Visualization of energetics and conformations from molecular computer simulations*

Paul K Weiner, Steven L Gallion, Eric Westhof† and Ronald M Levy

Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903, USA

The quantity of data generated from molecular dynamics simulations and energy minimizations of macromolecules is overwhelming. It is an arduous task to extract the relevant and interesting information from the numerous coordinate sets produced. To help solve this problem, the authors have developed a method to aid the visualization of the relevant information from the simulations. This approach combines animation of the results on a high performance graphics device, such as the PS300, with colour-coded atoms based on changes in energy or conformation. The method will be illustrated using as examples: the molecular mechanics minimization of a nonapeptide, the molecular dynamics simulation of the protein myoglobin, including the analysis of the motion of helices during a 300ps trajectory, and changes in sugar puckering that occur during the molecular dynamics simulation of a DNA oligomer. The method is also applicable for analysing energy components and conformational properties of a fixed conformation.

Keywords: computer applications + simulation, macromolecules, molecular dynamics simulations, energy minimizations, coordinate sets, Z_I-DNA

Received 27 January 1986
Accepted 24 February 1986

Computer simulations (including molecular mechanics, molecular dynamics, and Monte Carlo calculations) of proteins and nucleic acids provide the most detailed theoretical approach available for studying their structural and dynamical properties[1-3]. The recent availability of array processors and supercomputers has made it possible to carry out very long simulations of these macromolecules[4-6]. For example, the results of a 300ps molecular dynamics simulation of myoglobin[7] were recently reported.

This long trajectory consisted of 300 000 steps; a new coordinate set for the molecule corresponds to each step. For the myoglobin simulation, 1 200 coordinate sets were saved on tape for numerical analysis. The numerical analysis of the results of molecular dynamics simulations of biopolymers is a time consuming, computationally intensive task. The complementary analysis of these trajectories using real-time graphics methods can greatly aid in the structural interpretation of the results.

One of the original methods of graphical analysis of macromolecular simulations consisted of watching an animation sequence of the coordinate sets. By studying the motions of the atoms, the viewer could easily see large scale movements of the molecule, if they happened, in a relatively short time period. Since high performance graphics terminals were very expensive and scarce, a film was usually made at some remote site and brought back to the research lab for viewing[8]. This technique had limited usefulness, since the viewer was forced to always observe the motion from a fixed perspective. Next came greater user control through the realtime use of a graphics system[9]. Here the user could view the animation with real-time control of the playback speed and the viewpoint.

Current research efforts deal with experimentation with different representations of the molecule in an attempt to find those that best help to interpret the animated sequences. Simplification of the structure through the use of virtual bonds[9] gave a clear idea of certain motions. Other representations of the molecule were animated in an attempt to aid the recognition of cooperative motions of the atoms. An example of this is the use of animated difference matrices[10]. Even with these tools, it is still an arduous task to extract the relevant information from the numerous coordinate sets produced from the computer simulations.

New approaches to the visual presentation of data from macromolecular simulations are needed. For example, we would like to be able to answer the following questions which arise in the course of analysing macromolecular simulations:

- How do high energy (strained) regions of the macromolecule affect motions of adjacent parts of the macromolecule during minimization?
- What are the most mobile parts of a molecule?
- Which correlated internal coordinate changes lead to the flexibility of the molecule?

The first of these questions could not be answered with traditional methods of representation. The last two questions could only be answered after much study for complex systems. The authors have developed a method to aid in the visualization of the relevant information

*Presented in part at the New York Academy of Sciences Conference on Computer Simulation of Chemical and Biomolecular Systems, New York, October 1985

†Permanent address: Institut de Biologie Moleculaire et Cellulaire du CNRS, 67084 Strasbourg, France

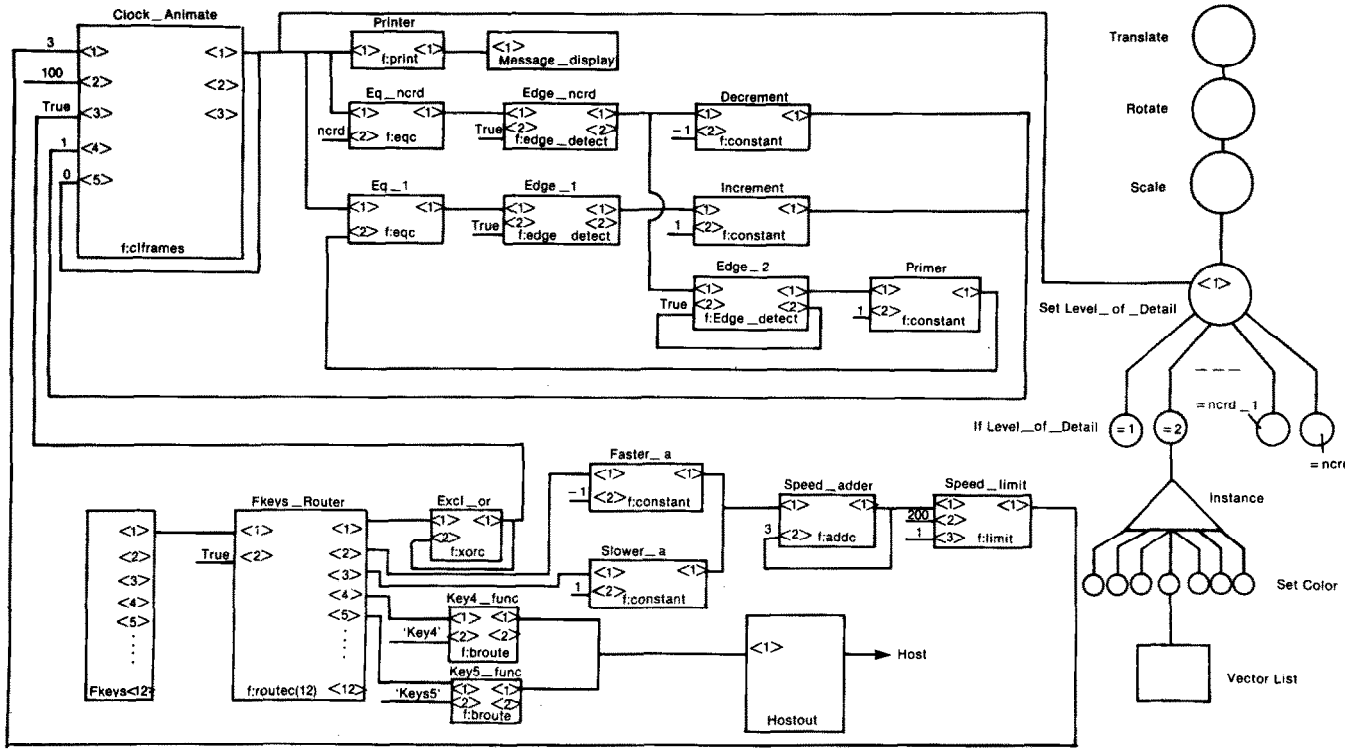

Figure 1. Design of the PS300 animation network used to incrementally traverse the LOD data structure. Rectangles represent functions with inputs and outputs, squares represent operation nodes and triangles represent instance nodes (groupings of other nodes). The lines connecting the nodes are the paths the structure is traversed in memory

from the simulations. The method combines animation of the results on a high performance graphics device, such as the PS300, with colour-coded atoms based on changes in energy or conformation. The method will be illustrated with points from an energy minimization of a nonapeptide that has an initial geometry with some very close contacts. The movement of the molecule to relieve this strain and the redistribution of energy will be studied. A second application involves the study of conformational changes of the α-helices of myoglobin during a 300ps molecular dynamics simulation. The qualitative results obtained from the animation are compared with the results obtained from a numerical analysis of the coordinates from the dynamics trajectory. As a final application, the changes in sugar puckering that occur during a molecular dynamics simulation of a Z-DNA hexamer are studied graphically. The graphical analysis reveals dynamical differences in puckering between the sugars of guanine and cytosine residues.

METHODOLOGY

The first step in the animation sequence consists of a series of FORTRAN programs running on the host VAX 11-780 that converts a Protein Data Bank formatted file[11] to a binary graphics data file containing Cartesian coordinates, connectivity, atomic symbols, etc. Additional routines read in the binary file, transform the coordinates to PS300 screen space and convert the arrays to PS300 object data structures. These structures are downloaded into the PS300 microcomputer using the E&S-supplied Graphics Support Routines (GSRs). Thus, FORTRAN routines prepare the molecule for display by creating the data in the form of vector (primarily coordinate and connectivity information) and character (primarily atomic symbol information) lists.

This data constitutes the primary information base on which modelling operations (e.g., rotation, scaling, translation) may be performed and to which modelling attributes, such as colour and intensity, may be attached. Together, the data, operations, and attributes form a Hierarchical PS300 Data Structure (HDS). A graphical representation of this HDS is shown on the right-hand side of Figure 1.

The function of the PS300 command language is to manipulate the data structure passed to the PS300. These programs are assembled from individual functions, linked into networks and used to channel the analogue signals of the dials, tables, buttons, etc. into the appropriate items required for a particular function. For example, the translation, rotation and scaling nodes of the data structure shown in Figure 1 require matrices as input while the set colour nodes require integers. The function networks perform the various modelling operations on the data structure established by the FORTRAN routines. Although it is possible to code the function networks from the FORTRAN segment by utilizing the GSRs, it is far more cumbersome than using the native PS300 Command Language. The construction of the network usually proceeds by following a diagrammatic representation. Such a representation is shown in Figure 1 together with the connection to the FORTRAN-created hierarchical data structure.

The PS300 implementation of the animation program is based on the concept of 'level-of-detail' (LOD). LOD settings allow data to be conditionally referenced for subsequent display based on the current value of the Set LOD node. By using a LOD method, selected parts of a model (frames in the present case, although it could

be different contour levels, etc.) are displayed and removed in a predetermined sequence. To run an animation sequence comprising a series of separate picture definitions, one sequentially increments the value in the Set LOD node (an If LOD node lower down in the structure is allowed to test for the condition of equivalence to the current Set LOD value). Thus, frame 1 is displayed when the Set LOD value is 1, frame 20 if the value is 20 and so on. Up to 32 768 frames may be displayed in this manner.

To allow more control over the movie, pausing, speed control, and single frame viewing mechanisms were implemented in the network. Interactive control of sequence suspension and continuation is accomplished by utilizing input 3 of the clock which stops the clock if false and starts it if true. Control over the rate at which the sequence is presented is established by increasing or decreasing the value on input 1 of the clock (the rate being inversely related to the magnitude of the value). The ability to select a particular frame at random is implemented by suspending the clock, requesting the desired frame from the user, and then sending this value to the Set LOD node. Because of the hierarchical design of the structure, the ability to rotate, translate, or scale the molecule is retained throughout the session.

DISPLAY OF PROPERTIES

Colouring according to energetics

What happens to the energy in high energy (strained) regions of a protein during energy minimization or molecular dynamics simulation? At what rate does this energy flow to surrounding regions and how far does the strain propagate? The analysis of energy relaxation in proteins has been approached using the methods of computational statistical mechanics[12,13]. These questions are also of interest for the energy minimization of structures obtained from crystallographic refinement. If there is a region of high strain energy, one would like to show how this affects nearby regions as strain is relieved and whether enough energy is available for atoms to jump into other minimum energy wells.

As an example of a graphical approach to this problem, the authors analysed the minimization of a nonapeptide (Ala-Gly-Pro-Tyr-Val-Gly-Pro-Val-Gly-Lys). The initial structure was built using the internal coordinates in the Amber database[14]. This method sometimes results in bad contacts in the initial structure, especially when Pro residues are present. In this case, the starting structure has very close contacts between atoms of the Tyr ring and the amide linkage of Gly 6. In Colour Plate 1, each atom is colour-coded according to the magnitude of its gradient (corresponding to the forces on each of the atoms) — red (> 100); magenta (> 10, $\leqslant 100$); green (> 1; $\leqslant 10$); cyan (> 0.1, $\leqslant 1$), and blue ($\leqslant 0.1$); the structure at the beginning of each line search in the conjugate gradient minimization[15,16] is displayed on the PS300. Colour Plates 1(a)–(d), show several photographs of the nonapeptide during the minimization. The high energy red region involving the Tyr and Gly residues of the initial structure in Colour Plate 1(a) is worth noting. This region is highly strained and the structure in Colour Plate 1(b), taken at the next step, shows the Tyr ring bending in an effort to relieve this strain. Colour Plate 1(c), taken 30 minimization steps later, shows that there is still strain present in this region (note the magenta colour) and that the strain has not spread to the end atoms. The predominant green colour of Colour Plate 1(d), taken at the 90th step, shows that the molecule has relieved most of its initial strain. The green colour remains for many more frames and shows that during this portion of the calculation, the molecule is traversing a relatively level portion of the surface on its path to the minimum energy conformation. These results show pictorially that the large initial strain is quickly relieved locally, but that the calculation only slowly approaches the minimum energy structure (all components of the gradient > 0.1).

The ability to relieve strain by many small adjustments of internal coordinates in the high energy regions has also been observed in the analysis of molecular dynamics simulations of proteins[1-3]. This aspect of molecular flexibility has important implications for proposed mechanisms for the transmission of information from one part of a polypeptide or protein to another part. Notice that the initial and final conformations of the nonapeptide (Colour Plates 1(a) and (d)) are very similar.

Colouring according to displacement

The second example, to be considered here, deals with a study of the internal motions of myoglobin during a 300ps dynamical simulation[7]. Myoglobin is over 80% α-helical, consisting of eight α-helices that pack against each other over extended regions. A detailed numerical analysis of the 300ps myoglobin simulation revealed that both the structure of the α-helices and the overall folding pattern were maintained throughout the entire 300ps simulation.

A series of coordinates representing time steps 4.5ps apart were displayed starting from the initial equilibrated structure. Only the carbon atoms and atoms of the haem groups were displayed on the PS300 so that the backbone displacements could be clearly seen. The colouring of the atoms is based on the displacement of atoms in each structure from that in the initial equilibrated structure. The displacement differences are given by — red: > 3Å; magenta: > 2Å and $\leqslant 3$Å; green: > 1.5Å and $\leqslant 2$Å; cyan: > 1.0Å and $\leqslant 1.5$Å; and blue: $\leqslant 1.0$Å.

Colour Plate 2(a) shows Myoglobin shortly after the beginning of the trajectory; all the atoms remain close to their positions in the equilibrated structure. In Colour Plate 2(b) after approximately 300ps parts of the molecule have begun to drift from the equilibrated structure after 4.5ps and continue to do so during the simulation. However, there are no large scale conformational changes. This agrees with the finding[7] that the deviation between the coordinates obtained from the start of the trajectory and the average coordinates for the entire 300ps simulation is < 1.4Å for the backbone atoms.

Colour Plates 2(c) and (d) provide a view down the H helix and shows the G–H loop near the middle of the simulation (Colour Plate 2(c)) and after 300ps (Colour Plate 2(d)). It is very clear from the two pictures that the atoms in the interior of the helix remain close to the initial structure (they are coloured blue in Colour Plates 2(c) and (d), while the G–H loop moves by more

than 3Å (red in Colour Plate 2(d)) during the 300ps simulation.

Colour Plate 2(e) is the X-ray structure of Myoglobin and Colour Plate 2(f) is the equilibrated structure that was the starting structure for the 300ps simulation. Note that almost all the surface helices have moved from their initial positions. This is a result of the protein becoming more compact during the equilibration.

This technique is particularly valuable when the animation is stopped at a particular frame. Ordinarily there would be no assistance in the determination of what was different in that structure from that of an earlier frame. It is possible to place a dashed image of the earlier structure on the screen, but this uses up memory and it would be necessary to look for conformational differences amongst many superimposed structures. The graphical approach illustrated in Colour Plate 2 highlights the differences with a minimal amount of screen clutter.

Colouring according to conformational changes

The final example deals with the problem of determining which internal coordinates dynamically change the most during a simulation. All atoms in an animation sequence are moving in what appears to be a chaotic way. Therefore it is extremely useful, rapid, and precise to colour a given internal coordinate, such as a dihedral angle, either in one or all residues according to conformational ranges that it occupies during the simulation. In this way it is possible to recognize at a glance the highly flexible internal coordinates and residues. With this information, it is only necessary to make and study a few plots of the internal coordinate of interest *versus* the time step in order to better understand the dynamics.

Internal coordinate colouring was particularly useful for recognizing and characterizing the sugar puckers of the nucleic acids during a dynamics simulation. The pucker of the five-membered sugar ring can be either of the envelope type, with one atom displaced with respect to the other four atoms, or twist type, with two atoms displaced with respect to the other three but in opposite directions. Because the sugars are chiral, the direction of puckering is important. Therefore, there are ten envelope and ten twist sugar puckers. According to the pseudorotation concept[17], these twenty puckerings can be represented on a circle with one sugar type every 18° as in Figure 2. The phase angle of pseudorotation will thus represent the pucker type (P = 0° is by definition C(3′)-endo-C(2′)-exo, where endo means that the displacement occurs toward the C5″ carbon atom). The continuous passage from one conformation to another occurs by concerted changes in the ring torsion angles without the ring becoming planar. On a static structure, with experience, it is not too difficult to recognize the pucker type. However, on a dynamical structure this is utterly impossible. Colouring of the sugars according to the phase angle of pseudorotation allows an immediate recognition of the sugar puckers adopted during the simulation. The mobility of the furanose rings gives an approach for understanding the dynamics of the structural fluctuations and conformational changes in nucleic acids[18].

A 2ps dynamical simulation of a Z-DNA hexamer[19] with coordinate sets saved every 0.01ps was used in this

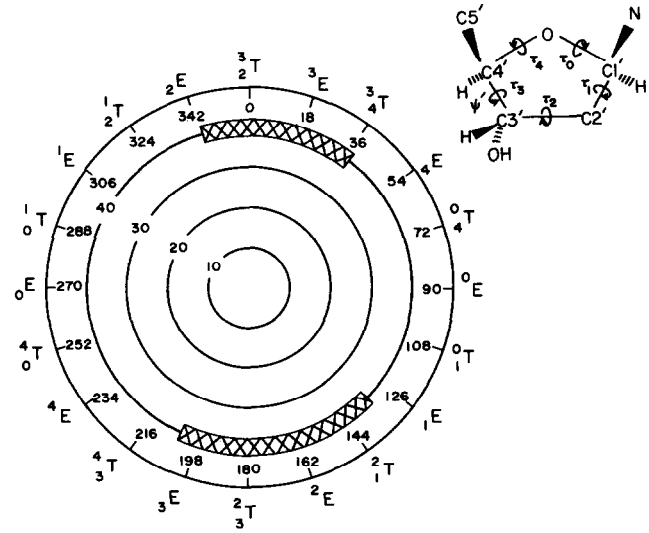

Figure 2. Pseudorotation wheel of furanose rings with the nomenclature and the preferred pucker domains (hatched regions). (Adapted from Reference 16)

study. Values of the phase angle represented by the colours are — red: (1–36); magenta: (37–100); green: (110–200); cyan: (201–250); blue: (255–350), and white (any value not in the previous ranges). Colour Plates 3(a)–(c) show that the sugars of the guanine residues change colour (red to magenta to green), while the sugars of the cytosine residues stay the same colour (Green). Thus the sugars of the guanine residue alternate between C(3′)-endo and C(2′)-endo, while the sugars of the cytosine residues stay approximately C(2′)-endo. It was readily apparent from the animation sequence that the sugars of the guanine residues were changing puckers much more frequently than the sugars of the cytosine residues.

CONCLUSIONS

The combination of animation on a high performance graphics device and colouring of atoms according to energetics and conformation appears to be an extremely useful tool for extracting relevant and interesting information from the massive amounts of data that a simulation can produce. Once interesting events are noticed, more traditional methods of analysis can be used to quantify the observations. This paper has presented several examples where the visual impressions of what was happening confirmed the interpretations arrived at by a numerical study of the data.

ACKNOWLEDGEMENTS

This work has been supported in part by grants from the National Institute of Health (GM-30580), the Petroleum Research Fund of the American Chemical Society, and the Charles and Johanna Busch Foundation, Rutgers University. RML is the recipient of a National Institutes of Health Research Career Development Award.

REFERENCES

1 **McCammon, J A** *Rep. Prog. in Physics* Vol 47 (1984) p 1
2 **Karplus, M** *Adv. Biophys.* Vol 18 (1984) p 165

3 **Levy, R and Keepers, J** 'Computer simulations of protein dynamics: theory and experiment' *Comm. Mole. Cell. Biophys.* Vol 3 (1986) p 273
4 **Van Gunsteren, W F et al.** *J. Comp. Chem.* Vol 5 (1984) p 272
5 **Gallion, S L et al.***Comput. Chem.* Vol 10 (1986) p 165
6 **Seibel, G L** *P.N.A.S.* Vol 82 (1985) p 6537
7 **Levy, R et al.** *Biophys. J* Vol 48 (1985) p 509
8 **Feldmann, R J and Levitt, M** *Molecular dynamics of bovine pancreatic trypsin inhibitor* (Movie) (1980) NIH Laboratory for Molecular Graphics, Bethesda, USA
9 **Todd, S and Gillett, J** *J. Mol. Graph.* Vol 1 No 2 (1983) p 39
10 **Morffew, A J** *J. Mol. Graph.* Vol 1 No 2 (1983) p 43
11 **Bernstein, F C et al.** *J. Mol. Biol.* Vol 112 (1977) p 535
12 **Mao, B et al.** *Biopolym.* Vol 21 (1982) p 1979
13 **Swaminathan, S** *Biochem.* Vol 21 (1982) p 5230
14 **Weiner, P K and Kollman, P A** *J. Comp. Chem.* Vol 2 (1981) p 287
15 **Fletcher, R and Reeves, C M** *Comp. J.* Vol 6 (1963) p 163
16 Minimization routine written by Paul Weiner, 1977
17 **Altona, C and Sundaralingam, M** *J. Am. Chem. Soc.* Vol 94 (1972) p 8205
18 **Westhof, E and Sundaralingam, M** 'Disorder and dynamics of five-membered rings in biological structures' in **Clementi and Sarma (Eds)** *Structure and dynamics: nucleic acids and proteins* Adenine Press, USA (1983) p 135
19 **Westhof, E et al.** 'Temperature dependent molecular dynamics and restrained X-ray refinement simulations off a Z-DNA hexamer' *J. Mol. Biol.* to be published