Amino Acids (1994) 7: 175–202

Amino
Acids

Modelling of peptide and protein structures*

Review Article

S. Fraga[1] and **J. M. R. Parker**[2]

[1] Department of Chemistry and [2] Alberta Peptide Institute and Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada

Accepted August 12, 1993

Summary. The modelling of protein structures (whether isolated, in solution, or involved in recognition processes) is reviewed, free of any mathematical apparatus, to provide an overview of the concepts as well as leading references. A general feeling for this field of work is first established by a sampling of some impressions on its difficulties and chances of success. Then, the main body of this work examines the information available (databases and parameters), presents the theoretical foundations for the modelling procedures (with emphasis on the potential energy functions), surveys the existing simulation techniques and prediction methods, and discusses the problems still to be faced. For completeness, a representative list of existing software packages is presented in the Appendix.

Keywords: Amino acids – Peptides – Proteins – Modelling – Simulations

Introduction

The *a priori* prediction of peptide and protein structures will have far-reaching implications in academic as well as biotechnology research. Recently, this challenging research problem has been the centre of considerable attention. Success in this endeavour will complete the description of the genetic code and will have a social as well as an economic impact, as it will open the way for the development of new drugs, synthetic vaccines, and industrial enzymes. Protein engineering will endeavour to design new proteins or to change the structural and/or functional characteristics of existing peptides and proteins for specific purposes (van Gunsteren, 1988).

---

* Presented at the Third International Congress on Amino Acids, Vienna, August 23–27, 1993.

A staggering amount of original work has been carried out, resulting in a wealth of information and establishing *most* of the necessary formulations, which have been incorporated in a variety of software packages. The use of such packages as *black boxes* has expanded dramatically and it is now appropriate to present a conceptual review of the field. This information may help experimental researchers, interested in complementing, expanding or interpreting their work with simulations, in approaching such a task with realistic expectations.

A feeling for the state of this field may be obtained from a sampling of the impressions of active researchers. The advent of fast, large-capacity computers induced an early euphoria ['*Many are now racing to see who can be the first to calculate correctly the three-dimensional structure of an enzyme from its sequence*' (Bradley, 1970)] but the difficulties soon became evident ['*Even very short peptides (5–30 amino acids) are difficult to model accurately in the absence of structural data. There are simply too many degrees of freedom*' (Wilson and Klausner, 1984], to the point that the goal almost seemed unattainable ['*The prediction of the three-dimensional structure of a protein from its amino acid sequence remains one of the fundamental unsolved problems in molecular biology*' (Thornton, 1988)]. Many will be the rewards of solving this problem ['*Deciphering the rules through which amino acid sequences determine protein folding will be a major scientific and technological advance*' (King, 1989)] and therefore one should reexamine the fundamental considerations when tackling it ['*The computational task of protein structure prediction is believed to require exponential time, but previous arguments as to its intractability have taken into account only the size of a protein's conformational space. Such arguments do not rule out the possible existence of an algorithm, more selective than exhaustive search, that is efficient and exact*' (Ngo and Marks, 1992)]. Work nevertheless continues and progress is made, and although doubts still linger and refinements must be made ['*It remains to see whether some of the apparent advances made in the past few years are real or illusory and whether the protein folding problem is well on its way to being solved. One problem that clearly remains is the elucidation of the magnitude of the various forces that determine the delicate free energy balance between the folded and unfolded states of the protein* (Honig et al., 1993)], the problem will perhaps be finally solved ['*As is evident in the many efforts to characterize the structures and stabilities of various folding intermediates and their roles in folding reactions, it seems very likely that substantial progress will be made in solving the folding problem in the next few years. The development of a folding code, analogous to the genetic code, will allow us to expand the central dogma of molecular biology to the prediction of the 3-D structure adopted by a given amino acid sequence*' (Lecomte and Mathews, 1993)].

The description presented below, with a review of the various components in protein modelling, will confirm the above hopeful, but realistic, expectations. No effort has been made in presenting a complete survey of all the existing literature, but we are confident that the most important references are included or may be traced back from the ones given in this work. The reader entering this field for the first time should, first of all, get acquainted with the work of Richardson (1981), Jaenicke (1987), and Chothia (1990), describing folding and associations in proteins.

The organization of the topics is as follows:

1. Protein structures
   (a) Forces in protein folding
   (b) Databases and parameters
       Structure databases
       Physicochemical parameters
2. Theory of modelling protein structures
   (a) Potential energy functions
3. Simulation techniques for isolated proteins
   (a) Temperature-independent simulations
       A priori energy minimization
       Construction from fragments: Prediction of the secondary structure
       Homology-based methods
       The use of distance maps
       Preferred- or restricted-conformation database predictions
   (b) Temperature-dependent simulations
       Molecular dynamics simulations
       Monte Carlo simulations
4. Molecular associations
5. Prediction of interaction sites
6. Peptide mimetics

Protein structures

Proteins are molecules and their study, whether isolated, in solution, and/or in interaction with other systems, should be carried out as for any other molecule: that is, within the framework of Quantum Mechanics and Quantum Statistical Mechanics. Quantum chemical calculations are appropriate for temperature-independent studies but consideration of the temperature requires a Statistical Mechanics treatment.

At present, computational limitations restrict such a study to smaller molecules and the preceding statement must be viewed as the expression of an ideal. Therefore, in practice, the study of proteins must be carried out within the framework of various approximations, using macroscopic information and parameterizations. Many results obtained in this way are excellent and have proved to be extremely useful but it must always be remembered that they have been obtained by means of approximations. In such treatments, the calculations are no longer quantum chemical but the concept of energy minimization and the use of (semiempirical, parameterized) potential energy functions are retained.

The description below will focus on how the available, experimental information may be used in order to overcome the difficulties.

Forces in protein folding

The amino acid sequence of a protein, denoted as its primary structure, determines the final, spatial conformation that it will adopt when isolated (Anfinsen, 1973). Interaction with other systems may, naturally, result in some change. In

an *a priori*, quantum-chemical study, only the sequence would be required as input and the calculations would yield the spatial conformation with minimum energy.

Experimental observation has confirmed the existence of well-defined, structural units (α-helices, β-sheets, β-turns, γ-turns, etc.) in proteins. A knowledge of such units defines the secondary structure (SS) of the protein. These units may adopt certain relative conformations [such as helix-to-helix packing (Chothia et al., 1981), face-to-face β-sheet packing (Cohen et al., 1981; Lesk and Chothia, 1982), etc.], which represent the super-secondary structure.

The tertiary structure (TS) of the protein is the final, spatial conformation adopted by the isolated (solvated) protein, while the quaternary structure denotes the structure of a molecular complex.

The problem to be solved is to determine how the information contained in the sequence is translated into the final TS (Chothia, 1984; Kim et al., 1990; Mathews, 1991; Lecomte and Mathews, 1993). At the microscopic level, the driving forces for the folding are the interactions embodied in the corresponding Hamiltonian operator (containing electron kinetic, electron-nucleus attraction, and electron-electron and nucleus-nucleus repulsion energy terms). At the macroscopic level, the total interaction existing in a (solvated) protein is interpreted as consisting of electrostatic and van der Waals interactions (arising from the electron-nucleus and electron-electron interaction terms in the Hamiltonian operator) as well as of various effects, such as H-bonding (Baker et al., 1984), salt bridges, solvation, and hydrophobic forces, which are distinct manifestations of the true, existing interactions. (H-bonding, if existing, would appear in an accurate calculation, without having to impose any additional condition; salt bridges denote the strong Coulombic interactions between charged groups; and the designation 'hydrophobic forces' is used in order to express the competition between the internal interactions and those with the surrounding hydration shell.)

Within the framework of this terminology, these are the interactions that guide the folding and stabilize the final TS (Weisman and Kim, 1991) and the folding might be envisaged to occur as follows (King, 1989). First, local, weak interactions between hydrophobic residues adjacent in the sequence might decide the early steps in the folding, generating SS units. These units may be stabilized by association, involving H-bonding and additional hydrophobic interactions as well as interactions between the main-chain dipole moments of the SS units (Hol et al., 1981). Alternatively, one can envisage a hydrophobic collapse taking place first, followed by formation of the SS (Dill, 1990) but probably the hydrophobic collapse and the formation of SS units are strongly coupled (Honig et al., 1993) and the artificiality of their interpretive separation should be accepted as a way of trying to explain what is a rather complicated process. The folding evolves towards a conformation with tightly-packed internal residues (with a possible reorganization of the SS units), with a final reorganization of the whole structure in order to attain the correct conformation, both stable and active.

The fact that internal residues involved in cluster interactions are conserved during evolution substantiates the role of the hydrophobic domains in the

folding process (Plochocka et al., 1988). The existence of close-packed hydrophobic cores (Murzin and Finkelstein, 1988) leads to a useful, simplified interpretation of the folding process in terms of only the close packing (Ngo and Marks, 1992). Such a requirement rules out the majority of conformations available to a protein through dihedral angle variation and may be invoked to determine the side-chain conformation in the protein core (Lee and Subbiah, 1991). Thermodynamic studies have also investigated how to incorporate the effect of solvation in the folding/unfolding process of proteins (Makhatadze et al., 1990; Murphy et al., 1991).

Databases and parameters

As mentioned above, *a priori* calculations would require only the sequence of the protein. Approximate methods, however, need, or may benefit from, additional information, particularly experimentally determined TS but also diverse parameters.

Structure databases

X-ray crystallography has been the main source of data, but it must be emphasized that far more sequences of proteins are known than TS. The known 3D-structures, collected in the Brookhaven Protein Data Bank (PDB), are an essential prerequisite for some of the methods (such as the prediction by homology) and very valuable in the determination of interaction centres by molecular dynamics (MD) simulations.

The PDB is redundant, insofar as it includes several entries for very similar structures and consequently, it may be appropriate instead to use a selection of representative structures (Boberg et al., 1992). It may be used for the development of both relational databases (RDB) and object-oriented databases (ODB), with tables containing information that quantifies aspects of protein structure. Sequential search is faster using an RDB (Islam and Sternberg, 1989) but complex queries are faster with ODB (Gray et al., 1990).

Additional techniques [nuclear magnetic resonance (NMR), circular dichroism (CD), and Raman and Fourier transform infrared (FTIR) spectroscopy] are being increasingly used for the experimental determination of protein structures to enlarge the database of known structures (Gronenborn and Clore, 1990; Wüthrich, 1991; van Gunsteren et al., 1991; Dyson and Write, 1991). For peptides in solution, Raman and FTIR spectroscopy may give information on the population of the different conformers but the NMR data correspond to an ensemble of rapidly interconverting conformers.

Physicochemical parameters

A variety of parameters have been proposed and used over the years in order to quantify different characteristics of a protein. The most representative parameters are the hydrophilicity (Bull and Breese, 1974; Levitt, 1976; Hopp and Woods, 1981; Kyte and Doolittle, 1982; Parker et al., 1986), accessibility (Janin,

1979), flexibility (Karplus and Schulz, 1985), and recognition (Fraga, 1982). The hydrophilicity, accessibility, and flexibility factors are of experimental origin while the recognition factors were derived theoretically. These factors are given for the individual amino acids and the corresponding values for those amino acids, when forming part of a peptidic chain, are obtained according to given rules, which introduce the effect of the particular sequence under consideration. (These factors are used particularly for the prediction of the SS and antigenic determinants and the rules mentioned above may be found in the references given below when discussing those predictions.)

Theory of modelling protein structures

Computer simulations of protein structures should not be viewed as an end by themselves but rather as a complement to, or an extension of, experimental work. Within that context, they will serve a rather useful purpose, insofar as they may save time and efforts. In this connection it must be emphasized that the simulations should not be concerned exclusively with the structures but also with their functional characteristics, such as stability with respect to changes in pH, temperature, and solvent, binding capability with respect to its (modified) substrate, and catalytic properties (van Gunsteren, 1988).

A temperature-independent modelling constitutes the first step in a computer simulation, but it may be bypassed if an experimentally-determined TS is available. Such a simulation will yield a spatial conformation of the protein but very scarce information on its functional characteristics. That information will be obtained, however, through a dynamic modelling, with consideration of the temperature dependence.

In a temperature-independent simulation, the prediction of the tertiary structure is performed on the basis of a minimum-energy criterion, through an exhaustive search of the conformational space available to the peptidic chain by variation of its dihedral angles. Dynamic simulations, although not concerned with the prediction of stable conformations, involve nevertheless energy calculations. In both cases, the energy evaluation is carried out using an appropriate potential energy function (PEF). A representative list (see the Appendix) of the many PEF developed in the literature are those of Allinger (MM2), Clementi (MOTECC), Fraga (ALTA:maPS), Hagler (DISCOVER), Jorgensen (OPLS), Karplus (CHARMm), Kollman (AMBER), Scheraga (ECEPP), and van Gunsteren (GROMOS).

Potential energy functions

A satisfactory PEF should reproduce the interactions contained in the Hamiltonian operator while being simple and therefore efficient, in order that the computing cost of the considerable number of energy evaluations to be performed will not be prohibitive.

Consequently, a typical PEF should contain contributions from all those molecular components that may affect the energy of the conformation. Within the context of the terminology in use, an intramolecular PEF will include terms

for changes in bond lengths, bond angles, dihedral (torsion) angles as well as bending potentials and non-bonded interactions (i.e., interactions between non-bonded atoms). The latter, which are the only ones appearing in an intermolecular PEF, are the electrostatic interactions, van der Waals forces, three-body forces, and H-bonding. Depending on the desired level of accuracy and the computing capabilities available, some of the above terms may be disregarded; for example, bond lengths and bond angles may be maintained constant in the so-called fixed-geometry calculations (see below).

The formal expressions for all those terms, well established and reproduced abundantly in the literature, are common to all the existing PEF. The various PEF differ in the way in which the corresponding numerical parameters have been determined, in a variety of procedures ranging from semiempirical approximations to proper quantum-chemical calculations. Because approximations are inherent to all the PEF, their quality should be judged on the basis of the results that they yield.

While dynamic simulations should be performed with intramolecular PEF containing most, if not all, of the above terms, static calculations are usually carried out with fixed geometry (i.e., with fixed bond lengths and bond angles), with a PEF consisting of torsion angles, electrostatic, and van der Waals contributions and, in some cases, the H-bonding terms.

Because the bond-length, bond-angle, and dihedral-angle terms are formally simple, we will discuss here only the electrostatic and van der Waals terms. For simplicity in the calculations, these interactions are usually expressed as summations of atom-atom pair potentials, which are given by 1/R expansions (where R is the separation between the two interacting atoms). This type of approach is exemplified, in its simplest form, by the Lennard-Jones potential and may be considered to be supported by the theoretical development of Buckingham (1963). Nevertheless, a 1/R expansion represents an approximation.

The predominant term in the expansion is the one corresponding to the electrostatic Coulomb (repulsive/attractive) interaction, $q_i q_j / R_{ij}$, where q_i, q_j denote the effective charges on the two atoms. These charges, although very useful and of an intuitive character, are of a rather artificial origin. They depend not only on the method (semiempirical or *ab initio*) and additional approximations (basis set expansion, etc.) introduced in the calculation of the molecular wave function but also on the population analysis used to predict them. An additional problem in the use of effective charges stems out of the impossibility of performing preliminary molecular calculations for the system(s) under study, as already mentioned above. Consequently, recourse must be made to the use of average values and in this connection the definition of atom classes, taking into account their molecular environments, represents a considerable advance (Clementi, 1980), giving support to the use of average values. There is, in any case, an additional deficiency, regarding the need of an appropriate dielectric constant (Harvey, 1989). A positive aspect, however, of the use of a dielectric constant is that, when properly chosen, it may help in simulating the effect of the solvent (Zimmerman, 1985).

The second, predominant term is the one required in order to eliminate the possibility of a collision of the atoms (or the molecules, in the case of molecular

associations) under the influence of attractive forces. In proper quantum-chemical calculations such a problem does not exist because of the repulsion between the interacting electronic density distributions as well as between the nuclei. In an 1/R-expansion, a corresponding (repulsive) terms must be arbitrarily introduced. Customarily, it has been approximated as $1/R^{12}$, with an appropriate coefficient, which ensures a short-range contribution.

Additional terms which may (and should) be included are those in $1/R^6$ and $1/R^4$, both of them attractive, associated with the dispersion and induction interactions (see, e.g., Fraga, 1982).

The best approach for the determination of the coefficients of a 1/R expansion is the one adopted by Clementi and co-workers (Clementi et al., 1977, Scordamaglia et al., 1977; Sordo et al., 1986, 1987; Iglesias et al., 1991). The corresponding expansions still represent approximations but they have a well-defined, theoretical foundation.

Simulation techniques for isolated proteins

Once the central role of the PEF in protein modelling is established and the assumption is made of the availability of an acceptable PEF, it is necessary to decide on the simulation procedure to be adopted. Before proceeding with the analysis of the various techniques, it is convenient to examine first what are the questions to be answered, what are the appropriate techniques, and what are the inherent approximations (Osguthorpe, 1989).

First of all, it must be emphasized that if the temperature dependence of the molecular properties is to be evaluated, active sites are to be pinpointed, and/or the solvent around the protein is to be considered, a dynamic simulation should be contemplated. That is, one should perform either a Monte Carlo (MC) or a molecular dynamics (MD) calculation, or a combination thereof.

If no starting TS is available for such a simulation or if only a prediction of the TS is needed, a temperature-independent simulation may be required/sufficient. Ideally, a total *de novo* prediction, starting from only the sequence, should constitute the ultimate goal. Unfortunately, this problem has not been solved yet and approximations must be considered. No pejorative connotation should be ascribed to the term 'approximations' in this instance, because they may lead to very successful predictions.

The first simplification consists of regarding the protein as a set of SS units which are assembled together, under a variety of criteria, to form the TS. Such an approach may be carried out in two ways, depending on the *a priori* level which is possible. For example, in a purely theoretical approach one would perform first a prediction of the SS of the protein, proceed then to construct the corresponding SS units, and finally assemble them. Such an approach is dependent on the availability of a procedure for a satisfactory prediction of the SS, which is the reason for the considerable attention that this problem has attracted.

In a more semiempirical approach, the SS elements to be assembled might be obtained from homologous proteins (HP), of known structure. Such a procedure represents a simplified version of the modelling by homology (MH), whereby the TS of the protein is approximated from those of HP. MH requires the

availability of alignment techniques, which will identify the conserved regions in the family of HP.

At the end of the scale, in terms of the *de novo* level of the procedure, are the techniques based on the use of distance maps.

In all the cases of semiempirical modelling, the TS initially obtained may be relaxed/optimized under a minimum-energy criterion.

Temperature-independent simulations

(a) A priori energy minimization

Formally this is a very simple procedure but it suffers from two fundamental deficiencies. On one hand it is doubtful that a brute force search of the conformational space available to the protein through variation of the dihedral angles will be able to locate the global minimum, that is, the conformation with minimum energy. This statement must be qualified with the consideration that the search for the global minimum is based on the premise that it corresponds to the native conformation; such a premise is not completely true, as the experimental conformation of the protein is the result not only of its own, internal interactions but also of the interactions with its surroundings (which include other cellular components, such as the membrane, and the solvent in which it may be immersed). The other obstacle to be overcome is the considerable computing time required, which may be prohibitive for large proteins. This fact, by itself, encourages the use of simplifications in the modelling.

The reader not familiar with this subject is advised to consult the comprehensive work of Scheraga and co-workers [see, e.g., the early work of Anfinsen and Scheraga (1975) and the more recent one of Roterman et al. (1989a, b)], which offers the most illustrative example of the evolution in this field, as well as with the work of Coghlan and Fraga (1985), which provides an illustrative example of the type of computer program used in such simulations. In its simplest version, an energy-minimization procedure is performed through successive variation of all/some of the dihedral angles of the protein, with simultaneous evaluation of the corresponding conformational energy (on the basis of the PEF, which has been adopted). The search is continued until a minimum energy is reached. Unfortunately, because of the complicated character of the corresponding potential energy hypersurface, the procedure may lead to a local minimum and not necessarily to the global minimum. Even so, the corresponding TS may be appropriate as starting point for a MD calculation, which will change it abruptly if it is unstable. In fact, the equivalent of annealing (but without a temperature dependence) may be used in an energy minimization procedure: once an energy minimum has been reached, (a) strong change(s) may be introduced in some dihedral angles and the energy minimization continued; escape from a local-minimum conformation may be achieved in this way.

Some comments are worthwhile regarding the dihedral angles to be varied. Usually, the ω angle is left unchanged (at 180° for all natural amino acids except proline and 180° or 0° for proline). The angles ψ and ϕ should always be changed but the side-chain dihedral angles may be maintained fixed in a first approxima-

tion, varying them in a subsequent refinement of the resulting TS; in most cases it may suffice to vary only the first χ angle.

It must be added that the brute force search is appropriate, and may yield satisfactory results, for short peptides.

(b) Construction from fragments: Prediction of the secondary structure

As already mentioned, the brute-force approach inherent to the energy-minimization procedure may be made more efficient through the use of prebuilt fragments, which are then assembled together under a variety of criteria (see below).

Whenever the protein under study belongs to a family of HP with known TS, the corresponding fragments may be taken from those proteins. The assumption made here is that those regions with (almost) the same sequence will have the same SS in all the proteins of the family. In the extreme case of proteins belonging to a specific class (α-helix, β-sheet, or α-helix/β-sheet), this statement has been generalized by the suggestion that their TS are directly determined by their SS (Ptitsyn and Finkelstein, 1980).

In all other cases, when a total, *de novo* prediction is to be performed, the construction from fragments is based on the use of SS units. Consequently, the prediction of the SS constitutes the first step in this approach.

A wide variety of methods have been developed for the prediction of the SS and here only a very general outline will be presented. The reader is referred, for more details and additional references, to the review of Richardson (1981) on secondary structures and the work of Kabsch and Sander (1983), Rose et al. (1985), Thornton (1988), Yada et al. (1988), Pascarella et al. (1990), and Huang et al. (1990).

The various prediction methods may use probabilistic or physicochemical considerations, structures of HP, neural networks, or suitable combinations of the above. The most popular probabilistic methods (based on the properties of single amino acids) are those developed by Chou and Fasman (1974a, b, 1978a, b, 1979) and Garnier, Osguthorpe and Robson (GOR) (Garnier et al., 1978). These methods have been modified extensively throughout the years and adapted for use on microcomputers, a fact that has contributed to their wider usage. [See the review work of Pascarella et al. (1990) for details and references.] The physicochemical method of Lim (1974), based on the recognition of sequence patterns, is rather effective but it is hindered by the complicated set of rules on which it is based. Pattern recognition in general (Rooman and Wodak, 1988) and the related knowledge-based neural network procedures (Qian and Sejnowski, 1988; McGregor et al., 1989; Kaden et al., 1990; King et al., 1990; Kneller et al., 1990) probably hold the greater promise, subject to the availability of more protein structures. The combined method of Biou et al. (1988) makes use of the GOR and homology procedures, with a bit pattern method added whenever no agreement is found by the other two methods, while the procedure proposed by Fraga and co-workers (Thornton et al., 1991a, b; Fraga et al., 1990) is based on a combination of the Lim and GOR algorithms, supplemented by consideration of the recognition factors. Similarly, a procedure developed by Parker and Hodges (Parker et al., 1986; Parker and Hodges, 1991a, b) uses a combination of HPLC

hydrophilicity, accessibility, and flexibility parameters to predict surface/turn/loop regions; the secondary structural regions between the surface or break regions are then predicted using a combination of Lim and Chou-Fasman algorithms. Altogether, the general conclusion is that the problem is far from solved and that *the best approach consists of using all the available information* (Thornton, 1988).

Once the prediction of the secondary structure has been made, the corresponding α-helix and β-chain units are first constructed. This task is easily completed on the basis of the geometric characteristics of those units, without any need whatsoever of energy calculations. These units are then assembled, with the connecting joining loops built in an arbitrary way, and finally the resulting TS is optimized under a minimum-energy criterion; in this process, the geometry of the structural units is maintained unchanged and the only dihedral angles to be varied are those of the joining loops. Consequently, due to the considerable reduction in the number of variables, the calculation is far more efficient and less expensive than the corresponding *a priori* energy-minimization procedure. The quality of the predicted TS will depend, naturally, on the quality of the SS prediction.

(c) Homology-based methods

In these methods the TS of the protein under study is constructed (see below) from the structure(s) of homologous protein(s), on the basis that the TS of proteins are much less variable than the sequence (Osguthorpe, 1989).

The first step in this procedure consists of a homology search, which at present may be performed in a rather routine and efficient way. [See, e.g., the reviews of Argos et al., 1991 and Tyler et al., 1991 and the work of Bacon and Anderson, 1986; Sander and Schneider, 1991; Schuler et al., 1991; Orengo et al., 1992; Russell and Barton, 1992; Saqi et al., 1992.] With a sequence identity over 50%, the correspondence may be pinpointed unambiguously, but a smaller identity may lead to difficulties, especially if part(s) of one sequence is (are) missing in the other (Taylor, 1988). Solvent-inaccessible cores (Hubbard and Blundell, 1987) are more conserved within a family of homologous proteins than the proteins as a whole and therefore they may constitute a better initial framework on which to build the structure of the protein under study. But, in any case, it must be emphasized that *a structure prediction based on local sequence similarity may not he appropriate unless the homology is associated with an evolutionary or functional correspondence* (Sternberg and Islam, 1990).

Once the sequence alignment has been performed and the conserved regions have been located, one could proceed as follows (Osguthorpe, 1989). The aligned regions are ascribed the structure in the homogous protein(s) (if completely identical) or they are mutated (if not completely identical) to the new sequence, retaining the original structure. Those regions are then joined by loops (with the sequence in the protein under consideration), whose structure may be determined under different criteria, such as meeting the spatial requirement of fitting between the N- and C-termini of the two secondary structure sections to be joined (Jones and Thirup, 1986), or through a systematic search based on energy

and packing considerations (Moult and James, 1986; Dudek and Scheraga, 1990). Alternatively, one could determine the structure in a single try, by fitting the sequence of the protein under study onto the backbone of the known structure and performing an energy optimization; such a procedure may be automated through the use of a library of protein folds, all of which are tested for the sequence under consideration, with the most probable match selected under a minimum-energy criterion (Jones et al., 1992).

(d) The use of distance maps

Distance constraints may be used in various ways in the prediction of protein structures, using information from crystallographic or other experimental sources or from semiempirical considerations.

'Statistically' averaged distances between residues may be obtained from the known, crystallographic structures of proteins. Then, given a new protein, the corresponding contact map may be constructed from such distances and the map may be used in order to locate the compact regions of the protein (Kikuchi et al., 1988a, b). Contact maps may also be used in conjunction with an optimization procedure (Ycas, 1990; Godzik et al., 1992).

Distance geometry methods may gain wider acceptance as more and more 2D- and 3D-NMR studies of proteins become available. These studies yield interatomic distances, which may be used as constraints in an optimization procedure. Such a semiempirical approach will produce an energy-optimized structure that satisfies the known distance constraints and therefore one may expect a rather successful prediction (Crippen and Havel 1988; Havel and Snow 1991). The calculations, however, require considerable computer time and therefore the efforts in this area will be directed towards the search for more efficient algorithms (Glunt *et al.*, 1993).

(e) Preferred- or restricted-conformation database predictions

The use of a restricted- or probable-conformation database (Wu and Kabat, 1973; Lambert and Scheraga, 1989; Unger et al., 1990) reduces the large number of possible conformational searches and focuses these searches in a more probable conformational space. In addition, the use of a simplified rigid-geometry model, where the number of degrees of freedom is reduced, has the advantage of further reducing the computing time (Palmer and Scheraga, 1991; Rooman et al., 1991). These methods are used to generate a large number of conformations, in order to avoid locking a predicted conformation into a local energy minimum.

Temperature-dependent simulations

All the techniques described above, even if free of approximations, will not yield a complete description of a protein and may even fail in certain cases, simply because they do not include the temperature dependence. Thus, they might succeed in predicting the correct crystallographic structure of a given protein

but they may not be able to provide an insight into its functional characteristics. On the other hand, they may fail completely for rather flexible peptidic chains, unless the simulation is performed with extreme care (Thornton and Fraga, 1991; Fraga and Thornton, 1993; Fraga et al., 1993). This situation may be remedied by carrying out temperature-dependence calculations, such as those embodied in the MC and MD approaches. [See, e.g., the work of Clementi (1990) for complete details and availability of programs for these two techniques.]

(a) MD simulations

MD simulations for proteins (McCammon and Harvey, 1987; Brooks III et al., 1988) are based on the use of the classical Newtonian equations of motion and require, as initial input, the set of atomic coordinates and an estimate of corresponding velocities. The initial atomic coordinates may be those obtained from crystallographic data or those evaluated in temperature-independent calculations. The continued integration of the equations of motion generates the coordinates and velocities of the atoms as functions of time, at the chosen temperature. Usually, the calculations will proceed to completion in a rather routine way, but it may happen that local overheating will occur, requiring corrective measures (McCammon et al., 1979). *The information to be obtained in these simulations is not to be found in a single conformation but rather in the complete trajectory.*

The main limitations of these calculations are the computer-time requirements and the approximations that may be present in the potential energy function to be used. Some comments, in addition to those already made above, must be made concerning the latter. In a complete, unconstrained calculation, the equations of motion are expressed in terms of the Cartesian coordinates of the atoms and the potential energy function will include the terms corresponding to the variations in the bond lengths, bond angles, and dihedral angles (in addition to all the other terms mentioned above). A considerable reduction in the required computing time may be achieved by adoption of a fixed geometry, that is, neglecting the changes in both the bond lengths and the bond angles (Noguti and Go, 1983a, b; Gibson and Scheraga, 1989, 1990), but the inclusion of bond-angle constraints is not recommended (van Gunsteren and Berendsen, 1977; van Gunsteren and Karplus, 1982).

MD simulations are useful in refining crystallographic structures, determining protein structures from NMR data, and evaluating the free-energy changes resulting from mutations (Karplus and Petsko, 1990; Hermans et al., 1992). In addition, a graphic representation (see, e.g., Nilsson 1990 and references therein) may help in understanding the biological function of the protein; in particular, it will show which hydrophobic pockets, inaccessible in the crystal structure, may open up along the trajectory, thus identifying them as possible candidates for the active site. [Simplified calculations may be restricted to the subset of atoms involved in that site and its surroundings (Karplus and Petsko, 1990).] Another connection between MD simulations and NMR experiments is provided by the information obtained on H-bond stability, having been confirmed that the most stable H-bonds involve the H-atoms observed to exchange most slowly with the solvent (Levitt, 1981).

(b) MC simulations

Most of these simulations are based on the formulation, or variations thereof, of Metropolis et al. (1953). The procedure, which generates a sufficiently large number of random conformations, yields a sampling of the conformation space according to a Boltzmann distribution at the temperature considered. *The structural, statistical, and thermodynamical properties of the protein are then evaluated as weighted averages over all those conformations.* MC simulations are very appropriate for the study of flexible peptides, which cannot be characterized by a unique, or a few discrete, structure(s) (Hagler, 1985). These simulations suffer from the same practical limitations as the MD simulations (see above). A reduction in the computing time may be achieved through the constraint of fixed geometry (Noguti and Go 1985) and by restricting the optimization to the side chains (Holm and Sander, 1992), a procedure which is of interest in building by homology.

MC variants are the MC minimization (MCM) procedure (Li and Sheraga, 1987, 1988) for the determination of the global minimum and the MCM method coupled with a thermalization process (MCMT), which allows for the escape from local minima (Caflisch et al., 1992). [See also the work of Snow (1992) on a simulated-annealing algorithm for the search of the global minimum starting from random conformations and the modified MD method with MC sampling of Morley et al. (1992), in which trial conformations are generated by high-T dynamics. It must be mentioned, however, that the latter procedure has not been applied yet to proteins.]

Molecular associations

Whether in nature or in industrial processes, proteins are not found isolated and therefore their study must take into account their environment (such as the solvent and other cell components). In addition, many proteins are involved in recognition process (such as enzymatic processes and the recognition of antigens). The common characteristic of all these cases is the existence of molecular associations, with additional interactions to be taken into account (Jaenicke, 1987).

The conformation adopted by the protein is the one corresponding to the lowest free energy for the complete system: that is, with consideration of the interactions existing within the protein itself, the interactions of the protein with the solvent and whatever ligands (substrate, antibody) may exist, and the interactions between the solvent molecules themselves and between the solvent and the ligand(s).

Consequently, only MD or MC simulations, which will allow all the possible interactions to manifest themselves, are appropriate for these studies (Goodsell and Olson, 1990; Caflisch et al., 1992). Useful information may already be obtained, however, by simplified approaches. For example, in the case of interaction with ligands it may be sufficient to allow for variable geometry during the docking procedure (Leach and Kuntz, 1992). Even the hydration of a protein with fixed geometry will identify the hydrophilic/hydrophobic regions of the protein surface and thus help in locating the antigenic determinants. And the

folding and stability of proteins in non-aqueous solvents may be discussed through a set of rules obtained by modification of those applying to aqueous solutions (Arnold, 1988).

Prediction of interaction sites

The interaction sites of a protein may be denoted as receptor sites and recognition sites, respectively, on the basis of their role: ligands dock at the receptor site of the recognizing protein while recognition sites are those regions recognized by the interacting molecule. The best example for this distinction is provided by the major histocompatibility complex (MHC) molecules: thus, e.g., the class I MHC molecules have a receptor site that binds the processed antigen but they simultaneously offer (on their surface) antigenic determinants which will be recognized if cells presenting those molecules on their membranes are transplanted (whence their designation of transplantation antigens).

As mentioned above, identifying the receptor site of a protein requires a time-consuming MD/MC calculation. At a low level of approximation, however, a tentative identification may be attempted from an existing TS and additional semiempirical considerations [such as, for instance, the location of hydrophobic centres (Fraga et al., 1990)].

On the other hand, the prediction of antigenic determinants, which has received far more attention, has been approached rather successfully by means of a variety of approximate, semiempirical methods. The immune response of an organism to a foreign antigen consists primarily of a cell-mediated mechanism, involving molecules of the immunoglobulin supergene family, expressed on the surface of antigen-presenting cells (APC), antibody-producing B cells, and T4 (helper T lymphocytes, T_h) and T8 (cytotoxic T lymphocytes, CTL) cells. The antigen, processed by APC, is displayed in the form of a peptide (Bjorkman et al., 1987a, b; Brown et al., 1988), corresponding to an immunodominant region, in association with class I or class II of the MHC (Dausset, 1981; Kimball and Coligan, 1983; Kaufman et al., 1984; Klein, 1986). The T-cells, specific for either class of MHC molecules (class II for T_h and class I for CTL) use similar receptors (Rupp et al., 1985; Marrack and Kapler, 1986).

Antibody epitopes are believed to be non-sequential, i.e., conformational, being formed by non-adjacent sequence regions brought together by the folding of the active structure. Consequently, predictive methods have focused on surface characteristics, such as hydrophilicity (Hopp and Woods, 1981; Parker et al., 1986), accessibility (Janin, 1979), mobility (Westhof et al., 1984; Tainer et al., 1984), and protrusion (Thornton et al., 1986). A knowledge of the tertiary structure of the protein is needed, however, to completely identify the conformational epitopes.

T-cell epitopes, on the other hand, are linear, i.e. sequential, and may appear either on the surface or in interior regions of the protein. The MHC receptor site may accommodate an α-helical peptide of about 20 residues or an extended chain of about 8 residues; if the conformation of the bound peptide is bent, or only partly helical with unfolded ends, it may have an intermediate number of residues (Bjorkman et al., 1987a, b; Brown et al., 1988; Schwartz et al., 1985).

Because in an individual, either class of MHC molecules must bind a very large number of foreign peptides, it is believed that the latter may share a common conformation. Thus, the main criteria used for the prediction of T-cell epitopes are the amphipathic α-helical propensity (Watts et al., 1985; DeLisi and Berzofsky, 1985; Berkover et al., 1986; Berzofsky et al., 1987; Cease et al., 1987; Cornette et al., 1989) or the existence of general sequence patterns or motifs (Rothbard, 1986; Lamb et al., 1987; Sette et al., 1988) or MHC allele-specific consensus sequences (Guillet et al., 1987; Rothbard et al., 1988). The prediction of amphipathic α-helical regions in a protein may be attempted in a two-step procedure: first, a prediction of its secondary structure is made (see above) and then the predicted α-helices are examined for amphipathic character, using either hydrophilicity or pseudo-hydrophilicity factors, the latter obtained as averages of hydrophilicity and accessibility factors. The amphipathic character of an α-helix may also be predicted by application of discrete Fourier transforms (Eisenberg et al., 1984; Finner-Moore and Stroud, 1984), or by a least-square approximation (Margalit et al., 1987). [Some of these methods have been implemented on microcomputers; see the work of Pascarella et al. (1990) for references.]

Peptide mimetics

As mentioned above, one of the goals of protein engineering is the development of new drugs. For this purpose, an additional simulation step is required because of the desirability of an oral administration of the drugs. Naturally-occurring peptides are not appropriate in this respect and it is necessary to search for a(n) (organic) molecule which will mimick the function of the peptidic chain.

When the enzyme, receptor or substrate structures are known, the *de novo* design of conformationally constrained peptides or peptide mimetics and small organic molecules with high binding affinity based on this information can lead to new therapeutic compounds (Corey et al., 1991; Owens et al., 1991; Saragoui et al., 1991; Jenks, 1992).

The corresponding simulation must be performed by means of quantum chemical methods, even if at a low level of sophistication. As a rule, it is not difficult to design a new molecule in which given functional groups are positioned according to a given geometry. This optimistic assessment must, however be tempered as follows: it may be necessary to design a number of such molecules, with the final selection made on the basis of pharmacological and clinical considerations [see, e.g., Weaver (1992) for the general strategy in the rational design of drugs].

The main difficulty in this task is that the use of quantum chemical programmes is not as straightforward as that of the software packages for the modelling of proteins. Some knowledge of Quantum Chemistry and of its terminology is required.

Conclusions

The summary of quotes, presented in the introduction, together with the outline developed in the remainder of this work have already emphasized the two main

conclusions:

(a) The complex problem of peptide and protein modelling is far from solved, in spite of the great advances of the last few years.
(b) The best approach is to combine and make use of all the available information, whether experimental, theoretical, or computational.

In addition, some comments on the use of software packages may be appropriate:

(a) The theoretical/computational component of research in this field should be considered as an auxiliary tool, in order to interpret, complement, and extend the experimental research at a reduced cost of time and efforts.
(b) The users of modelling software packages should be acquainted with the approximations and limitations of the latter.
(c) The usage of packages for proper quantum chemical calculations, in the case of peptide mimetics, may require some knowledge of computational chemistry.

References

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181: 223–230
Anfinsen CB, Scheraga HA (1975) Experimental and theoretical aspects of protein folding. Adv Protein Chem 29: 205–300
Argos P, Vingron M, Vogt G (1991) Protein sequence comparison: methods and significance. Protein Eng 4: 375–383
Arnold FA (1988) Protein design for non-aqueous solvents. Protein Eng 2: 21–25
Bacon DJ, Anderson WF (1986) Multiple sequence alignment. J Mol Biol 191: 153–161
Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 44: 97–179
Berkover I, Buckenmeyer GK, Berzofsky JA (1986) Molecular mapping of a histocompatibility-restricted immunodominant T cell epitope with synthetic and natural peptides: implications for T cell antigenic structure. J Immunol 136: 2498–2503
Berzofsky JA, Cease KB, Cornett JL, Souge JL, Margalit H, Berkover IJ, Good MF, Miller LH, DeLisi C (1987) Protein antigenic structures recognized by T cells: potential applications to vaccine design. Immunol Rev 98: 9–52
Biou V, Gibrat JF, Levin JM, Robson B, Garnier J (1988) Secondary structure prediction: combination of three different methods. Protein Eng 2: 185–191
Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987a) Structure of the human class I histocompatibility antigen HLA-A2. Nature 329: 506–512
Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Willey DC (1987b) The foreign antigen binding site and T-cell recognition regions of class I histocompatibility antigens. Nature 329: 512–518
Boberg J, Salakoski T, Vihinen M (1992) Selection of a representative set of structures from Brookhaven Protein Data Bank. Proteins Struct Funct Genet 14: 265–276
Bradley DF (1970) Recognition polymers. J Macromol Sci-Chem A4: 741–755
Brooks III CL, Karplus M, Pettit BM (1988) A theoretical perspective of dynamics, structure and thermodynamics. Wiley, New York
Brown JH, Jardetzky T, Saper MA, Samraoui B, Bjorkman P, Wiley DC (1988) A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. Nature 332: 845–850
Buckingham AD (1963) Permanent and induced molecular moments and long-range intermolecular forces. Adv Chem Phys 12: 107–142

Bull HB, Breese K (1974) Surface tension of amino acid solutions: a hydrophilicity scale of amino acid residues. Arch Biochem Biophys 161: 665–670
Caflisch A, Niederer P, Anliker M (1992) Monte Carlo minimization with thermalization for global optimization of polypeptide conformations in Cartesian coordinate space. Proteins Struct Funct Genet 14: 102–109
Cease KB, Margalit H, Cornette JL, Putney SD, Robey WG, Ouyang C, Streicher HZ, Fischinger PJ, Gallo RC, DeLisi C, Berzofsky JA (1987) Helper T-cell antigenic site identification in the acquired immunodeficiency syndrome virus gp 120 envelope protein and induction of immunity in mice to the native protein using a 16-residue synthetic peptide. Proc Natl Acad Sci USA 84: 4249–4253
Chothia C (1984) Principles that determine the structure of proteins. Annu Rev Biochem 53: 537–572
Chothia C (1990) The classification and origins of protein folding patterns. Annu Rev Biochem 59: 1007–1039
Chothia C, Lewitt M, Richardson D (1981) Helix to helix packing in proteins. J Mol Biol 145: 215–250
Chou PY, Fasman GD (1974a) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochemistry 13: 211–222
Chou PY, Fasman GD (1974b) Prediction of protein conformation. Biochemistry 13: 222–245
Chou PY, Fasman GD (1978a) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol 47: 45–148
Chou PY, Fasman GD (1978b) Empirical predictions of protein conformation. Annu Rev Biochem 47: 251–276
Chou PY, Fasman GD (1979) Prediction of β-turns. Biophys J 26: 369–384
Clementi E (1980) Computational aspects for large chemical systems. Springer, Berlin Heidelberg New York, 184 pp
Clementi E (editor) (1990) Modern techniques in computational chemistry: MOTECC-90. Escom, Leiden, pp 805–808
Clementi E, Cavallone F, Scordamaglia R (1977) Analytical potentials from 'ab initio' computations for the interaction between biomolecules. 1. Water with amino acids. J Am Chem Soc 99: 5531–5545
Coghlan B, Fraga S (1985) Determination of proteinic structures: an experimentation program. Comput Phys Commun 36: 391–399
Cohen FE, Sternberg MJE, Taylor WR (1981) Analysis of the tertiary structure of protein β-sheet sandwiches. J Mol Biol 148: 253–272
Corey RI, Altman KH, Mutter M (1991) Protein design: template-assembled synthetic proteins. Ciba Foundation Symp 158: 187–212
Cornette JL, Margalit H, DeLisi C, Berzofsky JA (1989) Identification of T-cell epitopes and use in construction of synthetic vaccines. Meth Enzymol 178: 611–634
Crippen GM, Havel TF (1988) Distance geometry and molecular conformation. Wiley, New York
Dausset J (1981) The major histocompatibility complex in man. Science 213: 1469–1474
DeLisi C, Berzofsky JA (1985) T-cell antigenic sites tend to be amphipathic structures. Proc Natl Acad Sci USA 82: 7048–7052
Dill KA (1990) Dominant forces in protein folding. Biochemistry 29: 7133
Dudek MJ, Scheraga HA (1990) Protein structure prediction using a combination of sequence homology and global energy minimization. I. Global energy minimization of surface loops. J Comput Chem 11: 121–151
Dyson HJ, Wright PE (1991) Defining solution conformations of small linear peptides. Annu Rev Biophys Biophys Chem 20: 519–538
Eisenberg D, Weiss RM, Terwilliger TC (1982) The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature 299: 371–374
Finner-Moore J, Stroud RM (1984) Amphipathic analysis and possible formation of the ion channel in an acetylcholine receptor. Proc Natl Acad Sci USA 81: 155–159

Fraga S (1982) Theoretical prediction of protein antigenic determinants from amino acid sequences. Can J Chem 60: 2606–2610

Fraga S, San Fabian E, Thornton S, Singh B (1990) Prediction of the secondary structure and functional sites of major histocompatibility complex molecules. J Mol Recogn 3: 65–73

Fraga S, Thornton SE (1993) Theoretical studies of peptidic structures. Environmental effects. Theoret Chim Acta 85: 61–67

Fraga S, Thornton SE, Singh B (1993) Elucidation of peptide conformation involved in the recognition of $(EYA)_5$, $EYK(EYA)_4$ and $EYAEAA(EYA)_3$ peptides by MHC class II molecules and T-cell receptor. J Mol Struct (Theochem) (in press)

Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120: 97–120

Gibson KD, Scheraga HA (1990a) Variable step molecular dynamics: an exploratory technique for peptides with fixed geometry. J Comput Chem 11: 468–486

Gibson KD, Scheraga HA (1990b) Dynamics of peptides with fixed geometry: kinetic energy terms and potential energy derivatives as functions of dihedral angles. J Comput Chem 11: 487–492

Glunt W, Hayden TL, Raydan M (1993) Molecular conformations from distance matrices. J Comput Chem 14: 114–120

Godzik A, Kolinski A, Skolnick J (1992) Topology fingerprint approach to the inverse protein folding problem. J Mol Biol 227: 227–238

Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. Proteins 8: 195–202

Gray PMD, Paton NW, Kemp GJL, Fothergill JE (1990) An object-oriented database for protein structure analysis. Protein Eng 3: 235–243

Gronenborn AM, Clore GM (1990) Protein structure determination in solution by two-dimensional and three-dimensional NMR. In: Nall BT, Dill KA (eds) Conformations and forces in protein folding. American Association for the Advancement of Science, Washington, pp 125–149

Guillet JG, Lai MZ, Briner TJ, Buus S, Sette A, Grey HM, Smith JA, Gefter ML (1987) Immunological self, nonself discrimination. Science 235: 865–870

Hagler AT (1985) Theoretical simulation of conformation, energetics, and dynamics of peptides. In: Hruby UJ (ed) The peptides: conformation in biology and drug design. Academic Press, New York, pp 213–300

Harvey SC (1989) Treatment of electrostatic effects in macromolecular modelling. Proteins Struct Funct Genet 5: 78–92

Havel TF, Snow ME (1991) A new method for building protein conformations from sequence alignments with homologues of known structure. J Mol Biol 217: 1–7

Hermans J, Yun RH, Anderson AG (1992) Precision of free energies calculated by molecular dynamics simulations of peptides in solution. J Comput Chem 13: 429–442

Hol WGJ, Halic LM, Sander C (1981) Dipoles of the α-helix and β-sheet: their role in protein folding. Nature 294: 532–536

Holm L, Sander C (1992) Fast and simple MC algorithm for side chain optimization in proteins: Applications to model building by homology. Proteins Struct Funct Genet 14: 213–223

Honig B, Sharp K, Yang A-S (1993) Macroscopic models of aqueous solutions: biological and chemical applications. J Phys Chem 97: 1101–1109

Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78: 3825–3828

Huang MC, Seyer JM, Kang AH (1990) Comparison and accuracy of methodologies for analysis of hydropathy, flexibility and secondary structure of proteins. J Immunol Methods 129: 77–88

Hubbard TJP, Blundell TL (1987) Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. Protein Eng 1: 159–171

Iglesias E, Sordo TL, Sordo JA (1991) Molecular associations from ab initio pair potentials. Comput Phys Commun 67: 268–284
Islam SA, Sternberg MJE (1989) A relational database of protein structures designed for flexible enquiries about conformation. Protein Eng 2: 431–446
Jaenicke R (1987) Folding and association of proteins. Prog Biophys Molec Biol 49: 117–237
Janin J (1979) Surface and inside volumes in globular proteins. Nature 277: 491–492
Jenks S (1992) Mimetics may one day replace peptide antibodies. J Natl Cancer Inst 84: 79–80
Jones DT, Taylor WR, Thornton JM (1992) A new approach to protein fold recognition. Nature 358: 86–89
Jones TA, Thirup S (1986) Using known substructures in protein model building and crystallography. EMBO J 5: 819–822
Kabsch W, Sander C (1983) How good are predictions of protein secondary structure? FEBS Lett 155: 179–182
Kaden F, Koch I, Selbig J (1990) Knowledge-based prediction of protein structures. J Theor Biol 147: 85–100
Karplus M, Petsko GA (1990) Molecular dynamics simulations in biology. Nature 347: 631–639
Karplus PA, Schulz GE (1985) Prediction of chain flexibility in proteins. Naturwissenschaften 72: 212–213
Kaufman JF, Auffray C, Korman AJ, Schackelford DA, Strominger J (1984) The class II molecules of the human and murine major histocompatibility complex. Cell 36: 1–13
Kikuchi T, Nemethy G, Scheraga HA (1988a) Prediction of the packing arrangement of strands in β-sheets of globular proteins. J Protein Chem 7: 473–490
Kikuchi T, Nemethy G, Scheraga HA (1988b) Prediction of probable pathways in globular proteins. J Protein Chem 7: 491
Kim PS (1990) Intermediates in the folding reactions of small proteins. Annu Rev Biochem 59: 631–660
Kimball ES, Coligan JE (1983) Structure of class I major histocompatibility antigens. Comp Topics Molec Immunol 9: 1–63
King J (1989) Deciphering the rules of protein folding. Chem Eng News: 32–54
King RD, Sternberg MJE (1990) Machine learning approach for the prediction of protein secondary structure. J Mol Biol 216: 441–457
Klein J (1986) Natural history of the major histocompatibility complex. John Wiley and Sons, New York, pp 291–607
Kneller DG, Cohen FE, Langridge R (1990) Improvements in protein secondary structure prediction by an enhanced neural network. J Mol Biol 214: 171–182
Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132
Lamb JR, Ivanyi J, Rees ADM, Rothbard JB, Howland K, Young RA, Young DB (1987) Mapping of T cell epitopes using recombinant antigens and synthetic peptides. EMBO J 6: 1245–1249
Lambert MH, Scheraga HA (1989) Pattern recognition in the prediction of protein structure. I. Tripeptide conformation probabilities calculated from the amino acid sequence. J Comput Chem 10: 770–797
Leach AR, Kuntz ID (1992) Conformational analysis of flexible ligands in macromolecular sites. J Comput Chem 13: 730–748
Lecomte JTJ, Mathews CR (1993) Unraveling the mechanism of protein folding: new tricks for an old problem. Protein Eng 6: 1–10
Lee C, Subbiah S (1991) Prediction of protein side-chain conformation by packing optimization. J Mol Biol 217: 373–388
Lesk AM, Chothia C (1982) Evolution of proteins formed by β-sheets. II. The core of the immunoglobulin domains. J Mol Biol 160: 325–342
Levitt M (1976) A simplified representation of protein conformations for rapid simulation of protein folding. J Mol Biol 104: 59–107

Levitt M (1981) Molecular dynamics of hydrogen bonds in bovine pancreatic trypsin inhibitor protein. Nature 294: 379–380
Li Z, Scheraga HA (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc Natl Acad Sci USA 84: 6611–6615
Li Z, Scheraga HA (1988) Structure and free energy of complex thermodynamic systems. J Mol Struct (Theochem) 179: 333–352
Lim VI (1974) Algorithms for prediction of α-helical and β-structural regions in globular proteins. J Mol Biol 88: 873–894
Makhatadze GI, Privalov PL (1990) Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. J Mol Biol 213: 375–384
Margalit H, Spouge JL, Cornette JL, Cease KB, DeLisi C, Berzofsky JA (1987) Prediction of immunodominant helper T cell antigenic sites from the primary sequence. J Immunol 128: 2213–2229
Marrack P, Kappler J (1986) The antigen-specific major histocompatibility complex-restricted receptor on T-cells. Adv Immunol 38: 1–10
Mathews CR (1991) The mechanism of protein folding. Current Op Struct Biol 1: 28–35
McCammon JA, Harvey S (1987) Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge
McCammon JA, Wolynes PG, Karplus M (1979) Picosecond dynamics of tyrosine side chains in proteins. Biochemistry 18: 927–942
McGregor MJ, Flores TP, Sternberg MJE (1989) Prediction of beta-turns in proteins using neural networks. Protein Eng 2: 521–526
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21: 1087–1092
Morley SD, Jackson DE, Saunders MR, Vinter JG (1992) DMS: a multi-functional hybrid dynamics/Monte Carlo simulation algorithm for the evaluation of conformational space. J Comput Chem 13: 693–703
Moult J, James MNG (1986) An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins Struct Funct Genet 1: 146–163
Murphy KP, Gill SJ (1991) Solid model compounds and the thermodynamics of protein unfolding. J Mol Biol 222: 699–709
Murzin AG, Finkelstein AV (1988) General architecture of the α-helical globule. J Mol Biol 204: 749–769
Ngo JT, Marks J (1992) Computational complexity of a problem in molecular structure prediction. Protein Eng 5: 313–321
Nilsson O (1990) Molecular conformational space analysis using computer graphics: going beyond FRODO. J Mol Graph 8: 192–200
Noguti T, Go N (1983a) Dynamics of native globular proteins in terms of dihedral angles. J Phys Soc Japan 52: 3283–3288
Noguti T, Go N (1983b) A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. J Phys Soc Japan 52: 3685–3690
Noguti T, Go N (1985) Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. Biopolymers 24: 527–546
Orengo CA, Brown NP, Taylor WR (1992) Fast structure alignment for protein databank searching. Proteins Struct Funct Genet 14: 139–167
Osguthorpe D (1989) Molecular modelling of biochemical systems. Biochemist 11: 4–9
Owens RA, Gesellchen PD, Houchins BJ, Dimarchi RD (1991) The rapid identification of HIV protease inhibitors through the synthesis and screening of defined peptide mixtures. Biochem Biophys Res Commun 181: 402–408
Palmer KA, Scheraga HA (1991) Standard geometry chains fitted to X-ray derived structures: validation of the rigid geometry approximation. I. Chain closure through a limited search of loop conformations. J Comput Chem 12: 505–526
Parker JMR, Guo D, Hodges RS (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted

surface residues with antigenicity and x-ray derived accessible sites. Biochemistry 25: 5425–5432

Parker JMR, Hodges RS (1991a) HPLC hydrophilicity parameters: prediction of surface and interior regions in proteins. In: Mant CT, Hodges RS (eds) High performance liquid chromatography of peptides and proteins. CRC Press, Boca Raton, Florida, pp 737–749

Parker JMR, Hodges RS (1991b) Prediction of surface and interior regions in proteins – Part I: Linear tripeptide sequences identify structural boundaries in proteins. Peptide Res 4: 347–354

Parker JMR, Hodges RS (1991c) Prediction of surface and interior regions in proteins – Part II: Predicting secondary structure in regions bound by surface exposed regions. Peptide Res 4: 355–363

Pascarella S, Colosimo A, Bossa F (1990) Computational analysis of protein sequencing data. In: Fini C, Floridi A, Finelly VN, Wittman-Liebold B (eds) Laboratory methodology in biochemistry. CRC Press, Boca Raton, Florida, pp 109–128

Plochocka D, Zielenkiewicz P, Rabczenko A (1988) Hydrophobic microdomains as structural invariant regions in proteins. Protein Eng 2: 115–118

Ptitsyn OB, Finkelstein AV (1983) Theory of protein secondary structure and algorithm for its prediction. Biophysics 22: 15–25

Qian N, Sejnowski IJ (1988) Predicting the secondary structure of globular proteins using neural network models. J Mol Biol 202: 865–884

Richardson JS (1981) The anatomy and taxonomy of protein structure. Adv Protein Chem 34: 167–339

Rooman MJ, Kochar JPA, Wodak SJ (1991) Prediction of protein backbone conformation based on seven structure assignments. J Mol Biol 221: 961–979

Rooman MJ, Wodak SJ (1988) Identification of predictive sequence motifs limited by protein structure database size. Nature 335: 45–49

Rose GD, Gierasch LM, Smith JA (1985) Turns in peptides and proteins. Adv Protein Chem 37: 1–109

Roterman IK, Gibson KD, Scheraga HA (1989a) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. I. Conformational predictions for the tandemly repeated peptide $(Asn\text{-}Ala\text{-}Asn\text{-}Pro)_9$. J Biomol Struct Dynam 7: 391–419

Roterman IK, Lambert MH, Gibson KD, Scheraga HA (1989b) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. II. ϕ-ψ maps for N-acetyl alanine N′-methyl amide: comparisons, contrasts and simple experimental tests. J Biomol Struct Dynam 7: 421–453

Rothbard JB (1986) Peptides and the cellular immune response. Ann Inst Pasteur 137E: 518–526

Rothbard JB, Lechler RI, Howland K, Bal V, Eckels DD, Sekaly R, Long ED, Taylor WR, Lamb JR (1988) Structural model of HLA-DRI restricted T-cell antigen recognition. Cell 52: 515–523

Rupp R, Acha-Orbea H, Hengartner H, Zinkernagel R, Joho R (1985) Identical V_β T-cell receptor genes used in alloreactive cytotoxic and antigen plus I-A specific helper T-cells. Nature 315: 425–427

Russell RB, Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. Proteins Struct Funct Genet 14: 309–323

Sander C, Schneider R (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins Struct Funct Genet 9: 56–68

Saqi MAS, Bates A, Sternberg MJE (1992) Towards an automatic method of predicting protein structure by homology: an evaluation of suboptimal sequence alignments. Protein Eng 5: 305–311

Saragoui HV, Fitzpatrick D, Raktabotr A, Nakaniski H, Kahn M, Greene MI (1991) Design and synthesis of a mimetic from an antibody complementary-determining region. Science 253: 792–795

Schuler GD, Altschul SF, Lipman DJ (1991) A workbench for multiple alignment construction and analysis. Proteins Struct Funct Genet 9: 180–190

Schwartz RH, Fox BS, Fraga E, Chen C, Singh B (1985) The T lymphocyte responses to cytochrome C. Determination of the minimal peptide size required for stimulation of T cell clones and assessment of the contribution of each residue beyond this size to antigenic potency. J Immunol 135: 2598–2608

Scordamaglia R, Cavallone F, Clementi E (1977) Analytical potentials from 'ab initio' computations for the interaction between biomolecules. 2. Water with the four bases of DNA. J Am Chem Soc 99: 5545–5550

Sette A, Buus S, Colon S, Miles C, Grey HM (1988) I-A^d-binding peptides derived from unrelated protein antigens share a common structural motif. J Immunol 141: 45–48

Snow ME (1992) Powerful simulated-annealing algorithm locates global minima of protein-folding potentials from multiple starting conformations. J Comput Chem 13: 579–584

Sordo JA, Probst M, Chin S, Corongiu G, Clementi E (1986) Non-empirical pair potentials for the interaction between amino acids. In: Clementi E, Chin S (eds) Structure and dynamics of nucleic acids, proteins, and membranes. Plenum, New York, pp 89–111

Sordo JA, Probst M, Corongiu G, Chin S, Clementi E (1987) Ab initio pair potentials for the interaction between aliphatic amino acids. J Am Chem Soc 109: 1702–1708

Sternberg MJE, Islam SA (1990) Local protein sequence similarity does not imply a structural similarity. Protein Eng 4: 125–131

Tainer JA, Getzoff ED, Alexander H, Houghton RA, Olson AJ, Lerner RA (1984) The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. Nature 312: 127–133

Taylor WR (1988) Pattern matching methods in protein sequence comparison and structure prediction. Protein Eng 2: 77–86

Thornton JM (1988) The shape of things to come? Nature 335: 10–11

Thornton JM, Edwards MS, Taylor WR, Barlow DJ (1986) Location of 'continuous' antigenic determinants in the protruding regions of proteins. EMBO J 5: 409–413

Thornton S, San Fabian E, Fraga S, Parker JMR, Hodges RS (1991a) Prediction of protein secondary structures using a combined method based on the recognition, Lim and Garnier-Osguthorpe-Robson algorithms. J Mol Struct (Theochem) 232: 321–336

Thornton S, San Fabian E, Fraga S, Parker JMR, Hodges RS (1991b) Theoretical prediction of secondary structures of proteins using recognition factors. J Mol Struct (Theochem) 226: 87–97

Thornton SE, Fraga S (1991) Theoretical studies of peptidic structures. Conformation of the tetrapeptide N-acetyl-Asp-Glu-Lys-Ser-NH-CH3. Can J Chem 69: 1636–1638

Tyler EC, Horton MR, Krause PR (1991) A review of algorithms for molecular sequence comparison. Comp Biomed Res 24: 72–96

Unger R, Harel D, Wherland S, Sussman JL (1990) Analysis of dihedral angles distribution. The doublet distribution determines polypeptide conformations. Biopolymers 30: 499–508

van Gunsteren WF (1988) The role of computer simulation techniques in protein engineering. Protein Eng 2: 5–13

van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 34: 1311–1327

van Gunsteren WF, Gros P, Torda AE, Berendsen HJC, van Schack RC (1991) On deriving spatial protein structure from NMR or X-ray diffraction data. In: Protein conformation. Ciba Foundation Symposium 161. Wiley Interscience, Chichester, England, pp 150–166

van Gunsteren WF, Karplus M (1982) Effect of constraints on the dynamics of macromolecules. Macromol 15: 1528–1544

Watts TH, Gariepy J, Schoolnik GK, McConnell HM (1985) T-cell activation by peptide antigen: effect of peptide sequence and method of antigen presentation. Proc Natl Acad Sci USA 82: 5480–5484

Weaver DF (1992) Applications of molecular physics 'biotechnology' to the rational design of an improved phenytoin analogue. Seizure 1: 223–246

Weissman JS, Kim PS (1991) Reexamination of the folding of BPTI: predominance of native intermediates. Science 253: 1386–1393
Westhof E, Altschuh D, Moras D, Bloomer AC, Mondragon A, Klug A, Van Regenmortel MHV (1984) Correlation between segmental mobility and the location of antigenic determinants in proteins. Nature 311: 123–126
Wilson T, Klausner A (1984) Computers reveal proteins' mysteries. Biotechnol 2: 511–519
Wu TT, Kabat EA (1973) An attempt to evaluate the influence of neighboring amino acids (N − 1) and (N + 1) on the backbone conformation of amino acids in proteins. Use in predicting the three dimensional structure of the polypeptide backbone of other proteins. J Mol Biol 75: 13–31
Wüthrich K (1991) Six years of protein structure determination by NMR spectroscopy: what have we learned? In: Protein conformation. Ciba Foundation Symposium 161, Wiley Interscience, Chichester, England, pp 136–149
Yada RY, Jackman RL, Nakai S (1988) Secondary structure predictions and determination of proteins – a review. Int J Pept Protein Res 31: 98–108
Ycas M (1990) Computing tertiary structures of proteins. J Protein Chem 9: 177–200
Zimmerman SS (1985) Theoretical methods in the analysis of peptide conformation. In: Hruby U (ed) The peptides, vol 7. Academic Press, New York, pp 165–212

Appendix

Software packages

ALTA:maPS/POETA

Fraga S (1982) A semiempirical formulation for the study of molecular interactions. J Comput Chem 3: 329–334
Coghlan B, Fraga S (1985) Determination of proteinic structures: an experimentation program. Comput Phys Commun 36: 391–399
Thornton S, San Fabian E, Fraga S, Parker JMR, Hodges RS (1991) Prediction of protein secondary structures using a combined method based on the recognition, Lim and Garnier-Osguthorpe-Robson algorithms. J Mol Struct (Theochem) 232: 321–336

AMBER

Weiner PK, Kollman PA (1981) AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. J Comput Chem 2: 287–303
Weiner SJ, Kollman PA, Case DA, Singh KC, Ghio C, Alagona G, Profeta Jr S, Weiner PK (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106: 765–784
Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 7: 230–252

BOXSEARCH/CLUSTER

Hart TN, Read RJ (1992) A multiple start Monte Carlo docking method. Proteins Struct Funct and Genet 13: 206–222

CHARMm/QUANTA

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMm: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4: 187–217
Momany FA, Rone R (1992) Validation of the general purpose QUANTA 3.2/CHARMm force field. J Comput Chem 13: 888–900

CONGEN

Bruccoleri RE, Karplus M (1987) Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers 26: 137–168

COSMIC

Vinter JG, Davis A, Saunders MR (1987) Strategic approaches to drug design. I. An integrated software framework for molecular modelling. J Comput Aided Mol Design 1: 31–51

DAISY

Koca J, Carlsen PHJ (1992) DAISY, a computational method: a novel tool for the study of the conformational behavior of flexible molecules. J Mol Struct (Theochem) 257: 105–130

DAPMATCH

Walls PH, Sternberg MJE (1992) New algorithm to model protein-protein recognition based on surface complementarity. Applications to antibody-antigen docking. J Mol Biol 228: 277–297

DG-II

Havel TF (1991) An evaluation of computational strategies for use in the determination of protein structures from distance constraints obtained by nuclear magnetic resonance. Prog Biophys Mol Biol 56: 43–78

DIANA

Güntert P, Braun W, Wüthrich K (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOSA. J Mol Biol 217: 517–530

DISCOVER/INSIGHT

Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT (1988) Structure and energetics of ligand binding to proteins. Protein Struct Funct and Genet 4: 31–47

DMC

Morley SD, Jackson DE, Saunders MR, Vinter JG (1992) DMC: A multifunctional hybrid dynamics/Monte Carlo simulation algorithm for the evaluation of conformational space. J Comput Chem 13: 693–703

DOCK

Des Jarlais RL, Sheridan RP, Seibel GL, Dixon JS, Kuntz ID (1988) Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. J Med Chem 31: 722–729
Shoichet BK, Kuntz ID (1991) Protein docking and complementarity. J Mol Biol 221: 327–346

ECEPP

Momany FA, McGuire RF, Burgess AW, Scheraga HA (1975) Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the natural occurring amino acids. J Phys Chem 79: 2361–2381

Dunfield LG, Burgess AW, Scheraga HA (1978) Energy parameters in polypeptides. VIII. Empirical potential energy algorithm for the conformational analysis of large molecules. J Phys Chem 82: 2609–2616

Nemethy G, Pottle MS, Scheraga HA (1983) Energy parameters in polypeptides. IX. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. J Phys Chem 87: 1883–1887

Sippl MJ, Nemethy G, Scheraga HA (1984) Intermolecular potentials from crystal data. VI. Determination of empirical potentials for O–H ··· O=C hydrogen bonds from packing configurations. J Phys Chem 88: 6231–6233

Lambert MH, Scheraga HA (1989) Pattern recognition in the prediction of protein structure. III. An importance-sampling minimization procedure. J Comput Chem 10: 817–831

Gibson KD, Scheraga HA (1990) Variable step molecular dynamics: An exploratory technique for peptides with fixed geometry. J Comput Chem 11: 468–486

Palmer KA, Scheraga HA (1991) Standard-geometry chains fitted to X-ray derived structures: Validation of the rigid-geometry approximation. I. Chain closure through a limited search of 'loop' conformations. J Comput Chem 12: 505–526

GROMOS

van Gunsteren WF, Berendsen HJC, Hermans J, Hol WGJ, Potsma JPM (1983) Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data. Proc Natl Acad Sci 80: 4315–4319

van Gunsteren WF, Berendsen HJC (1987) Groningen Molecular Simulation (GROMOS) Library Manual. BIOMOS B.V., Groningen, The Netherlands

IPSA

Schulze-Kemer S, King RD (1992) Inductive protein structure analysis. Protein Eng 5: 377–390

mdFRODO

Nilsson O (1990) Molecular conformational space analysis using computer graphics: Going beyond FRODO. J Mol Graph 8: 192–200

MM2/MMP2

Allinger NL (1977) Conformational analysis. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J Am Chem Soc 99: 8127–8134

MMPEP/MMPEN

Wolfe S, Weaver DF, Yang K (1988) MMPEP: Development and evaluation of peptide parameters for Allinger's MMP2 (85) programme, including calculations on crambin and insulin. Can J Chem 66: 2687–2702

Wolfe S, Khalil M, Weaver DF (1988) MMPEN: Development and evaluation of penicillin parameters for Allinger's MMP2(85) programme. Can J Chem 66: 2715–2732

MOTECC-90

Clementi E (ed) (1990) Modern Techniques in Computational Chemistry. ESCOM, Leiden

OPLS-AMBER

Jorgensen WL, Tirado-Rives J (1988) The OPLS potential function for proteins. Energy minimization for crystals of cyclic peptides and crambin. J Am Chem Soc 110: 1657–1666

ORAL

Zimmermann K (1991) ORAL: All purpose molecular mechanics simulator and energy minimizer. J Comput Chem 12: 310–319

PREDICT7

Carmenes RS, Freije PP, Molina MM, Martin JM (1989) PREDICT7, a program for protein structure prediction. Biochem Biophys Res Commun 159: 687–693

PROTEUS

Pabo CO, Suchanek EG (1986) Computer-aided model-building strategies for protein design. Biochemistry 25: 5987–5991

RAMBLE

Perkins TDJ, Barlow DJ (1990) RAMBLE: A conformational search program. J Mol Graph 8: 156–162

ROSE

Koca J, Carlsen PHJ (1992) ROSE, a computer program for automatic generation of probable conformations. J Mol Struct (Theochem) 257: 131–141

SETTLE/SPASMS

Miyamoto S, Kollman PA (1992) SETTLE: An analytical version of SHAKE and RATTLE algorithm for rigid water molecules. J Comput Chem 13: 952–962

WHAT IF

Wriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8: 52–55

Some commercial software packages are:

CAMSEQ

Molecular mechanics and molecular display
Weintraub Software Design Associates, Inc. P.O. Box 42577, Cincinnati, OH 45242, USA

CHEMLAB

Molecular mechanics, quantum mechanics, and molecular display
Molecular Design Ltd. 2132 Farallon Drive, San Leandro, CA 94577, USA

CHEM-X

Molecular mechanics, molecular dynamics, and molecular display
Chemical Design Ltd. 200 Route 17 South, Mahwah, NJ 07430, USA

CHEM3D

Molecular mechanics and molecular dynamics
Cambridge Scientific Computing Inc. 875 Massachusetts Ave., Cambridge, MA 02139, USA

DISCOVER/FELIX/INSIGHT/NMRchitect

Molecular mechanics and nuclear magnetic resonance
Biosym Technologies Inc. 9685 Scranton Rd., San Diego, CA 92121-2777, USA

GRAMPS

Molecular display
Abbott Laboratories. Abbott Park, L 60064, USA

maPSI

Modelling and analysis of protein structures and interactions
Protein Software Inc. Box 60499, University of Alberta Postal Outlet, Edmonton, AB, Canada T6G 2S7

PPSP

Secondary structure, homology, and preferred-conformation data base predictions
Synthetic Peptide Inc. Medical Sciences Bldg (Rm 355), University of Alberta, Edmonton, AB, Canada T6G 2H7

SYBYL/MENDYL/ALCHEMY

Molecular mechanics and molecular display
Tripos Associates Inc. 6548 Clayton Rd., St. Louis, MO 63177 USA

X-PLOR

Nuclear magnetic resonance
Molecular Simulations Inc. 16 New England Executive Park, Burlington, MA 01803-5297, USA

Authors' address: Dr. S. Fraga, Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2.

Received March 3, 1993