Molecular Modeling with Transputers

David N.J. White, J. Noel Ruddock, Paul R. Edgington and Guy H. Grant
Molecular Modelling Laboratory, Chemistry Department, Glasgow University, Glasgow, G12 8QQ, UK

A number of the key techniques in molecular modeling, such as molecular mechanics calculations, molecular dynamics calculations, and conformational search procedures, are extremely computationally demanding. Within the last year or two, parallel processing in general, and transputers in particular, have emerged as potential low-cost (relative to a vector supercomputer) solutions to the computational bottleneck. However, before any new technology can be useful in the computer-aided molecular design context, it must be supported by an adequate range of applications software.

Now that multiuser, multitasking operating systems and completely FORTRAN parallel programming environments are available for the transputer, the development of entirely transputer-based, parallel molecular graphics code is proceeding apace.

We have completed the development of an extremely comprehensive molecular graphics package for small, medium and macromolecules that runs on a transputer-based workstation with 64 megabytes of memory and a high-performance, high-resolution color raster scan display. When the molecular mechanics option is invoked from the workstation, the subsequent calculations are performed transparently, and in most cases in real time, on an attached transparallel minisupercomputer with 32 transputers and 128 megabytes of memory. Similarly, when conformational search procedures or molecular dynamics calculations are required, the transparallel minisupercomputer is dynamically reconfigured and reloaded with the appropriate parallel code for the task at hand. It has been common experience that 32 transputers deliver the same compute power on molecular graphics code as the Cray-1 or Cray XMP-12.

The transparallel minisupercomputer is also being used to attack one of the major difficulties with molecular mechanics calculations, that of obtaining high-precision force fields for a wide range of functional groups. At the present time, almost all force field development is done essentially by a time-consuming trial-and-error process, mainly because the computational optimization of force constants is very complex and very time-consuming. We have developed a parallel computational force constant optimization program that runs on our transparallel minisupercomputer and allows fast, accurate extension of force fields to encompass new functional groups. A new version of the White–Bovill force field and its extension to include new functional groups will be presented.

The very considerable calculations and database searching necessary for accurate computational protein model building are ideally suited to the workstation/minisuper/software described above. We will present brief details of some of our experiences involving model building of a serine protease, which required several large insertions to mutate the X-ray structure of the homologous protein into the target structure.

Model Building Studies of the C2 Component of Complement

G. H. Grant and D. N. J. White
Department of Chemistry, The University of Glasgow, Glasgow, G12 8QQ, UK

The complement system is a multimolecular biological system whose principal role is as the effector mechanism of the immune system against infection by microorganisms. The system itself consists of 11 proteins that interact via two possible pathways to provide an immune response. In the classical pathway the proteins can be grouped into three units: recognition, activation and membrane attack. Activation of C2 represents the initial stage in the second of these.

C2 is a single chain serine protease that shows considerable homology with the protein factor B. During complement activation, C2 is cleaved to form C2b (223 residues) and C2a (509 residues). Computational alignment of the C-terminal half of C2a has taken place and has shown this part of the protein to be around 35% homologous with another serine protease of known structure, trypsin, with most of the sequence conservation occurring in regions of regular secondary structure. The alignment has also shown that a series of significant gaps (> seven residues) exists in the trypsin sequence, with respect to C2a, but these are all found to occur in loop areas on the proteins's surface.

A methodology for protein model building has been developed and implemented in a modeling system, facilitating the conversion of a known structure into a model of a homologous protein where only the sequence is known. The residue substitutions are carried out such that the side chain is placed, where possible, in a similar conformation to that of its precursor. Insertions and deletions can be carried out automatically, or a fragment search can be initiated to find specific features from the Protein Data Bank. The modifications can be annealed after each change, using energy minimization of a zone surrounding the new residue or by Monte Carlo or dynamics calculations to search conformational space.

The structure of C2a has been modeled using the above approach, with loop insertions of up to 15 residues being made on the basis of grafting known loop structures onto the protein, conserving main chain interatomic distances where possible. In the case of the smaller loops, these can be compared to automatically generated loop structures where distance constraints have been applied to a conformational search procedure.