

DockPro: A VR-Based Tool for Protein-Protein Docking Problem



Serdar Cakici, Selcuk Sumengen, Ugur Sezerman and Selim Balcisoy

Sabancı University

Abstract—Proteins are large molecules that are vital for all living organisms and they are essential components of many industrial products. The process of binding a protein to another is called protein-protein docking. Many automated algorithms have been proposed to find docking configurations that might yield promising protein-protein complexes. However, these automated methods are likely to come up with false positives and have high computational costs. Consequently, Virtual Reality has been used to take advantage of user's experience on the problem; and proposed applications can be further improved. Haptic devices have been used for molecular docking problems; but they are inappropriate for protein-protein docking due to their workspace limitations. Instead of haptic rendering of forces, we provide a novel visual feedback for simulating physicochemical forces of proteins. We propose an interactive 3D application, DockPro, which enables domain experts to come up with dockings of protein-protein couples by using magnetic trackers and gloves in front of a large display.

Index Terms—Medicine and Healthcare, Protein-Protein Docking, Virtual Reality

I. INTRODUCTION

Proteins are organic compounds that are essential for proper functioning of the body as a whole. They take place in every action in the metabolism. Proteins are made of building blocks called amino acids. There are 20 different standard amino acid types. A protein's functions are defined related to which other protein(s) it interacts. One has to understand protein-protein interactions in order to understand all kinds of cellular events. The question of how proteins bind to other proteins is a hot topic since the problem of protein-protein interaction is at the heart of many different industrial products, such as biofuel industry, starch industry, and detergent industry.

Docking is the process in which at least two molecules bind to each other, in a specific position and orientation, and create a molecular complex. Knowing bound configurations of interacting proteins requires protein-protein docking, which enables us to understand:

- How two proteins interact with each other,
- Spatial configurations of possible protein-protein complexes,
- Specific properties of interactions on the surface (i.e. protein interface), where binding takes place.

It is important to have a stable concatenation of proteins in order to have a successful docking. In the process of protein-protein docking, two aspects should be taken into account: Physicochemical properties of proteins, and their shape complementarity. A protein can have different physicochemical characteristics on different surface regions (e.g. one region is attracted by water molecules, while another is not). Shape complementarity should also be considered since proteins have curved surfaces containing large number of cavities and knobs.

Several algorithms have been proposed for protein-protein docking. Fully automated applications are first introduced in early 1990s [9]. Depending on the complexity of the input proteins, docking process can take up to several hours, and may compute false positives.

Virtual Reality (VR) has been used to take advantage of an expert's domain knowledge and experience. Several VR tools on attacking docking problem have been proposed. However, each of them has serious usability drawbacks [6].

We propose an easy-to-use application, DockPro (Fig. 1), addressing the two issues of protein-protein docking: i) Shape complementarity. DockPro employs direct manipulation interaction technique, allowing a biologist to explore possible spatial configurations in real time. ii) Physicochemical properties. We provide a novel visual feedback for simulating physicochemical forces of proteins. Unlike fully automated systems that find relatively successful configurations of protein-protein couples in hours, similar results can be obtained in minutes with DockPro.

Since our application provides the means for figuring out the mechanism of protein-protein docking process, it can be used for i) educational purposes, ii) manufacturing industrial products, and iii) drug design.

There is a large class of proteins that governs important roles in many industrial areas. These proteins are called enzymes. Some areas that make use of enzymes are: Photographic industry, biofuel industry, starch industry, detergent industry. Our application is designed to be used during the course of protein engineering. During a protein design process, the protein's functionality can be evaluated via DockPro.

Drug design relates to ligand (small molecule)-protein docking. Since the main principles of protein-protein docking and ligand-protein docking are the same, our application can also be used for ligand-protein docking.

In section two, an outline of previous algorithms and applications is given. In section three, the importance of taking advantage of expert's knowledge is discussed. In section four,

we present our application, DockPro. In section five, we discuss force representation, and visual feedback, and then conclude our paper.

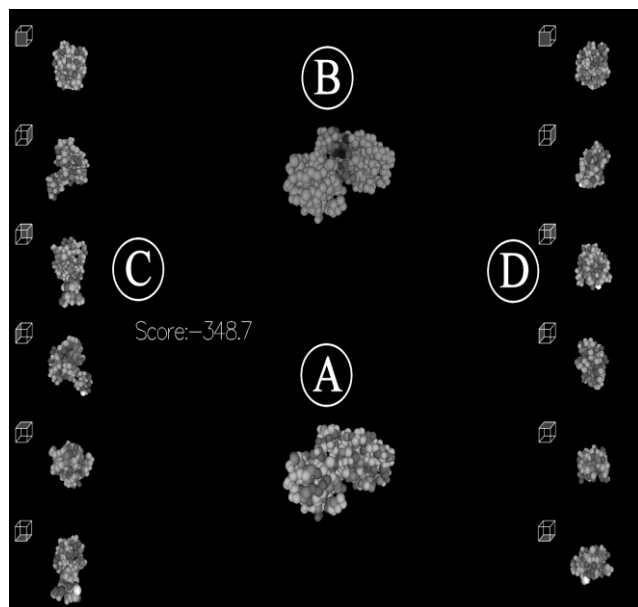


Fig. 1. DockPro's main window.

II. PREVIOUS WORK

There are two competing approaches to attack protein-protein docking problem: i) Automated, ii) VR-based.

Several research groups are working on automated docking, which can be reformulated as an optimization problem [8], [11]. The goal is to minimize binding energy of the molecular complex, however finding the best bound configuration requires checking unlimited spatial combinations of proteins. Consequently, heuristics are used with the goal of finding "good" results. The problem with these approaches, as indicated by [6], is that finding notable solutions is not guaranteed and this process might take up to several hours.

VR approaches have also been used for docking. Beginning from 1990s, several VR techniques have been proposed [1], [2], [10]. More recently, haptic interfaces have been introduced. [13] developed a haptic approach for ligand-protein docking problem. [6] have designed a haptic-based user interface with user needs in mind. Although haptic devices can be used for ligand-protein docking, they are inappropriate for protein-protein docking because of their highly limited workspace.

III. IMPORTANCE OF EXPERT KNOWLEDGE

In specific systems, there are certain general constraints that must be satisfied for a successful docking to happen. These constraints are usually known by molecular biology experts. With their knowledge, search space can be limited and hence a more accurate and faster solution could be expected.

We can give the role of hydrophobicity (water-fearing) as an example. If the expert using the system thinks that hydrophobicity plays a great role for input proteins, she can

come up with a good solution faster if she is given the chance to group amino acids accordingly (e.g. hydrophobic, polar and charged) and then give highly negative scores between the hydrophobic group and other groups. These tasks, among other important aspects, can be accomplished by using our application.

IV. DOCKPRO

We developed an interactive 3D application, DockPro. Humans are good at "put-the-block-into-the-gap" type of problems. A molecular biology expert can come up with a successful docking by changing translation and orientation of the proteins. Chemical and physical characteristics of atoms also play a key role in the docking problem. It is not a sufficient aspect to have good surface complementarity, on its own. In addition, care must be taken to concatenate atoms, which like to be near to each other, to have a stable concatenation.

For the purpose of interaction in protein-protein docking, we use magnetic trackers and gloves. User wears magnetic tracking sensors that are attached to the top of each glove on both hands (Fig. 2). Direct manipulation of proteins is done with hands. The virtual environment is displayed on an immersive workbench. The system provides a natural and easy way to work in front of a large display. It is natural; because at the time of docking, expert uses her hands as if she is trying to concatenate plastic protein models. It is user-friendly; because expert can carry out each necessary step while standing in front of the large display.

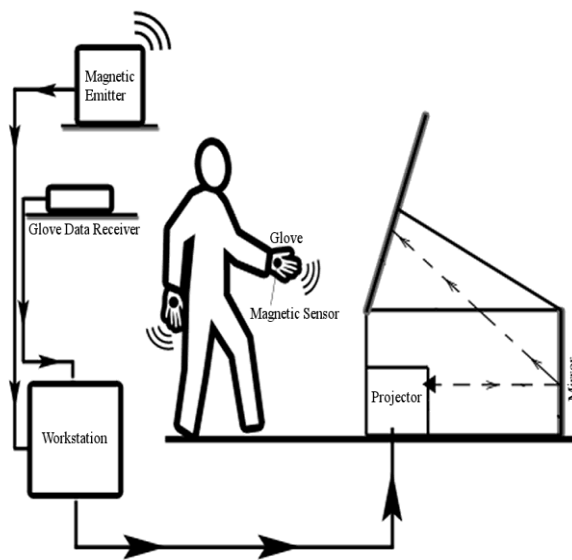


Fig. 2. System overview.

The system runs on Intel Pentium D 3.2 GHz CPU and NVIDIA Quadro FX1500 GPU. We use Flock of Birds magnetic sensor system to track user's hands [7]. To switch between different tasks, we developed a hand gesture recognition scheme on CyberGlove [5], which considers only six basic hand postures. Input files of proteins of interest are provided in PDB file format [3]. From these files, c-alpha atoms

(i.e. alpha carbon atoms) of each protein are extracted to be rendered later on.

DockPro consists of four windows, which are displayed to user one after another. These windows are used in order to:

- a) Group amino acids,
- b) Assign score (force) relations among groups,
- c) Assign colors to groups,
- d) Assist user to perform protein-protein docking.

4.1 Grouping

The grouping window enables the expert to group amino acids in any order. A group contains amino acids that have similar properties. Each group has distinctive characteristics which arise from their amino acids' physicochemical attributes. Every amino acid should be assigned exactly to one group.

In this window, user begins by choosing amino acids of the first group to be assigned. In order to accomplish this task, she moves the cursor to the related amino acids' icons by moving her right hand onto which a tracker is attached. Then, she chooses the amino acid (on which the cursor stands) by a hand gesture that is captured via the glove. Undoing is possible; she can remove any amino acid from the group that amino acid was previously assigned. Assignment of amino acids to the first group ends when a unique hand gesture is made.

These steps are done for other groups yet to come, if there is any. Expert can see the groups created so far at the bottom of the window (i.e. related legends are present). Groups are numbered, in the order they are created, from 0 to n-1, where n is the total number of groups. When there is no amino acid left that is uncovered, it means that we are done with the grouping phase and hence we move into the second window.

4.2 Score Relations

In the score relations window, we are able to determine the scores (i.e. physicochemical forces) between any pair of groups, which we have created in the grouping window. Each score between two groups shows how much amino acids in one group like to be around amino acids in the other group. Using these scores, we try to mimic physicochemical forces that play key roles in docking. Assigning score relations between amino acid groups constitutes binding energy functions, which determine the binding strength of the complex. Higher overall score means higher binding strength (i.e. better docking). Initially, all possible pairs have a score value of 0. User can assign, and also reassign, scores in any order. User is not obliged to appoint a score to each relation; untouched relations will remain to have values of 0. There are icons of group numbers and a scale bar. There are also informative legends about members of each group and scores assigned.

4.3 Color Assignment

In the color assignment window, unique colors are assigned to groups. These colors are set to be used in the docking window to mark which amino acid group each atom belongs to. In this window, we have icons of group numbers and also 20 colors. User is expected to match each group with a color.

4.4 Docking

In figure 1, docking window of DockPro, where main actions take place, can be seen.

In figure 1.a, DockPro's regular docking view is shown. There are three view modes. In the default view, both proteins are fully visible, and they can be docked as anticipated. Each atom takes up space relative to the Van Der Waals volume of the amino acid it belongs to. Manipulations on translations and rotations are nonisomorphic: Both of them are scaled down. As well as the primary aspect of providing shape complementarity, the overall score of the system also plays an important role in the docking process. The score of the system is shown on the corner. Our score relations are dynamic and computed in real time: Their magnitudes change inversely with the distance between atoms. Expert can understand how good a docking is by checking the overall score. When any two atoms of proteins collide, those atoms are highlighted: They blink at a constant frequency. Colliding proteins cannot penetrate into each other. Instead, user should adjust proteins' orientation/translation to relax the collision.

Apart from the default view, we have one general view mode for visualizing the protein complex, and one for fine-tuning. The former enables expert to see previous dockings and examine structures by treating them as a single entity rather than two different proteins. This view has two options. When a collision occurs, one protein's collision surface is drawn semi-transparent. Hence, expert can see totally opaque protein through semi-transparent amino acids of the other protein. This enables an in-depth analysis of the collision surface.

We have one more view mode for fine-tuning. In this mode, expert chooses a previous docking's collision surface and tries to increase the score continuing from this configuration. Both manipulations (translating and rotating) are nonisomorphic like in the default view; but they are scaled down further.

For each view mode and its sub modes, a symbolic hand gesture is assigned. By this way, user can switch from one view to another without any interruption. In addition, expert can also halt the system by a hand gesture. In our application, all gestures are done via left-hand glove.

When user finds a configuration noteworthy, she can save the data about resulting complex by a gesture. This data contains:

- a) Atom coordinates of both proteins,
- b) Score of the complex.

If there is data about the complex in a protein-protein docking benchmark (e.g. [4]), user can compare her docking with the benchmark and check how successful her docking is.

In 1.b, we can see the inside story of 1.a. This part differs from 1.a in that we are able to visualize the force relations that contribute to the total score. Here, no group coloring is present. Every atom has the same initial color: gray. Gray is the neutral color in our scale, meaning that the net force on an atom is zero. When net force on an atom changes, that atom's color changes too. Color of an atom signifies the charge and magnitude of force that is exerted. There are two endpoints in this color scale. One end is blue, and the other end is red. Blue color corresponds minimum score value, and red color corresponds to maximum

score value out of all relations. Depending on the magnitude, intensity of color (i.e. red or blue) changes accordingly. Current sub screen enables us to see how the overall score is constructed. Having the aim of maximizing the overall score in mind, user gets auxiliary information to come up with a successful docking. So it can be proposed that force feedback by color coding helps user by reducing the search space further.

Part 1.c and 1.d are dynamic legends to help user see the colliding atoms (if any) with the aim of making the docking process easier. In 1.c, there are six views of the protein that is controlled by user's left-hand. Each view plane is orthogonal to each other: Imagine that this protein is surrounded by a transparent cube. Each view is taken from one of the six sides of a cube. Auxiliary information about the view is provided by legends of imaginary cube: It shows from which side we are looking at our protein. Part 1.d is the dynamic legend for the other protein. At the time of docking, any colliding atom pairs are highlighted at part 1.c and 1.d as in 1.a. Consequently, we are able to see each possible collision which may not be seen from 1.a's default view. These legends, along with part a's collision surface view, enhance user's understanding of the process.

V. DISCUSSION AND CONCLUSION

5.1 Force Feedback

1) *Haptics*: Haptic devices have been used for ligand-protein docking problem [12], [13] to simulate electrostatic potential energy that plays an important role during the process. In haptic applications of ligand-protein docking problem, user moves the ligand with a haptic device in 3D space, and tries to position it on the protein. The force on the ligand is calculated and rendered to the haptic device as if it is emitted from one point. However, every atom of the ligand has a force interaction with every atom of the protein and hence each atom of the ligand has its own force. Since ligands are very small molecules, this approach does not give rise to any drastic errors.

While such aggregation is tolerable in the case of ligand-protein docking problem, it is inapplicable to protein-protein docking problem. It is not sufficient to calculate the total force between two proteins since there can be several amino acid groups that we should give feedback on their force relations. Consequently in this case force aggregation for the whole protein is not possible. Hence, it is better to calculate all force relations between each atom of two proteins, and render them in a visually appealing way.

2) *DockPro Visual Feedback*: Since we cannot use haptics for protein-protein docking problem, we propose a possible solution.

$$F_{p_i^1} = \sum_{j=1}^n f(p_j^2, p_i^1) \quad (1)$$

Total force exerted on i^{th} atom of protein p^1 by atoms in protein p^2 is shown in (1). Expert can define and adjust the scoring function f .

To provide visual feedback, we use the method of color coding. Colors stand for the magnitude of forces. Color of a given surface indicates the magnitude of total force exerted on that surface.

User can assign scoring functions (physicochemical forces) between each group. In DockPro, user creates her own intergroup score table (i.e. by creating groups and assigning scores accordingly). Usage of expert's own knowledge enables her to favor interactions observed between the examined proteins.

5.2 Workspace

1) *Haptics*: An important problem with haptic devices is their highly restricted workspace due to hardware constraints. In ligand-protein docking problem, there is a large molecule (protein) and a relatively small one (ligand). For this reason, one must "zoom" into the area of interest on the protein by a factor that is enough to fit the ligand easily. [13] developed a technique called Active Haptic Workspace, which enables zooming and panning. However, this method is inapplicable to protein-protein docking since both are large molecules requiring large workspace.

2) *DockPro Environment*: In our application, we are using magnetic trackers rather than haptic devices, and our workspace is only limited with the reach of our arms if we stand still. Moreover, the user can move back and forth to zoom in or zoom out further. Hence, our system does not suffer from workspace limitations on protein-protein docking.

3) *Conclusion*: In this paper, we have presented an interactive protein-protein docking application, DockPro. With its use of magnetic trackers and gloves, and use of large display, it provides an easy-to-use system. The application addresses several aspects of protein-protein docking:

a) It enables an expert to create her own combination of amino acid groups. In addition, expert can define force relations among groups. Unlike common practices of force calculations, this method allows problem specific force adjustments.

b) At the time of a collision between two proteins, the collision surface is extracted and one of the proteins is rendered semitransparent. This helps the user to analyze the collision area, which would most probably be occluded and unable to be seen due to the dense formation of atoms.

c) We propose a force aggregation scheme and render its results color coded on both molecules. This provides an efficient force representation alternative to haptics.

d) For each protein, there are six different orthogonal views which helps user see colliding atoms with the aim of making the docking process easier. They are dynamic, and we are able to see each possible collision which may not be seen from the default view. These views enhance user's understanding of the process along with collision surface view.

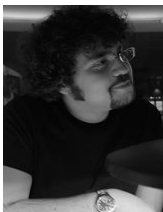
As a future work, we will focus on perceptual issues to improve the usability. Currently, we are integrating this system into a graduate level bioinformatics course at Sabanci University.

VI. ACKNOWLEDGMENTS

This project is partially supported by TUBITAK project 106E213.

REFERENCES

- [1] N. Akkiraju, H. Edelsbrunner, P. Fu and J. Qian. Viewing geometric protein structures from inside a CAVE. *IEEE Comput. Graphics Applications* 16, 58–61, 1996.
- [2] A. Anderson and Z. Weng. VRDD: Applying virtual reality visualization to protein docking and design. *Journal of Molecular Graphics and Modelling* 17, 3, 180–186, 1999.
- [3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne. The protein data bank. *Nucleic Acids Research* 28, 235–242, 2000.
- [4] R. Chen, J. Mintseris, J. Janin and Z. Weng. A protein-protein docking benchmark. *Proteins* 52, 88–91, 2003.
- [5] Cyberglove. <http://www.cyberglovesystems.com/>.
- [6] N. Ferey, G. Bouyer, C. Martin, P. Bourdot, J. Nelson and J. M. Burkhardt. User needs analysis to design a 3d multimodal protein-docking interface. *IEEE Symposium on 3D User Interfaces*, 125–132, 2008.
- [7] Flock of Birds. <http://www.ascension-tech.com/realtime/RTflockofBIRDS.php>.
- [8] T. N. Hart and R. J. Read. A multiple-start Monte Carlo docking method. *Proteins: Structure, Function, and Genetics* 13, 3, 206–222, 2004.
- [9] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo and I. A. Vakser. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proceedings of the National Academy of Sciences* 89, 2195–2199, 1992.
- [10] D. Levine, M. Facello, P. Hallstrom, G. Reeder, B. Walenz and F. Stevens. Stalk: An interactive system for virtual molecular docking. *Proceedings of IEEE Computational Science and Engineering* 4, 2, 55–65, 1997.
- [11] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* 19, 1639–1662, 1999.
- [12] H. Nagata, H. Mizushima and H. Tanaka. Concept and prototype of protein-ligand docking simulator with force feedback technology. *Bioinformatics* 18, 140–146, 2001.
- [13] E. Subasi and C. Basdogan. A new haptic interaction and visualization approach for rigid molecular docking in virtual environments. *Presence: Teleoperators and Virtual Environments, MIT Press* 17, 73–90, 2008.



Serdar Cakici completed his BSc degree with scholarship at Sabanci University, Istanbul, Turkey in 2007. He is a teaching assistant and MSc candidate with scholarship at Computer Graphics Laboratory at the same university. He is expected to complete his MSc degree in June 2009.

His current research interests are Bioinformatics and Virtual Reality in a broad sense. As his MSc thesis project, he is working on developing novel algorithms and applications related to

protein-protein docking problem and drug design.



Selcuk Sumengen is a Ph.D. candidate, working as a research and teaching assistant at the Computer Graphics Laboratory of Sabanci University (Istanbul/Turkey). He received his B.S. and M.S. (with TUBITAK fellowship) in computer science and engineering from Sabanci University in 2004, and in 2006.

His main research interests are mostly in the area of computer graphics, particularly physically based simulation, multi-resolution modelling and rendering. Recently, he has been working on parallel and distributed computing which might have very promising applications in the fields of computer graphics.



Osman Ugur Sezerman graduated from Bogazici University, Istanbul, (B.Sc. Elect. Eng. 1985, M.Sc. Biomedical Eng. 1987) and received a Ph.D. in Biomedical Engineering (1993) from Boston University, MA, USA. Previously he worked at Boston University and Bogazici University as a researcher and an instructor. He has been at Sabanci University Biological Sciences and Bioengineering Program since 1999. He has established the

Computational Biology Laboratory at Sabanci University. His current research interests are molecular modeling, synthetic vaccine and drug design, protein engineering, DNA chips, functional genomics, systems biology and biomarker detection and biosensors. He has over 50 peer reviewed conference and journal publications.



Selim Balcisoy obtained his B.S. in Electrical Engineering from Swiss Federal Institute of Technology, Zurich (ETH) in 1996. He received his PhD on Computer Science in 2001 from Swiss Federal Institute of Technology, Lausanne (EPFL). Between 2001 and 2004 he was Senior Research Engineer at Nokia Research Center USA, where he conducted research on mobile graphics. His research interests include Augmented Reality,

Virtual Environments, Cultural Heritage and Mobile Graphics. Dr. Balcisoy (co)authored over 30 publications at refereed international journals and conferences, and has been granted with one U.S. patent.