Molecular conformational space analysis using computer graphics: Going beyond FRODO

Olle Nilsson

Department of Physical Chemistry, University of Uppsala, Uppsala, Sweden and Department of Molecular Biology, University of Stockholm, Stockholm, Sweden

The molecular graphics program FRODO has been modified to support analytical animation of molecular dynamics trajectories. The enhanced program, mdFRODO, supports all features available in FRODO and is interfaced to GROMOS. A variety of analytical animation modes is included. Extensive coloring and atom selection features are implemented to aid the user in distinguishing features of interest in a set of conformations. Molecular conformational space can be analyzed efficiently and comprehended. Animations may be viewed in stereo, and the animated object can be overlaid with any of the standard FRODO objects. The mdFRODO program is of wide use in molecular dynamics, X-ray crystallography and two-dimensional NMR work. Examples illustrating various aspects of collective motion in protein molecules are given and discussed.

Keywords: computer graphics, molecular dynamics, conformational space analysis, animation, FRODO, GROMOS, surface plasticity

INTRODUCTION

Interactive computer graphics is established as a unique tool to rapidly analyze the huge amount of data generated by molecular dynamics (MD) simulations of biomolecules.[1-3] In this paper FRODO is modified to a conformational space analyzer interfaced to GROMOS. FRODO is a well known and extensively used crystallographic and model building program.[4,5] GROMOS is a well established molecular dynamics simulation program package.[6,7] The enhanced version of the program is called mdFRODO. It includes various interactive animation, difference and fluctuation, overlay and color display modes. Subsystems may be dissected out of the full molecular system, and these may be represented at different levels of detail, facilitating construction of composite molecular objects that contain only information of relevance for a particular analysis. Other molecular graphics programs, such as QUANTA,[8] Chem-X,[9] Insight[10,11] and HYDRA[12], contain only some of the features reported here, such as frame-by-frame animation. An important limitation with these programs is that source codes are not available. While this may be acceptable for industrial laboratories, it is a severe drawback for researchers in university environments. The objective of mdFRODO is to provide these investigators with an open code, where enhancements and modifications can easily be made to fulfil their individual scientific goals.

ANALYTICAL GRAPHICS TECHNIQUES

Animation is one method to assess conformational space. A frame-by-frame animation is produced by displaying one graphical representation of a molecular conformation (frame) for a given time, replacing it with the next one in a time series of frames, and so on. Such an animation will indeed visualize the dynamical evolution of a simulated system, but it is limited in one aspect. It does not contain any information about the history of the dynamics: At each time only the instantaneous conformation is seen; what has occured in the system before the current time is not remembered. Further complications are linked to animation. A displayed frame is often too complex to be readily memorized. A fairly small protein molecule of 100 residues contains about 1000 atoms, requiring about 2000 bonds to be displayed per frame in a full atomic representation. Because frames are shifted successively throughout the animation, the time of exposure to a single frame may be too short for full perception, or the frame complexity may render full perception impossible.

Complexity and lost history problems can be reduced by various techniques. Color encoding of bonds and the selec-

Address reprint requests to Dr. Olle Nilsson at the Department of Physical Chemistry, University of Uppsala, Box 532, S-751 21 Uppsala, Sweden.

Received 27 March 1990; accepted 15 May 1990

tion of only relevant atoms and bonds to be displayed in each frame make it easy to distinguish parts of the molecule and reduce the complexity of a frame. One may record the history of a system by statically displaying all frames on top of each other, but at the loss of the sensation of dynamics and an increase in complexity. This method has been used by default due to its simplicity since the dawn of molecular display systems. It is implemented, for instance, in the FRODO MOL subprogram.

In the next subsection we define a set of display techniques implemented in mdFRODO that allow graphical analysis of dynamical, history and conformational space properties of a molecule submitted to an MD simulation. The following four subsections define options to modify the resulting images from the display modes, and the remaining subsections describe different representations of objects. Examples where interactive graphical conformational space analysis is of use are given from a 2.2-ns, 277 K *in pseudo vacuo* MD simulation[13] of a small, rather globular protein molecule, the model built chloroplast carboxy-terminal fragment of the ribosomal L7/L12 protein, cCTF.[14]

Display modes

A time series of conformations may be assembled to a display set that can be processed by the following display techniques: overlay, animate, additive, sweep, borderwalk, relate, atomdist, trail and bordertrail.

A frame is a composite graphical object $\mathbf{F}(t)$, t denoting the conformation in a set of m structures $\langle 1 \ldots m \rangle$, from which the frame is constructed. Normally a frame is a set of vectors representing the bonds between atoms or atomic displacements. Usually the set is a time series of conformations where t denotes the time corresponding to a frame. A screen is defined as the display of one or several frames. A movie is the consecutive display of screens in real time. The atomic coordinates of one conformation are denoted by $\mathbf{r}(t)$.

Overlay: Structure comparison

Overlaying, or displaying, all frames from $\mathbf{F}(t_1)$ to $\mathbf{F}(t_m)$ on top of each other without animation keeps all history, at the expense of dynamics and simplicity. The overall geometrical conformational space accessed by a system can be inspected. Rotational and translational drift may be induced by changes in shape for a molecule simulated *in vacuo*. These drifts can be inspected by an overlay display. The internal fluctuations of a molecule can be visualized with the overlay mode if rotational and translational drifts are removed from each conformation by a least-squares fitting[15] of each conformation to a reference conformation in the trajectory or to the X-ray structure. For the fitting it is essential to select as reference points only atoms in the structure having small fluctuations to obtain a well defined reference state that is undisturbed by events like loop transitions and large side-chain fluctuations. The α-carbon atoms of α-helices and β-sheet regions of a protein molecule usually serve well as such reference points, because they are normally structurally well preserved during an MD simulation.

Animate: frame-by-frame animation

Animation is the simplest way of displaying the dynamics of a system. Here one frame, $\mathbf{F}(t_n)$, is displayed on screen at each time, thus neglecting a treatment of the system's history. It should be used if the frames are too complex to be displayed comprehensively with history. Also, if the frames contain too many graphical objects to be displayed rapidly enough or to be manipulated interactively, animation may provide an acceptable compromise solution to, for instance, an overlay display. Unfortunately, it is not possible to present the animated display mode with static pictures.

Additive

If the full accumulated history is of interest, frames can be displayed additively. In this mode the first screen in the movie contains frame $\mathbf{F}(t_1)$; the second screen $\mathbf{F}(t_1)$ and $\mathbf{F}(t_2)$ overlaid; the third screen $\mathbf{F}(t_1)$, $\mathbf{F}(t_2)$ and $\mathbf{F}(t_3)$ overlaid; and so on, building up a picture containing the history of conformational change. Some sensation of dynamics is maintained as the screen changes during time. The full history of the system is shown at any time t_n because all frames $\mathbf{F}(t_1)$ to $\mathbf{F}(t_n)$ are displayed.

Sweep: measuring conformational change

The sweep mode produces an animation with limited history. A set of N consecutive frames is displayed overlaid on screen at each time. The N-window of frames is displaced forward in the time series of frames producing an animation with a short memory. Only the preceding N frames are remembered. Thus, with a window size of three, frames $\mathbf{F}(t_n)$, $\mathbf{F}(t_{n-1})$ and $\mathbf{F}(t_{n-2})$ are displayed at t_n. At time t_{n+1} frames $\mathbf{F}(t_{n+1})$, $\mathbf{F}(t_n)$ and $\mathbf{F}(t_{n-1})$ are displayed. A complication occurs at the beginning of the series of frames where no two previous frames exist. These can be mapped to the last frames in the series, i.e., to $\mathbf{F}(t_m)$ and $\mathbf{F}(t_{m-1})$ if a circular sweep animation is desired. A window of three frames gives a good animated representation with memory at a limited expense of computational overhead for the display system. The sweep animation facilitates easy judgement of magnitudes and directions of conformational changes, because the eye is helped by having previous frames with which to relate the current frame. Reasonably large systems can be analyzed interactively with this mode as only three frames are shown at a time.

Borderwalk

Border states are ignored by the previous modes. Although the sweep mode keeps some memory, it does not help the observer to put conformational changes in context of the initial and final conformations. The remedy is to statically display the first and the last frame, $\mathbf{F}(t_1)$ and $\mathbf{F}(t_m)$, overlaid with the sweep animation. This is the borderwalk mode.

Relate

This mode relates one structure, for instance an MD average or a crystallographic structure, to the dynamics or the con-

formational differences of a set of structures. Here, one frame $\mathbf{F}(t_s)$ is selected from the set of structures and displayed statically on the screen along with a frame-by-frame animation of the frames in the set $\mathbf{F}(t_1)$ to $\mathbf{F}(t_m)$.

Atomdist: atomic displacements

Only atom-to-atom bonded representations of molecules have been used in the previous display modes. Individual atomic trajectories, distance deviations and fluctuations cannot be described comprehensively by such molecular representations. To accomplish this, the atomic displacements between each consecutive pair of conformations in the series are displayed overlaid on each other, with displacement vectors $\mathbf{r}(t_2) - \mathbf{r}(t_1)$, $\mathbf{r}(t_3) - \mathbf{r}(t_2)$, ... $\mathbf{r}(t_n) - \mathbf{r}(t_{n-1})$ for each atom. This static representation is the atomdist mode.

Trail: correlated motion, fluidity, elasticity and plasticity

The trail mode displays an animation of only three consecutive displacement vectors, $\mathbf{r}(t_n) - \mathbf{r}(t_{n-1})$, $\mathbf{r}(t_{n-1}) - \mathbf{r}(t_{n-2})$ and $\mathbf{r}(t_{n-2}) - \mathbf{r}(t_{n-3})$, for all selected atoms per screen. Fluctuations in atomic position and distance deviations are assessed with a short history. This is the atomic distance representation analogy to the sweep animation mode for bonded representations of conformations. Correlations in local and global motion can be inspected. Correlations in collective motion can be analyzed if high-frequency fluctuations are filtered out by structure averaging. The characteristics of atomic fluctuations, such as anisotropy and anharmonicity, can be judged in the dynamic display. Material fluidity, elasticity and plasticity properties are assessed.

Bordertrail

The border states, where atoms begin and end their trajectorial journey in an MD simulation, can be displayed statically in their bonded representations overlaid with a trail animation. This is the atomic distance representation analogy to the borderwalk mode.

Molecular movies

A molecular movie is produced by consecutively displaying screens constructed by one of the modes described above. The first screen is displayed on the computer screen for a specified time, whereafter it is replaced by the next screen in a time series of screens. This procedure is repeated for all screens. When the last screen is reached the animation is restarted immediately from the first screen, giving a circular display of screens.

Idle times

An idle time may be specified to reduce the speed with which a movie is running. In some cases if the molecules are small or if very schematized representations of the molecules are ordered, the screens may be displayed for too short a time. It can also be useful to hold the display of first object on screen for a longer time than the rest i circular movie, to indicate that a new cycle of the mo has been launched.

Manipulating time in a movie

Resampling Once a movie is produced, it may be that c formations have been sampled from an MD trajectory v too short a time step, that changes are too slow on scre or that one may wish to see only a rough estimate of molecular conformational space or dynamics. This can done by excluding frames from the display list. For instan every fifth or tenth frame can be selected for display fr the series of frames.

Masking There may be other reasons to reduce the size the series of frames. If an event of interest occurs only the latter part of a movie, it might be desirable to omit first part. If some screens in the middle of a movie undesirable, these can be excluded by a masking procedu

Cutting and pasting a time series of data causes a probl Time is no longer linear. This is a severe problem if tempting a physically valid evaluation of the time dep dence of a process. If time is to be kept linearly runni excluded screens must appear empty, which produces unwanted flicker on the screen. One solution is to all time to be nonlinear and to skip excluded frames. If o a few frames out of a large set are excluded, the effec small. However, if many frames are excluded and if frame set is small, the effects become significant. Use option to delinearize time with caution.

Coloring

Static, dynamic or zoned coloring may be applied for a of the display modes. A static color encoding gives objects in all frames the same color, which may be selec by the user. The dynamic color mode assigns one color each frame from a color list and varies the color from fra to frame. A molecule can be divided into subobjects zones. If a zoned color mode is selected these subobje will be given different colors, keeping the same color each subobject in all frames. The zone colors are picl from a color list that ensures that sequentially adjacent zo are given colors with good contrast.

Molecular representations and subobject partitioning

Molecules can be represented as composite objects in wh each subobject or zone may be created with a different le of detail by choosing the atoms to be displayed there Alpha-carbon, backbone or all-atom representations are p sible. The α-carbon and backbone representations draw bo sequentially between α-carbon or backbone atoms as tl occur in the coordinate list without consideration to bc distances. The all-atom representation creates bonds eit by a distance criterion or by topology information. Parts a system may be excluded from display; nonbonded ato

are displayed as stars in the all-atom representation. Any number of molecules may be present in a conformation. It is possible to manually guide the bonding of selected atoms using subobject partitioning and different representations.

Construction of bonds

Bonds in the all-atom representation are drawn following a distance criterion or by a bond list supplied by GROMOS topology files. The bond cutoff distance for the distance criterion is 1.9 Å by default, but can be varied. The search window for bonded neighbors is 38 sequential atoms. It is recommended that the topology option be used if a topology file is available. The topology contains a bond list that explicitly specifies the atoms between which bonds are drawn, disregarding interatomic distances in specific conformations.

However, using the same bond list for all of the conformations in a set may under some conditions be inappropriate. For instance, if an MD simulation is performed under periodic boundary conditions (PBC),[7] some of the atoms in a molecule may pass out through one side of the periodic boundary box and in through the opposite side, thereby apparently disrupting the covalent chain of the molecule. If the atomic coordinates or the bond list are not corrected for such border transitions before the bonds are drawn, erroneous bonds will appear with unphysical bond lengths. Corrections can be made by using the distance criterion to build bond lists and by updating the bond list for each conformation. An atom moving out of the PBC box will then appear as a cross in the opposite border region of the box, and it will be bonded only to other atoms following it. However the covalent connectivity of the molecule will be disrupted and it may in some applications be more appropriate to correct the coordinates for PBC transitions before transferring them to mdFRODO.

Creating bonds following a distance criterion is sometimes less successful. The three N_ζ hydrogens of the lysine residue usually appear bonded to each other due to their small separation ($d_{HH} \approx 1.6$ Å and $d_{NH} \approx 1.0$ Å). Disulphide bridges are not drawn due to the length of the disulphide bond ($d_{SS} \approx 2.0$ Å). Furthermore, nonoptimized MD average structures usually have nonoptimal interatomic distances. These problems are circumvented by drawing the bonds using topology. The topology option makes it possible to check an MD topology file graphically if extensive modifications have been made to it, as is the case when perturbation simulations are performed.

Hydrogen bonds

Hydrogen bonds and sets of hydrogen bonds can be displayed as dashed lines in combination with a structure in any atomic representation with any of the display modes available in mdFRODO. The selection of hydrogen bonds to be displayed can be done by bond occurencies. Color encoding can be done statically, dynamically or by bond occurencies. For instance, it is possible to display only those hydrogen bonds that are present 60–90% of the time in a simulation. The selected bonds may be shaded by strength with color levels. Weak bonds appear in blue color and strong bonds in white. Bonds with intermediate strength appear with other colors between blue and white.

Anisotropic RMS fluctuations

Atoms bonded together in a molecular structure show anisotropic behavior in their fluctuations due to interactions with other atoms. The orthogonal principal axis systems of these fluctuations can be computed from an MD trajectory and used to construct thermal ellipsoids.[16] The principal axis systems show the preferred orientations of fluctuations in space for individual atoms. The first principal axis corresponds to the direction with the largest amplitude fluctuation. The fluctuation ellipsoids describe the anisotropic distribution for individual atoms.

EXAMPLES

The consistency of MD simulations can be verified; this is the most obvious, and perhaps most frequent, use of mdFRODO. The interactive graphical approach complements the traditional theoretical analysis that makes use of averaging, RMS fluctuation and RMS deviation calculations, correlation functions and spectral analysis. Collective motion and secondary structure motion can be studied.[13] Domain, interaction-surface, side-chain and backbone dynamics can be studied separately and together. Active site dynamics can be analyzed. The subobject partitioning concept, in addition to the zoned color mode, allows one to focus attention on the dynamics of interesting regions and to cut out disturbing parts of a system. An active site may be described by the closely surrounding side chains, ions and solvent molecules with an all-atom representation; these subobjects are colored differently to highlight natural or arbitrary groupings and the rest of the peptide chain(s) of a protein may be described with a backbone representation. Hence it is possible to keep a background dynamics around a site of interest. Color Plate 1 displays a split vision stereogram of the detailed atomic structure of the cCTF helices *A*, *B* and *C* and the α-carbon structure of the other secondary structure elements in the zoned color mode during 300 ps of MD simulation. It is easy to distinguish the various segments of the polypeptide chain. The system fluctuates around one average structure and is well preserved and equilibrated.

Sampling a trajectory

In vacuo molecular dynamics simulations of protein molecules at room temperature sample conformational space with a rate that makes it adequate to select a conformation from the trajectory each 1–2 ps to visualize backbone dynamics, and each 0.05–0.1 ps if side chain mobilities are of interest.

Collective motion

Global rotational and translational modes of motion may be picked up by a molecule simulated *in vacuo*. No solvent damping is present. Normally rotational and translational motion whose origin coincides with the center of mass of the full molecule is removed before an MD simulation.

Nevertheless such modes of motion may be induced during the *in vacuo* simulation.† It takes normally no more than a few tens of picoseconds to build up the global rotational mode of a small globular protein, whereas the translational mode does not become significant until a time scale of hundreds of picoseconds is reached. Hence it is necessary to use least-squares superpositioning of MD conformations to analyze fluctuations in structure. Color Plate 2a displays the global rotational and translational modes of the cCTF backbone. These modes are removed in Color Plate 2b and the fluctuations around the average structure are seen. The fluctuations in the 10-ps average structure are displayed in Color Plates 2c and 2d. Average structure shifts are small in comparison with the instantaneous fluctuations in Color Plate 2b. The backbone structure basically fluctuates around one equilibrium conformation on the 300-ps time scale, and a significant number of the fluctuations appear to have frequencies higher than 10 ps^{-1} because the diversity in Color Plate 2b is larger than in Color Plate 2c. Some, usually charged, surfacial side chains occupy different average locations even in the 10-ps average structures (Color Plate 2d), indicating that they sample broad regions of conformational space.

It is evident that α-helices and the β-strands in β-sheet structures possess rigid bodylike modes of fluctuation.[13] The trail representation of an MD simulation verifies this statement. Correlated modes of low-frequency motions can be found. Due to the constraints present in the form of through space, noncovalent, main-chain hydrogen bonds (in addition to covalent constraints), these structures are forced to keep their shape with only relatively small fluctuations around their average structures at room temperature. Classical mechanical models for secondary structure fluctuations based on hydrogen bonding considerations have been outlined by Chou.[17] The structural integrity of the α-helices and the β-sheet region of the cCTF can be observed in Color Plates 2b and 2c. Loop regions, on the other hand, are generally not kept in place by extensive main-chain hydrogen bonding patterns, although stabilizing hydrogen bonds may be present.[18] They have fewer restrictions to move about in conformational space. Loop transitions are more likely to occur than are changes in the regular secondary structure of a protein molecule. The cCTF is a poor example of loop mobility because the loop regions are rather short in this molecule. The effect is more pronounced in MD simulations of proteins with long loops, such as bacteriophage T4 glutaredoxin. The hydrogen bonding pattern that maintains the integrity of the secondary structures of the cCTF is shown in Color Plates 3b and 3a using the hydrogen bonding display option of mdFRODO. By removing the side-chain atoms a clear view of the hydrogen bonding pattern is obtained (Color Plate 3b). The fluctuation vectors of the individual backbone atoms of the cCTF are plotted in Color Plate 4a (instantaneous fluctuations) and Color Plate 4b (average fluctuations) with the atomdist mode. It is clear that the structure fluctuates around a well defined state and that the fluctuations in 10-ps average structure are small.

The principal axis systems are useful to detect collective preferred orientations of fluctuation in regions of a molecule during part of a simulation. The degree of similarity of the direction and amplitude of the first principal axis of fluctuation between atoms can be judged. Anisotropy and changes in it can be estimated. This information is not only of importance for the short-time dynamics of the molecule but also for more time consuming processes, such as loop transitions, defolding of secondary structural elements and the folding process of the whole molecule. Color Plate 5a displays the principal axis systems for the some atoms of the cCTF during a 10-ps part of the trajectory. The RMS fluctuation ellipsoids describe the distribution of the atoms better than do the principal axis systems, although they contain the same information. Color Plate 5b displays the fluctuation ellipsoids for the same part of the trajectory as Color Plate 5a. Multiple peaks in the atomic distribution functions are clearly revealed if the ellipsoids are animated. Ellipsoids that change the direction of the first principal axis or that change shape have multiple peaks or sites of preference in space. Invariant ellipsoids, or ellipsoids that change only size, show that the atom has only one peak in its distribution function for the analyzed time. Interactive graphical analysis of protein molecule simulations shows that most of the surfacial atoms have multiple peaks in their distribution functions. Core atoms move more isotropically; they are encaged by a bulk protein environment. Color Plate 5c displays three consecutive frames of α-carbon thermal ellipsoids for the cCTF. Most of the α-carbons do not change their distribution functions. This is also valid for longer times. Exceptions are some of the loop α-carbons that change both the direction of the first principal axis of fluctuation and the degree of anisotropy.

The trail mode is perhaps the most powerful animated analytical tool. It allows a dynamical analysis of individual and collective atom fluctuations without the obscuring presence of bonds. If the bonded representation† of a molecule is present, the eye is fooled into examining variations in bond locations rather than atomic fluctuations, giving rise to artificial impressions. The atomic motion in a molecule is determined by individual masses and charges, modified by constraining bonds. This is valid in reality as well as in simulation. When studying the dynamics of a multiparticle

†Momentum and angular momentum are physically conserved properties. The induction of rotation does not conflict with the conservation of angular momentum: Rotation can be induced by a redistribution of the mass in a system if any of the global moments of inertia are altered during the simulation. The angular momentum is the product of the moment of inertia for a particular axis of rotation and the angular velocity around that axis. If the moment of inertia decreases due to changes in shape, the angular velocity must increase. Thus, the induction of major global rotational modes illustrates the fluidlike nature of a protein molecule. Small alterations in the angular momentum may originate in rounding errors in the simulation, causing minor rotation. However, if translational motion is gained by the molecule, energy must be introduced or removed from the system. There is only one possibility for this. Numerical errors are introduced by discrepancies in precision in the computer numerical library subroutines. The gain in global rotational motion for some systems may be drastically larger than the gain in translational motion, in agreement with theory. It is expected that the numerical effect is small, whereas the redistribution of mass may be significant.

†Actually, the bonded representation of a structure does not give a true picture of the molecular bonds. It is the distances between the atomic mass centra (nuclei) of a molecule that are shown. However, it does show the occurring atom-to-atom covalent constraints.

object, it is not the bonds that are important; they do not change. It is the average position and fluctuations of the individual atoms that give the object its dynamical and biological properties, not the shifts in bond locations. Hence the focus should be directed toward atomic fluctuations. It is more appropriate to visualize a protein as an internally fluctuating body than as a structure. The trail mode is the key to understanding the elasticity, plasticity and fluidity‡ properties of a molecule. Generally, a protein molecule consists of a solidlike central core, where atoms are caged and move elastically. The core supports a surface of fluidlike atoms with plastic behavior. Surfacial atoms are easily displaced by thermal motion or by approaching bodies, with the restriction that they cannot move far away from their original positions unless the backbone structure to which they are attached makes a conformational transition. The surface plasticity properties of a molecule ensure that it continuously changes its exterior shape and electric charge distribution. It is important to take this into account in ligand binding studies.

Comparative animation

We have found the use of animated representations of sets of comparable structures to locate differences in complex structural patterns most advantageous. This application is called "comparative animation." Consider the situation where a molecule in an MD simulation passes through a number of relatively stable structural substates. It is possible to locate these states and to compute the average conformations of atomic positions and hydrogen bonds for each substate. To characterize these substates and to find differences between structures, and differences in the location and strength of hydrogen bonds, an analysis set of the average conformations can be created with an analysis set of the hydrogen bond conformations. These data sets can be used to produce an animation. Unless the structures are very simple, an overlay display will be too complex to be readily comprehended. The animated representation resolves the complexity by displaying only one structure (with its corresponding set of hydrogen bonds) at a time. Regions that differ among the structures will show changes in location of bonds. The breakage and formation of hydrogen bonds can be monitored. Changes in hydrogen bond strength can be observed by variations in bond color.

Comparative animation is also of assistance when relating very similar structures, such as conformations before and after some cycles of energy optimization. If most of the bonds in the compared structures superimpose, a dynamically colored overlay will be messy. If the set of structures is animated with the dynamic color mode the resulting image will be a single static structure that changes its color without motion, except for the few locations that differ. These regions are easily spotted.

DISCUSSION

An enhanced version of FRODO with an interface to GROMOS has been developed. FRODO can now make use of topology and hydrogen bond data and animations can be produced. A variety of display techniques have been developed and implemented to facilitate the efficient analysis of dynamics, conformational space and structural differences in sets of structures. The usefulness of animated analysis techniques has been put into a biological context illustrated by surface plasticity analysis, principal axis of fluctuation analysis and comparative animation. We have demonstrated with the trail mode that alternative molecular representations can provide additional and complementary information of the dynamical behavior of a molecular system. Source code availability ensures that modifications can be done to fit individual needs. The program is not restricted to the animation of molecular objects. Any object that can be described as a set of vectors can be animated with mdFRODO.

Using static photographs like those shown in the plates for this paper, it is not possible to convey the assistance that a visual three-dimensional, animated macromolecular conformational space analysis gives. Here only the static part of the conformational space analysis can be accurately represented. Properties like fluidity, surface plasticity and core elasticity, the speed of conformational transitions, fluctuations and drifts can only be visually comprehended through animation.

To analyze sets of molecular conformations is of interest not only in the MD field. In X-ray crystallography and two-dimensional NMR sets of trial structures are generated and it is essential to rapidly grasp structural differences between conformations. The ability to gather these structures in an analysis set and examine various selections of dozens of conformations in a static display, to color encode them differently and to work on different levels of resolution by selecting different representations of atoms may prove helpful for workers in these fields. The comparative animation part is of use although it has no longer a direct physical time dependence coupling as in MD. It may be of assistance to estimate differences between structures in a sweep animation, for instance, without resorting to the standard static overlay display, in analogy to what we found in the analysis of hydrogen bond patterns. It would be possible to feed NOE connectivity data or distance constraints to mdFRODO instead of hydrogen bonds to examine and color encode violations. In the field of drug design it is important to analyze surface fluidity and plasticity, in addition to the static steric possibilities, to fit ligands to active surfaces. The animated display modes of mdFRODO in combination with Connolly or van der Waals surface representations may be of assistance for these workers.

The methods presented in this paper are of interdisciplinary use. In the fields of molecular biology, physical chemistry and material physics the analysis of dynamics in materials is important. Alternative graphical approaches to studying compounds may lead to discoveries and the elucidation of processes involving atomic motion in materials.

The program is fast enough to run animations of fairly complex systems that take minutes of real time to display; one obvious application is the production of animated video molecular movies. It is possible to feed a video recorder with the signal produced by the Silicon Graphics workstation through an RGB to PAL/NTSC converter with fair quality, or simply to copy the screen with a video camera, with impaired quality.

‡The surface atoms of a molecule are not truly fluid. They do not have the freedom of diffusion. "Fluidlike" is a more appropriate term.

PROGRAM AVAILABILITY

The program mdFRODO is copyrighted and distributed under the following conditions: Program mdFRODO, including sources, documentation and examples, and the documented source code subroutine library NICE are available from us at nominal administrative cost for university users, and by explicit licence agreement to commercial users.

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish National Board for Technical Development (grants no 87-02139P and 87-00694P). Profs. Orlando Tapia and Britt-Marie Sjöberg are gratefully acknowledged for their support. The author is indebted to Prof. Wilfred van Gunsteren, Prof. Alwyn Jones, Dr. Mercedes Campillo, Dr. Eduardo Horjales, Dr. Jeramy Walton, Malin Wästlund and Rob Sok for discussions, software and useful suggestions. The warm hospitality of the department of physical chemistry of the University of Groningen (The Netherlands) is greatly acknowledged.

APPENDIX

Implementation in the FRODO shell

The following paragraphs assume some familiarity with the FRODO program. To obtain a high degree of portability, the subroutines needed to perform the conformational space analysis are kept separate from the FRODO main program. The distance dependant bond generation routine is separate from FRODO's own routine. The analytical subroutines build up module MD. Basically, coordinates of sets of conformations are imported to module MD as formatted files, and exported to the FRODO display shell via a few routines that construct graphical objects. Module MD contain no provision for the interactive display and manipulation of conformations. These functions are provided by the FRODO shell. This method of implementation ensures that module MD can be incorporated in any other molecular graphics program with reasonable ease.

General features of mdFRODO

The display modes, coloring, atom selection and other options described in this paper have been implemented in the Silicon Graphics IRIS version of FRODO. All MD subroutines and most of the FRODO subroutines are written in FORTRAN 77. Module MD offers high flexibility because all modes, coloring and options can be mixed together to produce the desired analytical movie or static picture. The interface with FRODO is also very flexible. All other options in FRODO are active while module MD is used. The static picture or movie can be examined interactively using all of the standard FRODO options, such as rotation and zooming. Split vision stereo three-dimensional animations can be produced. MOL objects, van der Waals surfaces and foreground objects can be displayed and manipulated while the animation is running. Distance measurement and picking on the foreground objects are active during animation. An option to make the foreground object invisible has been added as a CHAT command. This makes it possible to use the mouse-click based FRODO options on one of the animated objects, if it shares coordinates with the foreground object. The MD subprogram can be turned on and off, enabling use of the standard FRODO without module MD whenever required.

Construction and display of movies

The principles of the MD animation driver are the following:

1 Conformations are transferred to the display system as separate formatted coordinate files.
2 Names of all the files to be used in the animation are gathered together in a formatted pointer file with one file name per record, including directory locations if necessary. Uninteresting conformations may be commented away in the pointer file with an exclamation mark at the first character position of each record before the file name.
3 The name of the pointer file is specified to module MD.
4 Selections of atomic representations, color mode, display mode and other options are entered to module MD.
5 As module MD is exited the conformations are read from the files and stored in arrays.
6 As the FRODO/CHAT mode is exited for the interactive display mode, the coordinates are used to create the objects needed for the animation. If a continuous update of the bond list is ordered, then the bond list is recomputed before each frame is constructed.
7 When all of the objects have been constructed mdFRODO enters the interactive display mode and begins to display frames as specified by the display mode and other options.

The pointer file is a primitive yet powerful concept that allows the user to reshuffle and update the sequence of conformations. It can be replaced with a database interface in the future. With a suitable database manager it would be possible to greatly increase the flexibility of data input and selection. The mdFRODO program is constructed with the database approach in mind.

Interface to MD software

The current version of mdFRODO is interfaced to the MD simulation program package GROMOS. Interfacing to other simulation packages is straightforward. Only modules for reading file formats specific for each package need to be added. The mdFRODO program reads standard formatted GROMOS conformation files of GSF† type and performs automatic conversion of units from nanometers, which is the GROMOS standard unit, to angstroms if necessary. The automatic scale option can be turned off and operated manually.

NICE

The scanning of MD trajectories and the preparation of coordinates for display is done outside mdFRODO. A structured and well documented FORTRAN 77 subroutine and sample program package NICE, designed to ease manipu-

†GSF, GROMOS standard file format.

Table 1. Comparison of performance between the Silicon Graphics 4D/20G and the 4D/80GT running the mdFRODO application with a set of 19 structures of cCTF. Three different representations of the molecule have been tested. Bonds are drawn between alpha carbons (CA), backbone atoms, and all atoms. READ FILES is the time in seconds to read the test set of data files into memory, scale them to Ångström and calculate one bond list using a 1.9 Å distance criterion. MAKE OBJECTS is the time in seconds to construct all the 19 graphic objects in the test set. MAKE UPDOBJECTS is the same as MAKE OBJECTS but with the bond list recomputed for each frame. ANIMATE is the time in seconds to run one movie in animate display mode of the test set of frames with no idle time. FRAMES PER SECOND is the maximun number of test frames per second that can be displayed in ANIMATE mode with mdFRODO

	CA atoms		Main chain atoms		All atoms	
Benchmark	20G	80GT	20G	80GT	20G	80GT
Number of vectors per frame	68		206		1362	
Total number of vectors in display set	1296		3914		25878	
READ FILES(s)	18	12	18	12	18	12
MAKE OBJECTS(s)	~1,5	<1	~1,5	<1	36	24
MAKE UPDOBJECT(s)	36	24	36	24	70	50
ANIMATE(s)	1,11	0,63	1,27	0,63	1,90	0,96
FRAMES PER SECOND	17	30	15	30	10	20

lation of MD data, has been developed. It offers a full GROMOS interface regarding file formats. It is possible to add compatibility to other simulation packages by including other file formats in the I/O routines of NICE. Extraction of conformations from MD trajectory files for mdFRODO and least-squares fitting is done by the NICE program PROEXT. Input GROMOS trajectory files may be binary, in single or double precision, or formatted. NICE now contains more than 140 subroutines that are documented both with extensive comments in the source codes and in an indexed manual.

The startup files generated by the GROMOS dynamics simulation program PROMD can be used directly by module MD because they are in the GSF format. Hydrogen bonds with time occurrence data can be read directly from GROMOS PROAHB files. Average structures and anisotropic RMS fluctuation data can be read from GROMOS PROAVX files. Bond lists can be read from formatted GROMOS topology files.

Performance

The mdFRODO program was tested on a Silicon Graphics Personal IRIS 4D/20G (the smallest model) with the *z*-buffer graphics card and the standard 8MB of RAM memory, and on a more powerful middle range machine, the Silicon Graphics Professional IRIS 4D/80GT, likewise with 8MB of RAM memory. The 4D/80GT is normally about two times faster than the 4D/20G running this application. An improvement by a factor of two may not seem like much, but it is very noticeable when working interactively with mdFRODO. Schematized representations of all backbone atoms of the cCTF with 639 atoms, 69 residues and 206 backbone bonds, are displayed too quickly on both machines to be comprehended in all animated modes of the program if no idle time is specified (about 15 frames per second on the 4D/20G in animate mode). The all-atom representation of the same molecule, with 1362 bonds, is displayed with a rate of 10 frames per second in animated mode on the 4D/20G and is also too fast to be readily comprehended. The construction of frames is done before the display loop of the program, ensuring fast performance in the animation. The construction of frames for the cCTF backbone example takes about 80 ms per frame. A normally sized analysis with 40 frames takes about 3 s in the construction of frames. A comparison of the software's performance in the two Silicon Graphics machines is listed in Table 1 and relative speed increase factors between the machines for different operations are listed in Table 2.

Table 2. Measured relative difference in performance between the Silicon Graphics 4D/20G and the 4D/80GT running the mdFRODO application. The abbreviations are the same as in Table 1

Benchmark	Speed increase factor using the 4D/80GT
READ FILES	1,5
MAKE OBJECTS	1,5
MAKE UPDOBJECTS	1,5
ANIMATE	2,0

Limitations

The current architecture and dimensions in mdFRODO allow a maximum of 300 conformations with a maximum of 8000 atoms to be included in the analysis set, without loss of performance caused by the virtual memory system. This is enough to produce minute-long uninterrupted molecular movies at a comprehensive speed. The hydrogen bond analysis set is limited to 40 conformations of a maximum of 2000 bonds each. It is possible to extend these dimensions further, if necessary. The RMS fluctuation analysis set is limited to 300 conformations for average structures and to 40 conformations for anisotropic RMS fluctuation data.

REFERENCES

1 Swanson, S.M., Wesolowski, T., Geller, M. and Meyer, E.F. *J. Mol. Graph.* 1989, **7**, 240–242
2 Hilmer, R.M. *J. Mol. Graph.* 1989, **7**, 212–216
3 Weiner, P.K., Gallion, S.L., Westhof, E. and Levy R.M. *J. Mol. Graph.* 1986, **4**, 203–207
4 Jones, T.A. *J. Appl. Crystallogr.* 1978, **11**, 268–272
5 Cambillau, C., Horjales, E. and Jones, T.A. *J. Mol. Graph.* 1984, **2**, 53–54
6 van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J. and Potsma, J.P.M. in *Proc. Natl. Acad. Sci. USA.* 1983, **80**, 4315–4319
7 van Gunsteren, W.F. and Berendsen, H.J.C. *Groningen Molecular Simulation (GROMOS) Library Manual.* BIOMOS B.V. Groningen, The Netherlands (1987)
8 Polygen (Europe) Ltd., 29 Queen Anne's Gate, London, SW1H 9BU, UK
9 Chemical Design Inc., Ste. 120, 200 Rte. 17, South Mahwah, New Jersey 07430, USA
10 Dayringer, H.E., Tramontano, A., Sprang, S.R. and Fletterick, R.J. *J. Mol. Graph.* 1986, **4**, 203–207
11 BIOSYM Technologies, Inc., 10065 Barnes Canyon Rd., Ste. A, San Diego, CA 92121, USA
12 Hubbard, R.E. in *Proc. Computer-Aided Design Conf. Oyez UK.* (1984) 99–106
13 Tapia, O., Nilsson, O., Åqvist, J., Horjales, E. and Campillo, M. in *Structure and Methods. DNA Protein Complexes* (R.H. Sarma and M.H. Sarma, Eds.) Adenine P., New York (1990) **2**, 147–170
14 Leijonmark, M., Liljas, A. and Subramanian, A.R. *Biochem. Int.* 1984, **8**, 66–76
15 McLachlan, A.D. *J. Mol. Biol.* 1979, **128**, 49–77
16 Willis, B.T.M. and Pryor, W. *Thermal Vibrations in Crystallography.* Cambridge Univ. P., Cambridge (1975)
17 Chou, K.-C. *Biophys. Chem.* 1988, **30**, 3–48
18 Tramontano, A., Chothia, C. and Lesk, A.M. *Proteins* 1989, **6**, 382–394