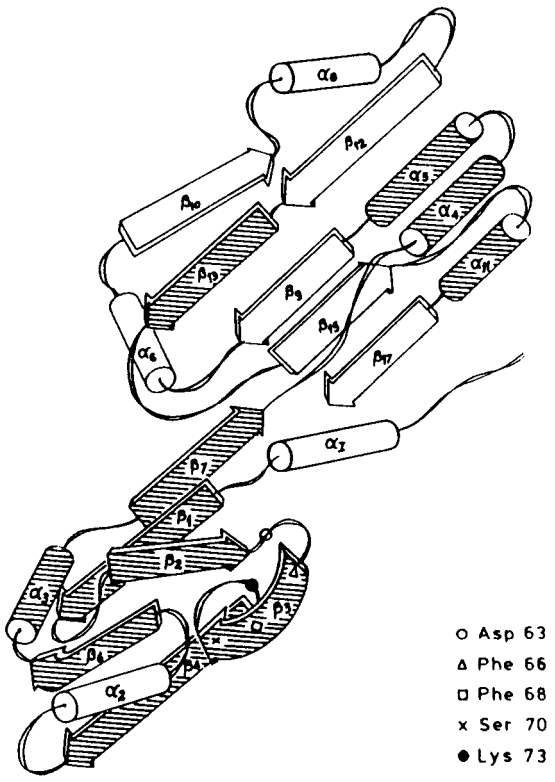

Fig 7. A model for the structure of *B. cereus* ß-lactamase I, obtained from secondary structure prediction studies, in the cylinder and arrow representation. Stippled secondary structure elements appear as conservative in four class A ß-lactamases. The location of the amino acid residues proposed to be part of the active site is indicated. (From Bunster & Cid, 1984).

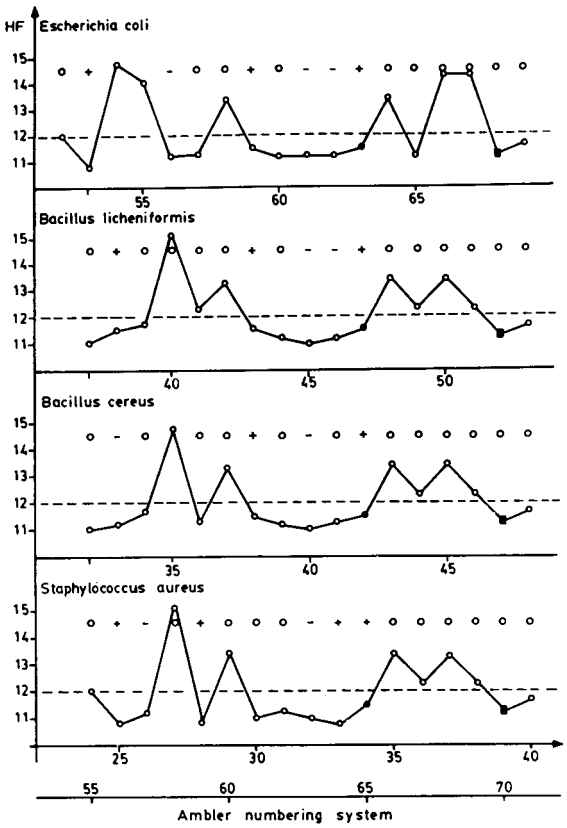

Fig 8. Hydrophobicity profiles for four class A ß-lactamases in the vicinity of the active site. Note the similarity between the profiles, even though the only invariant residues in the four enzymes are Arg 65, marked with a filled circle and Ser 70, a residue involved in the catalytic mechanism, marked with a filled square. The numbering system of Ambler is used. (From Bunster & Cid, 1984).

domain, even though domain II could still be part of the catalytic site, since the flexibility of the inter-domain region would allow a close contact. Also, in the model proposed, the active-site would include a charged loop following $ß_2$ (Fig 8). The presence of charges near Ser 70 coincides with the reported existence of carboxylic groups in the active-site (Durkin & Viswanatha, 1980). This loop contains Ser 70 which participates in the catalytic reaction (Waxman & Strominger, 1980). Both Tyr 105 and His 112, which should be near the active-site (Durkin & Viswanatha, 1980) are also located in this neighborhood.

Fourth, the existence of "small" ß-lactamases with a molecular weight similar to that of domain I, reported for the enzymes obtained from Pseudomonas aeruginosa plasmids R 157 and Rms 149 (Matthew,

1978; Ambler, 1980), and from *Streptomyces* UCSM-104 (Campos *et al*, 1979), also supported such a model.

Test of the validity of the model

To test the validity of the model, two aspects were considered: (a) if the activity of the enzyme is restricted to domain I, it should be maintained in the absence of domain II; and (b), a three-dimensional model of domain I should reveal a site compatible with the substrate and with the catalytic reaction.

(a) Search for an active domain I. To prove the first point, *B. cereus* ß-lactamase (molecular weight 27800) was reacted with CNBr under very mild conditions, for times ranging from 4 to 12 hours. CNBr reacts primarily with methionines to cause cleavage, and reacts secondarily with cysteines. The molecule chosen has no cysteines and only four methionines, none of them in domain I, so this reaction seemed adequate to separate domain I. Even though we obtained 'successful' results in a couple of experiments, these were not reproducible. To prevent a possible denaturation by BrCN, trypsin TPCK was used for a controlled proteolysis of the enzyme. If the molecule had two continuous domains, a preferential cleavage could be expected at some stage of the proteolysis, by rupture of the peptide bond after a lysine or arginine at the more exposed region of the polypeptide chain, the interdomain region. No preferential cleavage was observed at any time or with any ratio of protease/ß-lactamase, thus excluding the presence of continuous domains (Bunster *et al*, 1990).

(b) *Location of a possible active-site.* A three-dimensional model of domain I of the ß-lactamase from *B. cereus* was built using Kendrew skeletal models, following the pattern of the proposed structure represented in Figure 7. In this model a cavity, presumably the active-site, was clearly visible. This site could accommodate a molecule of benzyl penicillin as substrate, and stabilize it in a unique orientation. The benzene ring would interact hydrophobically with the aromatic rings of two phenylalanines, whereas the carboxyl group of the five-member ring could make a hydrogen bond with a lysine (Fig 9). The substrate oriented in this way was sufficiently close to Ser 70 to allow molecular interaction with it, and the close presence of Asp 63 and of Lys 73 even allowed the proposal of a catalytic mechanism, similar to that of serine proteases (Bunster, 1984).

The predicted versus the X-ray determined structures of ß-lactamases class A

In recent years the crystal structures of 3 of the four ß-lactamases considered in the previous studies have been determined: the structure of the *S. aureus* ß-lactamase at 2.5 Å resolution (Herzberg & Moult, 1987), that of the *B. licheniformis* enzyme at 2 Å resolution (Moews *et al*, 1990), and the structure of the *E. coli* TEM1 enzyme, refined to 1.8 Å resolution (Jelsch *et al*, 1993). Only preliminary X-ray studies of the *B. cereus* ß-

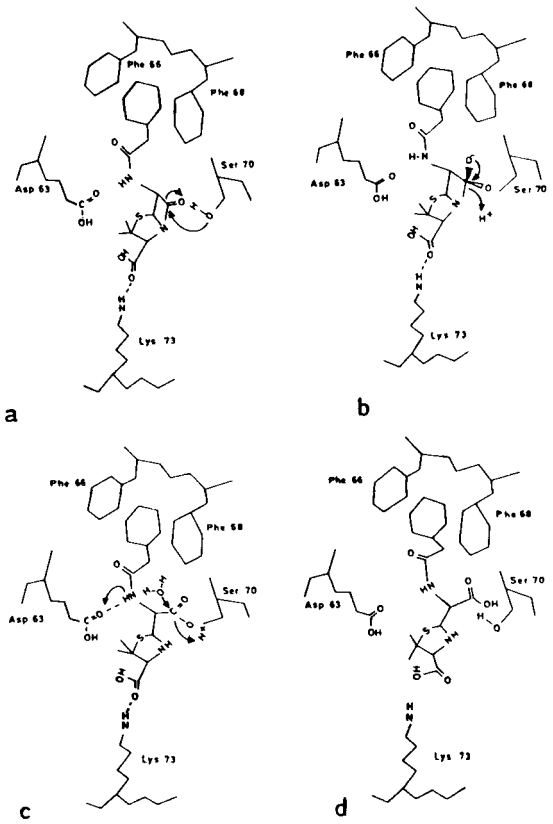

Fig 9. Scheme of the 'active site' proposed in the predicted model. Some of the amino acid residues and the tentative location of the substrate, that agrees with the mechanism proposed (a-d), are indicated. (From Bunster *et al*, 1990).

lactamase I have been published so far; they support the similarity of this structure with that of R61 carboxypeptidase, and with other class A ß-lactamases. (Phillips *et al*, 1987).

Structure coincidence for class A ß-lactamases. As suggested by the secondary structure prediction studies, all class A ß-lactamases share a very well conserved 3-dimensional structure (Fig 10).

Secondary structure prediction accuracy. Seven out of the 11 helical regions and 5 out of the 5 ß-strands were correctly predicted (Fig 11). As pointed in the discussion of the prediction methods, the main differences are given by the prediction of buried helical structures as ß-strands (Bunster *et al*, 1990).

Folding of the polypeptide chain. Two domains can be recognized in the determined structures, but these are discontinuous domains, formed by residues from the N- and C-terminals, connected by three strands of the polypeptide chain. This possibility was not considered in the proposed model.

The catalytic mechanism. The catalytic mechanism described for the three ß-lactamases whose structure has been determined to high resolution is very similar to that proposed for the *B. cereus* enzyme (Bunster *et al*, 1990). However, the position of the residues involved are not the same: the role proposed for Lys 73 in anchoring the substrate is performed by Lys 234, and the role proposed for Asp 63 in the deacylation step, is really performed by Glu 166 in the three enzymes.

From these results it is clear that, even with an optimum secondary structure prediction, a correct folding cannot be assured; some extra information is required. In modeling methods, this extra information is provided by the knowledge of the tertiary structure of related proteins.

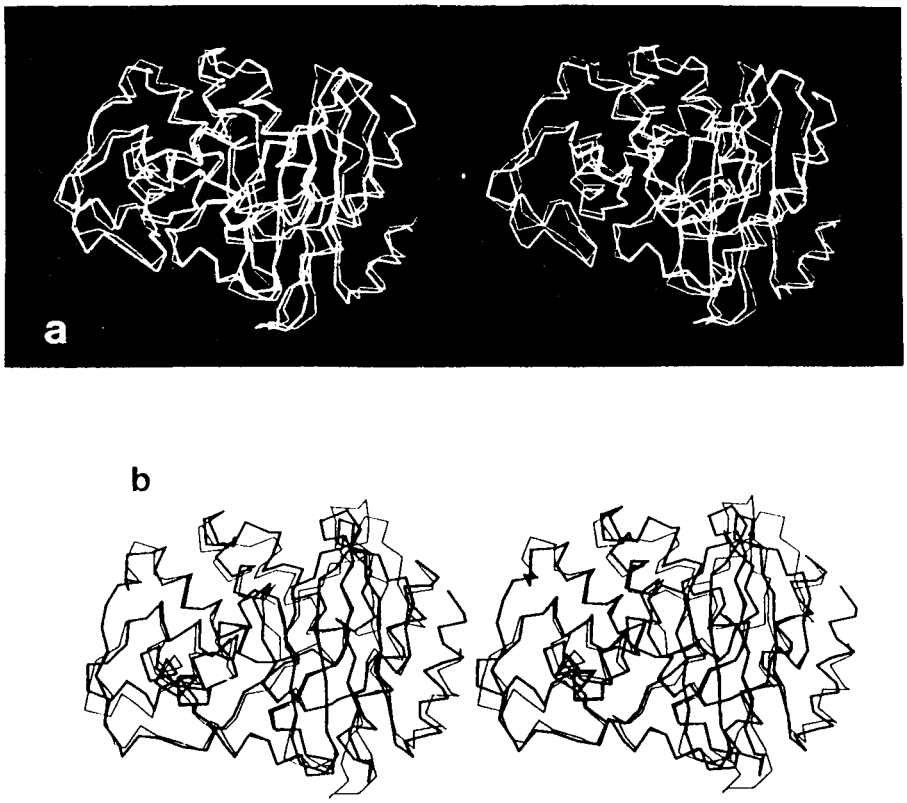

Fig 10. Superposition of the Cα backbone of the X-ray determined structure of Class A ß-lactamases : a) *S. aureus* PC1, in thin line and B. licheniformis 749/C in heavy line (From Moews *et al.*, 1990). b) E coli TEM 1 in thick line and *S. aureus* PC1 in thin line. Cα atoms from each domain (except loops and regions with insertions or deletions) were handled independently. (From Jelsch *et al*, 1993; reprinted by permission from John Wiley & Sons, Inc.)

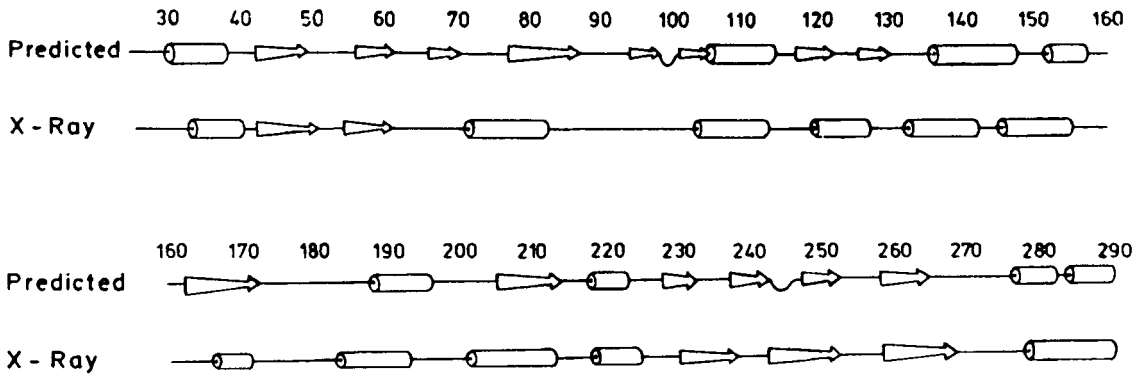

Fig 11. Alignment of the predicted secondary structure of *S. aureus* PC1 ß-lactamase with that determined by X-ray diffraction methods. Helical regions and ß-strands are represented as cylinders and arrows respectively. (From Bunster *et al*, 1990).

Modeling the three-dimensional structure of ß-lactamase I from B. cereus 569H

A three-dimensional model of the structure of the ß-lactamase from *B. cereus* was obtained by molecular modeling techniques, as described above. The model was built using the coordinates of the *S. aureus* enzyme, stored at the Brookhaven Protein Data Bank (Bernstein *et al*, 1977), and the sequences given by Ambler (1980).

1. The published structures of the *S. aureus*, *B. licheniformis* and *E. coli* RTEM1 enzymes gave the necessary information to locate the structurally conserved regions (SCRs) (Fig 10).
2. The alignment of the sequence of the *B. cereus* enzyme with the conserved regions was done using the hydrophobicity profiles, since the sequence invariance was low (Fig 8).
3. Model building started with the substitution, in the determined SCRs, of the amino acid residues of the *B. cereus* enzyme, in the corresponding sites of the sequences of the *S. aureus* enzyme. Figure 12 shows the coincidences of the C_as after the substitution was done. The VRs, such as strand 237-266 were built ab initio. The program INSIGTH II was used at this stage.
4. Energy minimization using the AMBER force field, in the Discover version, was employed to correct the built model. The course of energy minimization of the initial model, after 4 processes of 2000 minimization cycles each, is given on Table II.

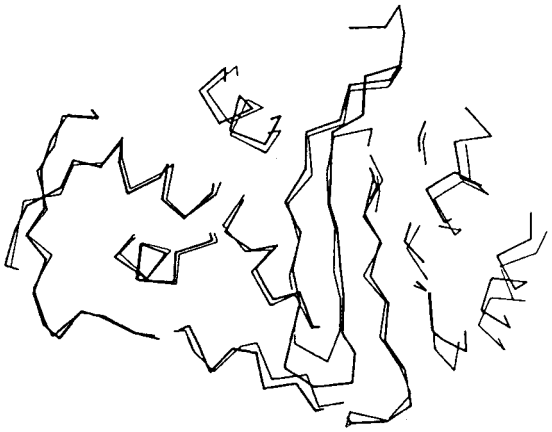

Fig 12. Cα backbone superposition of the SCRs of the *S. aureus* (thin line) and of the model for the *B. cereus* (heavy line) enzymes. (From M Bunster, M Canales & H Cid, 1994, unpublished results).

Table II

Refinement of the initial model of the *B. cereus* 569/H ß-lactamase I

Refinement cycle	Energetic component	Average absolute derivative	RMS deviation from *S. aureus* enzyme
0	+5068 kcal	2.408 kcal/Å	8.15 Å
4 x 2000	- 641 kcal	0.0345 kcal/Å	0.69 Å

Model evaluation. At this stage of refinement, the model for the *B. cereus* enzyme closely resembles the general pattern of all class A ß-lactamases, with a 5-stranded antiparallel ß-sheet , and a bundle of 8 helical structures (Figs 12 and 13). The Kabsh and Sanders program revealed that the secondary structure content of the model is 28% helical structure and 9% ß-structure, in

Fig 13. Superposition of the Cα backbone of the *S. aureus* enzyme (thin line) and of the refined model for the *B. cereus* enzyme (heavy line). Insertions and/or deletions were not included. (From M Bunster, M Canales & H Cid, 1994, unpublished results).

good agreement with CD measurements (M Bunster, M Canales & H Cid, 1994, unpublished results). All the amino acid residues involved in the active site of the *S. aureus* enzyme are conserved in the model, and they keep their accessibility and close neighbors. The only exception is Ala237, which is replaced in the *S. aureus* enzyme by Gln; this residue shows a lower accessibility, as would be expected for an hydrophobic amino acid residue. The SURFMW value of 0.413 $Å^2$, calculated for the model, lies within the 0.3-0.5 $Å^2$ range accepted for 'well-folded' protein structures .

EXAMPLE 3: ACTIVE-SITE STUDIES OF FRUCTOSE-1,6-BISPHOSPHATASE FROM PIG KIDNEY

Fructose-1,6-bisphosphatase catalyzes the hydrolysis of fructose-1,6-bisphosphate to fructose-6-phosphate and inorganic phosphate, a reaction which is a key regulatory step in gluconeogenesis. The enzyme is inhibited allosterically by AMP and also by fructose-2,6-bisphosphate.

Even though the enzyme, isolated from different gluconeogenic tissues, has been extensively studied (Pontremoli & Horecker, 1970), structural data regarding the active-site and regulatory sites were far from complete until the three-dimensional structure of the enzyme complexed with the substrate and AMP were solved in 1989 and 1990, respectively. Some amino acid residues involved in the active-site and in the AMP binding site have been previously identified in the enzyme obtained from rabbit liver and, by sequence homology, this identification was extended to the pig kidney enzyme (Marcus *et al*, 1982).

The complete primary structures were determined for the enzymes from sheep liver (Fischer & Thompson, 1983) and pig kidney (Marcus *et al*, 1982). When the first version of this article was published, only partial sequences were known for enzymes obtained from rabbit, chicken, turkey and mouse liver, and rabbit kidney (Pontremoli *et al*, 1983). Only one complete amino acid sequence for another mammalian enzyme has now been added : the fructose 1,6-bisphosphatase from rat liver (El-Maghrabi *et al*, 1988).

When the model shown below was proposed, no tertiary structure had been reported for any of these enzymes, even though crystals had been obtained for the chicken (Anderson & Matthews, 1977) and rabbit liver (Soloway & McPherson, 1978), and for the pig kidney enzymes (Seaton *et al*, 1984).

A model for the structure of fructose-1,6-bisphosphatase from pig kidney

This enzyme is a tetramer formed by subunits of molecular weight 36534, as calculated from the 335 amino acid sequence. A secondary structure was proposed by the methods already discussed, and a model built, based on this structure (Martinez & Cid, 1986) is shown in Figure 14.

The enzyme subunit is organized in two domains, as it would be expected for a globular protein with more than 300 amino acids (Levitt & Chothia, 1976). Each domain can be described as a b-barrel, formed by parallel and antiparallel ß-strands alternated with helical structures. In the proposed model, domain I would be formed by the first 141 residues, followed by a 17-residue strand of random coiled structure and by domain II including amino acids 157 to 335. In the packing proposed, the ß-strands of both domains would be oriented forming an angle near 90 degrees (Fig 15).

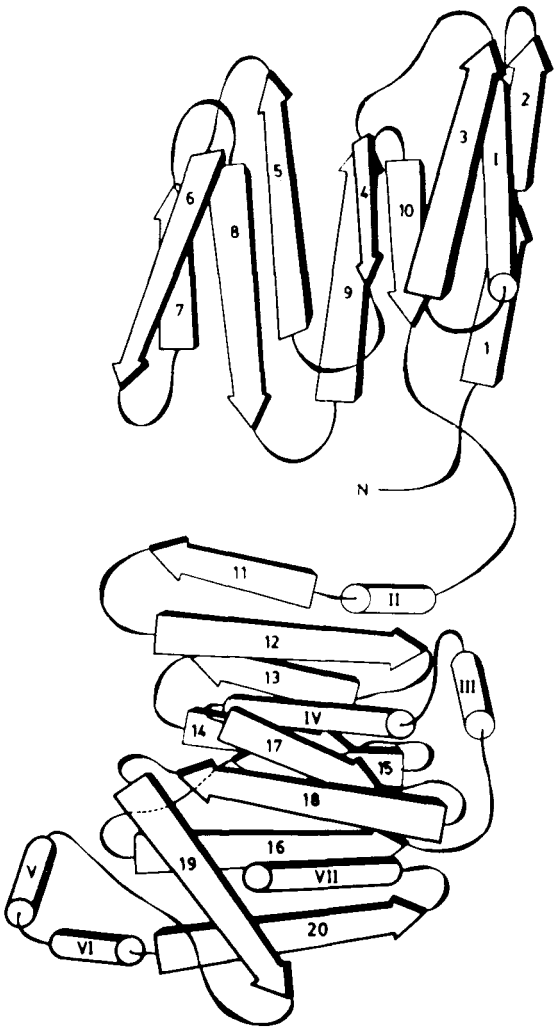

Fig 14. Two-domain model for the structure of fructose-1,6-bisphosphatase from pig kidney, in the cylinder and arrow representation. (From Martínez & Cid, 1986).

The location of the active-site, and of the AMP site in this model was established by secondary structure homology, made by aligning the hydrophobicity profiles, with the enzyme obtained from rabbit liver, where the sequences in the neighborhood of these sites are known (Xu *et al*, 1982; Suda *et al*, 1982), as shown on Figures 16 and 17. The region from amino acid residues -79 to -39 of the rabbit enzyme which has been postulated to involve the active-site is aligned with the sequence 249 to 290 of the pig kidney enzyme. The negative numbers of the rabbit enzyme sequence, proposed by Xu *et al* (1982), describe the position of the amino acid residues with respect to the COOH terminus of a fragment of MW 9850. There is a 75% homology in the sequence and nearly a 100% conservation of the secondary structure, since the amino acid replacements have been made by hydrophobically equivalent residues. Lys 274 in the pig kidney enzyme should correspond to Lys -54 in the rabbit enzyme, which has been reported to participate in the catalytic mechanism.

The location of the AMP binding site has been made by aligning the sequences 51 to 83 in the rabbit enzyme with 135 to 166 of the pig kidney enzyme, as seen in Figure 17. Even though there is only a 50% conservation of the primary structure, the similarity between the profiles is evident. Lys 141 would be involved in AMP binding in the pig kidney enzyme, the analog to Lys 58 in the rabbit protein (Suda *et al*, 1982).

Domain I, in addition to the AMP binding site, would contain the hyper-reactive SH group of Cys 128 (Chaterjee *et al*, 1984) and the region sensitive to proteolysis. This region would be the loop including residues 57 to 67 (Nakashima & Horecker, 1971) which is well exposed in the model proposed. The 17-residue strand that joins both domains is flexible enough so as to

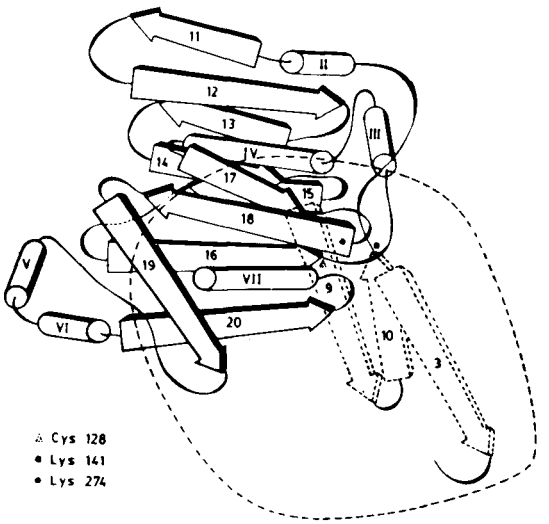

Fig 15. Superposition of the two domains in the proposed model structure of fructose-1,6-bisphosphatase from pig kidney. For the sake of simplicity, only the contour of domain I and three important ß-strands, $ß_{10}$ preceding Lys 141, $ß_9$ containing Cys128 and $ß_3$ preceding the loop sensitive to proteolysis were represented. Lys274 and Lys141 are marked with filled circles at the beginning of $ß_{18}$ and at the end of $ß_{10}$, respectively; Cys128 is marked with a triangle in the center of $ß_9$. These amino acid residues are close to each other according to NMR studies. (From Martínez & Cid, 1986).

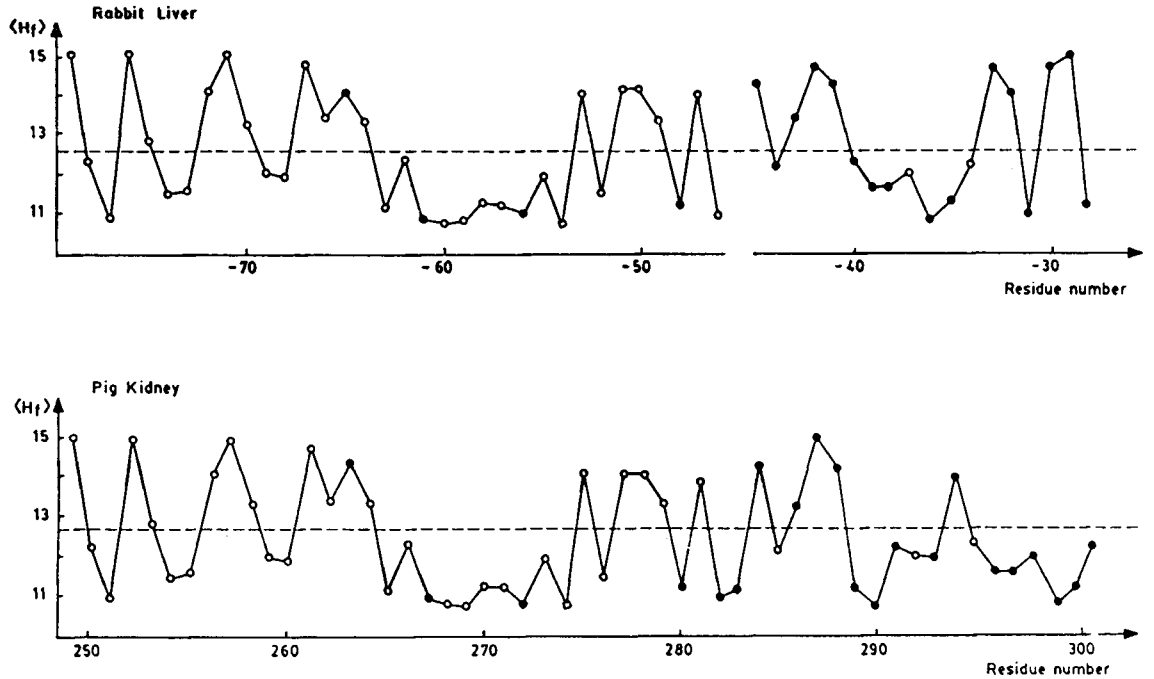

Fig 16. Hydrophobicity profiles for the sequences -79 to -27 of rabbit liver and 249 to 301 of pig kidney fructose-1,6-bisphosphatases aligned for best fit. <H_f> values for non-conserved amino acid residues are represented by filled circles. Lys-54 and its analog Lys274, which should be part of the respective active-sites, are marked with an open square. Note that the identity of the profiles is maintained up to residues -39 and 290, respectively. See text for explanation of negative sequence numbers. (From Martínez & Cid, 1986).

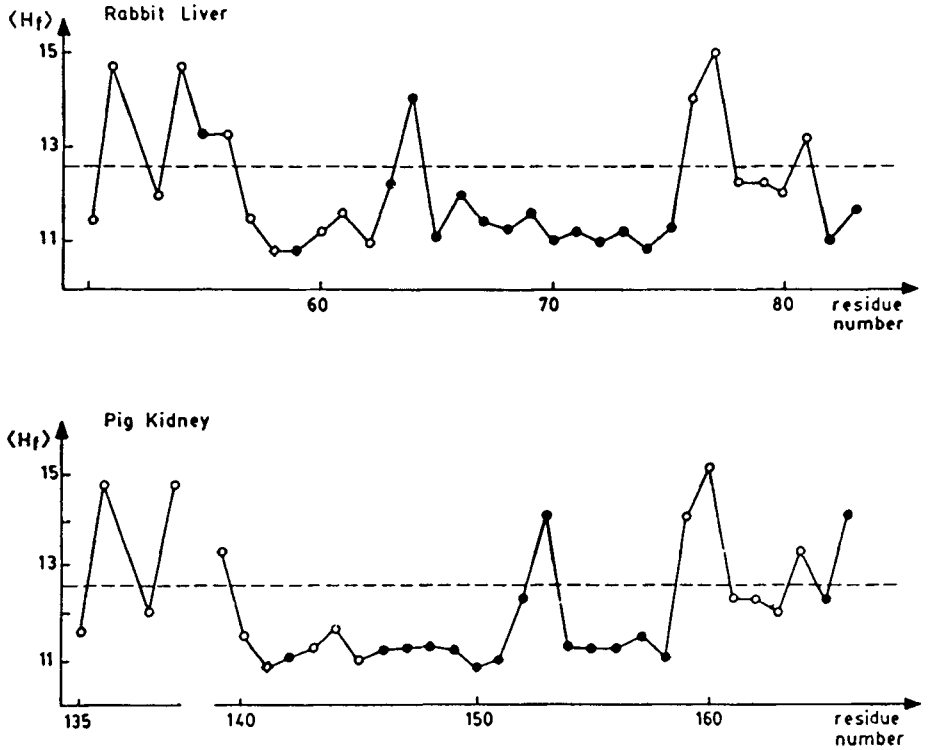

Fig 17. Hydrophobicity profiles for sequences 51 to 83 of rabbit liver and 135 to 166 of pig kidney fructose-1,6-bisphosphatases aligned for best fit. Note the similarity between the profiles, even though the conservation of the primary structure is low. Lys58 and its analog Lys141, involved in the AMP binding site, are marked with open squares. The nomenclature of Figure 16 has been used. (From Martinez & Cid, 1986).

allow their superposition in a way that permits the proximity of the proposed AMP-binding site, the active-site and the hyper-reactive SH group (Fig 15).

It must be emphasized here that the location of the active-site, and of the AMP site by secondary structure homology, agreed with that proposed by Marcus *et al* (1982), based on the primary structure homology found between pairs of hexapeptides in the pig kidney and rabbit liver enzymes.

Mechanism of inhibition by fructose-2,6-bisphosphate

The role played by fructose-2,6-bisphosphate in the inhibition of fructose-1,6-bisphosphatase has been controversial. Some authors (Van Schaftingen & Hers, 1981; Reyes *et al*, 1985) have reported the existence of an allosteric binding site for this metabolite, whereas others (Pilkis *et al*, 1981; Kitajima & Uyeda, 1983) favor a direct interaction with the active-site, and consider it a competitive inhibitor. A third model (Corredor *et al*, 1984) postulates a biphasic behavior for fructose-2,6-bisphosphate, depending on the substrate concentration: low concentrations of fructose-1,6-bisphosphate would favor the interaction of fructose-2,6-bisphosphate with the active-site, whereas, at high concentration of the substrate the metabolite would bind to its own allosteric site. A biphasic behavior of the metabolite is consistent with the proposals of positions 1 and 2. This model can explain, for example, the results of NMR studies on the bovine liver enzyme which suggested that fructose-2,6-bisphosphate binds at the active-site, since these studies were made in the absence of the substrate (Ganson & Fromm, 1985).

The existence of an allosteric site for fructose-2,6-bisphosphate, close enough to the active site, could also be postulated in the model structure proposed for fructose-1,6-bisphosphatase (Martinez & Cid, 1986). This site would include the hyper-reactive SH group of cysteine 128, considering that modification experiments of this group both in the pig kidney (Reyes *et al*, 1985) and in rat liver enzymes (Meck & Nimmo, 1983) have shown that it is protected by fructose-2,6-bisphosphate.

The X-ray determined tertiary structure of fructose 1,6-bisphosphatase

The X-ray structures of fructose-1,6 bisphosphatase and of its complex with its regulatory inhibitor, fructose 2,6-bisphosphate, have been solved and refined to 2.6Å resolution (Ke *et al*, 1989; Liang *et al*, 1992). Also, the structures of the complexes of the enzyme-fructose 6-phosphate-AMP-Mg^{+2}, and of the enzyme-fructose 6-phosphate, have been solved and refined to 2.1Å resolution (Ke *et al*, 1990, 1991). These studies have determined the exact position of the enzyme's active site, the location of the product and of the regulatory inhibitor, also at the active site, and the location of the AMP site.

Secondary structure prediction accuracy. The determined secondary structure of the enzyme, as compared to the predicted (Martinez & Cid, 1986) is shown on Figure 18 (Ke *et al*, 1989). With the exception of B10 and B12, most observed ß-strands overlap fairly well with the predicted ones. Helices H2, H6, H7 and H9 also coincide with regions of sequences predicted as helical structures. Helices H4 and H5 are short 3_{10} helices, detected only when the X-ray structure resolution was improved to 2.1Å (Ke *et al*, 1991). H1, H3 and H8 were predicted as ß-strands.

General description of the structure. The X-ray determined structure describes a monomer of the enzyme as formed by two domains joined by a single strand of the polypeptide chain. Both domains are formed by antiparallel ß-strands with a few parallel strands, alternated with helices (Fig 19). Domain I includes amino acid residues 6 to 200 (the predicted domain I had only 141 amino acids), and the orientation of its ß-strands and helices are nearly 90° from those of domain II. When Figure 19, that represents the determined structure, is compared with Figures 14 and 15, the overall similarity of the determined and predicted structures is evident.

Binding site of the regulatory inhibitor fructose 2,6-bisphosphate and of the allosteric inhibitor AMP. X-ray structure determination of complexes of the enzyme with its inhibitor (Fru-2,6-P_2) and with its

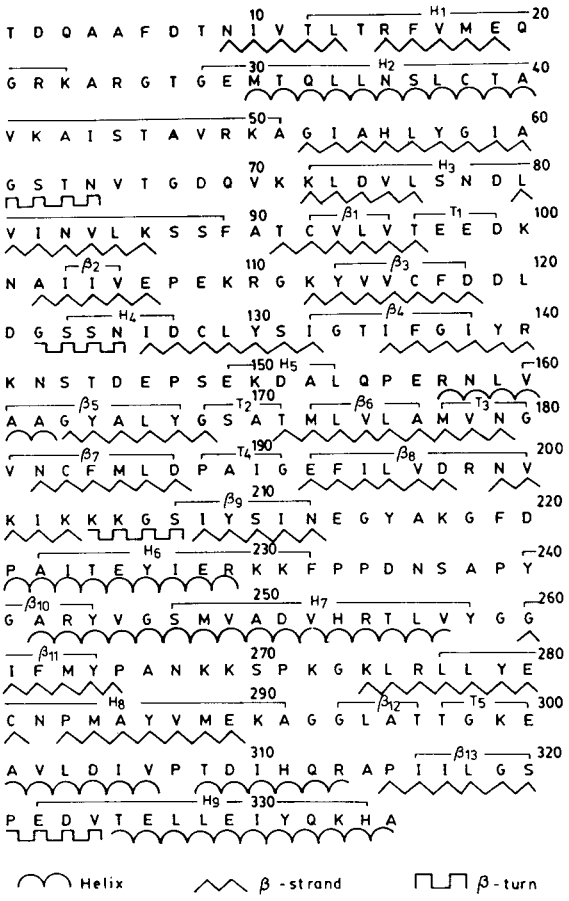

Fig 18. Comparison of the X-ray determined and predicted secondary structures of pig kidney fructose-1,6-bisphosphatases. One letter representation for amino acids is used. The letters H, ß and T (under the sequence) indicate the X-ray determined helical zones, ß-strands and turns, respectively. Symbols ⌒, ⩘ and ⊓⊔ represent the predicted regions of helical, ß-strands and ß-turn, respectively.

product (F6P) have definitely established that the regulatory inhibitor binds at the active site (Ke *et al*, 1989, 1990). Even more, the 6-phosphate and the sugar groups of the product (F6P) bind at the active site in positions that correspond to the 6-phosphate and ribose of the regulatory inhibitor (Fru-2,6-P_2).

The allosteric site for AMP binding, determined by X-ray methods is about 28 Å from the active site (Fig 19). The phosphate group of AMP interacts with the backbone atoms of Glu 29 and Met 30, and with the side chains of Thr 27 and Lys 112. The side chains of Tyr 113 are in contact with the AMP phosphate group and with its sugar ring. These results are not comparable with those of nuclear magnetic resonance studies in the rabbit and bovine liver enzymes, which indicated distances of the order of 7 Å (Cunningham *et al*, 1981; Ganson & Fromm, 1985). Also, no special relationship between any cysteine residue and the binding of AMP or Fru-2,6-P_2 was detected from the X-ray determined structures. The folding of the two domains in the predicted model was partially based on the NMR studies.

CONCLUSIONS

It is clear that the final word on the determination of the active-site, and in the knowledge of the catalytic mechanism of an enzyme, will be given by the complete determination of the tertiary structure of the protein by X-ray diffraction methods.

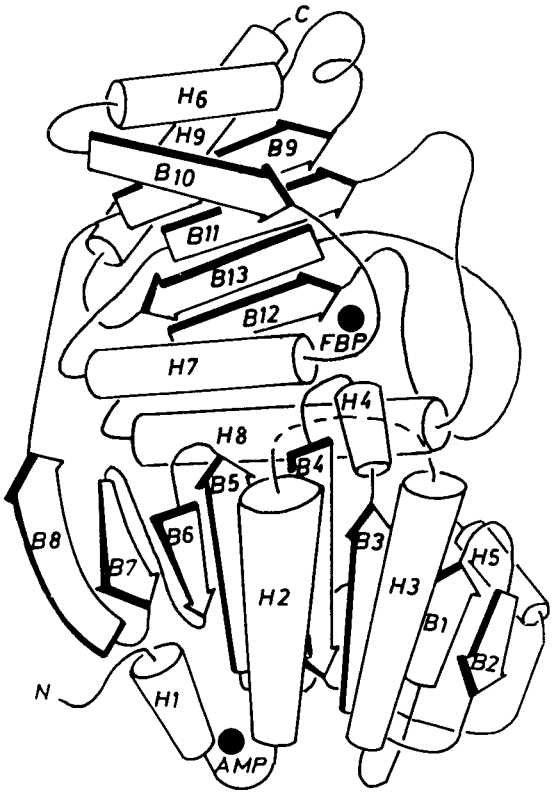

Fig 19. Secondary structure, as deduced from the X-ray structure, of one subunit (C1) of fructose-1,6-bisphosphatase from pig kidney in the cylinder and arrow representation. The letters N and C represent the N and C termini, respectively. The dotted line between H2 and H3 represent the missing loop of Ala55 to Gly 71. Two circles labeled FBP and AMP indicate the location of the active site and the AMP site, respectively. (From Ke *et al*, 1991; reprinted with permission). Note the similarity with Figure 15.

However, secondary structure studies and model building provide new means for the study of active and regulatory sites of enzymes in spite of their limitations, as shown on the examples here presented.

A reasonable good prediction of the secondary structure is not sufficient to obtain good results; a knowledge of the polypeptide chain is also needed. This problem can be overcome if the three-dimensional structure of related proteins are known. This information makes possible the use of molecular modeling, methods that have proved their validity, not only on unknown protein structures, but also in the determination of the mechanism of action of several enzymes of known structure.

ACKNOWLEDGMENTS

The support of research projects 020/83 and 1085/85 from CONICYT, 09/1987 from FONDECYT, and 2.09.81, 20.33.16, 20.31.12, 2.031.34, 91.31.43-1 and 93.31.51-1 from Dirección de Investigación, Universidad de Concepción, is gratefully acknowledged.

REFERENCES

AMBLER RP (1980) The structure of ß-lactamases. Phil Trans R Soc London B 289: 321-331

ANDERSON WF, MATTHEWS BW (1977) Crystallographic data for chicken liver fructose bisphosphatase. J Biol Chem 252: 556-557

ARRIAGADA E, CID H (1989) Search for a "toxic site" in snake venom phospholipases A_2. Arch Biol Med Exp 22: 97-105

BERNSTEIN FC, KOETZLER TF, WILLIAMS GJ, MEYER EF Jr, BRICE MD, BRUNIE S, BOLIN J, GEWIRT D, SIGLER PB (1985) The refined crystal structure of a dimeric phospholipase A_2 at 2.5 Å Access to a shielded catalytic center. J Biol Chem 260: 9742-9749

BUNSTER M (1984) A mechanism of action for ß-lactamases. Arch Biol Med Exp 17: R121.

BUNSTER M, CID H (1984) A model for the secondary structure of ß-lactamases. FEBS Lett 175: 267-274

BUNSTER M, CID H, VARGAS V, MARTINEZ J, BARROS L (1990) On the structure of Bacillus cereus ß-lactamase I: experimental test of a model. Med Sci Res 18: 729-733

CAMPOS M, GARCES E, MONTECINOS M, RUIZ J, WARD P (1979) Structural studies of a ß-lactamase from Streptomyces UCSM 104. Arch Biol Med Exper 12: 262

CHATTERJEE T, EDELSTEIN I, MARCUS F, EBY J, REARDON I, HEINRIKSON R (1984) Identification of the highly reactive sulfhydryl group of pig kidney fructose-1,6-bisphosphatase at cysteine 128. J Biol Chem 259: 3834-3837

CHOU PY (1989) Prediction of protein structural classes from amino acid composition. In: FASMAN GD (ed) Prediction of Protein Structure and the Principles of Protein Conformation. New York & London: Plenum Press. pp 549-597

CHOU PY, FASMAN GD (1974) Conformational parameters for amino acids in helical, ß-sheet, and random coil regions calculated for proteins. Biochemistry 13 211-245

CHOU PY, FASMAN GD (1978) Empirical prediction of protein conformation. Annu Rev Biochem 47: 251-276

CID H (1987) Active-site studies by secondary structure prediction. In: EYZAGUIRRE J (ed) Chemical Modification of Enzymes: Active-Site Studies. England: Ellis Horwood. pp 97-118

CID H, BUNSTER M, ARRIAGADA E, CAMPOS M (1982) Prediction of secondary structure of proteins by means of hydrophobicity profiles. FEBS Lett 190: 247-254

CID H, BUNSTER M, CANALES M, GAZITUA F (1992) Hydrophobicity and structural classes in proteins. Protein Engineering 5: 373-375

CITRI N, POLLOCK MR (1966) The biochemistry and function of ß-lactamase (penicillinase). Adv Enzymol 28: 237-323

CORREDOR C, BOSCA L, SOLS A (1984) Biphasic effect of fructose 2,6-bisphosphate on the liver fructose -1,6-bisphosphatase: mechanistic and physiological implications. FEBS Lett 187: 199-202

CUNNINGHAM BA, RAUSHEL FM, VILLAFRANCA JJ, BENKOVIC SJ (1981) Distances between structural metal ion, substrates, and allosteric modifier of fructose bisphosphatase. Biochemistry 20: 359-362

DALE JW, GODWIN D, MOSSAKOWASKA D, STEPHENSON P, WALL S (1985) Sequence of the OXA2 ß-lactamase. Comparison with other penicillin reactive enzymes. FEBS Lett 191: 39-44

DIJKSTRA B, KALK KH, HOL WG, DRENTH J (1981) The structure of bovine pancreatic phospholipase A_2 at 1.7 Å. J Mol Biol 147 : 97-122

DI NOLA A, ROCCATANO D, BERENDSEN HJC (1994) Molecular dynamics simulation of the docking of substrates to proteins. Proteins 19: 174-182

DUFTON PY, HIDER RC (1977) Snake-toxin secondary structure predictions. Structure activity relationships. J Mol Biol 115: 177-193

DUFTON PY, HIDER RC (1983) Classification of phospholipases A_2 according to sequence. Evolutionary and pharmacological implications. Eur J Biochem 173: 545-551

DUFTON PY, EAKER D, HIDER RC (1983) Conformational properties of phospholipases A_2. Secondary-structure prediction, circular dichroism and relative interface hydrophobicity. Eur J Biochem 137: 537-544

DURKIN JP, VISWANATHA T (1980) Bacillus cereus 569/H ß-lactamase. Structure-function relationship. Carlsberg Res Commun 45: 411-421

ELLIS RJ, VAN DER VIES SM(1991) Molecular chaperones. Annu Rev Biochem 60 321-347

EL-MAGHRABI MR, PILKIS J, MARKER AJ, COLOSIA AD, D'ANGELO G, FRASER BA, PILKIS SJ (1988) cDNA sequence of rat liver fructose-1,6-bisphosphatase and evidence for down-regulation of its mRNA by insulin. Proc Natl Acad Sci USA 85: 8430-8434

FISCHER WK, THOMPSON EDP (1983) Amino acid sequences studies on sheep liver fructose-bisphosphatase. II. The complete sequence. Aust J Biol Sci 36: 235-250

GANSON NJ, FROMM HJ (1985) Nuclear magnetic resonance studies of fructose 2,6 bisphosphate and adenosine 5'-monophosphate interaction with bovine liver fructose-1,6-bisphosphatase. J Biol Chem 260: 2837-2843

GREER J (1990) Comparative modeling methods: Application to the family of the mammalian serine proteases. Proteins 7: 317-334

HERZEBERG O, MOULT J (1987) Bacterial resistance to ß-lactam antibiotics: Crystal structure of ß-lactamase from *Staphylococcus aureus* PC1 at 2.5 Å resolution. Science 236: 694-701

HOLLEY LH, KARPLUS M(1991) Neural networks for protein structure prediction. Meth Enzymol 202: 204-224

JELSCH C, LENFANT F, MASSON JM, SAMAMA J-P (1992) ß-lactamase TEM1 of *E. coli*. Crystal structure determination at 2.5 Åresolution. FEBS Lett 299: 135-142

JELSCH C, MOUREY L, MASSON J-M, SAMAMA J-P (1993) Crystal structure of *Escherichia coli* TEM1 ß-lactamase at 1.8 Å resolution. Proteins 16: 364-383

KE H, THORPE C, SEATON BA, LIPSCOMB, WN, MARCUS F (1989) Structure refinement of fructose-1,6-bisphosphatase and its fructose 2,6-bisphosphate complex at 2.8 Å resolution. J Mol Biol 212: 513-539

KE H, ZHANG Y, LIPSCOMB WN (1990) Crystal structure of fructose-1,6-bisphosphatase complexed with fructose 6-phosphate, AMP, and magnesium. Proc Natl Acad Sci USA 87: 5343-5247

KE H, ZHANG Y, LIANG J-Y, LIPSCOMB WN (1991) Crystal structure of the neutral form of fructose-1,6-bisphosphatase complexed with the product fructose 6-phosphate at 2.1 Å resolution. Proc Natl Acad Sci USA 88: 2989-2993

KENDREW JC, DICKERSON RE, STRANDBERG BE, HART RG, DAVIES DR, PHILLIPS DC, SHORE VC (1960) Structure of myoglobin. Nature, London 185: 422-427

KITAJIMA S, UYEDA K (1983) A binding study of the interaction of ß-D-fructose 2,6-bisphosphate with phosphofructokinase and fructose-1,6-bisphosphatase. J Biol Chem 258: 7352-7357

KNOX JR, KELLY JA, MOEWS PC, MURTHY N (1976) 5.5 Å crystallographic structure of penicillin ß-lactamase and radius of gyration in solution. J Mol Biol 104: 865-875

LEVITT M, CHOTHIA C (1976) Structural patterns in globular proteins. Nature, London 261: 552-558

LIANG J-Y, HUANG S, ZHANG Y, KE H, LIPSCOMB WN (1992) Crystal structure of the neutral form of fructose 1,6-bisphosphatase complexed with regulatory inhibitor fructose 2,6-bisphosphate at 2.6 Å resolution. Proc Natl Acad Sci USA 89: 2404-2408

LIFSON S, SANDER C (1979) Antiparallel and parallel ß-strands differ in amino acid residue preferences. Nature, London 282: 109-111

LIM VI (1974) Algorithms for prediction of α-helical and ß-structural regions in globular proteins. J Mol Biol 88: 857-872

MARCUS F, EDELSTEIN I, REARDON I, HEINRIKSON RL (1982) Complete amino acid sequence of pig kidney fructose-1,6-bisphosphatase. Proc Natl Acad Sci USA 79: 7161-7165

MARTINEZ J (1984) Estudios de un modelo de estructura de ß-lactamasas. Biochemist's Thesis, University of Concepci≤n, Chile. (In Spanish)

MARTINEZ J, CID H (1984) Structure of ß-lactamases. Experimental test of a model. Arch Biol Med Exp 17: R156

MARTINEZ J, CID H (1986) A model for the structure of fructose-1,6-bisphosphatase from pig kidney. Arch Biol Med Exp 19: 77-83

MATTHEW M (1978) Properties of the ß-lactamase specified by the Pseudomonas plasmid R151. FEMS Microbiol Lett 4: 241-244

McCAMMON JA, HARVEY SC (1987) Dynamics of Proteins and Nucleic Acids. London: Cambridge University Press

MEEK DW, NIMMO HG (1983) The interaction of fructose 2,6-bisphosphate with an allosteric site of rat liver fructose 1,6-bisphosphatase. FEBS Lett 160: 105-109

MOEWS PC, KNOX JR, DIDEBERG O, CHARLIER P, FRERE J-M (1990) ß-lactamase of Bacillus licheniformis 749/C at 2 Å resolution. Proteins 7: 156-171

NAKASHIMA K, HORECKER BL (1971) Modification of the catalytic properties of rabbit liver fructose diphosphatase by a particulate fraction from liver. Arch Biochem Biophys 146: 153-160

NIMMO H, TIPTON K (1975) The effect of pH on the kinetics of beef- liver fructose bisphophatase. Eur J Biochem 58: 567-574

NOSTRAND B (1974) The high resolution structure of human carbonic anhydrase, form B. PhD Thesis, University of Uppsala, Sweden

PARKER W, SONG P-S (1990) Location of helical regions in tetrapyrrole-containing proteins by a helical hydrophobic moment analysis. J Biol Chem 265: 17568-17575

PHILLIPS DC, CORDERO-BORBOA A, SUTTON BJ, TODD RJ (1987) Crystallographic studies of the ß-lactamases from Bacillus cereus. Pure Appl Chem 59: 279-286

PILKIS SJ, EL-MAGHRABI MR, MC CRANE Mm, PILKIS J, CLAUS TH (1981) The role of fructose 2,6-bisphosphate in regulation of fructose-1,6-bisphosphatase. J Biol Chem 256: 11489-11495

PONNUSWAMY PK, PHRABHAKARAN M, MANAVALAN P (1980) Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. Biochim Biophys Acta 623: 301-316

PONTREMOLI S, HORECKER BL (1970) Fructose-1,6-diphosphatases. The Enzymes (3rd edn) 4: 611-646

PONTREMOLI S, MELLONI E, HORECKER BL (1983) The liver aldolase-fructose bisphosphatase complex. Biochem Soc Trans 11: 241-244

RENETSEDER R, BRUNIE S, DIJKSTRA BW, DRENTH J, SIGLER PB (1985) A comparison of the crystal structures of phospholipase A_2 from bovine pancreas and Crotalus atrox venom. J Biol Chem 260: 11627-11634

REYES A, HUBERT E, SLEBE JC (1985) The reactive cysteine of pig kidney fructose 1,6-bisphosphatase is related to a fructose 2,6-bisphosphate allosteric site. Biochem Biophys Res Commun 127: 373-379

SEATON BA, CAMPBELL RL, PETSKO GA, ROSE DR, EDELSTEIN I, MARCUS F (1984) Preliminary X-ray crystallographic studies of pig kidney fructose-1,6-bisphosphatase. J Biol Chem 259: 8915-8916

SESSIONS RB, DAUBER-OSGUTHORPE P, CAMPBELL MM, OSGUTHORPE DJ (1992) Modeling of substrate and inhibition binding to phospholipase A_2. Proteins 14: 45-64

SOLOWAY B, MC PHERSON A (1978) Molecular symmetry of fructose-1,6-bisphosphatase by X-ray diffraction analysis. J Biol Chem 253: 2461-2462

SUDA H, XU G-J, KUTNY RM, NATALINI P, PONTREMOLI S, HORECKER M (1982) Location of lysyl residues at the allosteric site of fructose 1,6-bisphosphatase. Arch Biochem Biophys 217: 10-14

SUMMERS NL, KARPLUS M (1990) Modeling of globular proteins. A distance-based data search procedure for the construction of insertion/deletion regions and Pro/non-Pro mutations. J Mol Biol 216: 991-1016

TSERNOGLOU D, PETSKO GA (1976) The crystal structure of a post synaptic neurotoxin from sea snake at 2.2 Å resolution. FEBS Lett 68: 1-4

VAN SCHAFTINGEN E, HERS HG (1981) Inhibition of fructose-1,6-bisphosphatase by fructose 2,6 bisphosphate. Proc Natl Acad Sci USA 78: 2861-2863

WALKINSHAW MD, SAENGER W, MAELICKE A (1980) Three-dimensional structure of the "long" neurotoxin from cobra venom. Proc Natl Acad Sci USA 77: 2400-2404

WAXMAN DJ, STROMINGER JL (1980) Sequence of active-site peptides from the penicillin sensitive D-alanine carboxypeptidase of *Bacillus subtilis*. Mechanism of penicillin action and sequence. J Biol Chem 255: 3964-3976

WESTERLUND B, NORLUND P, UHLIN U, EAKER D, EKLUND H (1992) The three-dimensional structure of notexin, a presynaptic neurotoxic phospholipase A_2 at 2.0 Å resolution. FEBS Lett 301: 159-164

WHITE SP, SCOTT DL, OTWINOWSKI Z, GELB MH, SIGLER PB(1990) Crystal structure of cobra-venom phospholipase A2 in a complex with a transition-state analog. Science 250: 1560-1563

XU G-J, NATALINI P, SUDA H, TSOLAS O, DZUGAJ A, SUN SC, PONTREMOLI S, HORECKER BL (1982) Rabbit liver fructose-1,6-bisphosphatase: Location of an active lysyl residue in the COOH-terminal fragment generated by a lysosomal proteinase. Arch Biochem Biophys 214: 688-694