PROTEAN: A New Method for Deriving Solution Structures of Proteins

Bruce Duncan+, Bruce G. Buchanan*, Barbara Hayes-Roth*,
Olivier Lichtarge+, Russ Altman*, James Brinkley*,
Micheal Hewett*, Craig Cornelius*, Oleg Jardetzky+

+Stanford Magnetic Resonance Laboratory
Stanford University Medical Center
Stanford, CA 94305

*Knowledge Systems Laboratory
Computer Science Department
Stanford University
Stanford, CA 94305

Abstract

Nuclear Overhauser Enhancement Spectroscopy (NOESY) is a powerful NMR technique for obtaining structural information on proteins in solution. 2DFT NOESY experiments readily provide information on dipole coupled protons, an indication of close spatial proximity. Unfortunately, quantitative interpretation of NOE measurements in terms of molecular structure is problematic because a rigorous interpretation requires prior knowledge of the structure. Nevertheless, a careful approximate analysis of the spectroscopic information can yield meaningful structural information if adequate precautions are applied. Here we outline a new strategy to take into account these limitations of the data. We also describe the PROTEAN system, a computer program under development to implement this strategy. Finally we provide a simple example of the use of PROTEAN to determine the solution structure of the lac-repressor headpiece.

Introduction

X-ray crystallography is the primary method for obtaining detailed information about protein structure. This method however, is limited to proteins which can be crystallized and provides no information on variations in the structure which may occur in solution. Evidence from NMR indicates that these variations may be important [1,2].

Various biochemical techniques are available to provide information about proteins in solution. One NMR technique, Nuclear Overhauser Enhancement Spectroscopy (NOESY), identifies protons coupled by a dipole-dipole relaxation mechanism which can occur when the protons are in close spatial proximity (2-4 Angstroms). NOE derived distance constraints from 2DFT NMR experiments have been used in numerous studies of protein structure [3-13].

Unfortunately, limitations on the precision and abundance of NOE derived distance constraints do not allow the specification of a single unique structure. The imprecision of structural parameters that can be derived from an the NOE experiment can be shown as follows. The transfer of magnetization between any two spins i and k in an NOE experiment is described by the generalized Bloch equations [14]:

$$1)\quad \frac{dM_i}{dt} = -\rho_i(M_i - M_i^o) - \sum_j \sigma_{ij}(M_j - M_j^o) + \sigma_{ik}M_k^o$$

where M is the magnetization vector, i the observed, k the irradiated and j any other interacting spin. This description of the NOE phenomena is an adequate approximation to the more rigorous density matrix formulation [5]. The molecular interpretation of the relaxation parameters ρ and σ is given by:

$$2) \quad \rho_i = K f_{ij}(\tau) \sum_j \frac{1}{r_{ij}^6}$$

$$3) \quad \sigma_{ij} = \frac{K f_{ij}(\tau)}{r_{ij}^6}$$

where K is a product of atomic constants, r_{ij} the interatomic distance and $f_{ij}(\tau)$ is a spectral density function for any pair of protons *ij*. A simple relation between the measured magnetization transfer (NOE) and the internuclear distance therefore exists only if (a) the two spin approximation applies, i.e. only direct and no indirect magnetization transfer is occurring in the experiment and (b) $f_{ij}(\tau)$ is accurately known. In protein NMR typically neither is the case [14]. This prevents the calibration of NOEs in terms of precise interatomic distances and presents a basic dilemma: To interpret NOE data correctly it is necessary to know the structure which is to be determined from them. Thus NOE data allow, at best, the determination of only approximate structures.

We have recently shown [15] by solving the Bloch equations (1) for a variety of specific structures that even at the shortest experimentally feasible mixing times in an NOE experiment (20-50 msec) indirect magnetization transfer may play a significant role. Thus there is no universally valid relationship between the magnitude of an NOE and internuclear distance. This conclusion also follows from earlier model calculations of Bothner-By [5]. In addition, $f_{ij}(\tau)$ is not known accurately [14]. Other effects such as signal-to-noise ratio limitations also restrict the precise interpretation of NOE measurements but to a much lesser extent, so that the apparent precision of the calculated distances can be much higher than their inherent accuracy. To retain accuracy, distance ranges have to be used for structural determinations, and these distance ranges depend on the specific conditions of the NMR experiment. We have shown elsewhere that both direct and indirect long range NOEs obtained at mixing times of 50 msec or less imply an upper distance limit of about 4 Angstroms. For mixing times of 100 msec, the upper distance limit may be up to 6 Angstroms [15-17]. The precision of the interpretation of NOE experiments can be substantially improved by using refinement methods based on the generalized Bloch equations and taking all indirect pathways of magnetization transfer and internal motion into account. However these calculations require at least an approximate starting structure.

In addition to their being imprecise, there are rarely sufficient distance constraints to allow specification of a unique structure. An estimate of the quantity and quality of distance constraints required to determine a protein structure to a certain degree of precision are given in [18].

A generally valid structural interpretation of NOESY experiments can be made if the data interpretation method does not ignore the uncertainties inherent in the experimental data or the fact that the structure is underdetermined.

We require a procedure for structure determination which relies exclusively on experimental data and does not impose arbitrary theoretical constraints. Such theoretical constraints are implicit in distance geometry [19,20] or energy minimization [7,21] procedures which assume that the correct structure corresponds to a minimum of an error function and that a single structure satisfies *all* of the experimental constraints simultaneously. Since the structure is underspecified, it is possible for other structures, also minima of an error function, to be consistent with the input data. Recently, the determination of the structure of the E. coli ST 1 peptide illustrates this point. The constraint satisfaction program PROTO [22] was used to derive structures of ST 1 using standard bond angles, van der Waals radii as well as experimental coupling constant and NOE distance constraints. Various runs of the program produced several structures satisfying the constraints but not all of them faithfully reproduced the experimentally observed magnetization transfer as predicted from the generalized Bloch equations [8]. Thus a structure computed from the minimum of an error function may satisfy the "straightforward" interpretation of the spectroscopic data, but may not satisfy a more rigorous interpretation.

Data interpretation methods based on distance geometry, or energy minimization may produce structures which correctly predict the experimental data from the Bloch equations. However these methods cannot guarantee that there are no other structures that would also satisfy the constraints and conversely may produce structures that do not satisfy the constraints as shown in the example of the ST 1 peptide. We would like a procedure that can represent all compatible structures, thus giving some indication of the imprecision of the data and providing initial structures for further refinement. This is the goal of the PROTEAN system.

The Strategy of the PROTEAN system

A computer program called PROTEAN has been developed to determine protein solution structures [23-26,40]. PROTEAN seeks to determine a representative sample of all structures consistent with available experimental data. These structures can be checked by the Bloch equation refinement procedure to ensure that a rigorous interpretation of the experimental data is maintained.

PROTEAN at present relies primarily but not exclusively on NMR derived NOE distance constraints. Since NMR data alone are insufficient to define a unique structure, PROTEAN supplements NMR information with other experimental data to reduce this indeterminacy. Hydrodynamic, light scattering, and small angle X-ray scattering experiments yield information about the gross shape and size of the protein. Photo-chemically induced dynamic nuclear polarization (photo-CIDNP), fluorescence quenching, and paramagnetic perturbation experiments can determine some of the atoms that lie on or near the surface of the molecule. Each of these techniques weakly restricts the space of possible conformations.

The kinds of information available to PROTEAN are summarized below.

- *General protein data:* PROTEAN has available general protein information such as standard bond lengths, bond angles, amino acid conformations and van der Waals radii.
- *Primary Structure:* Since protein sequencing methods are well established, PROTEAN requires that the primary structure is known.
- *Secondary Structure:* The identification of secondary structures in peptide chains by a combination of NMR measurements of accurate exchange rates and NOEs along the peptide backbone are reasonably reliable [14,27-29]. PROTEAN requires the protein secondary structure or data from which the secondary structure can be derived.
- *Experimental Data:* PROTEAN also can make use of short range distance constraints derived from NOE experiments, maximum distance information from small angle X-ray or neutron diffraction, shape constraints, and surface constraints. The software architecture of the PROTEAN system allows other experimental or theoretical constraints to be used as well.

PROTEAN systematically searches for protein conformations that are consistent with the input data and uses two novel techniques to make this task computationally feasible. These techniques are called refinement of descriptive detail and refinement of accessible volumes.

Refinement of descriptive detail allows protein structural elements to be represented in three different ways called solid level, superatomic level, and atomic level representations. Solid level representations treat large groups of atoms as single units. For example, helices and beta strands are treated as rigid cylinders. This representation is also commonly used for cartoon drawings of protein structure [30,31]. By manipulating these aggregates of atoms, the program is able to avoid the computational cost of manipulating individual atoms. Once the spatial locations of these solid level objects are approximately defined, they can be refined to the superatomic level. At this level, smaller groups of atoms (peptide units, methyl groups, aromatic rings, etc.) are treated as single units. Finally, at the atomic level, individual atoms are manipulated. Currently, the entire solid level and parts of the superatomic and atomic level have been

implemented.

The second technique, refinement of accessible volumes, enumerates all *allowed* conformations of the major structural elements (helices and beta strands) without generating all *possible* conformations and testing each conformation for compatibility with the applied constraints. Refinement of accessible volumes proceeds as follows. Constraints are applied to define the region of space that one structured domain (helix, beta strand) can occupy with respect to another fixed domain and still satisfy the given constraints. This region of space, called the "accessible volume", is restricted by the introduction of additional constraints. More domains are added to the evolving solution and are checked for mutual compatibility with the previously placed structures. Detection of incompatibilities further restricts the positions of the domains.

Initially, the program defines the primary and secondary structures, specifies constraints, and chooses one domain (e.g., a helix) as a fixed "anchor". For a helix as anchor, a convenient choice of coordinate system is to define the axis of the helix as the z-axis of a right-handed Cartesian coordinate system and the x-axis as a perpendicular axis passing through the nucleus of the first (N-terminal) alpha-carbon in the helix. The program then selects a second structured domain to position relative to the first by considering which of the unpositioned domains has the most and the strongest constraints to the initial anchor. Then the second domain is positioned relative to the first by discretely sampling all of space and retaining positions in which the second domain can be located and still satisfy each constraint between them. This region, the "accessible volume", represents the limits of our knowledge about the structure defined by a subset of the experimental constraints.

After an anchor is chosen and a second domain is positioned relative to it, the next step is to choose either another constraint or a new domain to work on, or to choose to describe part of the structure at a finer level of detail. The choice of what to do next may be made by the user or automatically by PROTEAN. With the introduction of each constraint, the allowed locations of domains relative to one another are further restricted. With the introduction of each new domain, the partial structure is expanded. And with increased detail in the description of any part, locations of individual atoms are more closely determined.

Because we save only the locations that satisfy all the constraints considered so far, at every point in the procedure we have a partial definition of a structure that is accurate with respect to constraints considered, even if lacking in precision. If the accessible volume of a domain is reduced to nil with the introduction of new constraints, we can conclude that the constraints cannot be satisfied simultaneously, and therefore that no single structure exists. This may be the case, for example, in a protein structure fluctuating between two or more conformations. Otherwise, we know that the constraints can be satisfied simultaneously and that a single structure *may* exist. We then have a choice: (1) to *assume* that the constraints *are* satisfied simultaneously, and thus start with the reduced family of positions when exploring new constraints; or (2) to acknowledge that there may be important conformations for which all the constraints are *not* satisfied at the same time, and so work with the full set of positions implied by *each constraint*, reasoning with combinations of them to define plausible families of conformations. The latter procedure is considerably more time consuming, but is more cautious. PROTEAN allows either choice.

Placement of domains relative to the fixed domain is a procedure whose computational complexity is linear in the number of domains but a complicated function of the number and "strength" (constraining power) of the constraints. Checking mutual compatibility is a procedure whose computational complexity depends critically on the number of objects and the accessible volumes of each object [25,32,33]. When all the domains of a protein have been checked for compatibility, the resulting structure is one of the possible "instances" of the conformation. Representations of all the instances is the result sought by PROTEAN. In spite of this computational cost, approximations to the topology of the myoglobin structure based on simulated NOE data take 20 hours to compute [34].

These two techniques, refinement of

descriptive detail and refinement of accessible volumes, describe the strategy of the PROTEAN system. How this strategy is currently implemented is described below.

The PROTEAN system

PROTEAN consists of three major components; a computational system, a reasoning system, and a display system.

The computation component, known as the geometry system, performs the calculations required for placing objects in space and checking that constraints are satisfied [25]. The geometry system, written in the C programming language and executed under the UNIX operating system, systematically searches all of space by sampling discrete locations at some predefined resolution. Sampling resolutions for solid level placement are typically 2 Angstroms for positional degrees of freedom and 20 degrees for rotational degrees of freedom. These sampling intervals can be modified as the structure positions become more precise. The geometry system checks that NOE distance constraints, surface constraints, and volume constraints are satisfied. At the solid level, amino acid sidechains are not explicitly represented, consequently to check NOE distance constraints the geometry system calculates the equivalent distance range between $C\alpha$ atoms taking the specific NOE proton location and sidechain flexibility into account. The geometry system checks shape constraints by computing the approximate shape of the solid level objects and comparing it to the shape determined from hydrodynamic or spectroscopic measurements. The shape is modeled as an ellipsoid whose dimensions are determined from the moment of inertia of the $C\alpha$ points. Surface constraints at the solid level are checked by requiring that surface atoms are not in the buried "core" of this ellipsoid. The geometry system also checks for van der Waals collisions between the cylindrical objects. Each step of the calculations errs in favor of a liberal estimate of the accessible volume. Thus each step of the calculation represents an upper bound of the uncertainty of the conformation. The final goal is to represent the *least* upper bound by refining the degree of descriptive detail and further restricting the accessible volumes of the substructures.

At any particular stage of the structure determination process, there are several possible tasks that can be performed, such as; add a new domain to the emerging solution, add a new constraint to the emerging solution, refine a structure to an atomic representation, check a surface constraint, etc. PROTEAN's reasoning system dynamically determines which tasks can be performed and which specific task is best to perform at any particular time. (For example PROTEAN favors placing structured elements such as helices and strands first, before unstructured elements such as random coils are placed.) PROTEAN's reasoning system is based on the reasoning paradigm known as the blackboard model which developed from research in Artificial Intelligence (AI) [35,36]. Variations of this model have been used in other computation systems requiring intricate symbolic reasoning [37-39], and it provides a powerful framework within which PROTEAN performs symbolic as well as numerical computations.

Blackboard systems consist of a global database called the "blackboard" where problem specifications and emerging solutions are stored, and "knowledge sources", which are program modules that monitor the blackboard and become "active" when conditions for their applicability are satisfied. For example, after a helix is posted on the blackboard, the knowledge source ANCHOR-HELIX becomes available for execution. If the protein has more than one helix, multiple ANCHOR-HELIX options are activated. Each of these options is placed on a list possible tasks. Control heuristics evaluate the desirability of performing each task at each stage of problem solving. The control strategy is heuristic in the sense that the path to the solution is not uniquely determined. The heuristics determine the efficiency of obtaining a solution but do not change the characteristics of the solution found. This method insures that the results (within small error bounds introduced by the resolution of the sampling intervals) are independent of the order of the actions taken. PROTEAN's reasoning system is implemented in the AI architecture called BB1 [36]. The reasoning system is written in the LISP programming language and runs on XEROX

1100 series workstations.

The display system of PROTEAN allows the visualization of emerging solutions. It differs from most other molecular modeling systems (MMS, Midas, MOGLI) in that it can display protein structures in various representations (for example, cylinders or atomic representations) as well as displaying numerous conformations simultaneously. Representations of protein shape modeled as ellipsoids can also be displayed. The display system, also written in C, runs on Silicon Graphics IRIS 3020 workstations.

Table 1
Lac-Repressor Headpiece
Input Data

PRIMARY STRUCTURE

MET1	LYS2	PRO3	VAL4	THR5
LEU6	TYR7	ASP8	VAL9	ALA10
GLU11	TYR12	ALA13	GLY14	VAL15
SER16	TYR17	GLN18	THR19	VAL20
SER23	ARG22	VAL23	VAL24	ASN25
GLN26	ALA27	SER28	HIS29	VAL30
SER31	ALA32	LYS33	THR34	ARG35
GLU36	LYS37	VAL38	GLU39	ALA40
ALA41	MET42	ALA43	GLU44	LEU45
ASN46	TYR47	ILE48	PRO49	ASN50
ARG51				

SECONDARY STRUCTURE

Coil-1:	MET1-THR5
Helix-1:	LEU6-GLY14
Coil-2:	VAL15-SER16
Helix-2:	TYR17-ASN25
Coil-3:	GLN26-ARG35
Helix-3:	GLU36-LEU45
Coil-4:	ASN46-ARG51

NOES

VAL4-TYR17	ALA10-TYR17
VAL4-LEU45	TYR12-ALA32
VAL4-TYR47	TYR12-ALA41
THR5-TYR47	TYR12-MET42
LEU6-TYR17	TYR12-GLU44
LEU6-VAL24	TYR12-LEU45
LEU6-MET42	ALA13-VAL38
LEU6-TYR47	ALA13-VAL41
TYR7-TYR17	VAL15-TYR47
ASP8-LEU45	TYR17-MET42
VAL9-MET42	VAL20-VAL38
VAL9-LEU45	VAL24-TYR47
VAL9-TYR47	VAL30-MET42
ALA10-VAL20	MET42-TYR47

Results and Discussion

Results from PROTEAN's calculation of the lac-repressor headpiece are presented below. More detailed results from PROTEAN are given in [23,34,40].

The lac-repressor headpiece is a protein with fifty-one amino acids, whose complete solution structure is not known and which is the subject of current structural studies [6,7]. The crystal structure of the lac-repressor headpiece is not known, but crystal structures are available for some homologous repressor molecules [41].

Table 1 shows the primary structure, secondary structure, and distance constraints inferred from NOEs that are used as input data for PROTEAN.

PROTEAN was run at the solid level using the input data shown. The accessible volumes for HELIX-2 and HELIX-3 with respect to the fixed position of HELIX-1 are shown in Figure 1.

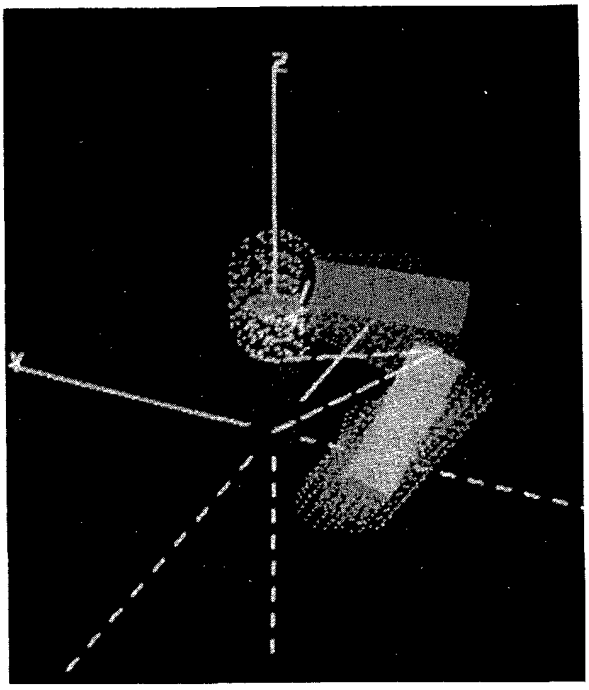

Figure 1
Solid Level Representation
of the Lac-Repressor Headpiece

This result defines the main topological features of the molecule, and the accessible volumes within which elements of the structure remain uncertain. These volumes indicate the extent to which the structure

can be specified from the existing solution data at this level of detail. It is seen from the figure that the accessible volumes are sufficiently large to preclude a unique definition of a single structure, but also sufficiently small to recognize major structural relationships. The helix locations are continuous, and we can also see that there are no other major structural families compatible with the constraints used.

The topology corresponds well with other repressor molecules whose crystal structures are known. A lac-repressor structure has been computed from the optimization method PROTO and NOE distance constraints [22]. A lac-repressor structure has also been determined from molecular dynamics computations and NMR distance constraints, but atomic coordinates are not available for comparison [7]. A detailed comparison of these structures and the results computed by PROTEAN will be reported when the implementation of the atomic level of the PROTEAN program is completed.

The accessible volume of HELIX-2 has an average root-mean-squared (rms) deviation of 2.8 Angstroms from the average $C\alpha$ positions. [The average $C\alpha$ positions are defined as the collection of the centroids for each individual alpha-carbon in the accessible volume of the structure.] The accessible volume of HELIX-3 has an average rms deviation of 2.9 Angstroms from the average $C\alpha$ positions. The maximum rms deviation between two locations of HELIX-2 is 6.3 Angstroms, and the maximum rms deviation between two locations of HELIX-3 is 11.3 Angstroms. These rms deviations are computed from the $C\alpha$ points only. Positions for the unstructured domains (random coils) are specified less precisely. For this example, the total computation time is about 3 hrs on the IRIS workstation.

At PROTEAN's current state of development, the solid level has been completely implemented. At the solid level, distance, surface, and shape constraints can be used. Solid level representations, though rather crude, have been shown to accurately define protein topology based on experiments with the known crystal structures of domain 1 of bacteriophage T4 lysozyme and sperm whale myoglobin [40,34]. In all cases, the solid level exploration of the conformational space gives a remarkably good and accurate definition of the main topological features, and the accessible volume of each secondary structure contains the crystal structure.

A structure of domain 1 of bacteriophage T4 lysozyme computed by PROTEAN is shown in Figure 2. The figure shows the solid level abstractions for helices and beta strands, and the crystal structure positions of these solid level objects (line segments). The figure also shows the bounding ellipsoid used to check surface and volume constraints.

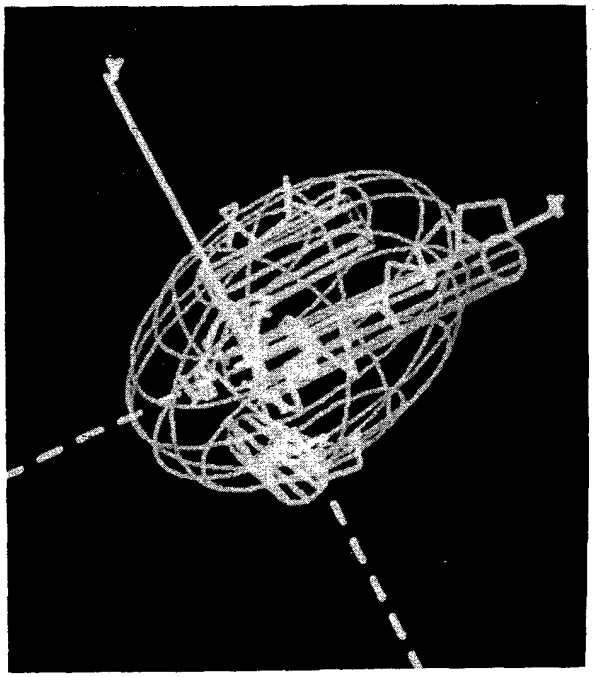

Figure 2
Solid Level Representation of
Domain 1 of Phage T4 Lysozyme

With myoglobin and simulated NOE datasets of 96 NOEs representing hydrogen atoms less than 3.3 Angstroms apart, PROTEAN derives structures with rms deviations (for the $C\alpha$ carbons of helical elements only) of 5 Angstroms [34]. This variation can be improved by examining the structure at the superatomic and atomic levels, a task under active development.

An example of atomic level results is shown in Figure 3. At the atomic level, the accessible volume of individual sidechains is determined by enumerating sidechain

conformations that satisfy the distance, surface, volume, and van der Waals constraints. PROTEAN can produce the coordinates of the sidechain and backbone atoms of helices and beta strands in Brookhaven Protein Databank format. These coordinates can then be used for molecular graphics or optimization procedures.

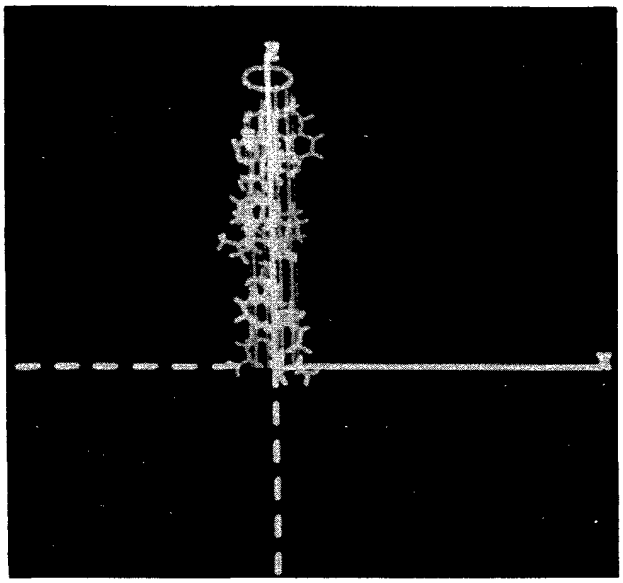

Figure 3
An Atomic Level Representation

Once atomic level descriptions are complete, the resulting variation represents the inherent imprecision of interpreting NOEs as distance constraints of 2-4 Angstroms. Bloch equation refinement may improve the precision further. There is no limit to the precision that can be achieved by adding more experimental data (surface, shape, packing constraints).

Conclusions

The key features of the structure determination problem -- (1) the fact that a one-to-one correspondence between experimentally measured and molecular structural parameters does not hold for solution methods as it does in crystallography, (2) the possibility of using constraints that are more easily expressed symbolically than numerically, (3) the availability of different kinds of constraints, (4) the incompleteness and irreducible uncertainty of the data -- make it desirable to follow more than one option in the analysis of the data and suggest that AI methods as implemented in the PROTEAN system are appropriate.

One of the key features in the design of PROTEAN is its flexibility in incorporating different kinds of information. A second key feature is that the method does not overinterpret the data, but defines the *family* of structures compatible with a conservative interpretation of the data. Results from using only constraints inferred from experimental NMR data indicate that additional kinds of constraints are necessary to define the structure precisely. These may come from additional experimental data or from theoretical considerations. The family defined by PROTEAN can then provide approximate structures for several different refinement procedures. A rigorous comparison between structures obtained from different refinement procedures still needs to be made.

Acknowledgments

We thank A.N. Lane and J.F. Lefevre for helpful discussions; F. Frayman for his PROTO results; Jeff Harvey, Alan Garvey and John Brugge for programming assistance; Grace Smith and Anthea Waleson for manuscript preparation.

This work was funded in part by: NIH Grant #GM07365, NIH Grant #RR-00785, NIH Grant #RR-00711, NSF Grant #DMB84-02348, NASA/Ames Grant #NCC 2-274, Boeing Grant #W266875, DARPA Contract #N00039-83-C-0136.

References

1. Roberts, G.C.K. and Jardetzky, O. (1970) *Adv. in Protein Chem.* **24**, 447-545.
2. Wuthrich, K. (1976) "NMR in Biological Research: Peptides and Proteins", North Holland, Amsterdam.
3. Brunger, A., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1986) *Proc. Natl. Acad. Sci.* **83**, 3801-3805.
4. Wagner, G. and Wuthrich K. (1979) *J. Mag. Res.* **33**, 675-679.
5. Bothner-By, A.A. and Noggle, J.H. (1979) *J. Am. Chem. Soc.* **101**, 5162-5170.
6. Jardetzky, O. (1984) IN: Progress in Bioorganic Chemistry and Molecular

Biology, (Yu. A. Ovchinnikov ed.), 55-63, Elsevier Science Publishers B.V., Amsterdam.
7. Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M. and Boelens, R. (1985) *J. Mol. Biol.* **182**, 179-182.
8. Gariepy, J., Lane, A., Frayman, F., Wilbur, D., Robien, W., Schoolnik, G.K. and Jardetzky, O. (1987) *Biochemistry*, in press.
9. Kline, A.D., Braun, W. and Wuthrich, K. (1986) *J. Mol. Biol.* **189**, 377-382.
10. Williamson, M.D, Havel, T.F. and Wuthrich, K. (1985) *J. Mol. Biol.* **182**, 295-315.
11. Braun, W. and Go, N. (1985) *J. Mol. Biol.* **186**, 611-626.
12. Braun, W., Wagner, G., Worgotter, E., Vasak, M., Kagi, J.H.R. and Wuthrich, K. (1986) *J. Mol. Biol.* **187**, 125-129.
13. Ohkubo, T., Kobayashi, Y., Shimonishi, Y., Kyogoku, Y., Braun, W., and Go, N. (1986) *Biopolymers*, in press.
14. Kalk, A. and Berendsen H. J. (1976) *J. Mag. Res.* **24**, 343-366.
15. Jardetzky, O., Lane, A.N., Lefevre, J.-F., Lichtarge, O., Hayes-Roth, B. and Buchanan, B. (1986) IN: NMR in the Life Sciences, Plenum Publishing Corporation, New York, in press.
16. Lane, A.N., Lefevre, J.F. and Jardetzky, O., unpublished results.
17. Jardetzky, O. and Lane, A. (1986) IN: Proceedings of the International School of Physics Enrico Fermi, Varenna, Italy, in press.
18. Wako, H. and Scheraga, H. (1981), *Macromolecules* **14**, 961-969.
19. Havel, T.F., Crippen, G.M. and Kuntz, I.D. (1979) *Biopolymers* **18**, 73-81.
20. Braun, W., Bosch, C., Brown, L.R., Go, N. and Wuthrich, K. (1981) *Biochim. Biophys. Acta* **667**, 377.
21. Clore, G.M., Gronenborn, A.M., Moss, D.S., Tickle, I.J. (1985) *J. Mol. Biol.* **185**, 219-226.
22. Frayman, F. (1985) "PROTO." Ph.D. Thesis (Computer Science), Northwestern University.
23. Lichtarge, O., Buchanan, B., Cornelius, C., Brinkley, J., Jardetzky, O. (1986) *Knowledge Systems Laboratory Technical Report KSL-86-12*, Stanford University.
24. Hayes-Roth, B., Buchanan, B., Lichtarge, O., Hewett, M., Altman, R., Brinkley, J., Cornelius, C., Duncan B. and Jardetzky, O. (1987) IN: Insight Series in Applied AI, (Englemore, R. and Morgan, A. eds.) Addison-Wesley, New York, in press.
25. Brinkley, J.F., Cornelius, C., Altman, R.B., Hayes-Roth, B., Lichtarge, O., Duncan, B., Buchanan, B. and Jardetzky, O. (1986) *Application of Constraint Satisfaction Techniques to the Determination of Protein Tertiary Structure* Technical Report KSL-86-28, Stanford University.
26. Hayes-Roth, B., Buchanan, B., Lichtarge, O., Hewett, M., Altman, R., Brinkley, J., Cornelius, C., Duncan, B. and Jardetzky, O. (1986) IN: Proceedings of AAAI-86, Fifth National Congress on Artifical Intelligence, **2**, 904-909 Morgan Kaufman Publ, Los Altos, CA.
27. Wuthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) *J. Mol. Biol.* **155**, 311-319.
28. Wemmer, D. and Kallenbach, N.R. (1983) *Biochemistry* **22**, 1901-1906.
29. Zuiderweg, E.R.P., Kaptein, R. and Wuthrich, K. (1983) *Proc. Natl. Acad. Sci.* **80**, 5837-5841.
30. Richardson, J.S. (1981) *Adv. Protein Chem.* **34**, 167-339.
31. Levitt, M. and Chothia, C. (1976) *Nature* **261**, 552-558.
32. Mackworth, A.K. (1977) *Artificial Intelligence* **8**, 99-118.
33. Mackworth, A.K. (1985) *Artificial Intelligence* **25**, 65-74.
34. Lichtarge, O. (1986) Ph.D. Thesis (Biophysics), Stanford University.
35. Rich, E. (1983) *Artificial Intelligence* New York, McGraw-Hill.
36. Hayes-Roth, B. (1985) *Artificial Intelligence.* **26**, 251-321.
37. Erman, L.D., Hayes-Roth, F., Lesser, V.R. and Reddy, D.R. (1980) *Computing Surveys* **12**, 213-253.
38. Nii, H. P., Feigenbaum, E.A., Anton, J.J. and Rockmore, A.J. (1982) *AI Magazine* **3**, 23-35.
39. Terry, A. (1983) Ph.D. Thesis (Information and Computer Science), University of California, Irvine.
40. Altman, R.B. and Jardetzky, O. (1986) *J. Biochem*, in press.
41. Ohlendorf, D.H. and Matthews, B.W. (1983) *Annu. Rev. Biophys. Bioeng.* **12**, 259-284.