Using Blender for molecular animation and scientific representation

Monica Zoppè[1*], Yuri Porozov[1,4], Raluca Andrei[1,4], Stefano Cianchetta[1], Maria Francesca Zini[1], Tiziana Loni[2], Claudia Caudai[1], Marco Callieri[3]
[1]Scientific Visualization Unit, LTGM. Institute of Clinical Physiology, CNR Pisa. [2]Big Bang Solutions, Polo Tecnologico di Navacchio, Pisa. [3] Institute of Information Sciences and Technologies, CNR Pisa. [4] Scuola Normale Superiore, Pisa, Italy.

[*]Corresponding Author: mzoppe@ifc.cnr.it www.scivis.ifc.cnr.it

ABSTRACT

The inside of a cell is a nanoworld in which life happens at tiny scale and high speed: hundreds of protein characters (minute as ants or large as elephants) play their roles in several different and fascinating environments, as varied as forest and desert, ocean and metropolis. Biologists have revealed the events of inner life using experimental quantitative and qualitative techniques and describe them in (difficult) scientific reports. We want to bring the beauty of biology to everyone's reach, using Blender in a rigorous but at the same time creative way.
We have produced scripts to import crystallographic and NMR information from scientific databases, and are in the process of developing algorithms to obtain animations from still (3D) images. We also intend to use creative renderings in order to convey information about the physical and chemical properties of the subcellular environments and of the surface of each component in the cell.
Our work has the primary aim of enabling scientists to observe directly the objects of their study, by showing the kinetics and conformational changes as animations, but it can be of great value for teaching, for the spread of scientific knowledge and thinking, and provide artists with inspiration as representation of life, well beyond the DNA spiral.

1. INTRODUCTION

The last few years have seen impressive advances in the development of computer graphics and animation, and also in the description of biological processes at the cellular and molecular level.
We think that the time is now mature to propose a vision of life at the nanoscale in 3D animation, and that this effort will bring important benefits to the biological research community, to the development of 3D techniques, to the teaching sector and to the spread of scientific knowledge for a wider public.
In this presentation, we will briefly expose some principles of cellular biology which are relevant to the animation, and discuss some of the issues that we face and the advances that we obtained using Blender 3D software (1) to import and animate proteins.

1.1 The cell

Cells, the fundamental unit of life, are microscopic entities that contain a full microscopic world. A good idea of the relations that take place within cells, can be obtained by 'exploding' a typical cell to a size with which we are familiar (see Table 1).
As reported, if we look at a medium size cell of about 10 μm size, and we compare it to a village or to a lake, not very big but very deep, all internal components can be represented accordingly, and we see that the nucleus, which can occupy up to half the volume of a cell is the major internal object. Objects of this size (including the Endoplasmic reticulum, the Golgi apparatus, mitochondria chloroplasts and some other structures) can be seen with microscopic techniques that allow us to visualize their shape and (sometimes) their dynamic activity. It is important to note that, in contrast with the human-size world with which we are familiar, the entire volume is occupied, such that it might be easier to imagine a water body rather than one filled with air. Furthermore, we have to notice that gravity is irrelevant at this size (the mass of objects is too small to be significantly affected by the Earth gravity field), and the movements of cellular components is mostly driven by thermal agitation.
The boundary of the cell, as well as the walls delimiting internal volumes, is made of membrane, a soft, flexible and (relatively) thin layer that mediates communication and transport of material and information between the inside and the outside of cells. This is an extremely important structure that deserves more detailed description, which is out of the purpose of this document, and which can be studied in any of the many excellent textbooks dedicated to cellular and molecular biology (Refs 2 and 3).
Going smaller, we meet nucleic acids: DNA and RNA. Everyone is familiar with the double helix of DNA, but few people realize that in relative size, if the diameter of the helix is 2 cm (a very thick rope or hi tension cable), its length is 20.000 km, about half the Earth circumference. DNA is packed in a

Table 1

CELL	5 – 50 μm	x 10^7	**Village, Small lake**	50 –500 m
Internal Structures				
Nucleus	3 – 15 μm		Sports field, large (10 floor) building	30 –150 m
Golgi Apparatus	1 – 5 μm		Medium building (3-6 floors), Airplane	10 – 50 m
Membrane (thickness)	5 – 7 nm		Wall (internal), Front door	5 – 7 cm
Ribosome	30 nm		Cat	30 cm
Proteins				
GFP, Actin	3 – 4 nm		Apricot	3 – 4 cm
Spectrin	100 nm		Snake	1 m
NFkB complex	10 – 12 nm		Grapefruit	10 – 12 cm
DNA				
double helix diameter	2 nm		Small pipe,	2 cm
lenght	2 m		From North to South pole	20.000 km
Other molecules				
ATP	1,5 nm		Cherry	1,5 cm
Ca^{++} ion (without water)	0,2 nm		Flea	2 mm
Ca^{++} ion (with 1 shell of water)	1,2 nm		Hazel nut	1.2 cm
Water	0,28 nm		Small ant	2.8 mm
Sugar (glucose)	0,6 nm		Pea	6 mm
Cholesterol	2 nm		Bee	2 cm
Virus (HIV)	100 nm		5-6 y human	1 m

TABLE 1
Comparative size of a cell and its internal strucutres. The Factor 10^7 is selected so that 1 A is equivalent to 1 mm

very efficient organization that allows access to it both for retrieving information and for replicating it every time a cell divides. This organization is accomplished thanks to the involvement of proteins, the major players of cellular life, and the most immediate subjects of our animation efforts (see below).
From this overview, it should be clear that we can observe cellular life at many different levels of focus, spanning 4 or 5 orders of magnitude. However, if we can easily recognize the size of familiar sights (a valley or mountain, a building, a tree or an insect), there are no immediate references for attributing dimensions to objects that we have never seen before, such as ribosomes and actin (see fig. 1 and Ref. 4). One of the tasks we face is to provide the observers with clues indicating the scale of the objects that we model.

1.2 Proteins

Because proteins are the major characters of cellular life, and in fact are a major subject of scientific studies, we developed first a system to import them in Blender. It is necessary to describe some details of their general structure to understand how they are built (in nature and in Blender) and how they can move.
Proteins are constructed as a linear sequence of amino acids, which are small assemblies of atoms that share some features that allow them to be linked directionally one after the other. There are 20 different types of aminoacids, distinguished by the number and nature of the atoms attached to the linkable part, called Main Chain (for a more scientific and detailed description, see any molecular or cellular biology textbook).
Each protein contains from few hundreds to few thousands aminoacids, and, despite being a linear sequence, each one of them immediately after synthesis folds in space to acquire a 3D structure which is remarkably stable, although flexible.
The structure of proteins can be determined experimentally, and is stored in the Protein Data Bank as a .pdb file, which contains a number of information about the sequence of the molecule, the details of experimental procedures to obtain

Fig. 1

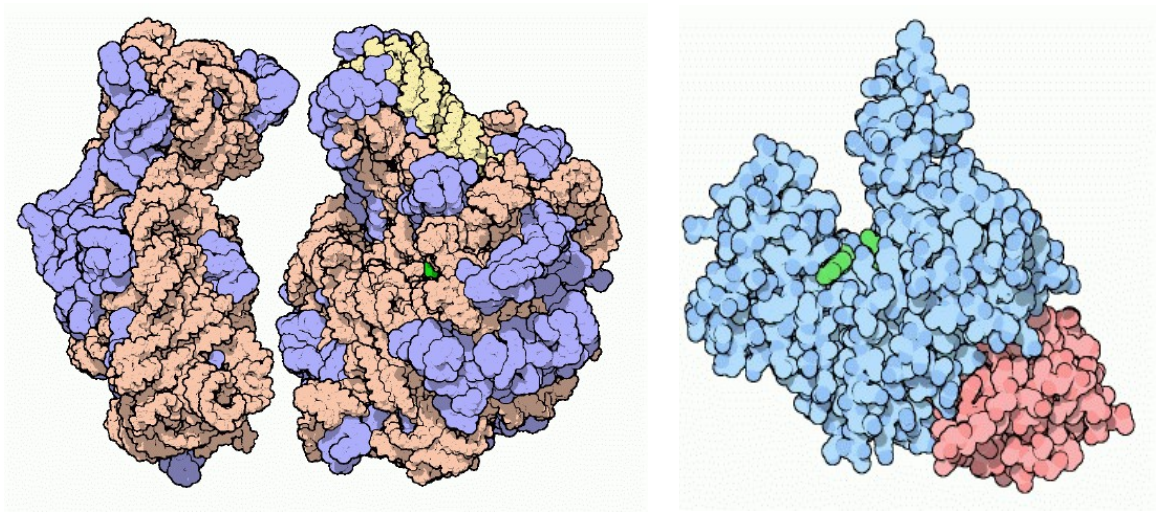

FIGURE 1

Images of a Ribosome (Left) and of an Actin monomer (Right), from the Molecule of the Month in the PDB website, by David S. Goodsell (ref 5).

The two are not in scale: their relative size is comprable to a cat (ribosome), and a small fruit (actin).

the structure, and the list of all atoms of the proteins and their XYZ coordinates. With this information and including the chemistry of aminoacids (that is, how atoms are connected), it is possible to build in the 3D environment the complete structure of any protein.
While X-ray crystallography results in determination of the position of all atoms with good resolution, for a single conformation, other types of techniques such as Nuclear Magnetic Resonance can yield a collection of coordinates, corresponding to a number of positions that the protein can assume. To obtain motion, all we have to do is find the path that every atom follows to go from one conformation to another, taking into account also the limitations and constraints imposed by chemistry and physics.
We describe here the initial results of our work to produce such molecular motion.

2 RESULTS

2.1 PDB Importer and Animator

Starting from our previous work in Maya (6), we wrote a program to read .pdb files and build the molecule in Blender. The .pdb file of interest is fetched and read line by line. Atoms are identified for their nature (Carbon, Oxygen, Nitrogen etc.), their position and the aa to which they belong. This information is elaborated using a library that stores atomic connections for all aminoacids.
Through the interface, shown in Fig. 2, the user can select the .pdb file, which atoms are to be imported (main chain only, main and side chains, or all atoms including hydrogens), what kind of object is to be built (empties, spheres, metaballs), in which order the different conformations are to be imported (the .pdb file has no specific order) and the transition time between different conformations. Note that in the .pdb file every conformation is called MODEL.

Fig. 2

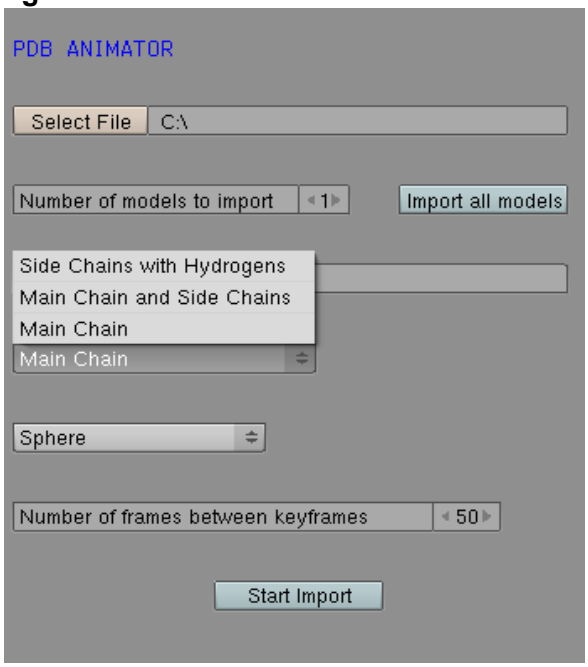

FIGURE 2.
The PDB Importer and Animator user interface.

Atoms are instanced to spheres, the chemical bonds are built as rigid body joints, and a keyframe is assigned to every conformation in the list. The spheres corresponding to different atoms are sized according to the atomic Van der Waals radius (0.1nm=1 Blender Unit) and have a texture, for visualization and rendering, and a spherical collision radius (bounding box) for evaluation of motion.
Once all models of interest are imported, Blender will have an IPO curve for every atom (as a consequence of having keyframes), that interpolates directly between positions at subsequent conformations, as shown in Fig. 3, left. However, these will not consider the joints (that maintain fixed distance between connected atoms) nor collisions. To obtain a trajectory that includes both former features, it is necessary to play the scene with the Game Engine(7). The scene also contains a recorder (8), that records the position of atoms during the game and inserts a key frame to the atomic IPOs for every frame (right panel in Fig 3).
At this point the motion is set for re-playing without further calculations, and we can retrieve the position of all atoms at intermediate frames (as new .pdb files) and use them to evaluate the quality of the structure in physical and chemical terms, using specialized programs. Preliminary results give us encouraging indications.

2.2 Rendering

The actual aspect of molecules beyond the resolution limits of our sight is something that does not exist. Nevertheless, it is possible to represent the space occupied by the atoms that compose the molecule, and to attribute to its surface visual properties to indicate some of the behavioral features of the surface. As mentioned before, at nanometer scale concepts such as color, or brilliance, or roughness, opacity and so one have no meaning; instead we face properties such as pH (acidity, or proton concentration), electric potential, hydropathy, oxidizing or reducing power and others.
One of the most relevant properties that affect molecular behavior is the Electric Potential (EP) determined by their atoms. This feature is so important that almost all protein visualization programs have developed a way to display it, typically as a color code applied to the protein surface.
In an effort to display the behavior associated with EP generated by the molecules, we have performed some steps that permit to import values in Blender, as schematized in Fig. 4.
The .pdb file is first converted to .pqr, that is, values of partial charges are associated to every atom, according to its properties as inferred from the .pdb information. This step is performed once for all conformations attributed to a molecule (i.e., the electric value associated to each atom does not change with its position).
This file is sent to VMD (9), and the module APBS electrostatics (10) is executed. This module solves the Poisson-Boltzmann equation on a discrete grid which extends around the molecular surface (See fig. 5). The surface is saved as a VRML file and the EP calculated in each cell of the regular grid, is

Fig. 3

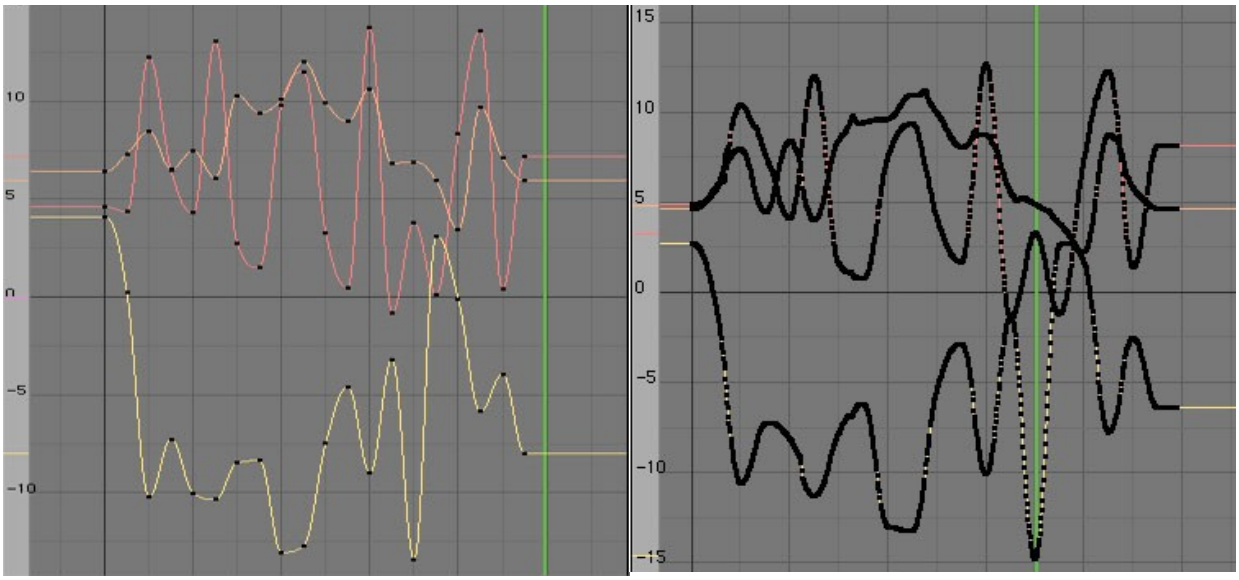

FIGURE 3
One representative IPO curve for a single atom before and after Game Engine evaluation.

saved in a simple ascii file (.dx). The molecule surface mesh is saved as a VRML file and the EP, calculated in each cell of the regular grid, is saved in a simple ascii file (.dx). Theses data are used with a home made program (grid mapper) to map values from the grid to the surface, and are then stored in a new file tha can be easily read by Blender: we chose the texture coordinate channel (the U component, leaving the V for future use), since the values to be stored are signed and can be quite large.
Once the potential data has been read inside blender and mapped on the surface, it can be used for rendering using s shader PyNode.
The visualization of the forces generated by electric charges, and exerted to the surrounding medium (water, which is dipolar itself, some ions and eventually other proteins), is an issue which is still in development. The solution to this visualization problem should include a number of features, including those reported in table 2.
A number of possible solutions (still under study) might include:

- Sparks, from and to surface to indicate positive and negative values.
- Waves on surface
- Water clue: particle system reporting the orientation of water molecules surrounding the protein
- Fur twisting: clock- or counterclockwise for direction (- or +) and length to indicate strength of field
- Surface bumps / holes
- Fog or other volumetric effect around molecule

Table 2

Electrostatic potential visualization
Not associated with other (human) feels
Associated with the behavior of molecules
Deliver positive and negative forces (attractive and repulsive according to the nature of the interacting protein)
Computationally affordable for real time (for navigation of set scene) real time (for visualization of interactions) rendered scene
Aesthetically pleasing
No use of colour (possibly)

Fig. 4

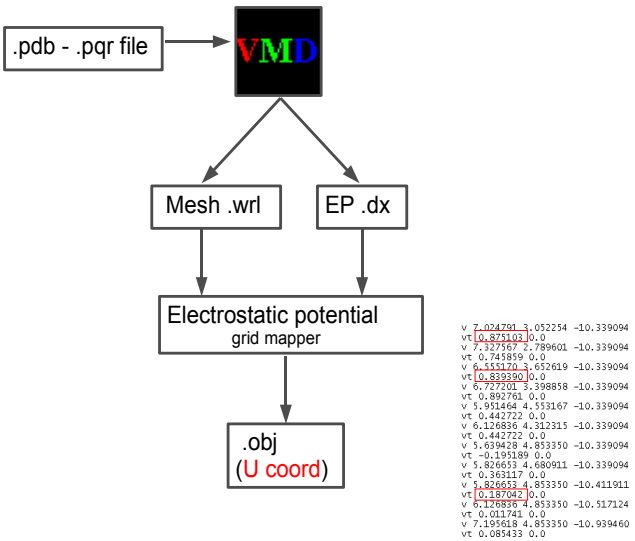

FIGURE 4
Data flow to import Electric Potential values int Blender. A similar flow can be used to import Hydropathy data

Fig. 5

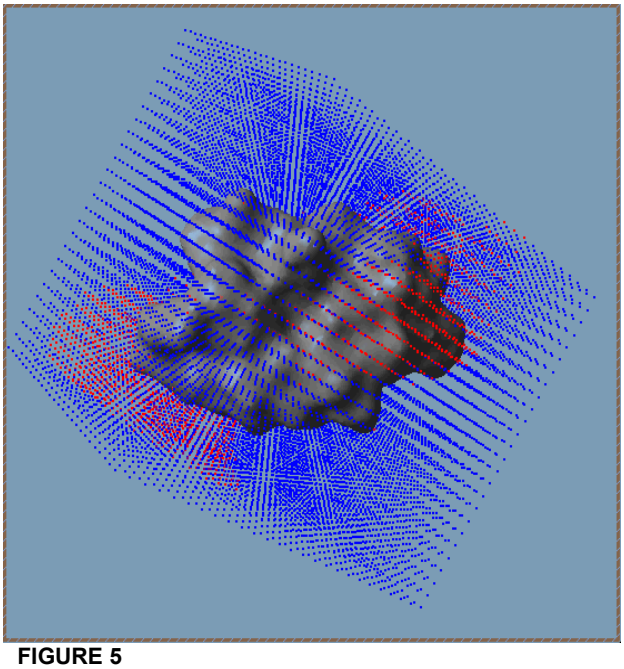

FIGURE 5

Example of visualization of the EP values in a volume surrounding a protein. Values are entered as numbers, and are converted into colors according to the sign of the value (blue positive, red negative).

3 DISCUSSION

The task of showing the workings of minute elements that perform the operations of life implies a number of issues new to both biology and 3D computer graphics. In this report, we detail how we address two of them: import of molecules and representation of some of their features.
Among other problems that still await solution (and which we are studying) are the extraction of

biologically meaningful movement, the conveying of dimensions and timing, and the forces that underlie the process.
With the pdb Importer and Animator, we can represent in the 3D space of Blender the shape and structure of proteins. The importing process includes a method of building the molecule that already imparts to the protein the structural features that set the mechanical constraints associated with motion. Covalently linked atoms are connected by bonds that limit reciprocal motion to rotation along bond axis.
Game engine evaluation of possible motion, intended as transition between two set conformations, results in movements that avoid atomic compenetration. However, whereas this evaluation can be intended as a system that accounts for Van der Waals interactions, other forces are at this time not considered: most notably Hydrogen bonds and electrostatic interactions.
Introducing these forces would result in a more realistic interpolations, but would also slow the computation considerably, making it impossible to use our tool as a fast system for motion evaluation. Hydrogen bonds internal to proteins are typically (with some notable exceptions) very stable, and are in fact among the components that provide stability to secondary and tertiary structures. H bonds on protein surface, on the other hand, are very important for intermolecular interactions and should be taken into account when proteins are considered in their environment. When considering single monomeric proteins, surface residues capable of making H bonds confer a property (called hydrophilicity) that, even if not quantitatively considered, should be represented, in order to convey to the observer the strict relation that these patches establish with water molecules and ions in the medium.
Similarly, electrically charged residues or atoms on the protein surface are very important for the relations that they have with the surrounding water layers and /or with other molecules nearby.
This is the focus of the rendering efforts that we face and of which we presented here some initial steps.
With these and other instruments that we are developing, we intend to offer to biologists, artists and other interested people a glimpse of how the meeting of different disciplines (biology, chemistry, physics, computer sciences and art) can reveal a better understanding of molecular life.
The process to show molecular and cellular biology is only at its beginning, and the contribution of scholars of all disciplines involved, once overcome the initial difficulties in communication, will induce advancement of all sciences, and new knowledge and insights also for the wider public.

Note: All programs and materials described in this report are distributed under GPL license and can be obtained form our website, or by request from Authors.

4 ACKNOWLEDGEMENTS

The work presented here is possible thanks to a grant form Regione Toscana assigned to MZ. YP and RA are supported by a fellowship and a studentship from Scuola Normale Superiore, respectively. CC is PhD student in the department of Mathematics, Univ. of Pisa.

5 REFERENCES

1. www.bender.org
2. Alberts, B, *et al. Molecular Biology of the Cell.*Garland Science, NY, USA.
3. Branden, C. and Tooze, J. *Introduction to Protein Structure.* Garland Publishing, NY, USA.
4. Protein Data Bank website: H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne: The Protein Data Bank. Nucleic Acids Research, 28 pp. 235-242 (2000).
5. Molecule of the Month page on PDB website: http://www.pdb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/alphabetical_list.html.
6. Porozov, Y and Zoppè, M. *Using Maya to visualize and proposing motion of biological molecules*. Manuscipt in preparation.
7. http://download.blender.org/documentation/NaN_docs/BlenderGameReference
8. Recorder provided by Osobná Stránka (http://ash.zendfree.com).
9. Visual Molecular Dynamics. Humphrey, W., Dalke, A. and Schulten, K. *VMD - Visual Molecular Dynamics*, J. Molec. Graphics, 1996, vol. 14, pp. 33-38. http://www.ks.uiuc.edu/Research/vmd/
10. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci USA 98, 10037-10041, 2001