Visualization of moving biomolecules: a new approach based on professional 3D animation software

Yuri Porozov[1,2], Raluca Andrei[1,2], Monica Zoppè[1,*].

[1]Scientific Visualization Unit, IFC – CNR, via Moruzzi 1, Pisa. Italy.
[2]Laboratorio di Biologia Molecolare, Scuola Normale Superiore, Pisa, Italy.
[*]Corresponding Author. mzoppe@ifc.cnr.it
www.scivis.ifc.cnr.it

ABSTRACT
We are setting up a system that enable us to visualize proteins and other biological molecules in a 3D virtual environment built according to scientific information and physico-chemical properties. This system will permit a novel view and understanding of the functioning of cells, of protein interactions and of dynamical relationships occurring in the small units of all living systems.

INTRODUCTION
The vast amount of knowledge accumulated on the structure of cells, the shapes and movements of its constituents, the interaction among participants and with the environment is at present in a form which is accessible only to experts of the fields. Moreover, this information is often difficult to interpret in terms of dynamic deployment of single events.
Our aim is to use available biological information to describe the inside working of a cell in 3D animated representation. To reach this aim, we are using Maya/Autodesk, one of the most powerful software developed by the industry of 3D animation and special effects (1). With data imported directly from the Protein Data Bank (PDB), we animate protein movements in virtual space, according to information and rules derived from physics, chemistry, biochemistry and other scientific sources.

Using our Maya script, atomic coordinates of proteins and other molecules are imported, together with chemical structures. If more than one conformation is present for a molecule, then these are imported and the program is run to interpolate intermediate position that transit the protein from one conformation to another.

The first examples we present are two small molecules (Triazine, and Bitucarpin) for which theoretical dynamic studies already revealed the energy landscape, which is used for validation of our system.
Most biomolecules however, and notably proteins, contain very large number of atoms, requiring different, more complex programs that can accomodate the large information content of such molecules.
We will present results obtained with our system and offer demonstration of how it can be applied to peptides (we used the V3 peptide of HIV-1 gp120) and the entire gp120 protein.

From the point of view of animation programs, animating means inferring intermediate steps between a start and an end position, i.e. moving the virtual object in space along time. The kind of movement can be governed by sets of rules defined as mathematical expressions, many of which are already present in Maya; other can be codified.
It is important to notice that all what our system does is to interpolate positions while avoiding prohibited ones, therefore the calculations are extremely rapid. In this respect, it is completely different from the most widely used Molecular Dynamics programs, which compute positions according to very complex energy calculations. On the other hand, it is possible that some interpolated movements are patently wrong, which introduce the need for human revision of every animation.

Another important aim of our project is the delivery physico-chemical information of molecules and of the environment such as pH, electronegativity, hydrophilicity and others that are of importance for the way biomolecules behave. This process, in Computer Graphics, is called Rendering, and we will also show some of the progress in this respect.

Small molecules

Animation of molecular structures implies that information relative to the identity of the atoms, their positions, their reciprocal relations, are first imported in the animation system. After this, different positions can be assigned to every single atom at different time-points and interpolation of atom's positions between time points can be calculated. To this aim, we have first used a few small molecules for which the energy landscape of different positions has been generated through Molecular Dynamics studies: Triazine (2), Bitucarpin (3) and Di-Ala dipeptide (4).

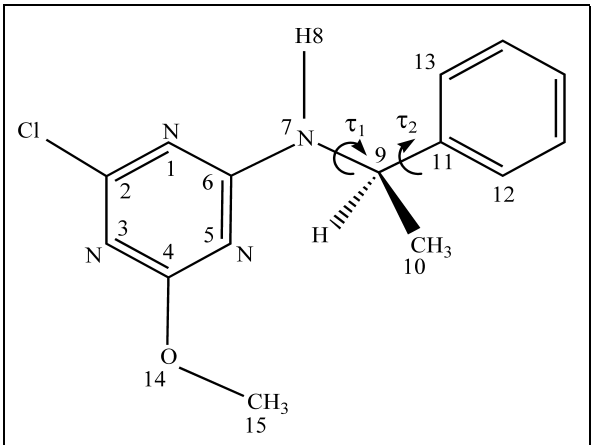

Fig. 1.

Triazine (2-chloro-4-methoxy-6-[(R)-1-phenyl-ethylamino]-1,3,5-triazine) is a small molecule composed of 31 atoms, with a relatively simple structure of two rigid disks connected by a C-N bridge (see Fig.1). The different conformational positions that Triazine can assume are basically variations of τ_1 and τ_2, i.e. rotations around the two chemical links that connect N7, C9 and C11.
Dynamical simulation studies by Alagona et al (2), have revealed the energy landscape for all possible conformations that Triazine can assume. For this reason we chose it as the initial test molecule of our bio-chemical Maya system.

The program we developed to import chemical data into Maya assigns every atom to a position. Atoms are linked through bones (see panel A in Fig. 2), which behave like chemical bonds, have fixed length, and are constrained by codified rules.
Four different conformers, three minimal energy positions and one intermediate (A-B-C and H, see Fig. 3) were imported, and assigned to time points in the animation (key-framed).
Coordinates for all atoms in some intermediate positions calculated by Maya along two possible pathways between C and H were retrieved and fed back into pdb-like files. Angles τ_1 and τ_2 were calculated and plotted entered into the energy map, allowing for physico-chemical evaluation of the path calculated by Maya.

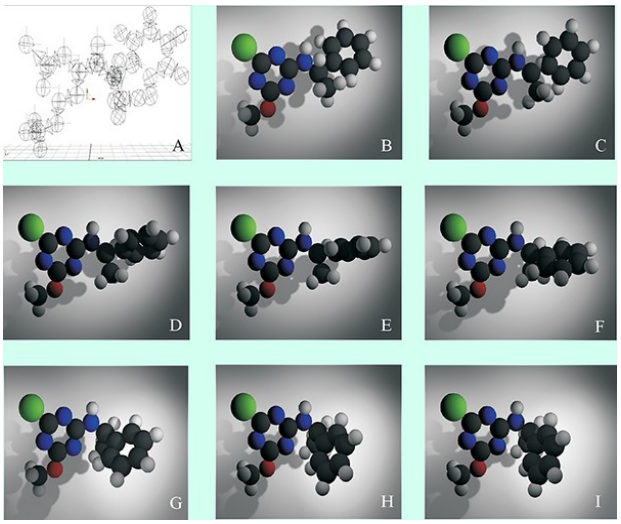

Fig. 2.

Fig. 3 reports the energy landscape, from ref. 2, with the paths calculated by Maya for transition between position C and position H, following the two possible trajectories.
Note that the path labelled with black dots, which includes an almost 180º rotation of the phenyl ring, spans the energy field more than once. This is because the energy is calculated on a chemical basis, where Carbons 12 and 13 are equivalent; however, in a topological view, each of them has its own identity, and the landscape is in fact twice as large.
These results show that Maya can calculate paths avoiding the energy peaks, i.e. describing

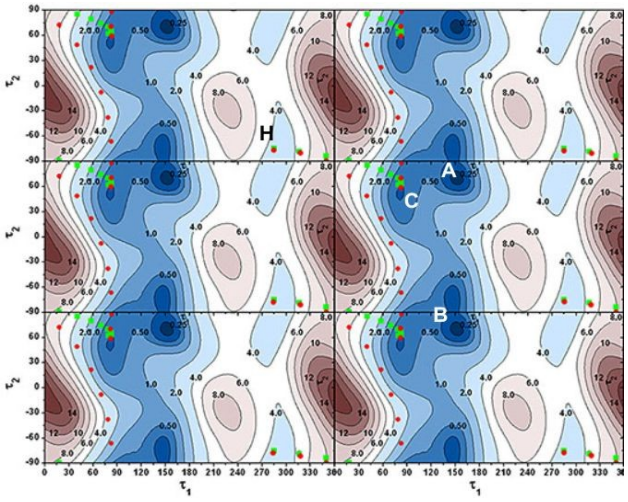

Fig. 3

a movement that is chemically acceptable, flowing naturally in the 'valleys' of the landscape.

Results for Bitucarpin, (a plant chemical for which the energy map has also been calculated) and for the Di-Ala dipeptide will be presented as demonstrations in real time at the meeting.

Peptides and proteins

Proteins can contain up to several thousands or tens of thousands atoms. For some of them, crystal structures have been determined in different conformations, allowing us to set two key-framed positions and to elaborate possible interpolations to transit from one conformer to the other. For other small proteins or peptides, NMR studies provide variable numbers of conformations that the peptide can assume in solution: we have taken advantage of this information to script an animation program that runs in Maya.

Fig. 4 shows the interface of our program. The user can upload the .pdb file to be used as source and set a number of features, including the kind of source (X-ray or NMR), the atoms to be represented (including or excluding hydrogens), the timing of animation and the atoms to be considered for the animation.

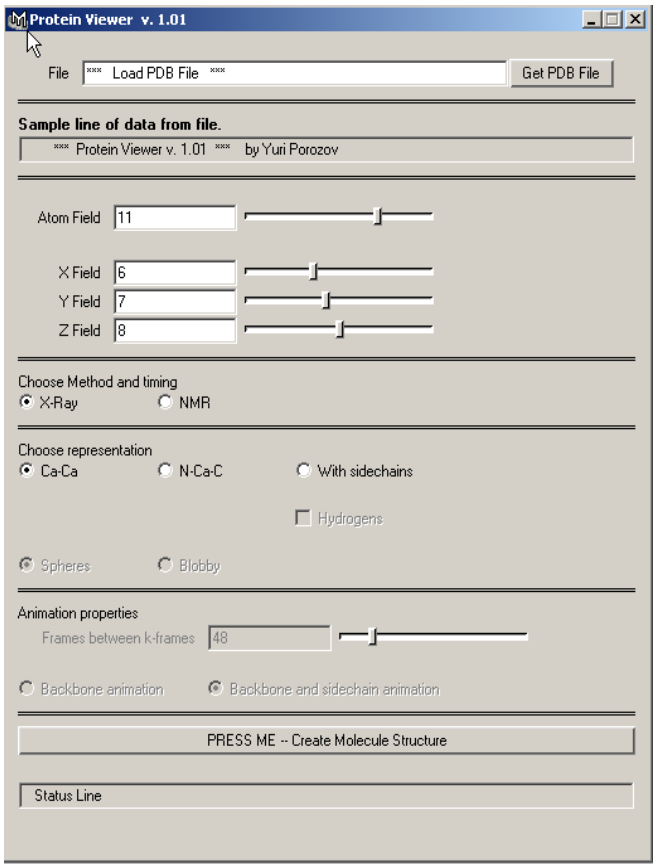

Fig. 4.

For proteins we have used the Particle feature to create them in the 3D space of Maya. Particles are 'light objects' in terms of processing power, and can be dealt with either as a single object that includes them all, or on a *per particle* basis. In other words, large movements (such as bends on a hinge) and relative movements (of the object in space) can be imposed and calculated very fast. The *per particle* attributes are used for rendering (where each kind of atom displays different), and for imaging detailed movements.

To test the system for NMR, we have used the 20 conformers of V3 (5), PDB entry 1CE4. All sructures, after being ordered using a statistical apporach, were imported in to Maya. Each conformation was assigned to a different time-point and animated. The resulting animations will be shown during the demonstration, and can be seen on our website www.scivis.ifc.cnr.it.

Gp120 structure has been solved in different conformations: either unbound (6) or bound (7) to CD4. Its interaction with the cellular receptor CD4 triggers a major movement of parts of the protein, in particular the V3 loop. V3 is implicated in the selection of co-receptor and in the subsequent step of co-receptor binding.

Rendering

Visual perception of the world is a very complex process that we perform automatically. When producing totally artificial images, to obtain the impression of realism, we have to introduce a large set of effects, such as light sources, casting shadows consistent with the illumination, assigning optical properties to materials, fixing the 'eye settings', i.e. the (virtual) camera properties and so on.
To assign texture to objects is a complicated task even for reproducing properties of 'real' objects, when all we have to do is to copy the 'visual feel'. The rendering process (i.e. assignment of visual properties to surfaces) in CG programs involves the setting of material (2d and 3d textures) color,

reflection, luminosity, lights, ambient light and camera movements. Proteins and other biological molecules are, in essence, chemical compounds with specific properties that are determined by the nature of their atoms and the way they are connected and organized in the 3D space.
These properties are defined, in physico-chemical terms, as potentials, typically expressed with complex equations and/or numerical values. One of the aims of our effort, is to convey the significance of these properties in a visual way. Chemical programs can calculate, for example, the electrostatic potential of a surface, or its hydrophobicity, and report it on the surface using a conventional code, typically a colour scale.
We report in Fig. 5 some images obtained while studying different ways to render applied to a form created with a random process or to a shape representing a branched complex sugar typically found on glycoproteins. Images were obtained using the RenderManForMaya plug-in from Pixar.

Fig. 5.

CONCLUSIONS

The initial work presented here is part of a large project that will bring to virtual (and visible) life the processes that occur in the real (but invisible) world of cells.
Because the understanding obtained through sight is much more direct than through word description or intellectual (mental) representation, this will permit researchers to get a more direct grasp of the phenomena under study. Providing a different vision, it should also enable the formulation of new questions, or an alternative way to formulate old, still unanswered ones.
Furthermore, the availability of a virtual cell might allow testing new hypothesis in the virtual cell before performing real experiments.

Also, a direct representation will greatly facilitate the teaching of cellular and molecular biology, at various levels, from secondary school to higher university, and it will also be available for museums, thus attracting new students to the fascinating field of biology.

REFERENCES

1. www.alias.com
2. Alagona, G., Ghio, C., Monti, S. (2006). A Test Case for Time-Dependent Density Functional Theory Calculations of Electronic Circular Dichroism: 2-Chloro-4-Methoxy-6-[(R)-1-Phenylethylamino]-1,3,5-Triazine Theor. Chem. Acc. (in press). Published online: Jan. 5, 2007.
3. Alagona, G., Ghio, C. and Monti, S. B3LYP/6-31G* conformational landscape in vacuo of some pterocarpan stereoisomers with biological activity. Phys. Chem. Chem. Phys., 2004, 6, 2849-2857.
4. Wang, Z. X. and Y. Duan (2004). "Solvation effects on alanine dipeptide: A MP2/cc-pVTZ//MP2/6-31G** study of (Phi, Psi) energy maps and conformers in the gas phase, ether, and water." J Comput Chem **25**(14): 1699-716.
5. Vranken, W. F., F. Fant, et al. (2001). *Conformational model for the consensus V3 loop of the envelope protein gp120 of HIV-1 in a 20% trifluoroethanol/water solution.* Eur J Biochem **268**(9): 2620-8.
6. Chen, B., E. M. Vogan, et al. (2005). *Structure of an unliganded simian immunodeficiency virus gp120 core.* Nature 433(7028): 834-41.
7. Kwong, P. D., R. Wyatt, Robinson, J., Sweet, R. W., Sodroski, J., Hendrickson, W. A. (1998). *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody.* Nature **393**(6686): 648-59.