of the peptide backbone have been smoothed out to form a continuous space curve. It takes about 2 h of VAX equivalent time to do the smoothing and about 2 min on the S1-100 array processor. When compared to the carbon α-representation which is too simple and harsh or the normal peptide backbone representation which is too complex, the smoothed space curve representation has the airy qualities of the original Richardson diagrams. Using Jane Richardson's monograph as a guide and the Protein Data Bank from Brookhaven National Laboratory, 125 proteins and/or domains have been extracted. Each space curve is coloured to show helices and sheets and is oriented to bring the N-terminus and/or the C-terminus of the protein to the front. Using stereo pairs of images it is possible to begin to think about the late stages of the folding of proteins. In reality what one does is imagine the unfolding of the protein from the crystal structure, hence the name of the kit. There are some real surprises. You can really imagine the sequence of motions by means of which the protein folded. The space curve representation makes it easy to think about 'tugging' loops, peptide strands and domains in 3D. In work to follow, I will systematically pull apart all of these proteins. The Protein Unfolding Kit will be distributed in stereo slides, video tape (3/4 inch industrial, VHS and Beta formats), and as coordinate datasets on magnetic tape.

5

Modelling of related proteins

A M Lesk and C Chothia
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The general response of protein structures to mutations, insertions and deletions is conformational change. A 'core' of the structure, that retains its basic folding pattern, comprises major elements of secondary structure and residues flanking them, including active-site peptides. The core may be as little as 40% of the structure for distantly related proteins, but is 90% or more for proteins with amino-acid sequence homologies of 50% or more. There is a direct relation between the structural difference of the main-chain atoms of the core residues of a pair of proteins and the overall amino-acid sequence homology. The deviation reflects the shifts and rotations of packed secondary structures with respect to one another. Successful model building of an unknown protein depends on knowing the structure of a reasonably close relative. We describe the prediction and test of a model for the V_L and V_H domains of the antilysozyme antibody D1.3.

6

An algorithm which predicts the conformation of short lengths of chain in proteins

J Moult and M N G James
MRC Group in Protein Structure and Function, Biochemistry Department, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada

Graphics aided model building of protein molecules homologous to those of known structure is often successful for the conserved core of the protein structure, but fails to predict correctly the conformation of regions where there is low homology between the amino-acid sequences. Since such regions are often the principle determinants of functional differences between proteins this is a serious deficiency. We report here an algorithm which can be used to examine a large number of possible conformations of such regions, using a series of filters so that only a few of these need be subjected to full energetic evaluation. The loop sequence to be considered is divided into two halves. Then each half is grown from its fixed end on the rest of the protein independently, using backbone atoms only, and generating a set of internally allowed conformations. All such conformations which penetrate more than a specified distance below the surface of the protein are rejected as the chain grows. Next, pairs are selected from the two sets of independent half chains for which the end-to-end distance is less than a specified amount, and their ends joined. Side chains are then built onto each of the accepted full loops, generating a set of internally allowed conformations, and again rejecting those for which there is excessive protein penetration. Finally, remaining conformations are energetically evaluated, including solvent effects by the use of image charges and the size of the exposed hydrophobic area. This procedure has been applied to the prediction of the conformation of surface loops of *Streptomyces griseus* trypsin, up to 6 residues in length. For these cases, the number of possible conformations (including the side chain combinations) is restricted by the filtering to a few thousand, and the lowest energy conformation amongst each set has nearly the lowest RMS to the X-ray structure. The best RMS values are between 0.5 and 1.1 Å.

7

Development of software tools for protein structure design

H Nakamura*, T Yamazaki*, H Abe, T Noguchi†, Y Seno††, H Wako‡ and N Go‡‡**
*Department of Applied Physics, Faculty of Engineering, University of Tokyo, 7-2-1 Hongo, Bunkyo-ku, Tokyo, 113 Japan
**Department of Physics, Nishinippon Institute of Technology, Japan
†Faculty of Engineering, Kyushu University, Hakozaki, Higashi-ku, Fukuoka-shi, Japan
††Department of Physics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan
‡Juntendo University, Japan
‡‡Department of Physics, Kyushu University, Hakozaki, Higashi-ku, Fukuoka-shi, Japan

Several software tools for designing new proteins by site-directed mutagenesis have been developed by our working group.
Molecular graphics tool: Mild (Molecular ILlustration and Design) is a molecular graphic program on a raster 3D graphic system (Seiko GR3000). Interactive usage with different models including solid models is possible. Real time amino-acid substitution is available on the

screen. Another graphic program GRAIP (GRaphics of Atomic Information Picture) used to draw a pixel image of the space filling model with its cross section is linked to Mild, using colours to indicate information.
Conformational energy calculation: FEDER (Fast ECEPP and its DERivatives) is a program used to calculate the conformational energy of proteins with respect to dihedral angles based upon the ECEPP function. It finds correct energy minima by a rapid calculation of the first and the second derivatives of the energy function using a new algorithm[1]. Its modified version for an array processor (S-810, Hitachi, 630MFLOPS) is much more effective.
Standard format to describe information of proteins: FLEXS (FLexible, EXtensible and Structural) format is a new format, which can describe any information of proteins with any type and any length, structurally. The Protein Data Bank (PDB) format is widely used, but there are many limitations, especially in describing protein structures by dihedral angles only. The FLEXS format links many program tools, such as between Mild and FEDER. Even when a relational database is available, such a format is still necessary for group work. Examples of the performance of these tools are shown by photos and a film, in which a Monte Carlo simulation of thermal fluctuation[2] of the ovomucoid and the manipulation of Mild are introduced.

References

1 **Abe, H et al.** *Comput. & Chem.* Vol 8 (1984) p 239
2 **Noguchi, Y and Go, N** *Biopolymers* Vol 24 (1985) p 537

8

Molecular dynamics simulations: techniques, experimental basis and determination of free energies

W F van Gunsteren
Department of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

The basic procedures and techniques of the computer simulation method of molecular dynamics (MD) will shortly be reviewed. Its predictive power will critically depend on the basic approximations that are made and the force fields that are applied. Therefore, results of MD simulations of various molecules will be compared to experimental data derived from X-ray, neutron diffraction of 2D NMR experiments. Second, the derivation of the thermodynamic properties, entropy and free energy for systems of interacting particles from simulations will be discussed. This is by no means a trivial task, but recently progress has been made. Examples of the calculation of the free energy of hydration of methanol will be given. Possible application of these techniques to calculate binding constants of inhibitors or substrates to enzymes will be evaluated. As an example the calculation of the relative binding constant of two different inhibitors of dihydrofolate reductase (DHFR) will be discussed.

9

Crystallographic approaches to nucleic acid structure and dynamics

E Westhof
Institut de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique, 15 rue R Descartes, F-67084 Strasbourg, France

Our knowledge of the tertiary structure of nucleic acids is derived from X-ray diffraction studies of crystalline fibres and single crystals. However, the fact that X-ray crystallography yields well-defined structures characterized by a set of coordinates fosters a rather rigid and static view of nucleic acids. Only recently was it shown that X-ray diffraction can contribute to our knowledge of the dynamic properties of macromolecules. This advance was made possible by the development of refinement methods, which allow the precise determination of atomic coordinates and of the atomic Debye–Waller factors, together with the impetus given by molecular dynamics simulations of macromolecules. Further, the availability of X-ray structures of DNA oligomers and t-RNA molecules at various resolutions has opened up a new era in the study of the structure and biological functions of nucleic acids and of their interactions with other ligands and proteins. The correct interpretation of crystallographic data, especially of low to medium resolution, depends on a thorough understanding of the effects of the refinement model upon the derived quantities, such as coordinates, thermal parameters, and positions of solvent molecules. Several examples of nucleic acid structures refined with molecular graphics, molecular mechanics, or molecular dynamics will be presented and discussed: at high resolution, a Z-DNA hexamer and its solvation; at medium resolution, a comparison between two refinement methods of a B-DNA dodecamer as well as a comparison between two tRNA molecules; at low resolution, a study of heavy metal and drug binding to Z-DNA oligomers.

10

MaxTwist: energy minimization of macromolecules in intelligent degrees of freedom

C Sander
Abteilung Biophysik, Max-Planck-Institut fuer medizinische Forschung, Jahnstrasse 29, 6900 Heidelberg, FRG

'Babysitting' irrelevant degrees of freedom is an unnecessary and very time-consuming activity of energy minimization programs. For example, bond stretching requires an energy several times kT, at room temperature, for a single quantum excitation. So there is usually no biological interest in variations in bond length. Yet the most widely used intramolecular minimizers vary all $3*N$ atomic position coordinates, i.e. more than a thousand degrees of freedom for a 58-residue protein. A lot of time is spent changing variables that physically do not change. MaxTwist differs in two respects. 1) The basic description of molecular conformation is in terms of internal coordinates, not atomic position coordinates.