# DNA *in Virtuo*
# Visualization and Exploration of 3D Genomic Structures

J. Hérisson, P.-E. Gros, N. Férey, O. Magneau and R. Gherbi
Bioinformatics team and VENISE action
LIMSI-CNRS
BP 133
F-91403 ORSAY CEDEX (France)
+33 (0)1 69 85 81 64
herisson@limsi.fr

**ABSTRACT**
In this paper, we address the potentialities offered by Virtual Reality and scientific simulation for 3D modeling and immersive visualization of huge genomic sequences. Advanced work on 3D data modeling and structuring is proposed. In Bioinformatics, DNA sequences are often represented within linear format. However, they also have a three-dimensional structure potentially suitable for genomic analysis. The representation of such 3D structure allows biologists to observe and analyze genomes in an interactive way at different levels: from gene to chromosome. We developed a powerful software platform that provides a new point of view for sequences analysis: *ADN-Viewer*. But, a classical eukaryotic chromosome of 40 millions base pairs requires about 6 Gbytes of 3D data. In order to manage these huge masses of data in real-time, we had to design various scene management algorithms and immersive human-computer interaction for *user-friendly* data exploration.

## 1. INTRODUCTION
Most biologists work on textual DNA files (sequence of the four known nucleotides A, C, G and T) that are limited to linear representation. Besides, such linearity offers only local and partial vision of these molecules (each letter in the sequence represents a nucleotide). However, within a cell, DNA is supercoiled as a double-helix, resulting into complex trajectory defined in space. Thanks to 3D conformation models, we are able to construct the 3D trajectory of a DNA molecule given its textual sequence. Some software as DNATools [1] represent the 3D DNA trajectory but they are limited up to several hundreds of nucleotides. In fact, DNA sequence contains generally several million of nucleotides. In this context, the visualization of a whole chromosome requires a lot of graphic and memory resources. Our main aim was to design a powerful and *user-friendly* visualization software tool that renders whole molecules in real-time, taking into account such mass of data.

The 3D information, resulting from modeling construction, could either be visualized or post-processed. The 3D visualization, in contrary to the textual one, offers a global vision of the molecules and is achieved through a software called *ADN-Viewer* [2] developed at LIMSI-CNRS. Nowadays, *ADN-Viewer* can load and visualize multiple sequences of about several tens of million of nucleotides (depending on memory size). *ADN-Viewer* was also interfaced with a genomic database containing all living organisms presently sequenced and annotated. The database gives us additional information about genes names, genes locations and other genetic objects within a chromosome. In addition, *ADN-Viewer* is also available within immersive virtual environment, providing stereoscopic visualization on large

screens and 3D *user-friendly* interface, using powerful navigation, gesture, speech commands... Such multimodal human-computer interface offers intuitive use and is very suitable when processing and managing huge amounts of data, as it is the case of complete chromosomes.

## 2. DNA *IN VIVO*

DNA molecule is composed of four elementary molecules (or nucleotides), represented by four letters: A (*Adenine*), C (*Cytosine*), G (*Guanine*) and T (*Thymine*). But, DNA has a double-helix shape and is formed by two complementary strands (Watson strand and Crick one) [4]: when a nucleotide *A* is on one strand, a *T* nucleotide must be on the other strand, and when a *C* nucleotide is on one strand, a G nucleotide must be on the other one. Two complementary nucleotides are linked by hydrogen bonds and thus form a base pair (bp) or a plate. A textual DNA sequence represents only one of the two strands.

## 3. DNA *IN SILICO*

### 3.1 3D Engine

The 3D engine of our *ADN-Viewer* software takes as input data both textual DNA sequence and 3D conformation model and outputs 3D co-ordinates of each nucleotide. The 3D conformation model (established by bio-physicists [5]) provides, for each di-nucleotide (i.e. each succession of 2 nucleotides in the textual sequence), three angle values and a raise translation. The first plate (base pair of two nucleotides) is placed at the origin of the graphical scene. By scanning the textual sequence in a linear way, *ADN-Viewer* computes the position and orientation of one plate by applying the Equation 1 on the previous plate position and orientation (Figure 1).

$$M_{xy} = T(-\frac{h}{2}) * R_z(\frac{\Omega}{2}) * Q(\sigma, \delta - 90) * R_z(\frac{\Omega}{2}) * T(-\frac{h}{2})$$

**Equation 1. The DNA 3D structure equation, where *h* = 3.39 angstroms, *T* is a translation along *Z*-axis, *R* is a rotation around *X*-axis and $Q(\alpha, \beta) = Rz(-\beta)*Rx(-\alpha)*Rz(\beta)$.**

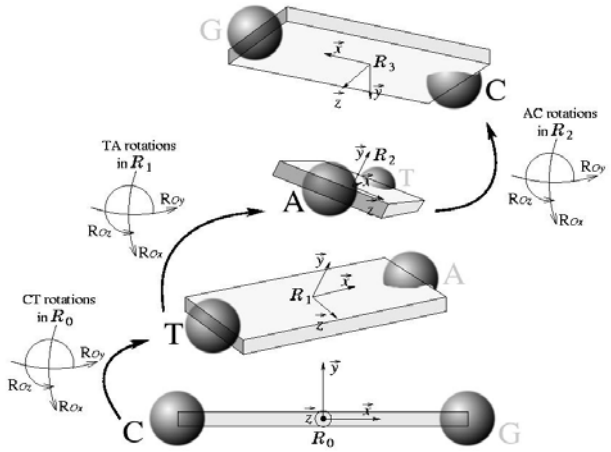


**Figure 1. Construction of the DNA 3D trajectory.**

### 3.2 Data Storage

For performance reasons, we have to store the 3D chromosome trajectory resulting from previous computation. Besides, in order to perform quantitative processing (compactness, curvature…) on these trajectories, we need to know the spatial position and orientation for each plate. We store, for each plate, a 3D point (3 double precision floating points) for position information, and a quaternion [6] (4 double precision floating points) for the orientation one. With this kind of data structure, *ADN-Viewer* can load about *n*/50 Mbp, where *n* is the memory size.

## 4. DNA *IN VIRTUO*

*ADN-Viewer* offers 3 types of 3D DNA sequence modeling. The sequence can be visualized as a whole, partially (for instance a gene) or at very local scale by displaying all atoms and atomic links of the considered nucleotides.

### 4.1 Visualization at Chromosome Scale

As already said, the visualization of whole chromosome offers us a global point of view about the DNA molecule. The user can identify various areas, where some of them show compactness characteristics and others exhibit more relaxed ones (Figure 2).

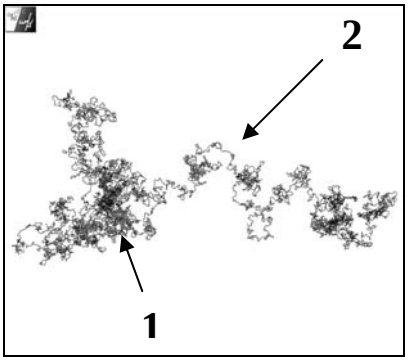


**Figure 2. Genomic visualization of *Saccharomyces cerevisiae chrIII* (~300 Kbp). Area 1 shows compact DNA while area 2 shows relaxed DNA.**

In this genomic modeling, we only display segments linking successive plates. But with a chromosome of 10 Mbp, 10 million segments per frame have to be displayed. In addition, about 15 frames per second (fps) are necessary to have visual fluidity, even if it is well-known that 25 fps rate is better. Consequently, $10.10^6$ Mbp * 25 fps = $250.10^6$ segments per second have to be displayed. Such a performance is not possible to obtain, even with powerful graphic cards. We will now describe how we tackled this problem by proposing a new filtering algorithm.

### 4.2 Visualization at Low Scale

The user has the possibility to select a smaller chromosome part (as gene) in order to observe the double-helix of this part (Figure 3).

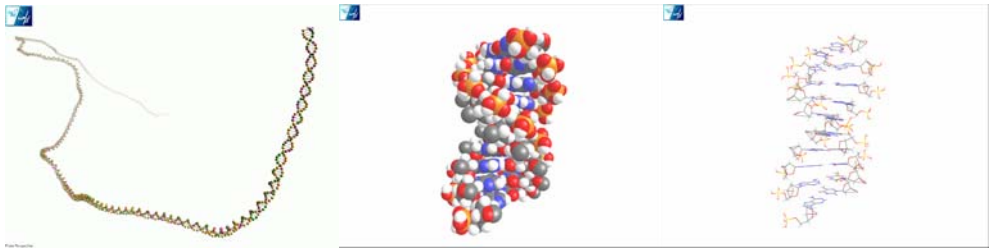

**Figure 3. On the left, a gene (~500 bp) with genic modeling where each nucleotide is represented by a colored sphere. In the center, the full nucleic representation of a gene part (10 bp) where each atom is represented by a colored sphere. On the right, the wired nucleic representation of the previous gene part.**

The non-uniform filtering described in section **Error! Bookmark not defined.** is applied here again to select which plate will be displayed. In addition, fog effect increases also the 3d structure perception of the gene.

## 5. GENOMIC EXPLORATION

The trajectory is the basic information of DNA sequence. However, a chromosome could be modeled by a collection of genes and intergenic areas. Moreover, several biological signals, patterns or motifs constitute the functional content of these chromosomes. We describe in the following how we represent these information by offering the user with genomic content-based exploration.

### 5.1 ADN-Viewer and a Genomic Database

We have developed a genomic database (*GenoMEDIA* [3]) that stores annotated DNA sequences: especially information about DNA sequences and genes positions and names. Thanks to SQL queries, *ADN-Viewer* can download any requested annotated sequences in order to augment the visualization (Figure 4). Such augmented representation is very useful for biologists in order to perform supplementary genomic studies. For example, two distant genes in textual DNA sequence could be closed in 3D space, due to the DNA rolling up.

**Figure 4. 3D Visualization of *Scerevisiae chrMT* (~50 Kbp). Genes are displayed in white and intergenic areas in black.**

The filtering process described in section **Error! Reference source not found.** does not care of genes positions within the chromosome. In particular, important information about gene start and stop could not be displayed. Genes are displayed with different colors than intergenic areas. So the filtering provides a graduation of colors at gene positions, and this could disturb the user when he wants to observe the exact genes positions and orientations in genomic visualization. To overcome this problem, we must force the display of the first and last points of each gene. In addition, with this representation, the user can select any gene by mouse pointing detection mapped in 3D space (Figure 5).

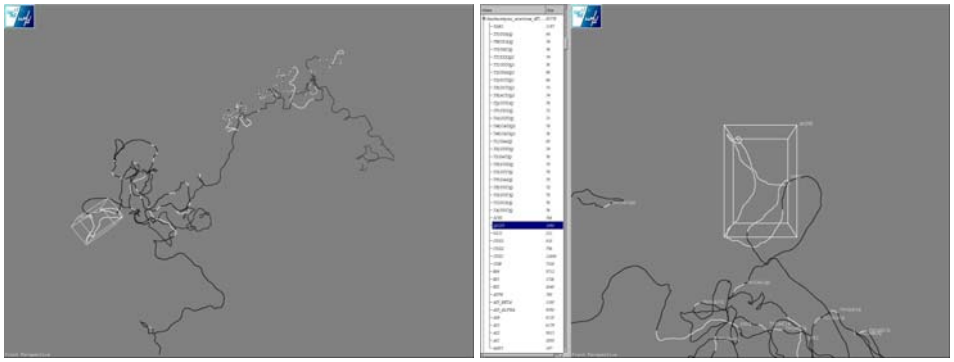

**Figure 5. Gene designation by mouse pointing (left) and Gene selection by mouse click (right).**

Biologists asked also to have the possibility to display several chromosomes at the same time, which is useful to compare all chromosomes between them of one eukaryote's organism (17 for *S. cerevisiae*), or to compare several chromosomes coming from various organisms. So, *ADN-Viewer* offers the visualization of several chromosomes. Each chromosome has its own model-view matrix and thus can be manipulated in an independent manner from others and display additional information. This multi-chromosomes visualization also allows bioinformatics comparison of chromosomes in terms of compactness or other geometrical features. However, this visualization is not exploitable and the genomic analysis becomes unavailable for biologists. Two limitations could be listed: on the one hand, a classical desktop screen has very limited space, and on the other hand, even if the objects are 3D modeled, only 2D projections are perceived on these screens. Stereoscopic perception could help us to overcome such limitations.

## 6. IMMERSIVE EXPLORATION

Of course, we could have some 3D perception when objects move (manipulations). Nevertheless, the main perceptual tool that provides 3D visualization is a stereoscopic one. Human beings perceive obviously 3D real objects, thanks to binocular vision. So, when the user observes 3D objects on 2D screen, the depth perception is appreciably affected. Besides, if the user interacts with objects with classical devices (keyboard and mouse), he cannot manipulate these 3D objects in an easy way. Classical human-machine interaction paradigms are well adapted for desktop environment. However, they are rapidly useless when 3D objects are manipulated in 3D space. The stereoscopy mechanism [10] and large visualization area make it possible to decrease significantly the gap between the virtual space and the real one [14],[15]. For these reasons, *ADN-Viewer* was integrated within the LIMSI Virtual Reality platform (VENISE[7]). This immersive environment platform is equipped with two components: hardware part and software architecture. For hardware one, we use two retro-projected orthogonal screens (2mx2m size) and an Onyx2® [8] as graphic computer. This hardware device is managed by a middleware developed at LIMSI-CNRS: *EVI3D* [9]. This middleware includes two different parts. The first part handles the geometric kernel that manages model-views and projections on screens. The second part (*VEServer*) manages various 3D and immersive devices as DataGlove for gesture recognition, speech recognition system, 3D position and orientation tracking system, immersive mouse… The DNA chromosomes exploration and perception are significantly increased (large screens, stereoscopy, 3D human-machine interfaces, multimodal interaction...) allowing the user to focus completely on his task. Another powerful interaction potentiality is the interfacing of the vehicle paradigm *HCNav* [11] with *ADN-Viewer*. It consists in navigating in the virtual world with free-hands interaction. Thanks to the tracking of head position and orientation (performed by Flock of Birds® [12]

fixed on stereoscopic glasses), the user drives the vehicle in virtual world by his head and body movements. This tracking is also used to adapt dynamically the stereoscopy according to the user's position and orientation. Such navigation offers us two advantages: the user does not need to use the mouse to observe a chromosome any longer, and he can use his hands for other tasks (Figure 6) as gene designation or selection.

**Figure 6. *HCNav* system allows free-hands interaction.**

For other interaction types (grasping, selection...), many interfacing devices could be used, such as speech for commands or DataGlove for manipulating objects. *ADN-Viewer* often uses a 3D pointing device: the wand [13]. It has proved very useful to draw a virtual ray having for origin the wand and for direction the orientation of the device. We can thus easily point or select any object while navigating.

## 7. CONCLUSION AND FUTURE WORK

In this paper, we addressed the potentialities offered by Virtual Reality and scientific simulation for 3D modeling and immersive visualization of huge genomic sequences. Advanced work on 3D data modeling and structuring was proposed. The representation of DNA 3D structure allows biologists to observe and analyze genomes in an interactive way at different levels: from gene to chromosome. We developed a powerful software platform that provides is used for sequences analysis: *ADN-Viewer*. This software handles huge masses of data in real-time, we designed various scene management new algorithms and immersive human-computer interaction for *user-friendly* data exploration.

For the future, a first work will deal with the modeling and visualization of whole DNA sequence annotated content (genes, promoters, enhancers, insertion sites, introns, exons…). The problem is how we could visualize intrinsic 3D information (DNA double-helix) with textual annotations that are not 3D (names, biological origin, function…). In immersive environment, all visualized objects have to be so in 3D in order to avoid biologist's cognitive charge if he must alternate between 2D flat text and 3D objects.

A second work will be to investigate the use a PC-based graphical cluster of PCs to display very huge DNA sequences (several hundreds of millions of base pairs, for plants, animals and human organisms). We will be here in front of some databases and applications distribution. We have to design a smart collaboration of the cluster nodes.

## 8. REFERENCES

[1] Søren Wilken Rasmussen. DNATools. http://www.dnatools.org.

[2] Gherbi, R. and Hérisson, J. *Representation and Processing of Complex DNA Spatial Architecture and its Annotated Content.* In proceedings of the International Pacific Symposium on Biocomputing (PSB 2002), january 2002, Hawaii, USA.

[3] Gros, P.E., Férey, N., Hérisson, J., Gherbi, R. *GenoMEDIA, a Midlleware Platform for Distributed Genomic Information*. In proceedings of IEEE International Conference on Information & Communication Technologies: from Theory to Applications (ICTTA'04), april 2004, Damascus, Syria.

[4] Watson, J.D. and Crick, F.H.C. *MOLECULAR STRUCTURE OF NUCLEIC ACIDS*. Nature, april 1953, vol. 171, p. 737-738.

[5] Shpigelman, E.S. and Trifonov, E.N. and Bolshoy, A. *CURVATURE: software for the analysis of curved DNA*. CABIOS, 1993, vol. 9, num. 4, p. 435-440.

[6] Sir William R. Hamilton. *On a new species of Imaginary Quantities connected with the Theory of QUATERNIONS*. Irish Academy Proceedings, november 1843, vol. 2, p. 424-434.

[7] Transversal Action VENISE. LIMSI-CNRS. http://www.limsi.fr/Recherche/ActionVenise.

[8] Silicon Graphics®. http://www.sgi.com.

[9] D. Touraine, P. Bourdot, Y. Bellik and L. Bolot, *A framework to manage multimodal fusion of events for advanced interactions within virtual environments*. In the proceedings of the workshop on Virtual environments 2002.

[10] Stereoscopy. http://encyclopedia.thefreedictionary.com/stereoscopy.

[11] Bourdot, P. and Touraine, D. *Polyvalent display framework to control virtual navigations by 6DOF tracking*. In the proceedings of the IEEE Virtual Reality, 2002.

[12] Flock of Birds®. Ascension Technology Corporation.

[13] V-WAND™ and WANDA™. Fakespace Systems.

[14] A. Pardi, D. Osguthorpe, S. Klingenstein and M. Dubin. *Iusing Immersive Virtual Reality for Visualization of Macromolecules*. BP Center for Visualization. University of Colorado.

[15] Schkolne, S., Hiroshi Ishii, Peter Schröder. *Tangible + Virtual = A Flexible 3D interface for Spatial Construction Applied to DNA.*