

Molecular Studio for Virtual Protein Lab

Background of the Invention

With the breakthrough advance in human genome project (HGP), Life
5 Science research and education have become ever more crucial. Bio-
molecular study is more challenging than some other fields due to its
complexity of bio-molecular concepts (especially in terms of structure), its
very small scale and high dimensionality of bio-molecular objects, and rapid
technological growth of in the bio-molecular field.

10 Traditionally, bio-molecular education is conducted through classroom
teaching and wet lab experiment. Visual aids of color photos or enlarged
models with material such as wood, plastics, etc, are used to help the
understanding of the structural concepts. While wooden or plastic models
are tangible and relatively cheaper, they demonstrate typical shortcomings
15 of static structures with limited complexity. In contrast, the real life bio-
molecular proteins are in constant change due to folding or interaction.
Furthermore, the number of known proteins through X-ray crystallography
or Nuclear Magnetic Resonance (NMR) imaging is increasing at an
exponential rate.

20 Patent US4812128 describes a modular way to simulate molecules with
plastic modules representing atoms and the bond connections among
atoms.

Patent US5884230 presents a computer system that models three-
dimensional structures of proteins based on the rotation of amino acid planes.
25 It is a trend using multimedia and computer software tools in bio-molecular
education.

Patent US6211868 describes a system to synchronize the educational
multimedia for different purposes of training and education.

Typically, multimedia or software based bio-molecular learning offers (sometimes online) several good features including repeatability of the learning lessons and labs, mass availability of the tools to users, and
5 capability for structural visualization. However, conventional multimedia and educational software often lead to passive learning. Therefore, little or no direct participation by the learners is available and thus hands-on experience like in wet lab experiment cannot be easily obtained.

Wet lab experiment is still a must in bio-molecular study. For instance in Life
10 Science DNA education, several places in the world have licensed the wet lab based "DNA learning lab" product developed by the Cold Spring Harbour Laboratory in New York, USA. Most of the biological wet lab work, however, is costly due to the material consumed and process involved, not to mention the establishment of expensive labs. Wet lab experiment is also
15 not reversible and any mistakes can lead to a fresh re-do. In some cases, bio-molecular experiment can be extremely dangerous. For instances, experimental study on "Coronavirus/SARS" or "HIV/AIDS" is strictly limited to very specialized people.

Computational bio-molecular simulation is increasingly used in molecular
20 biology research. Surely, it may never being able to replace totally the wet lab experiments. However, it has been increasingly recognized that software simulation can play not only a complement but also value-added role to the lab work. For instances, with data sets generated through wet lab experiment, software simulation can look into various possible solutions for
25 problems such as protein secondary structure prediction and drug screening. In fact, a wide range of bio-informatics software is available in the market nowadays. Examples include Insight™ from Accelrys and MolCAD™ from Tripos. Patent US6208942 describes a computer program called molecular hologram QSAR for discovering structure-activity relationships utilizing
30 weighted 2D fingerprints in conjunction with statistical methodology. Patent US6188965 illustrates an apparatus and method for automated protein

design. However, most of the software are designed for bio industrial or research uses thus having little concern for user-friendly interfaces and human-computer interactions for the purpose of educating of the bio-molecular knowledge.

- 5 Therefore a need exists for a solution to the problem of allowing education in bio-molecular sciences while at the same time providing an interesting learning environment.

Patent US1995000513488 illustrates an example of using VR technology such as image processing, 3D graphics and display, force-feedback, and
10 movement and positioning for simulation of a medical procedure. Although there are several patents on three-dimensional technological applications in education (US5707127, for instance), patent searching shows VR application for bio-molecular research and education is rather new.

The present invention provides a solution by design of a Molecular Studio
15 to serve the use in Virtual Protein Lab with various advantages over the prior art.

Summary of the Invention

The present invention relates to virtual reality (VR) technology which solves the above mentioned problem.

- 20 As a high-level interactive computer graphics/visualization technology, VR emphasizes the human-computer integration by employing various natural interfacing techniques. With VR, users can have a few different feelings of participations such as immersion, reach-in, hands-on, walk-through, and fly-over, etc., with the virtual environment. VR-enabled bio-molecular education
25 will be very different from conventional multimedia or software-based learning. As for bio-molecular simulation, VR will offer a new channel for active participation and interactive communication during the process of bio-molecular education and research. Multi-sensorial VR will make it feasible for knowledge exploring and creative learning of bio-molecular structure.

As an emerging technology, VR has been applied in engineering design, medical training and pre-treatment planning, military simulation, education, and entertainment, etc.

In one aspect, the invention may broadly be said to consist in a system for
5 instructing users in bio-molecular structure comprising:

means for modeling a molecular structure in three dimensions;

means for visualizing a molecular structure in three dimensional stereo virtual environment; and

means for interacting virtually with the modelled molecule structure to allow
10 an interactive user to interact with the structure in a VR environment.

Preferably the virtual environment allows six degrees of freedom.

Preferably the molecule is a bio-molecule or a protein molecule.

Preferably the virtual environment may associate the viewer's viewpoint with the protein structures in primary, secondary, tertiary and quaternary
15 levels.

Preferably the protein structure is derived from data held from protein crystallographic and NMR imaging investigations.

Preferably the data is held in a databank and the displayed protein structure is derived from algorithms acting against the data.

20 Preferably the molecule structure is represented in the form of NURBS (Non-uniform Rational B-spline Surface) or tessellation of quadrilateral and triangular meshes for display.

Preferably the molecule structure is visualized in the form of protein polypeptide atoms, protein surfaces, protein bio properties (electrostatic,
25 hydrophobic and hydrophilic), protein-ligand interactions, protein conformations, or a combination of the above.

Preferably the molecule structure is visualized with multiple protein features and properties by actively cross-sectioning of the protein surface and other structure with the section plane dynamically translated and with proteins and ligands dynamically rotated concurrently.

- 5 Preferably the protein structure may be interactively navigated and visualized by users or viewers.

Preferably the interactive navigation of the protein structure may be realized by a user moving on a virtual rider carrier.

- 10 Preferably the protein structure at both molecular and atom levels may be modified by the virtual user or viewer.

Preferably the motion control of the molecular navigation is by user using of a steering wheel and foot pedal and the grasping control of the molecular manipulation is by user using a sensor glove.

- 15 In an alternative aspect, the invention consists in a method of providing virtual interaction with a molecule comprising modeling a molecule in three dimensions, displaying the molecule in a stereo VR environment, and interacting with the molecule by virtual movement, modification, or replacement of protein structures and molecule atoms.

- 20 In a further alternative aspect the invention consists in a method of providing virtual interaction with a molecule comprising modeling a molecule in three dimensions, displaying the molecule in a VR environment, and providing for the virtual reality viewpoint and navigation to follow or not follow the path of the protein secondary structure.

- 25 These and various other features as well as advantages which characterize the present invention will be apparent upon reading of the following detailed description and review of the associated drawings.

Brief Description of the Drawings

Fig. 1 shows the major components of the molecular studio for the reach-in and hands-on use of virtual protein lab.

Fig 2 shows methods of stereo displaying the virtual reality molecular
5 images.

Fig 3 shows methods of navigating the protein structure by riding a virtual transport or carrier.

Fig. 4 shows methods of hands-on learning protein amino acids by a touching and modifying of the virtual amino acid balloons with a sensor
10 glove. Similar methods are used to learn other protein atoms such as carbon alpha using the hands-on glove touching.

Fig. 5 shows methods of reach-in learning the protein conformation by riding of the protein peptide plane with a virtual transport or a carrier.

Fig.6 shows methods of learning the multiple protein features by actively
15 cross-sectioning to allow users to see the relation between different protein structures or properties in an integrated and 3D protein view.

Detailed Description

As shown in Fig. 1 the major components of the current inventive Molecular Studio for the use in a virtual protein lab includes a stereo display unit 101, an
20 interactive control unit 103, and a computational engine 104 communicating with a database 105. The data on the molecule to be displayed is stored in the database, having been derived using standard techniques such as X-ray crystallography and NMR. The data is then displayed using known display techniques, for instance, that described by J Priestle in the Journal
25 of Applied Crystallography 21 (1988) at 572-576 as "Ribbon: a stereo cartoon drawing program for proteins". Improvements with molecular visualization are mainly made in terms of compatible support for various

stereo view functions in active mode, passive mode, auto mode, and anaglyph mode.

A user 102, viewing the display can then, using a VR environment, navigate the viewpoint with the control unit 103 through the protein molecule. The user may need to wear a goggle 106. In the case of active stereo, an emitter 107 is used to control the shutter glasses. The molecule is displayed with an algorithm visualizing the protein structure using ribbons, skins, balls, sticks, electric clouds, hydrophobic and hydrophilic color maps, or a combination of the above. The algorithm is also able to perform active and real time cross-section of the protein when displaying multiple protein features and properties such as in the case of protein-ligand interaction. The calculating algorithm in the engine 104 is fast enough to give a real time presentation of changes in the molecules orientation, and the location of the viewpoint within the molecule.

Note the display screen may be tailored to have a circular shape. In Fig 1, a viewer can see the screen through a cylindrical tube with certain depth and an opening at the top end. This design can mimic the microscopic view and also produce a darker viewing space to reduce the environmental light. The system can also attach to a projection screen 108 to allow many audiences to view the same molecular structure.

Stereo display unit

This unit as described in Fig. 2 consists of a screen 202, one or more projector 201 to emit the separate views of a stereo pair, and one or more goggles 203 to view the stereo images. OpenGL software library is used to support the graphic rendering and stereo viewing. In active stereo case, the goggle is a pair of shutter glasses in which the pair of stereo views are alternately shown by the projector, controlled by the emitter, and viewed alternately by left and right eyes as the synchronized shutters in the glasses alternate. Different options may be available depending on the projection type (rear or front), channel number (single or multiple), and screen type

(flat or curve). In the active solution, at least one projector and one emitter is required to work together with the stereo shutter glasses. In the lowest cost implementation, a 17" CRT monitor is used with a pair of liquid crystal shutter glasses and an emitter.

- 5 In the passive stereo case, the polarizing goggles and projection screens are used without the use of shutter glasses and emitters. In the anaglyph stereo, two color filtering goggles can be used. No goggles are need in the auto stereo view. Another goggle free stereo view needs viewers to be trained with the cross-eye skill so that two pair parallel (or side-by-side)
- 10 displayed images can be fused to form a stereo view.

The display screen as shown at the top of Fig 2 is a high standard rear projection (curved or flat) screen 204. At the second row of Fig 2 is shown a front projection screen with projector 205 and glasses 206. The third portion of Fig 2 shows a normal stereo monitor 207 with shutter glasses 208 and emitter

15 209. It should be understood that while a shutter system is shown other viewing systems such as polarized or bi-color stereo may be utilized. Also the anaglyph stereo can be realized on any normal CRTs, high definition plasma or LCD displays 210 (fourth row), or even projection on walls 212 (fifth row) using a normal projector (213). Only a bi-color (red/blue) filtering goggle 211

20 and 214 is needed with an individual viewer.

A set of Molecular Studios can be network connected to form a virtual protein lab. In this case or any time specially requested, all student screens can use CRT or LCD displays. An additional large size projection system such as a projection screen (rear or front, single channel or multiple

25 channels), a projection TV, a high-definition plasma, a video wall, and so on can be used by an instructor.

However, the virtual protein lab can have all the possible settings with the stereo display pending on the budget available, and the preference and requirement from the customer.

Interactive control unit

This unit is comprised of typical VR interactive devices. While the unit utilizes a standard or wireless keyboard, stylus and normal or wireless 3D mouse actuating buttons for various tasks, users may use a hands-on sensor glove sub-unit or a steering wheel navigation sub-unit. There are several 3D gloves available in the market. The basic functions required, however, are "picking up" graphic entities of the bio-molecular structures and "moving" or "modifying" the entities selected. The glove should also be able to work in a 3D stereo VR environment with enough functional keys (programmable) to support "deleting" and "inserting" bio-molecular entities selected. In a low cost solution, a P5 game glove is used. Navigating within bio-molecular structures such as protein surface is not an easy job. Useful navigation controls include view zooming, view shifting, view rotating, etc. As the protein surface is very complicated like an undiscovered and huge underground cave, lighting control, viewing angle control, direction (forward and backward) control and speed control are also important for navigation such as walk-through and fly-over. For a low cost solution, a steering wheel is used and its functions (programmable) are customized to fit the bio-molecular navigation purpose. The use of a navigation approach such as fly-over, roller-coaster, skiing, driving, skating and so on with a bio-molecular structure requires more visual, audio and speed control. Based on the protein polypeptide structure, the navigation idea is implemented with the consideration for the purpose of learning molecular structure such as protein conformation and protein-ligand interaction in the current inventive system. Users can see and grasp all the protein information such as protein surfaces, electrostatic clouds, amino acids, ligands during navigation along the backbone track or path of the proteins.

Computational engine

The computational unit is responsible for several tasks including (1) bio-molecular model representation: (2) bio-molecular graphic visualization: (3)

interactive bio-molecular operation or manipulation; and (4) VR device interfacing.

5 The main backbone path of the molecule is clearly evident and it is possible to either set a virtual reality engine to follow this path, or leave it to the user to attempt to follow this path, or even completely free move. The molecule appearance is represented by first calculating the backbone path using cubic B-splines or NURBS, and then moving a "frame" along the path to fill in detail to the required level. Resolution of the calculated appearance to triangular meshes allows use of OpenGL display algorithms.

10 The bio-molecular model may be represented in several different levels of information. At geometric level, coordinate values of atoms and radius values of von der Waals forces, etc, are recorded. Both backbone curve geometry and protein surface geometry are organized in this level. LOD (level of detail) is preferably also adopted here to improve the efficiency of graphical visualization and interaction. The data structure for geometric information is designed in a flexible fashion allowing dynamic change under user control (eg, folding simulation). At topological level, connectivity relations among the bio-molecular entities are recorded. Navigation within the bio-molecular or protein surface; and simulation of protein-ligand binding rely on both geometric and topological information of the relevant bio-molecular entities. At the biological level, information can be very comprehensive pending on the applications such as protein secondary structure prediction and active site binding identification.

25 The graphic visualization with the inventive system implements most of the common graphics display modes for bio-molecules and proteins. These include stick mode display, ball mode display, stick-ball mode display, space-filling mode display, wire-frame mode display, surface mode display, skin mode display, and ribbon mode display. For visualization of bio-molecular properties, electrostatics mode display and polarity mode display may also be used with the current system. The graphic visualization with the current inventive system emphasizes real time response and interactive

30

display. LOD technique is incorporated in visualization for optimized graphic display with the supports from both the OpenGL library and especially designed algorithms.

5 Using manipulative aids such as a sensor glove and a sufficient level of magnification in the VR environment it is possible to grasp individual atoms, or amino acids, in the molecule and either manipulate them into a different position within the constraints of the possible structure variations, or actually modify them by changing the locations. Both of these procedures may have cascading changes forced in the molecule which may drastically affect the contour shown in the display, depending on the level of sophistication of the computational engine. The software providing these functions uses standard data glove interface calls interacting with the software defining the three dimensional display of the molecule. In this way the approach of the users virtual hand to a protein amino acid or peptide atom can be modeled to allow modification or movement when "grasped".

Interactive operation or manipulation is one of the key parts of the computational engine. It allows a two-way communication between virtual and real worlds. A user may submit requests to perform operations on virtual entities, the bio-molecular entities in this case. The requests are accepted or rejected by the interactive devices through various sensorial channels and then translated digitally to actions for the virtual world to update. These modifications are then carried out by the computational engine by activating relevant functions or methods for corresponding update. In the current inventive system, the human-computer interaction is carefully designed with special concern for the interactive devices. For instance, in the roller-coaster ride requires that the graphic view change rapidly upon receiving commands from the front end devices such as a steering wheel in the low cost implementation.

Where the user requires a ride or navigation through the molecule the viewpoint may be associated with the main path of the molecule and steering of the user viewpoint set to follow the path. Since the computational engine

establishes the main path it is simple to require the viewpoint to be a specified distance above the path and to control the speed at which the viewpoint moves. Sound effects may be associated with the virtual reality ride experience too.

5 Fig. 3 shows a view of a molecule main path 301 with virtual transports or carriages 302 in place. The carriages are set to track the path at a required speed which can be controlled by the device such as steering wheel. By navigation, reach-in to protein structure becomes feasible which will make further hands-on based learning possible.

10 To have the interactive manipulations working properly, interfacing with VR devices should be addressed. At the initial stage, all device and their drivers are installed if necessary. APIs (application programming interface) of the devices are then used for the purpose of customization of the device functions through programming with DirectX or COM technologies.

15 Database

Information within the database is stored in a standard format, and classified in a standard manner. The primary source of the input data is the protein structure database PDB (Protein Data Bank). Additional information can be obtained from other sources such as SCOP (Standard Classification of
20 Proteins). Other information may come from signal transduction pathway databases, and metabolism pathway databases, and pharmaceutical databases, etc. The unit may be able to communicate with online biological database directly through an Internet connection.

25 Interactive Hands-on Learning of Protein Structure

Fig. 4 shows methods of hands-on learning the protein amino acids 401 or peptide atoms 402 by touching of the virtual amino acid balloons with a virtual hand in a VR sensor glove 403. Similar methods are used to learn other protein atoms such as carbon alpha using the hands-on glove

touching. At least one sensor glove is needed after users who have navigated along the protein path 404 to reach a particular site of a ligand, an amino acid, a peptide carbon alpha atom, etc. With the sensor glove, users can grasp, rotate and manipulate them. By this close hands-on, users
5 can learn protein structure in a 3D and interactive fashion.

Fig. 5 shows methods of reach-in learning the protein conformation by riding on the protein peptide plane 501 with a rider carrier 502. Protein conformation is important concept in bio-molecular education. With reach-in riding, users can reach closely into the peptide plane to see the co-planarity
10 of the peptide atoms. It is also feasible now to use the sensor glove to try the rotation of the peptide plane with respect to their twisting angles 503. By doing so, concept of Ramachandran plot can be easier understood.

Fig.6 shows methods of learning the multiple protein features by active cross-sectioning. This is to allow users to see the relation between different
15 protein structures or properties. To see the protein-ligand interaction, the protein surface and ligand structure should be displayed. Because the multiple structures or protein properties can be very complicated, an active sectioning allows the visualization of the internal structure 604 of the protein and ligands with a cross-section plane 601 cutting them. The cross section
20 plane can be translated along a given direction 602. This is further enhanced by applying a real time rotation 603 of the protein and ligand structure or properties to help the dynamic display of the relational information.

Industrial applicability

25

The invention has potential applications in several sectors including pharmaceutical and Life Science industry for virtual drug screening and interactive drug design. It can also assist in searches for surface features of molecules which may indicate pockets or cavities within the protein surface

for the applications of protein-ligand docking; and the identification of active sites from the electrostatic cloud for protein binding purpose.

It will be appreciated by those skilled in the art that the teachings of the present invention can be applied to other variations of the invention without
5 departing from the scope of the present invention.

What is claimed is:

1. A system for instructing users in molecular structure comprising
means for modeling a molecular structure in three dimensions;
means for visualizing modelled molecular structure in a stereo view; and
- 5 means for interacting virtually with the modelled molecule structure to
allow an interactive user to interact with the structure in a VR
environment
2. A system as claimed in claim 1 wherein the VR environment allows six
degrees of freedom.
- 10 3. A system as claimed in claim 1 wherein the molecule is a bio-molecule or
protein molecule.
4. A system as claimed in claim 2 wherein the VR environment may
associate the viewer's viewpoint with the major path of the protein
molecule and be constrained to follow this path.
- 15 5. A system as claimed in claim 4 wherein the major path is a protein
backbone of the secondary structure.
6. A system as claimed in claim 4 wherein the VR environment may model
virtual moving effects by riding on virtual carrier while following the path
of the molecule.
- 20 7. A system as claimed in claim 2 wherein the VR environment may
associate the viewer's viewpoint with the surface of the protein molecule
without being constrained by the major path of the protein.
8. A system as claimed in claim 7 wherein the VR environment may model
virtual moving effects by free navigating inside or outside the protein
25 surface.

9. A system as claimed in claim 1 wherein the displayed protein structure is derived from data held from protein crystallographic and nuclear magnetic resonance imaging investigations.
10. A system as claimed in claim 9 wherein the data is held in a databank
5 and the displayed protein structure is derived from algorithms acting against the data.
11. A system as claimed in claim 3 wherein the molecule structure is represented in the form of NURBS or tessellation of quadrilateral and triangular meshes for display.
- 10 12. A system as claimed in claim 1 wherein the virtual protein amino acids, peptide planes, or peptide atoms of the molecule may be modified, moved or replaced by the virtual user.
13. A system as claimed in claim 1 wherein the reach-in motion control is by user use of a steering wheel and foot pedal.
- 15 14. A system as claimed in claim 1 wherein the hands-on grasping control is by user use of a sensor glove.
15. A method of providing virtual interaction with a molecule comprising modeling a molecule in three dimensions, displaying the molecule in a VR environment, and interacting with the molecule by virtual
20 modification, movement or replacement of molecule amino acids, and peptide atoms.
16. A method of providing virtual interaction with a molecule comprising modeling a molecule in three dimensions, displaying the molecule in a VR environment, and providing for the virtual reality viewpoint to follow
25 or not follow the path of the molecule.
17. A method providing moving effects wherein following the molecule path is associated with of riding on a carrier.

18. A method providing learning the protein amino acids and peptide atoms by grasping and then modifying of them using a sensor glove after reach-in riding with a virtual carrier closer to them.
19. A method providing learning the protein conformation by reach-in riding
5 of the protein peptide plane with a virtual carrier.
20. A method providing learning the multiple protein features and properties by active cross-sectioning to see the relation between different protein structures or properties.