Simple protein model building tool

Kazunori Toma

Computer Science Department, Asahi Chemical Industry Co Ltd, 2-1, Samejima, Fuji, Shizuoka, 416 Japan

An interactive protein model building program, named Alpha, running on the Evans & Sutherland PS340, is presented. It has two prominent features: flexible construction and an informative display of the protein model. These characteristics arise from the adoption and analysis of the α-carbon representation of a protein. Although its concept and program are simple, Alpha is a useful tool for investigation of the 3D structure of a protein, whether or not it is elucidated.

Keywords: molecular graphics, protein structure, α-carbon model

Received 3 March 1987
Accepted 31 March 1987

Many computer graphics programs exist for the representation of protein molecules, and some of them have been described in this journal[1,2]. However, they are used almost exclusively for either the investigation of a known protein structure or as an aid in structure-elucidation studies, in combination with X-ray crystallography[3,4], nuclear magnetic resonance[5], or energetic calculations.

However, even in the study of a protein of undetermined structure, use of some models may be helpful, even if only to achieve a rough image. Although the recent development of gene engineering means that such models are now needed more than ever, few suitable protein graphics programs exist. Computer graphics software, which allows model building of a protein of unknown 3D structure, could be used in protein-structure prediction studies, protein engineering work, etc.

Some important features of such a program are: interactive response, flexibility in structure modification, and informative representation of protein structure. Alpha is devoted to these goals. It handles protein structure using the α-carbon model and illustrates structural features by means of colour-coded spheres which correspond to the characteristics of the amino acids.

IMPLEMENTATION

Alpha was developed using a Vax 11/780 as the host computer and an Evans & Sutherland PS340, which has both calligraphic and raster displays, as a graphic terminal connected by a DMR-11 communication line. The calligraphic display was used as a model-building terminal and the raster display as a CPK-like representation of the molecule in which the CPK modelling firmware of Evans & Sutherland was used. The program was written in VAX-FORTRAN and the PS340 programming language.

α-CARBON MODEL OF PROTEIN

The α-carbon model of protein, in which protein structure is represented by single-line virtual Cα–Cα bonds, is selected as a structure-building model not only because it is simple and has enough information about main-chain folding but also because it gives considerable flexibility.

The flexibility arises from the nature of the α-carbon model. Analysis of several protein structures in the Protein Data Bank[6] reveals that, except in some cis-peptide bonds, the length between neighbouring α-carbons is about 3.8 Å and that the angles formed by neighbouring groups of three α-carbons range from 60° to 160°, which are good indicators of the secondary structure types. These facts reflect the nature of the polypeptide structure in which α-carbons are connected mainly with trans-peptide groups. Taking these features into account, in Alpha the virtual bond length between α-carbons is set as constant, and the virtual bond angle formed by neighbouring three α-carbons is allowed to change freely. It makes model building much smoother than with the traditional dihedral angle control around chemical bonds, which is structurally valid but frustrating to handle.

COLOUR CODING OF AMINO-ACID TYPES

The 20 naturally occurring amino acids can be classified by some physicochemical properties which are important in the determination of tertiary structure. In Alpha, amino-acid residues are shown with the colour-coded spheres centred on the α-carbon positions, like the CPK representation of atoms, on the raster display. The colour indicates the classification, such as hydrophilic or hydrophobic, charged or noncharged, acidic or basic, small or large, etc.

In addition, sequence-dependent information, such as the conserved residues among related proteins, the result of secondary-structure predictions, the positions of introns etc., is also important in the study of the tertiary structure. Alpha treats such information in the same way as above, if adequate tables are prepared.

OVERVIEW OF ALPHA

Model building and transformation are performed on the calligraphic display and use both front and side

 0263-7855/87/020101-02 $03.00

views. Although the display has depth-queueing capability, it is sometimes very difficult to recognize the correct chirality and the two-eye point method is considered to be superior in the construction of the protein model. As in other systems, real-time rotation, translation and scaling by the dials are available (see Colour Plate 1).

To modify the structure, one must press the modelling function button, then touch an amino-acid point using the tablet pen. After that, both parts can be rotated freely around the chosen point using the dials. During modelling, a template structure can be used, which is helpful in the construction of standard secondary structures, such as the α-helix and β-strand, or in the incorporation of partial structures taken from known proteins, such as those in the Protein Data Bank (see Colour Plate 2).

To obtain the amino-acid-type-colour-coded CPK-like model on the raster display, one has only to press the function button. There are choices according to physicochemical and sequence-dependent classifications. If the protein model on the calligraphic display is moved forward, the cross-sectional view of the molecule can also be shown. This representation is useful not only in the evaluation of the building structure, but also in the analysis of the known structure (see Colour Plates 3 and 4).

CONCLUSION

A protein is made from amino acids, and an amino acid is made from atoms. This hierarchical nature of protein structure explains the usefulness of Alpha. The level of amino acids is adequate to investigate the rough image of protein tertiary structure, especially those that are undetermined, and to handle all atoms of a protein at one time is a difficult task and may be meaningless in most cases.

ACKNOWLEDGEMENT

The author would like to thank H Watanabe for comments and helpful suggestions. This work was supported by grants from the Science and Technology Administration.

REFERENCES

1 **Chevrier, B and Ripp, R** 'Program for the visualization and interactive study of molecules on a calligraphic display system' *J. Mol. Graph.* Vol 4 No 4 (December 1986) pp 217–219

2 **Burridge, J M and Todd, S J P** 'Protein secondary structural representations using real-time interactive computer graphics' *J. Mol. Graph.* Vol 4 No 4 (December 1986) pp 220–222

3 **Jones, T A** 'Frodo: a graphics model building and refinement system for macromolecules' *J. Appl. Crystallogr.* Vol 11 (1978) pp 268–272

4 **Diamond, R** 'Builder: a computer graphics program for biopolymers and its application to the interpretation of the structure of tobacco mosaic virus protein disc at 2.8 Angstroms resolution' in **Srinivasan, R et al (Eds)** *Biomolecular structure, function, conformation and evolution* Pergamon Press, UK (1981) pp 567–588

5 **Billeter, M, Engeli, M and Wüthrich, K** 'Interactive program for investigation of protein structures based on ^{1}H NMR experiments' *J. Mol. Graph.* Vol 3 No 3 (September 1985) pp 79–83

6 *Protein Data Bank* Chemistry Department, Brookhaven National Laboratory, New York, USA

7 **Loebermann, H, Tokuoka, R, Deisenhofer, J and Huber, R** 'Human α1-proteinase inhibitor: crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function' *J. Mol. Biol.* Vol 177 (1984) pp 531–556

Erratum

On page 44 of Volume 5 No 1 (the March 1987 issue) in the paper by S E Jakes, N Watts, P Willett, D Bawden and J D Fisher the algorithm (lines 36–45; right hand column), should have read:

```
CREATE_DISTANCE_TABLE;
FOR J := 1 TO Q DO
   BEGIN
      INTERSECTION := JLIST[1];
      FOR X := 2 TO Q-1 DO
         INTERSECTION := INTERSECTION
         AND JLIST[X];
      FOR X := 1 TO Q-1 DO
         JLIST[X] := JLIST[X] AND
         INTERSECTION
   END;
```