BRAGI: A comprehensive protein modeling program system

Dietmar Schomburg and Joachim Reichelt

GBF (Gesellschaft für Biotechnologische Forschung), Mascheroder Weg 1, D-3300 Braunschweig, FRG

A protein modeling program package has been developed. The user friendly system is implemented on high performance graphic devices and facilitates the modeling of a protein with an unknown three-dimensional structure out of that of a homologous one or the design of new variants. A wide range of features can be used by the researcher for this purpose. The system is written in FORTRAN 77 and E&S GSR functions or the HP implementation of the PHIGS standard, respectively.

Keywords: molecular modeling, protein design

Received 12 April 1988
Accepted 1 June 1988

INTRODUCTION

Since the first reports on the use of site directed mutagenesis in 1982,[1] protein engineering or — when rationally aimed — protein design has been recognized as a promising and fascinating field of research in many countries. In Japan (PERI) and the USA (CARB) research institutes have been founded with the focus on protein design. More and more research institutes in the United States, Canada, Japan and Europe have been starting broad research projects on protein design (UK: SERC, W. Germany: GBF, EMBL).

Possible prospects for applications of designed proteins with new activities or other new properties are very high in the areas of pharmacology, enzyme applications in food industry, waste treatment and chemical synthesis, vaccine design and biosensors, among other uses.[2-4] But encouraging results have so far been obtained for only a small number of cases, including insulin, proteases and peptidic protease inhibitors, and some others.[5,6] On the other hand, many unpredicted and surprising results of site-directed mutagenesis experiments are reported at scientific meetings and in the literature.[7,8] These developments show that our methods and tools in this area are still rather crude and urgently require improvement.[9,10]

In the course of designing and producing new proteins, many different steps are necessary (see Figure 1, for an overview[11]). Two very important ones involve protein modeling:

(1) Construction of the three-dimensional (3D) structure of a protein from that of a homologous one as an alternative to experimental structure determination (X-ray diffraction or two-dimensional n.m.r.).
(2) Identification of those positions in the protein sequence that should be changed, and prediction of the result this change in the primary structure has on 3D-structure, activity and stability of the protein.

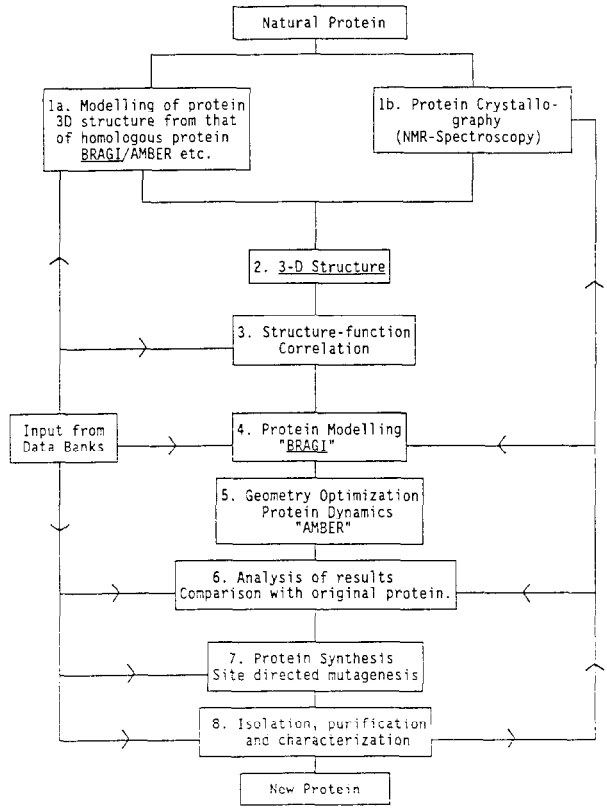

Figure 1. Steps involved in designing new proteins

In modeling a new protein three principal steps are involved:

(1) Starting from a given 3D structure a new protein is modeled with the help of molecular graphics. Possible sites for amino acid exchanges are identified and local changes are applied.
(2) The 3D structure of the modified protein is

predicted by force field calculations, including energy minimization and protein dynamics.

(3) The predicted 3D structure is compared with that of the original protein. By application of the knowledge about structure–function or structure–stability correlations, the new protein is evaluated and either synthesized or the cycle is started again at step 1.

With respect to step (2) a considerable improvement has already been achieved or is at least possible by using better force fields, including solvent force fields, by use of higher speed parallel computers and by new mathematical approaches to the local energy minima problem.

Step (1) has an especially great influence on the success of the modeling. Here the researcher must apply knowledge of the structure–function and structure–stability correlations together with knowledge about structural chemistry, molecular interactions, etc. He also needs access to a number of different data banks, including protein sequence data banks, the Brookhaven data bank of known protein 3D structures, the Cambridge Crystallographic data bank, and several others, depending on the project. The success of this step depends greatly on the computer software that should provide the researcher with flexibility in the display and in modifying proteins, but it should also help to avoid traps like the local energy minima problem. In addition the interactive start of protein force field calculations and an analysis and comparison of protein 3D structures should be possible.

In the course of the GBF activities of protein design, we developed a computer software package that is especially suited for performing steps (1) and (3) and that helps in step (2).

The program BRAGI was developed for the purpose of modeling proteins with new properties from the known 3D structure of existing ones. Although FRODO[12] is used in our group for the purpose of density fitting in protein crystallography, we felt a need for a program that was written especially for the use in protein design projects and

- can be used almost instantly by the average biologist who is not a computer specialist
- allows exchanges of amino acid residues, change and manual optimization of side chain and main change torsion angles, and the examination of the resulting protein in a very straightforward and fast way
- has an effective interface to a protein force field and protein dynamics program
- gives a realtime feedback about the repulsion energy during change of torsion angles
- offers several different coloring schemes (atoms, residues, molecules, different color schemes in different molecules, etc.)
- allows fast orientation in a protein by displaying secondary structure details

BRAGI is menu driven, but experienced users can also type in commands from the keyboard (Color Plate 1). It is implemented on an Evans & Sutherland PS 300 vector graphics screen connected to a DEC VAX (VMS) via DECNET, and on the Hewlett-Packard SRX high performance raster workstations.

BRAGI offers a comprehensive protein modeling environment. It allows the display and manipulation of proteins and small molecules in several different display modes. A van der Waals or solvent-accessible surface can be displayed as a dot surface, a solid image or transparent CPK surface. The program allows a high flexibility in performing changes in the sequence and geometry of a given protein. A quasi-realtime feedback about the repulsional term of the nonbonding energy is given while rotations around the bonds are performed by the user. This helps to avoid local energy-minima. Rms-fitting of proteins and the analysis of intra- and intermolecular contacts is also possible.

The final step of energy-minimization and/or protein-dynamics is made very easy by an interface to AMBER. BRAGI creates the necessary files, starts the AMBER programs and afterward rebuilds a Brookhaven-type file with the original names and residue numbers, so that the resulting 3D structure can easily be compared with the original structure.

The program is constantly being improved, according to needs and ideas arising from protein modeling tasks performed at the GBF. It has been used for several protein design projects in different groups. Color plates 1–4 and 6 and 7 show a complex between a human PSTI variant (3D structure modeled) and human leukocyte elastase (modeled 3D structure). The structures were modeled in the course of a successful protein design project that consisted of the design of an elastase inhibitor from a trypsin inhibitor (human PSTI).[13]

DISPLAY AND MANIPULATION OF PROTEINS

In the present version BRAGI is meant to be a valuable tool in protein modeling. We decided not to add an extra file format to the range of existing ones, but to use exclusively the protein data bank ("Brookhaven") format. BRAGI can also be used to display small molecules in the format of the Cambridge Crystollgraphic Data Centre or the "SHELX" format. These formats are internally changed to a simplified Brookhaven-type format and a file of that type is written to disk. Modified proteins can be written back to disk in Brookhaven format.

BRAGI can display up to ten "molecular clusters," each including one protein and other molecules. Using exclusively the Brookhaven format when reading or writing protein data, the program performs a connectivity check and asks the user which action should be taken when chemically inconsistent distances are found. Connectivities for amino acids are added; for other groups they have to be supplied. Any information about secondary structure is interpreted by BRAGI and can be displayed on the screen as standardized helices or "arrow sheets".

Several display modes, including perspective, orthogonal and several stereo modes, are possible. Depth cueing and clipping can be turned on or off. It is also

possible to change the foreground and background clipping planes.

Hardcopies of the current display can be obtained on a color printer or on a plotter (vectors).

Colors

Colors can play an essential part in the modeling process by providing an effective way to supply the researcher with a large amount of information about the chemical, stereochemical and electronic nature of the protein or parts of it.

By default in BRAGI the character of the amino acid side chain determines the color of that residue. Hydrophobic, hydrophilic, aromatic, positively and negatively charged amino acids are distinguished by color. Disulfide bonds, the backbone and non-peptide groups also are marked with different colors (see Color Plates 3 and 4). A different grouping of amino acid side chains can be set up by the user. The display of each of these groups can be turned on and off separately.

In a second possible color scheme the color is dependent on the atom type where carbon is displayed in white, nitrogen in blue, oxygen in red, for example (Color Plate 4).

Different molecules can be displayed with different colors in both of the described color coding schemes or in uni (Color Plate 1). The colors for the different parts can be changed for each group in each molecule individually.

Surface

The van der Waals or the solvent accessible surface can be calculated and displayed as a dot surface.[14] The previously calculated surface can also be written to and later read back from a file. For each protein, a different surface color (represented by an "angle" on the color circle) can be chosen by the user (Color Plate 3). For solvent accessible surface calculations ("Connolly surface") the radius of the probe can be chosen.[15]

On the raster graphic screen a CPK solid surface can be displayed. Several qualities of "Phong"-shaded sphere approximations are implemented. The surface can be made quasi-transparent so that the bond vectors can also be observed. These surfaces are colored according to the chosen coloring scheme (Color Plate 4).

Selective Display of Parts of Proteins

When only parts of a molecule are to be shown, the researcher can define a central atom around which all residues within a defined distance are displayed. Outside the sphere only the side chains or everything can be removed from the display. This is a useful option when design at an active center is performed.

For optimization of protein-protein interactions the groups within a certain distance of the interacting parts of the two protein surfaces can be displayed separately (Color Plate 2).

Display of hydrogen bonds.

When this option is selected, BRAGI displays all hydrogen bonds in a protein or between two proteins. In this case all distances smaller than 3.0 Å between polar "hydrogen atom donors" and "acceptors" are treated as hydrogen bonds with 1–4 interactions excluded.

Secondary structure display

If correctly marked in the Brookhaven file, the secondary structure elements of the protein can be displayed by BRAGI. This is especially helpful for manual alignment of different proteins. α-Helices are displayed as red regular helices on the vector screen or as red tubes on the raster display. β-Sheets are displayed as a series of parallel arrows according to the direction of the β-strand (Color Plate 5).

MODIFICATION OF PROTEIN STRUCTURE

There are a number of possible ways to modify the 3D structure of a protein. The modification may be a change in conformation of main chain or side chain torsion angles; an exchange of one or several amino acids; deletion of an amino-terminal or a carboxy-terminal part; an internal deletion or insertion; or addition of one amino acid or a group of amino acids at the ends of the sequence.

In this step the modeler has to make intelligent guesses about the local geometry of the modified protein before he can start the overall energy optimization and possibly a protein dynamics calculation later. A wrong orientation of the side chain or wrong modifications of the main chain angles may cause the force field program to optimize the structure into the wrong energy minimum. Therefore during (local) torsion angle change the modeler needs as much information about interactions as it is possible to have. In BRAGI the distances between atoms are internally controlled during torsion angle modification and lines on the display mark atoms that are significantly (0.5 Å) closer than the sum of the van der Waals radii.

In addition a number is calculated and displayed that represents the sum of the repulsional interactions of the selected group with the rest of the molecule. This number is updated in almost real time during conformational angle modifications of the protein. This proved to be an especially valuable feature during the modeling process.

Primary structure modification

Amino acid substitution

When an amino acid is to be exchanged the group is picked and the replacing group is chosen from a number of defined groups that are displayed in a menu. The new side chain is built into the protein in a standard geometry (Color Plate 6). At the same time BRAGI is set into the"geometry change mode," so that changes

of the side chain and main chain angles are possible in a very straightforward way without further input (see above). The orientation of the original side chain is preserved as a "shadow" until the orientation of the new group is optimized and the substitution is accepted. The current values of side chain and main chain torsion angles are displayed as numbers.

Insertions and Deletions — Building loops

The modeling of proteins that have insertions and/or deletions with respect to the original molecule is an especially difficult task, but the necessity may occur even during the modeling of protein structures from highly homologous ones.

For modeling loops we modified an intelligent idea that was originally published by Jones and Thirup.[16] From 69 proteins out of the Brookhaven data bank we have built a "loop data bank" in which "loops" of lengths between one and nine are stored. When an insertion or deletion has to be modeled, the modeler may decide, for example, to replace a loop of three amino acids in his protein by another loop of four amino acids from the data base. He picks the two groups between which the new loop has to be inserted, and BRAGI offers the twenty best-fitting loops from the internal data base with information about the RMS fit, the sequence and the "source protein" (Color Plate 7). The loops are fitted on the $C_\alpha - C(:O)$ bond of the last residue before the loop and on the $N(H) - C_\alpha$ bond of the first amino acid following the loop. The other main chain atoms of these groups are also displayed. The modeler then decides which loop he wants to replace his original one, or he can also try a loop of different length when no satisfying results were obtained in the first attempt.

"Hydrolysis" of proteins

A protein chain may be split into two. For modeling of a shorter protein one part may be removed or added to other peptide chains (see above).

Building proteins from the scratch or by combination of several building blocks

A protein can be built by adding one group after the other to the carboxy terminus or by extraction of "building blocks" (helices, turns, etc.) from several other proteins. The main chain torsion angles may be changed interactively by rotation of the dials or by setting them to predefined standard torsion angle combinations.

Geometry change

This mode is entered either automatically from the primary sequence change mode (see above) or by direct selection from the main menu.

In this mode all torsion angles of a particular side chain and also Φ and Ψ of that group can be changed by use of the dials. For proline a *cis*-trans change of ω is also possible. The main chain torsion angles can be set to standard values corresponding to α-helix, parallel or antiparallel β-sheet. This option may be especially helpful when a secondary structure is built from scratch.

During the change of torsional angles strong non-bonding interactions of the modified part with the rest of the molecule are displayed on the graphic screen by light yellow lines between the interacting atoms. The intramolecular part of the non-bonding repulsion energies — between this particular group and the rest of the molecule, and between the two parts of the molecule that are left and right of the affected group — is calculated and displayed as a number (dimensions have no real meaning). This makes it quite easy to localize the energy minima of the side chain location and therefore to avoid a local minimum in the following energy minimization (Color Plate 6). Current values of main chain and side chain torsion angles are displayed. The original position of the side chain is preserved as a "shadow image" on the display until the geometry change is accepted.

FORCE FIELD CALCULATIONS

After completion of the molecular graphics modeling, an energy minimization of the changed molecule or a protein dynamics run is necessary to remove any local close contacts and to localize the new energy minimum. Using one of the existing protein force field programs usually means the creation of a number of files with about 100 parameters. The output is generally not directly comparable with the starting molecule, due to a necessary renumbering etc. Creation of the necessary files for and interactive start of molecular force field calculations using the molecular mechanics program AMBER[17] is therefore included in BRAGI.

BRAGI creates all necessary input files for AMBER (2.0) and starts the AMBER jobs LINK, EDIT and PARM interactively. ANAL (geometry and energy analysis) can also be run interactively. Geometry minimization (AMBER program MINIM) can be started as a BATCH job. BRAGI offers standard parameters for the minimization and analysis. These can also be changed. If non-standard values are preferred for the minimization, the values are entered in BRAGI interactively. In this mode the researcher gives values for control of the amount of printed output, time limits, force field details like inclusion of solvent, dielectricity constant, periodic boundary conditions, cut-off radius, number of cycles, optimizer and several others. Any error messages produced by AMBER can be observed while running BRAGI.

The Brookhaven-type AMBER output file does not obtain information about connectivities, disulfide bridges, etc. These have to be built into the AMBER coordinate output file before the result can be displayed using BRAGI V4.0. Another problem is that in AMBER no free numbering of the groups is possible. They have to be numbered continuously starting with 1. This makes it often rather difficult to compare the structures before and after minimization. A resequencing program is

therefore included in BRAGI that after AMBER energy minimization automatically rebuilds all group numbers, chain identifications and disulfide bridges. This allows the direct comparison of the two 3D structures before and after energy minimization.

ANALYSIS OF INTERACTIONS

There are two different options for the calculation of interactions between groups. In the first, all intramolecular contacts of an atom up to a given value can be shown as lines on the display, and the list of contacts is displayed as text. In the second, all intermolecular distances between two selected molecules up to a given value are displayed as lines (the green lines in Color Plate 3). Distances to water molecules or other molecules can also be displayed. Special options that display hydrogen bonds are also included (see above).

COMPARISON AND PROTEIN FITTING

During the modeling process often two proteins have to be compared or aligned. BRAGI has a built-in routine that performs a RMS fitting[18] between proteins. It is possible to fit parts of the proteins including only C_α-atoms, all backbone atoms or all atoms of the two fitted sections. The parts of the molecules that are to be fitted can be picked.

HARDWARE

For protein modeling a high performance graphics device is essential. Until the end of 1986 a combination of a Evans & Sutherland PS 300 with a DEC VAX was the standard hardware configuration for molecular modeling. This was also the configuration in which the original versions of BRAGI were developed. Meanwhile, the program has been converted to one of the new UNIX raster graphic workstations, namely the Hewlett-Packard 350 SRX and the RISC workstation HP 825 SRX.

CONCLUSIONS

Though a huge step forward compared to random or "semi-random" mutagenesis, computer-aided protein design is still an area of research in which the methods still need development to improve the reliability of the predictions. The areas in which progress is most urgently needed involve the understanding of the protein folding, the development of new molecular mechanics force fields (and algorithms), the improvement of tools for the evaluation of "homologies" with respect to 3D structure and the development of molecular graphics and modeling packages that use the stereochemical knowledge we have (expert systems).

The protein modeling system described in this paper provides a comprehensive tool for the researcher who wants to design proteins with new properties. The program aims to facilitate the essential steps in the modeling of protein in which human knowledge and intervention are crucial. It provides an interface to an energy minimization and protein dynamics program package.

Work on BRAGI is continuing in parallel with modeling activities that are performed during a number of protein design projects in our group and in collaboration with others. Improvements of the manipulation and modification of small molecules are being implemented now.

REFERENCES

1 Winter, G. *et al. Nature,* 1987, **299**, 756
2 Ulmer, K. M. *Science*, 1983, **219**, 666
3 Carter, P. and Wells, J. A. *Science*, 1987, **237**, 394
4 Massey, R. J. *Nature*, 1987, **328**, 457
5 Leatherbarrow, R. J. and Fersht, A. R. *Protein Eng.*, 1986, **1**, 7
6 Cunningham, C. and Wells, J. A. *Protein Eng.* 1986, **1**, 319
7 Wells, J. A. and Powers, D. B. *J. Biol. Chem.*, 1987, **261**, 6564
8 Knowles, J. R. *Science*, 1987, **236**, 1252
9 Pabo, C. O. and Suchanek, E. G. *Biochemistry*, 1986, **25**, 5987
10 Blundell, T. L. *et al. Nature*, 1987, **326**, 347
11 Blundell, T. and Sternberg, M. J. E. *Trends in Biotechnology*, 1985, **3**, 228
12 Jones, T. A. and Pflugrath, J. in G. M. Sheldrick, C. Krüger and R. Goddard (eds.) *Crystallographic Computing 3: Data Collection, Structure Determination, Proteins and Databases*, Clarendon Press, Oxford, 1985
13 Schomburg, D. *et al.*, unpublished results
14 Pearl, L. H. and Honnegger, H. A. *J. Mol. Graphics*, 1983, **1**, 9
15 Connolly, M. L. *J. Appl. Cryst.* 1983, **16**, 548
16 Jones, T. A. and Thirup, S. *EMBO J.*, 1986, **5**, 819
17 Weiner, P. K. and Kollman, P. A. *J. Comp. Chem.*, 1981, **2**, 287
18 McLachlan, A. D. *J. Mol. Biol.* 1979, **128**, 49