Generation of 3D structures from distance information

University of York, York, UK, 8–10 January 1987

The three day meeting held in York was the first full conference on the topic of generation of 3D structures from distance information. The meeting was well attended (118 delegates, with 44 from overseas) and extremely buoyant throughout.

Dr Alan MacKay (Birkbeck College, London) introduced the general theme of the meeting; given a set of points and bounds on the distances between them, which 3D representations of these points are feasible? He described this as an example of a 'trapdoor' problem where the transformation in one direction is considerably harder than in the reverse direction. Generating a set of distances from a set of atomic coordinates is easy; the reverse transformation is much harder. An amusing illustration in his introduction was the demonstration that by using the air fare between European capital cities as a measure of the distance between them it is possible to generate, fairly accurately, the positions of these cities on a map of Europe!

In the distance geometry approach to generating 3D molecular structures from distance information, key steps are to obtain a metric matrix from a trial distance matrix and then to project this metric matrix into 3D. Dr Gordon Crippen (University of Michigan) introduced the basic elements of distance geometry and went on to describe his recent attempts to improve the methodology by taking account of energy in the projection into 3D, a process known as 'energy embedding'. He reported that the structures produced were always of low energy and in fact were often at the global minimum. At present this technique is limited to about 50–60 atoms but it will clearly become a powerful tool as this limit is extended. This presentation prompted much discussion on whether a single point representation for each amino-acid in a protein would be adequate in a search for the global minimum-energy structure. Only time will tell. Either way, it is already clear from the number of later references in the meeting to the use of 'distance geometry structures' as starting points for other techniques that distance geometry methods are an extremely valuable contribution to the general problem of providing 3D molecular structures from distance information.

Dr Wilfred Van Gunsteren (University of Groningen) introduced the second major technique dealt with at this meeting, restrained molecular dynamics. In this method the distance constraints are formulated as an extra term added to a standard molecular mechanics force field. Either molecular mechanics or molecular dynamics can be used but, using the GROMOS suite of programs, van Gunsteren showed that the ability of molecular dynamics to take a structure over barriers of the order of kT gave great advantages over energy minimization techniques.

NMR data on the *lac* repressor headpiece allowed van Gunsteren to explore the applicability of dynamics both as a standalone technique and in combination with distance geometry. He discussed the concept of 'non-NOE', meaning that when used judiciously the lack of a nuclear Overhauser enhancement (NOE) between two resonances can be taken to indicate that the atoms giving rise to these resonances are further apart than a defined distance. Although the quality of the structure was not obviously improved the convergence time was shorter.

Andrew Torda (University of New South Wales) described work on polypeptides of about 50 residues that are found in a northern Pacific sea anemone. These have a potent effect in increasing the force of contraction of mammalian heart muscle. Working from ^{1}H NMR assignments and a set of more than a hundred distance constraints from NOEs, this was a good example of where it had been possible to derive a 3D structure using restrained molecular dynamics. Torda is presently repeating the exercise but carrying out the dynamics in water.

Dr Analisa Pastore (Oxford University) reported studies on the 'Molecular dynamics of membrane proteins' in particular two cytolytic polypeptides, δ-haemolysin and melittin. She explored the use of restrained molecular dynamics, starting with both helical and extended structures, and used the coupling constants as an independent check on the structures produced.

Dr Lennart Nilsson (Karolinska Institute, Stockholm) demonstrated that these techniques are also applicable to the derivation of nucleic-acid structure. From NOE information and restrained molecular dynamics he was able to describe details of the 3D structure of the anticodon region of an RNA pentadecamer fragment from tRNA.

The second day of the meeting began with Dr Werner Braun (ETH, Zurich) comparing two programs written specifically for the determination of protein structures from NMR data. Disgeo (written by Dr T Havel and available from QCPE) uses the distance geometry technique described by Crippen. Braun's Disman uses a distance-based penalty function and minimizes this function allowing only torsion angles to vary. To overcome the problem of local minima Braun described how Disman varies the number of distance constraints included in the penalty function.

Initially data for residues close in sequence are included, allowing local conformational space to be sampled.

Subsequently the longer-range (in sequence) data are included, resulting in the formation of tertiary structure. In test data with BPTI both methods were shown to reproduce the backbone of the crystal structure to an RMS deviation of less than 2Å. Of the two methods, Braun's view was that Disman sampled the range of structures consistent with the input distance constraints better than did Disgeo. In fact the randomness of sampling of conformational space by the various methods was a topic of interesting debate throughout the meeting.

Using real NMR data on the alpha amylase inhibitor Tendamistat, Braun obtained structures that agreed well with the crystal structure obtained independently and in parallel by Huber (Munich).

Dr Marius Clore (Max Planck Institut für Biochemie, Martinsried bei München) further demonstrated the power of NMR for protein structure determination by describing the application of restrained molecular dynamics to a range of small proteins, including hirudin, a potent thrombin inhibitor for which no crystal structure is known. His start points for restrained molecular dynamics calculations were either the output of distance geometry calculations or were randomly chosen. In one case, crystallographers have been able to use Clore's NMR-based structure in their refinement calculations.

Clore described the protocol for his dynamics calculations. Initially only constraints between atoms less than five residues apart in sequence were included and the force constant was chosen to make the restraint relatively soft. As the calculation progressed, longer-range constraints were included and the force constant was increased. There is a similarity between this and the variable target approach in Disman. Using this protocol Clore said that in general one may be able to start from extended structures and to produce structures consistent with the input constraints. He showed this to be the case with model calculations on Crambin. However this is a protein containing several helical regions and the initial constraints are good for allowing the helices to nucleate. It remains to be seen how successfully proteins with no helices will fold in similar circumstances.

Dr Flemming Poulsen (Carlsberg Institute, Copenhagen) described the use of distance geometry to determine the solution conformation of a subtilisin inhibitor. Comparison between this structure and an X-ray structure of the inhibitor bound to subtilisin indicated that the gross conformational features of the two were similar.

Dr R Boelens (University of Groningen) described work on a fragment of the *lac*-repressor in which one headpiece domain was bound to its recognition sequence in a length of double-stranded DNA. Fortunately ^{1}H NMR signals from nucleotide bases are sufficiently well separated from aromatic amino-acid signals that the DNA structure can be monitored to some extent in the complex. Similarly aliphatic regions of the protein spectrum are still recognisable in the complex. Ordinarily an Mr = 14 000 would be beyond the range of 2D NMR, but because of 'fast exchange' it is possible to titrate resonance assignments from the protein in free solution to the protein in the complex. Noesy experiments then allow a few distance mappings to be made between recognisable hydrogen atoms in DNA and recognisable hydrogen atoms in the protein. With the protein having the helix–turn–helix motif now becoming familiar in the DNA-binding proteins, this allows a 'fix' on the binding position, at least in the fragments under study.

The final two speakers described their methodology for finding the conformations of small cyclic peptides consistent with NMR data. Dr Alan Tonge (Glaxo) described how he had used exhaustive conformational searching, factor analysis and restrained energy minimization to determine the solution conformation of Tyrocidine A. Dr Dominique Marion (CNRS and University of Orleans) included distances as Langrangian constraints in energy minimization calculations on two bacterial lipopeptides. The spectra of these lipopeptides were determined in pyridine solution and their rate of molecular tumbling meant that to obtain NOE information by conventional methods it is necessary to collect spectra at −20°C. An alternative approach, also described, was to collect spectra at room temperature from rotating frame experiments using the Roesy scheme.

The third day of the meeting was informal and was intended to give people an opportunity for further developing points which had arisen during the formal sessions. Over 70 people stayed for this final session and a lively discussion ensued, particularly in the following areas.

A point of considerable interest was that with distance geometry methods no-one has yet been able to provide a good algorithm for selecting trial matrices such as to ensure random sampling of conformational space. Disgeo uses a Gaussian distribution in its random sampling procedure. In a simple but interesting test case introduced by Dr van Heussel (Groningen) it was shown correspondingly that Disman adequately samples the boundaries of the allowed regions but not the allowed interior, although the likely consequences in a real problem are unclear.

The distance geometry procedure involves a loss of information and a reduction in molecular size during projection down from a space of high dimensionality to 3D. At some point, therefore, a rescaling is necessary. The scaling applied in Disgeo, for example, is based on the radius of gyration for the structure. A radius of gyration can be calculated from the trial matrix and, in the absence of any other good means of deciding a scaling factor, an assumption was made in Disgeo that the final structure should have the same value. This seems a reasonable empiric decision, but a better algorithm would clearly be valuable. Dr Sheek (Groningen) though there was promise in an alternative scaling procedure based on optimising the scaling factors via a penalty function relating the projected distance matrix to the initial trial matrix.

There was a lot of interest in discussions on the best start points for restrained molecular dynamics calculations. The meeting favoured as complete a sample of randomly generated structures as possible; in particular, model structures built at a graphics terminal were felt to be difficult to free from subjective bias. The use of distance geometry and Disman methods to present plausible start points for restrained molecular dynamics runs appears to be a good option although, as mentioned above, the adequacy of sampling of conformational space needs further investigation.

Discussion took place on what kind of protocol to use for restrained molecular dynamics. Van Gunsteren's advice was that in general, when the purpose is to generate 3D structures consistent with distance information from measurements for nuclear Overhauser enhancements, the simulation should be carried out *in vacuo*; with current cost/performance ratios in computing, inclusion of water will usually slow down the convergence more than is justified by the results.

The initial temperature and the force constant for the distance constraints are selectable variables. Ideally one should be able to find a structure which agrees with NMR data and is still stable when the penalty for violating distance constraints is removed. Starting from distance geometry produced structures, only 10-15ps of molecular dynamics simulation is required. Random start structures, on the other hand, need considerably longer periods of simulation.

A structure can be generated from the dynamics simulation in a number of ways. An average of the last few picoseconds of the simulation can be taken or, alternatively, the system can be gradually cooled down to 0 K. The most common method, however, is to energy minimize a structure selected from the simulation.

The criterion for choosing 'non-NOEs' as a constraint between two atoms was well discussed. As Marion noted, the absence of an NOE between two resonances does not necessarily mean that the two atoms giving rise to these resonances must be greater than a certain distance apart. The presence of a trace of a paramagnetic ion might eliminate many NOEs. Other reasons for being unable to detect specific NOEs include flexibility in the region of the protein involved, 'fast' exchange among different environments and line broadening due to 'intermediate' exchange processes or from a resonance being *J*-coupled to other resonances. Boelans described a program developed at Groningen which examines all pairs of resonances which do not exhibit NOEs and eliminates from consideration those for which reasons such as those given above may apply, and incorporates the remainder as lower bound constraints in distance geometry and restrained molecular dynamics simulations.

The question of 'handedness of fold' has been identified as a problem for distance based methods; perhaps more disconcerting was the suggestion by Braun that, for sections of a structure, distance data may not be able to distinguish between two solutions in which the local folds are almost mirror images of each other.

Another topic of interest was the group of issues surrounding the possible submission of NMR-based protein structures for incorporation into publicly accessible databases such as the Brookhaven collection. The general feeling of the meeting was that the structures being produced were of sufficient quality to be valuable to the general scientific community, particularly those involved in biotechnology.

Whether to submit several sample structures from a molecular dynamics run, or some kind of average, will still need debate. Since relatively small proteins are being studied at the moment, however, the amount of storage space is not an issue. 10 sets of coordinates for a protein of 50 residues (*M* about 6 000) would occupy no more disc space than a single protein of 60 000. In fact it would occupy less, since some of the data would not need to be replicated if the database were suitably constructed.

Altogether the meeting was extremely timely and it provided a valuable forum for exchange of ideas and information in a rapidly developing area.

B Sheard, D Timms and A J Wilkinson
ICI Pharmaceuticals, Macclesfield, UK