Rigid Molecular Docking in Virtual Environments with Haptic Feedback

by

Erk Subaşı

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Computational Science and Engineering

Koç University
June 2006

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Erk Subaşı

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee has been made.

Committee Members:

_____

Çağatay Başdoğan, Ph. D. (Advisor)

_____

Burak Erman, Ph. D.

_____

Attila Gürsoy, Ph. D.

_____

Özlem Keskin, Ph. D.

Date _____

ABSTRACT

In this thesis, we present computationally efficient methods for visualization and simulation of molecular interactions in virtual environments with haptic feedback. In our simulations, the haptic device is used to guide a rigid ligand molecule into a receptor site while the molecular forces acting on the ligand molecule are scaled and reflected to the user in real-time. We demonstrate that the presence of a haptic interface accelerates the binding process and reduce the binding errors if it is used as a precursor to estimate the initial configuration of the ligand molecule at the binding site. After placement and rough alignment of the ligand molecule inside the binding cavity with the help of a haptic device, we use a novel computational approach to determine the final binding configuration of the ligand molecule. In this approach, the ligand molecule is treated as a rigid body seeking for the lowest potential energy configuration. The rigid body configurations of the ligand molecule is calculated in a least square sense using the new positions of its atoms moving under the influence of molecular interaction forces. The proposed approach is computationally more efficient than the molecular simulation methods that utilize the standard rigid-body formulations. We also present new methods for haptic visualization of a protein surface interactively to search for potential binding sites. Our experimental studies with 6 subjects show that subjects can successfully identify the true binding site among the 5 potential binding sites using visual and haptic cues. In addition, we show that the proposed distance minimization approach can be used to find the final configuration of a ligand molecule inside the binding cavity after it is initially aligned by the subject.

ÖZET

Bu tez çalışmasında, moleküler etkileşimlerin sanal ortamlarda haptik geri-beslemeli olarak gösterilip simülasyonunun yapılması amacıyla verimli hesaplama yöntemleri geliştirilmiştir. Simülasyonlarımızda, haptik cihaz kılavuzluğunda katı ilaç molekülü protein yüzeyindeki kenetlenme bölgesine götürülür, ilaç molekülü üzerinde oluşan kuvvetler hesaplanıp ölçeklenerek kullanıcıya gerçek zamanlı olarak sanal ortamlarda yansıtılır. Bu çalışmada, haptik arayüzün, bağlanma bölgesindeki ilaç molekülünün başlangıç konumunun tespitinde bir ön aşama olarak kullanıldığında, bağlanma işlemini hızlandırdığını ve bağlanma hatalarını azalttığını gösteriyoruz. İlaç molekülünün haptik cihaz yardımıyla kenetlenme bölgesine yerleşimi ve kabaca hizalanmasını takiben, ilacın bağlantı bölgesindeki son konumunu yeni bir hesaplama yöntemi ile buluyoruz. Bu yaklaşımda, ilaç molekülünü en düşük potansiyel enerji seviyesini arayan katı bir cisim gibi görüyoruz. İlaç molekülünün katı cisim konumlarının hesaplanması moleküler etkileşim kuvvetlerinin etkisi altında kalarak hareket eden atomlarının yeni pozisyonları kullanılarak en küçük kareler yaklaşımıyla yapılmıştır. Önerilen yaklaşım standart katı cisim formülasyonlarını kullanan moleküler simülasyon yöntemlerine göre hesaplamalı açıdan daha verimlidir. Ek olarak, protein yüzeylerinin etkileşimli olarak görselleştirilmesi ve olası bağlanma bölgelerinin tespiti amacıyla yeni bir haptik görselleştirme yöntemi önerilmiştir. 6 denekle yürüttüğümüz deneysel çalışmalarımız göstermiştir ki, görsel ve haptik öğelerle beslenen denekler gerçek bağlanma bölgesini yanlış bölgelerden başarı ile ayırt etmişlerdir. Bunun yanında, önerdiğimiz hesaplamalı yöntemi kullanarak ilaç molekülünün bağlanma bölgesinde ilk hizalanmasını takiben molekülün en son konumu da başarıyla bulmuşlardır.

ACKNOWLEDGEMENTS

I would like to thank, first and foremost, my advisor, Assist. Prof. Dr. Çağatay Başdoğan for his guidance, suggestions and continuous support throughout my graduate study and completion of this thesis. I feel fortunate of being one of his research students throughout my study, which provided me the chance of being instructed by an insightful and expert academician and continuously supported also in my extra curricular life by a generous person.

Also, I would like to thank all the faculty and students in Center for Computational Biology and Bioinformatics in Koç University, for their passion to help and share knowledge. Special thanks go to Aydın, İbrahim, İhsan, Mert, Nesra and Yaşar, for their help in my thesis work.

Last, I am very grateful to my family, especially to that brave woman, my mother; and my special friends for their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Biomolecular interactions are the core of the entire regulator and the metabolic processes that together constitute the process of life. The application of computational methods to study the formation of intermolecular complexes has been the subject of intensive research during the last decade. With increasing processing power and growth in the known molecular structures database automated prediction of molecular interactions became the key to rational drug design. One of the most crucial ones of these interactions, molecular docking, is defined as the process by which two molecules interact with each other and bind in an orientation and position determined by their geometric shape and local physicochemical properties. The geometric shapes of molecules define how well the binding surfaces complement each other while the physicochemical properties define how well the binding strength of the interaction energies holds molecules together. Hence, the goal in molecular docking studies is to determine whether two molecules interact and if so, the position and orientation of molecules such that their binding surfaces are maximized while the total energy of the complex is minimized.

Molecular docking is a key to understanding cellular events and has applications to drug design. It is known that proteins involved in cellular signaling, gene regulation, metabolism, immunity, and many other vital processes perform their functions by binding to some substrate molecules. If we can determine which molecule(s) binds to a particular protein and how the protein interacts with the bonded molecule(s), we can possibly enhance or inhibit its activities. This information, in turn, can be used to develop new drugs that are more effective against diseases. In trying to find molecular compounds that combat diseases,

researchers have concentrated their efforts on two fronts: one involves experimental high-throughput screening of large libraries of compounds and the other involves virtual screening using computational approaches. The experimental approach is known to be labor-intensive and expensive and it is believed that the ability to computationally screen potential interactions between molecules can speed up the development of new drugs, while reducing development costs and improving efficiency.

Developing efficient computational methods to determine if and how a molecule binds to another one is a challenging task. Even if we neglect the physicochemical interactions between two molecular structures and only consider their geometric features for binding complimentary, the large search space already poses a significant challenge. Given that these structures are not completely rigid, especially when binding, the problem becomes even more complicated. In fact, this flexibility of the molecules results in hundreds to thousands of degrees of freedom and a huge number of possible binding conformations [1]. For this reason, most of the approaches to molecular docking have been limited to “rigid” docking to reduce the number of computations. In rigid docking, the binding molecules are considered as rigid objects that cannot change their spatial shape before and during the docking process. Furthermore, the majority of docking algorithms make the simplifying assumption that one of the molecules, typically the larger size molecule, does not move at all. Since the binding molecule, typically the smaller size molecule, can make 3 translational and 3 rotational movements in free space only, this assumption limits the search space to 6-dimensional configuration space.

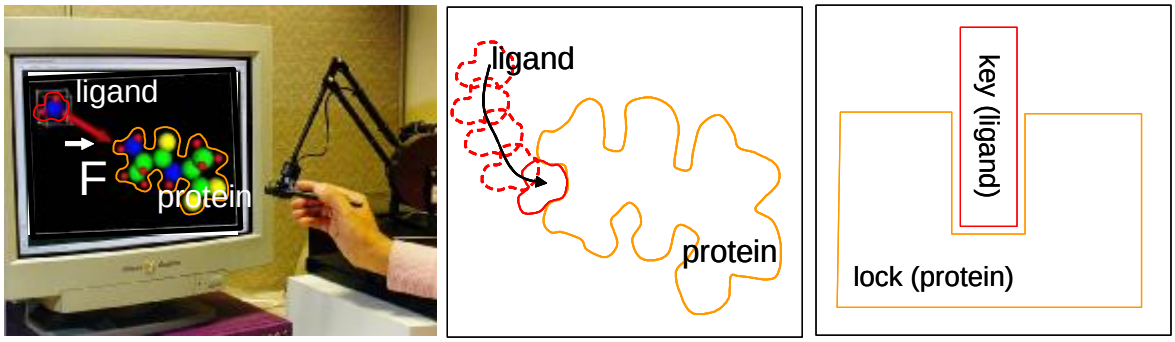


Figure 1.1: Our virtual reality set-up for simulating molecular docking in 3D environments with visual and haptic feedback to the user. The analogy of a lock and key is used to describe the process of molecular docking: the small drug molecule (key) must have a particular shape and physicochemical properties to bind an active site on the protein surface (lock).

In this thesis, computationally efficient new methods are proposed to contribute to the solution of the ligand-protein binding problem with the utilization of haptic devices. In our approach, the user explores the surface of the protein molecule using the haptic device to find the true binding site. The ligand molecule is then guided and inserted into the binding cavity with the help of force feedback (See Figure 1.1). Feeling the molecular interaction forces enables the user position and align the ligand molecule inside the binding cavity better. For the visualization of protein surfaces to search for potential binding sites and in order to simulate the molecular force interactions during the ligand insertion in real-time, we have developed efficient haptic visualization and rendering techniques. In particular, we propose a new concept, named as "*Active Haptic Workspace*" (AHW), for the haptic visualization of very large scale virtual objects. In the case of molecular docking, the protein surface has many irregularities (knobs and cavities), mandating large scale factors to be used in visualization and docking to capture the details. Since the workspace of a haptic device is typically limited by its physical dimensions, the visible part of the protein molecule in haptic workspace is updated in our approach, giving the impression to the user as if he/she is moving the haptic workspace actively over the surface of object. Following the initial

estimation of the ligand configuration with the help of visual and haptic cues, molecular dynamics simulations are performed using our new distance-minimizing algorithm to compute its final binding configuration in time-stepping iterations. The conventional approaches to perform the same task without a good initial guess typically suffers from being computationally too-intensive or from getting trapped in local minima more often [20]. In our approach, we calculate the intermediate rigid-body configurations of the ligand molecule at each time step of the molecular dynamics simulations using a new algorithm that minimizes the positional error between two sequential configurations in a least square sense. The proposed method is computationally more efficient than the algorithms employing conventional rigid-body equations for rigid molecular docking problems [18, 24].

The thesis is organized as follows; a literature review on protein-ligand docking problem and virtual reality applications governing it is given in Chapter 2. In Chapter 3 the forces involved during the protein-ligand interactions are discussed as well as the basic concepts on molecular dynamics are introduced. Our approach to rigid docking is discussed also in this chapter. Methods for rendering haptic interactions between the ligand and protein molecules and the haptic visualization of large protein surfaces are given in Chapter 4. Chapter 5 presents the design details and the results of our molecular docking experiments conducted with 6 human subjects. Finally, the last chapter concludes the study with a discussion of the performed work and the future research studies.

Chapter 2

LITERATURE REVIEW

The ligand-protein docking involves a large molecule (the protein - also called the 'receptor') and a small molecule (the ligand) while the protein-protein docking involves two proteins that are approximately the same size. Computational approaches developed to solve protein-ligand and protein-protein binding problems typically involve multiple stages, but different strategies in each stage [2 - 8]. Most algorithms execute a low resolution geometric search and then a high resolution refinement stage [5]. For example, if we consider the rigid ligand-protein binding problem, one begins by searching the knobs and cavities on the surface of the rigid protein molecule to identify potential candidate sites mostly based on shape complimentary. Then, molecular dynamics is used to bring the ligand and receptor molecules together in energetically the most favorable configuration. Success of the most docking algorithms depends on several factors including the search and optimization strategies, models of molecular dynamics and parameters, and scoring functions defined to find the optimum configuration. Cole et al [9] argue that comparing these docking algorithms is highly difficult and it is easy to produce results that are misleading or not generally applicable.

More recently, virtual reality techniques and devices have been applied to molecular docking [10 - 13] to bring the human operator into the loop for achieving better results [14]. The goal is to integrate human motor, perceptual, and cognitive abilities into the interactive molecular visualization and simulation environments for multi-modal exploration of the search space and interactive molecular docking. For example, it is now possible to explore and manipulate molecular models in virtual environments with the use of high fidelity haptic

devices and rendering algorithms. The first work in this area dates back to early 1970's. Brooks and his colleagues, in a project called GROPE, have demonstrated the simulation of molecular interactions with force feedback in virtual environments for the first time [15]. Despite the early discovery of the benefits of haptics in molecular simulation, significant progress has not been made until recently due to the lack of commercial haptic devices and rendering software libraries. Yet, there are still only a limited number of publications in this area, which are briefly reviewed below.

The use of haptics in molecular biology applications can be grouped into three. The haptic devices are used for a) teaching structural molecular biology and b) feeling interaction forces during molecular visualization, docking and interactive particle steering in virtual environments, and for c) interactive physical manipulation of actual molecular structures in real world. Sankaranarayanan et al. [16] have developed an augmented reality system supported by haptic feedback for teaching structural molecular biology to students [16]. In this system, graphical virtual models are superimposed on the physical models such that the students interact with these virtual enhancements through a haptic device while manipulating the physical model. Birmanns and Wriggers [12] utilize a haptic device for interactive registration of low-resolution electron microscopy data with high-resolution molecular structures. They use the gradient of a cross correlation function to calculate the interaction forces and torques which are conveyed to the operator via a haptic device for guidance in finding the optimum fit. To achieve real-time and stable haptic rendering rates (in the order of 1 kHz), they utilize vector quantization techniques. Lee and Lyons [17] present a new method for smooth rendering of haptic interaction forces generated by Lennard-Jones (LJ) potential field during the simulation of ligand-protein docking. When LJ potential is used and the ligand atoms are in close proximity of the receptor atoms, the magnitude of the interaction forces increase quickly, exceeding the limits of the haptic device and leading to force instabilities. In rendering these forces, Lee and Lyons (2003) keep the gradient of the potential field unaltered when the distance between the atoms of ligand and protein

molecules is greater than the sum of their van der Waals radii, but render a simple hard-surface wall when they are smaller. This approach eliminates the force instabilities even in the presence of strong force gradients. Bayazit et al. [22] present a new framework for the solution of ligand-protein binding problem based on the path planning approaches used in robotics. They use a probabilistic motion planning algorithm and investigate the effects of supplementary user input collected via a haptic device in identifying the low energy configurations. In their approach, the user manipulates the rigid ligand around a fixed protein molecule in a virtual environment and samples the configuration space for low-energy configurations using the haptic device. Then, the selected configurations are connected to each other via a probabilistic road map planner to find the accessibility of the binding site. In addition to helping a user to sample the configuration space, the haptic device also lets the user investigate the path generated by the motion planner. Stone et al [26] have integrated the haptic devices into NAMD, a public domain molecular dynamic simulation package for molecular steering. In their application, the haptic device enables the steering of molecules during molecular dynamics simulations with real-time force feedback to the user. The data communication between the graphics engine and NAMD is achieved through an efficient socket connection. The data transfer between visual and haptic displays is provided trough the VRPN protocol developed by UNC [33]. Using the haptic device, they have demonstrated that a molecular biologist could interactively steer a sodium ion through a gramicidin-A channel while feeling the interaction forces in virtual worlds. Another interesting application of the haptic devices to molecular biology is the nano-manipulator system developed at UNC. The nano manipulator system described in the work of Guthold et al. [47] integrates a haptic device with an atomic force microscope (AFM) for manipulation of molecular structures such as DNAs. Currently, direct manipulation of nano-scale objects using a standard AFM system is not possible since the scanning and manipulation processes are done in sequences using the same AFM probe. In nano-manipulator, the haptic interface enables the user to manipulate a DNA structure remotely and feel the interaction forces

during the manipulations while the changes, for example, in DNA shape and position can be observed in a simulated world in real-time. The actual shape and position of the DNA molecule is updated using the AFM scans once in every scan only, reducing the need for frequent scanning.

Chapter 3

BASIC MOLECULAR MECHANICS

When studying the structure of matter, the most thorough way of inspection is to apply quantum mechanics. In this case, the interaction between two macromolecules - the ligand and the protein - could be found out by solving the combined Schrödinger equation of both systems. Possible states of the combined system could be achieved through this method. However, quantum mechanical approach can not lead into a practical solution, since it is not always possible to find an explicit solution for this difficult problem. Of course, it is possible to find a numerical solution, but it soon turns out that even the numerically solvable quantum field models are computationally too heavy to produce truly exploitable results. It is far more productive to apply a bit more primitive, mechanic, model. This means that we need to study the quality and quantity of forces between the interactive particles. Depending on the computational method we may assign different weights to different kinds of forces. It is quite common to resort to certain simplifications, and some of the interacting forces are not used in the modeling.

3.1 Force field models

Molecular mechanics stem from the idea, that the electrons of the atom can be thought as fixed. Geometry of a molecule can be approximated effectively by taking all the interacting forces into account. Bonded interactions are described by spring forces, and non-bonded interactions are usually approximated by potentials resembling van der Waals interaction. The desired parameters are determined by experimental observations. Geometry

is further optimized by finding the energy minimum. Total energy is represented by set of potential energy functions. In addition to these functions, a set of parameters is also needed to compute the total energy. In addition to these two parts, information about atom types and atom charges is also required. We also usually need a set of rules to type atoms, generate parameters not explicitly defined and to assign functional forms and parameters. These methods together form a force field. Force fields are usually employed to generate accurate predictions to complex problems by interpolating and extrapolating from relatively simple experimental set of molecules. Examples of classical force field models which are used mainly in biochemistry include AMBER, CHARMM and CVFF. Throughout this thesis the CHARMM (Chemistry at HARvard Macromolecular Mechanics) model is incorporated. CHARMM is a program for macromolecular dynamics. In addition to performing MD using algorithms for time-stepping, long range force calculation and periodic images, it can be used for energy minimization, normal modes and crystal optimizations. There are several potential energy functions parameterized for protein, lipid and nucleic acid simulations. CHARMM also incorporates free energy methods for chemical and conformational free energy calculations. For our calculations we have made use of two CHARMM parameter files.

- For electrostatic parameters; CHARMM22 All-Hydrogen Parameter File for Proteins.
- For topology and Lennard-Jones parameters; combined CHARMM All-Hydrogen Topology File for CHARMM22 Proteins and CHARMM27 Nucleic Acids from CHARMM22.

3.2 Molecular Dynamics

Molecular dynamics is a computer simulation method for calculating the time evolution of each atom of a molecule. In most simple form in molecular dynamics, the equation of motion of an atom is derived using the Newton’s third law:

$$F_i = m_i \frac{d^2 r_i}{dt^2} \tag{3.1}$$

where, F_i is the total force acting on each atom of the molecule and typically derived from the potential energy function, r_i is the position of the atom, m_i is the mass of the atom. This equation can be integrated to calculate the new velocity and position of each atom of the molecule at each time step using standard numerical techniques. The trajectories of atoms carry several interesting structural and energetic properties and enable the molecular biologist to explore the relation between the structure and the function of the molecule. For example, by looking at the new configurations of atoms that have a similar and transition paths between them, molecular biologists can learn about the molecular bases of biochemical processes. Molecular dynamics simulations have also been used in studying protein-ligand and protein-protein docking. There are many programs to perform molecular dynamics (MD) simulations such as NAMD [44], MOLDY [43], AMBER [41] and CHARMM [42]. Using standard MD to find the global minimum energy of a docked complex is difficult since traversing the rugged hyper surface of a biological system is problematic. Often an MD trajectory will become trapped in a local minimum and will not be able to step over high energy conformational barriers. Thus, the quality of the results from a standard MD simulation is extremely dependent on the starting conformation of the system. So, our general approach makes its promise here, collecting initial position data from a human operator which does have grasping knowledge and intuition about molecular problem will yield better initializations compared to a blind automated guess. In our simulations, we follow the basic principles of mechanics for molecular dynamics computations and neglect the affect of several other factors such temperature, pressure, etc. Our goal in this chapter is to demonstrate the computational benefits of our proposed approach in molecular docking by comparing it with the simplified rigid molecular dynamics computations.

3.3 Rigid Protein-Ligand Docking

The protein-ligand docking problem can be considered as an optimization problem where the goal is to search for the configuration space of possible conformations for the lowest energy configuration of the ligand and protein molecules. Since, both the ligand and protein molecules are characterized by a collection of atoms and rotatable bonds between the pairs of atoms; this search is highly challenging [1]. Researchers have frequently utilized rigid-body assumption to reduce the number of computations. If the ligand and protein molecules are assumed to be rigid, then, the geometric approaches to molecular docking can be more easily implemented [25]. Moreover, the calculation of intrabody forces is not necessary and longer time steps are possible in molecular dynamics simulations due to the elimination of higher frequency motions [23]. If we further assume that the protein molecule is held fixed, only the rigid-body transformations (translation and rotation) of the ligand molecule must be considered for the docking computations. Under the light of these assumptions, we formulate the rigid docking problem as

> *Given rigid ligand and protein molecules (L and P, where P is fixed) with their atomic coordinates in R^3, find an optimum rigid transformation T: R^3-->R^3 such that the potential energy of L and P is minimized.*

The total potential energy of two interacting molecules can be calculated using the contribution of bonded and non-bonded terms. If the atoms of a molecule are modeled as spheres and the bonds between them as springs, the mechanics of spring deformation can be used to describe the effect of bonded terms. When the atoms change their position, this results in a flexible molecule, bonds stretch, bend, and twist. Since we treat the ligand and protein molecules as rigid bodies and neglect the internal interactions, the bonded terms do

not contribute to the total energy. Non-bonded terms include van der Waals and electrostatic interaction energies which can be described between two atoms *i* and *j* as

$$E_{ij} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + K\frac{q_i q_j}{Dr_{ij}} \tag{3.2}$$

The first two terms of this equation, also known as Lennard-Jones potential, represent the contribution of van der Waals energy and the last term represents the contribution of the electrostatic energy. A_{ij} and B_{ij} are the constants that depend on the atom type, r_{ij} is the distance between the atoms *i* and *j*, q_i and q_j are their point charges, D is the dielectric constant, *K* is the conversion factor. The interaction force between atoms *i* and *j* can be calculated using the gradient of the energy function as

$$F_{ij} = -\tilde{N}E_{ij} \tag{3.3}$$

During molecular interactions between a ligand and a protein molecules, the total force acting on atom *i* of the ligand molecule is calculated by adding all the interaction forces between atom *i* and the atoms of the protein molecule as

$$F_i = \sum_{j=1}^{P} F_{ij} \tag{3.4}$$

where, *P* is the number of atoms of the protein molecule. If the ligand molecule is modeled as a rigid body, the equations of rigid-body dynamics must be used to simulate its behavior. The ligand molecule can modeled as a group of linked multiple rigid objects if intermolecular bonding interactions are considered or as a single rigid object if they are not neglected as in our approach. The formulation of rigid-body molecular dynamics is well documented in the

literature [18],[43]. The behavior of a rigid ligand molecule under the influence of net forces, $F_{net}(t)$, and torques, $\Gamma_{net}(t)$, can be written in the form a differential system as

$$\frac{d}{dt}\begin{pmatrix} x \\ q \\ Mv \\ Iw \end{pmatrix} = \begin{pmatrix} v \\ \frac{1}{2} w q \\ F_{net} \\ T_{net} \end{pmatrix} \tag{3.5}$$

where, x is the position of the center of mass of the ligand, q is unit quaternion describing ligand's orientation, v and w are its linear and angular velocities respectively, M is the mass matrix and I is the inertia tensor. The net force and torque acting on the center of mass of the molecule at any time instant can calculated as

$$F_{net} = \sum_{i}^{L} F_i \tag{3.6}$$

$$\Gamma_{net} = \sum_{i}^{L} (r_i - x) \times F_i \tag{3.7}$$

where, L is the number of atoms of the ligand molecule and r_i is the position of a ligand atom in global coordinate frame. An implicit time integrator is suggested to obtain stable solutions, but it is not always applicable since getting explicit equations for molecular system is not trivial. Instead, a semi-implicit method, usually Verlet algorithm, is widely used in molecular dynamics literature.

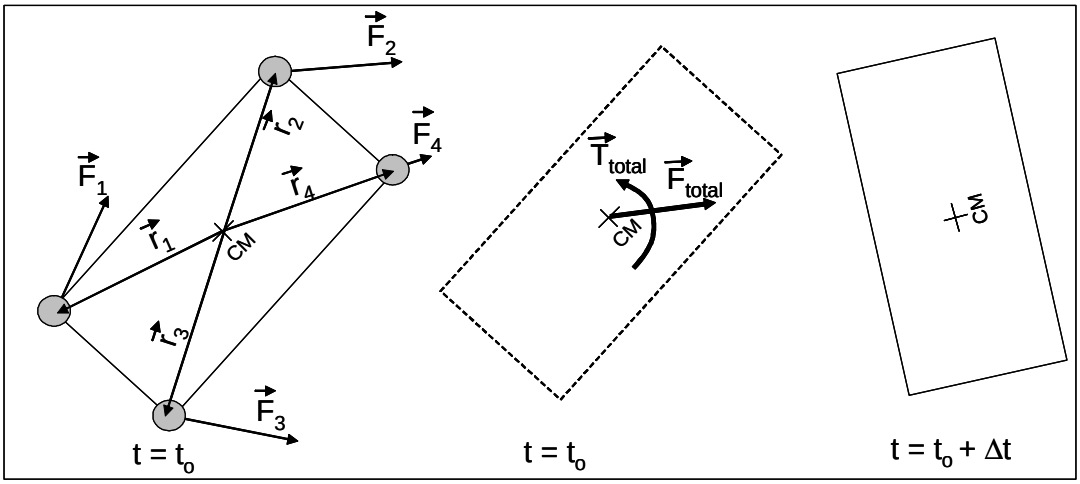


Figure 3.1: Rigid Body Simulation

3.3 Least Square Minimization Approach

We propose a new approach that is computationally more efficient than the standard integration of rigid body equations for simulating the dynamical behavior of the ligand atom. We initially relax the rigid-body assumption and calculate the new positions of the ligand atoms using the Newton's third law (3.1); where, m_i is the mass of the atom. This equation can be integrated to calculate the new velocity and position of each atom of the ligand molecule at each time step of the simulation using the numerical techniques. However, if the rigid-body assumption has to be maintained (i.e. no deformations is allowed), the ligand molecule should not change its shape as its atoms take new configurations. Hence, an optimum rigid-body transformation (translation and rotation) moving the ligand molecule from the previous rigid-body configuration to the new configuration has to be calculated. We define this problem as a least square minimization problem and state it as

Given two set of corresponding points, find the optimum rigid transformation, T_{4x4}, *that minimizes the distance between them.*

If $r_i^{t_0}$ represents the previous coordinate of a ligand atom at time $t = t_0$ and $q_i^{t_0}$ represents its virtual coordinates at time $t_0 + \Delta t$ due to the effect of molecular force F_i acting on it, then the optimum transformation between two sets of atoms can be calculated by minimizing the total distance error between them as

$$E = \sum_{i}^{L} \left\| q_i^{t_0} - T r_i^{t_0} \right\| \tag{3.8}$$

where, E is the total error and T_{4x4} is the transformation matrix.

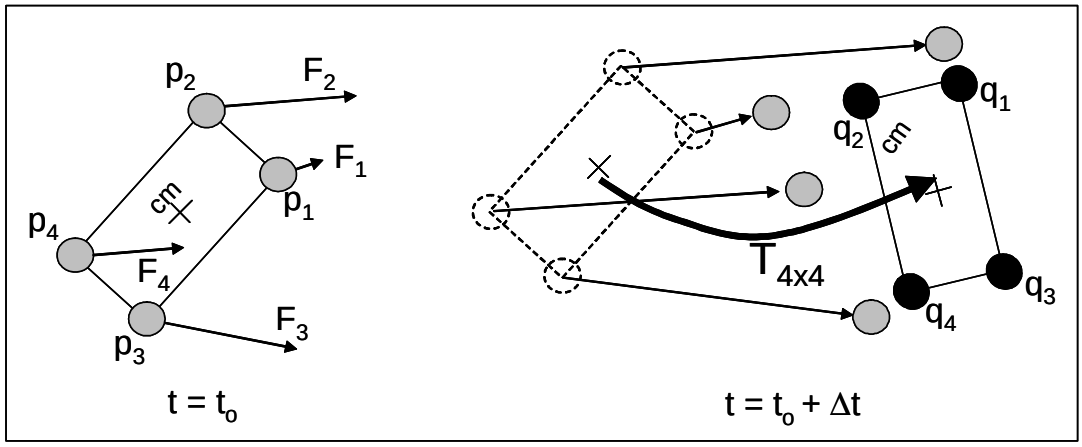


Figure 3.2: Least Square Minimization Approach

This optimization problem can be solved using the singular value decomposition method to obtain a translation vector and a rotation matrix. If p_{3x1} is the optimum translation of the center of mass of the ligand molecule and R_{3x3} is the optimum rotation about the center of mass, then $r_i^{t_0+\Delta t} = Rr_i^{t_0} + p$ represents the new coordinate of the atom that satisfy the rigid body assumption. To obtain the optimum transformation matrix, we first calculate the deviations of the actual and virtual coordinates at $t = t_0$ from their mean values

$$\hat{r}_i^{t_0} = r_i^{t_0} - \left(\frac{\sum_i^L r_i^{t_0}}{L} \right) \tag{3.9}$$

$$\hat{q}_i^{t_0} = q_i^{t_0} - \left(\frac{\sum_i^L q_i^{t_0}}{L} \right) \tag{3.10}$$

and then construct a coherence matrix

$$A = \hat{R}\hat{Q}^T \tag{3.11}$$

where, $\hat{R} = \left[\hat{r}_1^{t_0}, \mathbf{K}, \hat{r}_L^{t_0}\right]$ and $\hat{Q} = \left[\hat{q}_1^{t_0}, \mathbf{K}, \hat{q}_L^{t_0}\right]$. The singular value decomposition of this matrix ($[U,V,D] = SVD(A)$) enables us to calculate the optimum translation vector and the rotation matrix as

$$R_{3x3} = VU^T \tag{3.12}$$

$$p_{3x1} = \left(\frac{\sum_{i}^{L} q_i}{L} \right) - R \left(\frac{\sum_{i}^{L} r_i}{L} \right) \tag{3.13}$$

Finally, the homogenous transformation matrix can be constructed as

$$T_{4x4} = \begin{bmatrix} & R_{3x3} & & p_{3x1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.14}$$

Below is represented simple visual and numerical results obtained with the purpose of comparing standard rigid-body formulations with our SVD based approach. Both cases are implemented in MATLAB. A crucial point was to select the right integration scheme in order to obtain stable results for standard rigid body dynamics. Verlet Algorithm yielded best results when compared with Newton and Runge-Kutta. So, semi-implicit characteristics of Verlet algorithm produced more stable simulations against its standard explicit counterparts. Since our formulation of SVD approach contains semi-implicit behavior internally, we can speak of a more stable simulation generally independent from integration scheme.

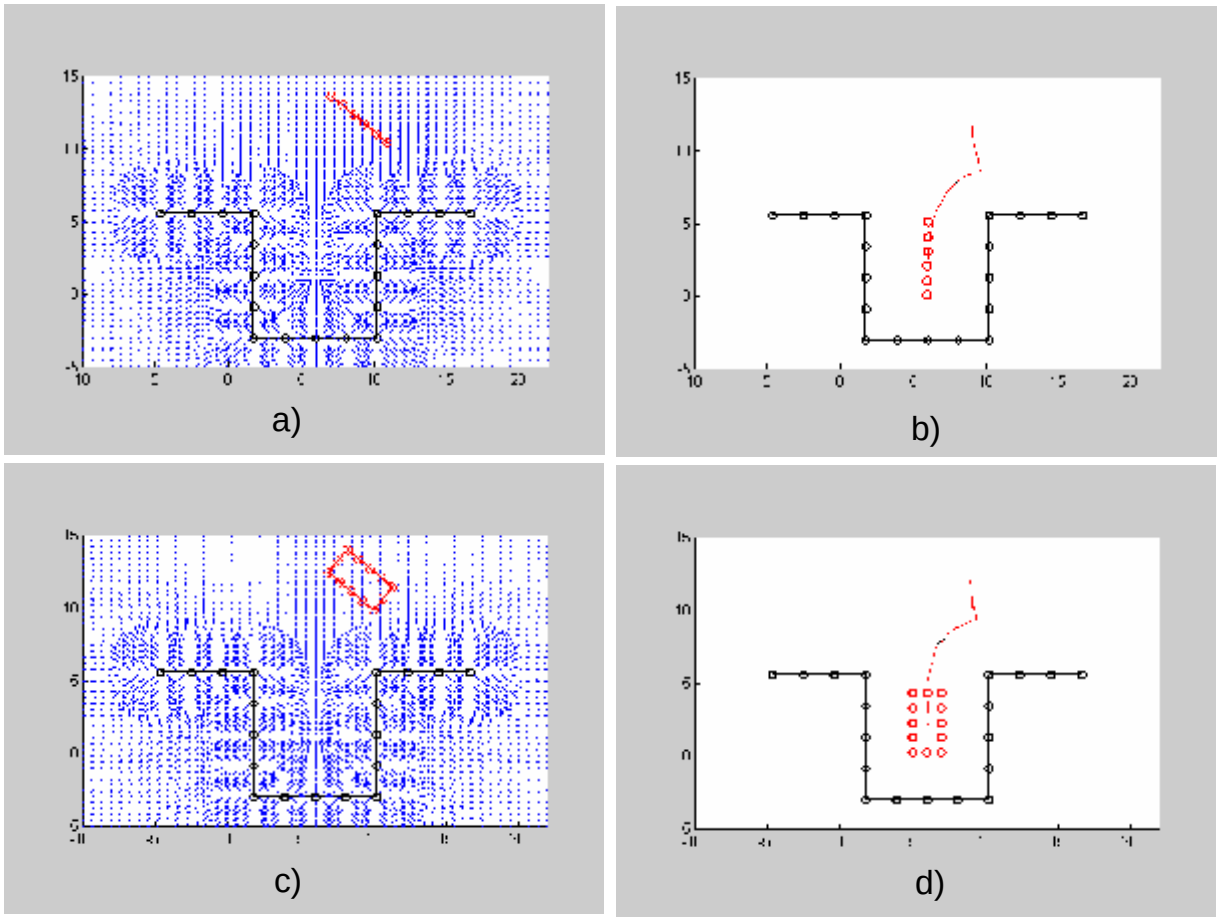


Figure 3.3: Visual comparison of our approach with the rigid body solution in 2D. A line segment and a rectangular box, made of carbon atoms are guided into an artificially created potential well (see Figures a and c respectively). The path followed by the center point of the line segment and the box were calculated using our approach (red line) and the rigid body solution (blue line) are shown in Figures b and d respectively (note that the difference in paths is not visible at all since they almost overlap).

Table 3.1: Numerical comparison of our approach with the rigid-body solution. The results were obtained using MATLAB.

LP Pair	**Number of Ligand Atoms**	**Number of Protein Atoms**	**Execution Time (sec)** ***Rigid-Body Dynamics***	**Execution Time (sec)** ***SVD***	**RMSE (Angs)** ***Rigid-Body Dynamics***	**RMSE (Angs)** ***SVD***
Line Segment	6	19	31.2810	25.5000	1.2897e-004	9.2609e-005
Rectangular Box	13	19	64.7190	49.4060	1.4752e-005	1.2034e-005

In our simulations, the user manipulates the ligand molecule in a 3d virtual environment using a haptic device and interacts with the protein molecule to explore the potential binding sites on its surface. We have worked initially with a public domain software package, Pocket in ProShape package of Stanford [30], to determine all the potential binding sites in advance and highlight them to the user in virtual environments for exploration with the haptic device. Once the binding site is determined and the ligand molecule is roughly aligned inside the binding cavity with the help of visual and haptic displays, the molecular dynamics calculations are executed to dock the ligand molecule to the protein molecule. In comparison to our approach, Bayazit et al. [22] sample the configuration space using the haptic device and then connect the low potential samples using a probabilistic road planner to calculate the docking path. They conclude that the user input via the haptic device helps the planner.

3.4 Least Square Minimization Approach with Monte Carlo Disturbance

In our approach, after it is being initially inserted into the binding cavity with the help of haptic device, the ligand molecule starts searching the configuration space of possible conformations for the lowest energy configuration using the Least Square Minimization approach. Least Square Minimization approach being a gradient method, i.e. involving the first derivative of energy function, is known to have the tendency to get stuck at local minima. Because of this characteristic in molecular dynamics literature it is common to use some other methods with tandem with gradient methods. In this thesis work, we have implemented 2 different Monte Carlo method variants for getting away from trapped local minima.

The most popular realization of the Monte Carlo method for molecular systems is the Metropolis method, which can be briefly summarized with a pseudo code as follows;

1. First, specify the initial atom coordinates.
2. Select some atom *i* randomly and move it by random displacement: ΔX_i, ΔY_i and ΔZ_i.
3. Calculate the change of potential energy ΔV corresponding to this displacement.
4. If $\Delta V < 0$ accept the new coordinates and go to step 2.
5. Otherwise, if $\Delta V > 0$, select a random number ***R*** in the range [0,1] and:

   A. if $e^{-\Delta V/kT} < R$ accept the new coordinates and go to step 2,

   B. if $e^{-\Delta V/kT} \geq R$ keep the original coordinates and go to step 2.

In our work, in order to embed Metropolis method into our algorithm, we have simply disturbed our force calculations by some random amount and after executing our Least Square Minimization method with these disturbed forces, made a decision based on Boltzmann probability about accepting this new conformation or not.

Analyzing the simulations carried out with this approach we have obtained although the new algorithm increases the quality of results, it is not a breakthrough. So we came up with a variation of standard Metropolis method tuned for our simulations. The idea is to bias the random displacement step such that the displacement is going to occur as a random move added with as a pure rotation around the principal axes of ligand molecule geometry. The main reasoning for this approach is to catch the conformations which are missed due to the symmetrical geometry of some ligand molecules. Since the Least Square Minimization algorithm already incorporating a Singular Value Decomposition step, it is an easy and cheap operation to obtain the principal axes of Ligand geometry.

Also note that, we have implemented our algorithms in such a way that a fixed number of pure gradient minimization steps are following every metropolis step, to guarantee that we are at the minimum energy conformation around the local neighborhood. Results regarding our simulations can be found in Section 5.2.2.2 and Appendix B.

Chapter 4

HAPTIC VISUALIZATION

4.1 System Overview

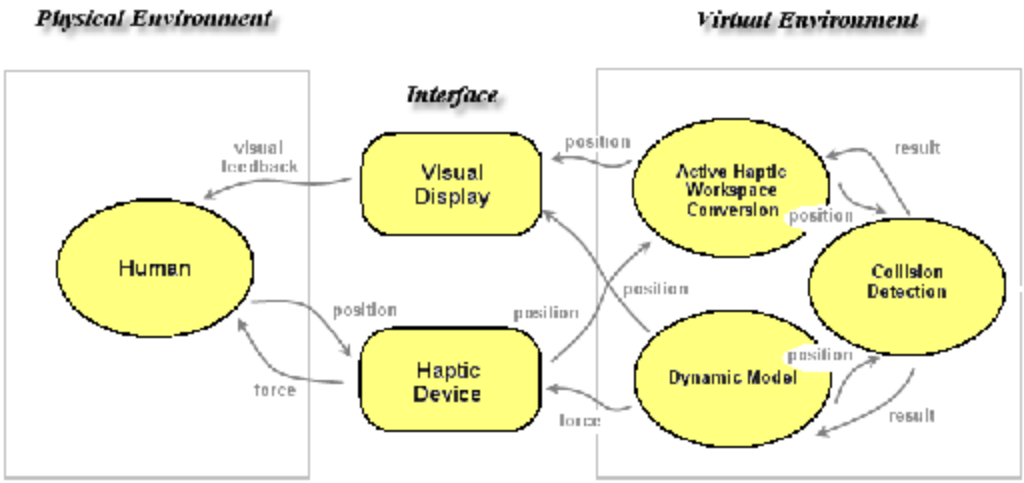


Figure 4.1: Overview of virtual docking system

The architecture of our system is described in Figure 4.1. The system is composed of a PHANToM haptic device (6 dof input, 3 dof output) [35], and a dual Intel Xeon processor workstation (for graphics display, position tracking, molecular calculations and collision detection). The operator uses the PHANToM to manipulate a single rigid object (ligand) in

the virtual scene. Haptic interaction consists of the following steps: (i) tracking the user's motion, (ii) detecting collisions between the ligand and protein, (iii) computing appropriate collision and molecular forces, and (iv) exerting forces on the user via the PHANToM. Our haptic-interaction application were developed using the C++ General Haptic Open Software Toolkit (GHOST SDK) [34]. For graphics programming we have used Open Inventor [38], an object oriented general purpose scene graph API, which is widely used especially among academic society. Following sections explain design considerations and details about the realization of haptic system.

4.2 Real-Time Stable Haptic Rendering of Molecular Interactions

The goal of haptic rendering is to enable a user to touch, feel, and manipulate simulated objects through a haptic device. During the last ten years, significant advancements have been made in haptic devices and rendering algorithms [36] and a new research area, what we call *computer haptics*, has emerged [39]. Just as computer graphics is concerned with synthesizing and rendering visual images, computer haptics is the art and science of developing software algorithms that synthesize computer generated forces to be displayed to the user for perception and manipulation of virtual objects through touch. Haptic rendering of a 3D virtual object generally involves two phases: collision detection and collision response. First, we check if the virtual model of a haptic probe (a point, a line segment, or 3D object) penetrates into the virtual model of the 3D object (collision detection), and then a mechanic interaction model (e.g. Hooke’s law: magnitude of the force is equal to the depth of penetration times a coefficient) is used to calculate the interaction force (collision response). This force is reflected to the user through the device for haptic exploration and manipulation of the 3D object. In this work where we model molecular interaction energies haptically, we render also non-zero forces in no-collision cases obviously due to the effects of non-bonded terms of molecular forces.

4.2.1 Real-Time Haptic Rendering

Non-bonded terms include van der Waals and electrostatic interaction energies which can be described between two atoms *i* and *j* as

$$E_{ij} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + K\frac{q_i q_j}{Dr_{ij}} \qquad (4.1)$$

The first two terms of this equation, also known as Lennard-Jones potential, represent the contribution of van der Waals energy and the last term represents the contribution of the electrostatic energy. A_{ij} and B_{ij} are the constants that depend on the atom type, r_{ij} is the distance between the atoms *i* and *j*, q_i and q_j are their point charges, D is dielectric function. *K* is a factor that converts the electrostatic energy into kilojoules per mole. The interaction force between two atoms *i* and *j* can be calculated using the gradient of the energy function as

$$F_{ij} = -\tilde{N}E_{ij} \qquad (4.2)$$

Hence, the total interaction force between two molecules can be calculated as

$$F_{total} = \sum_{j=1}^{N}\sum_{i=1}^{M} F_{ij} \qquad (4.3)$$

where, N and M are the number of atoms in the ligand and the protein molecules respectively.

As can be seen from the formulation, the computation of interaction forces (i.e. the forces due to non-bonded interactions) involves a large number of pair-wise calculations: we consider each atom of the ligand and loop over all the atoms of the protein to calculate the total force acting on the ligand. The brute-force approach requires $O(NM)$ calculations, which is not feasible for real-time implementation. To reduce the computation time, a grid-based approach is used. The 3D space occupied by the two molecules is divided into smaller grid cells and the force calculations are carried out off-line for each atom type as if there is a single ligand atom at the grid point interacting with the atoms of the protein molecule. During the simulations, the force acting on each ligand atom is determined using its cell location in the grid and a tri-linear interpolation of the pre-calculated force vectors at corners of the cell. Then, the total force between two molecules at any instant is calculated via the summation of the contribution of each type of atom.

4.2.2 Stable Haptic Rendering

In simulating force interactions between a ligand molecule and a protein molecule, the virtual model of the haptic probe becomes the 3D virtual model of the ligand molecule and interacts with the 3D virtual model of the protein molecule. The user holds and manipulates the haptic probe in physical space which results in changes in the position and orientation of the ligand molecule in virtual space. The net force and torque that arises due to the molecular interactions between the atoms of the ligand and protein molecules are calculated in real-time using the grid structure and reflected to the user. In fact, haptic simulation of collision and force interactions between two arbitrary shaped 3-D objects is not a trivial task and the research in this area is still very active. However, a 3D geometric model of a molecule is typically constructed from 3D spherical atoms with a characteristic radii and it has a certain topology. A common 3D surface representation of a molecule based on 3D spherical atoms is the Connolly surface [40]. To construct the Connolly surface of a molecule, a probe “water” atom is rolled over the surface atoms of the molecule defined by their van der Waals radii while bridging the gaps via smooth patches. While the Connolly

surface representation is used for graphical display of ligand and protein molecules in our simulations, we take advantage of the spherical representation of atoms for calculating haptic interaction forces between ligand and protein molecules.

To describe how we render the haptic interactions between ligand and protein molecules during docking, we first discuss the haptic interactions between two atoms, represented by spheres; in 2D space (see Figure 4.2). The spheres shown in the figure are $\mathbf{r}_{p_1}$ and $\mathbf{r}_{p_2}$ away from the origin and having a characteristic radii of R_1 and R_2. As mentioned earlier, the haptic rendering process involves two stages: collision detection and response. Detecting collisions between two spheres in real-time is quite easy. The spheres intersect if

$$\left\| \mathbf{r}_{p_1} - \mathbf{v}_{p_2} \right\| \leq (R_1 + R_2) \tag{4.4}$$

and the difference,

$$(R_1 + R_2) - \left\| \mathbf{r}_{p_1} - \mathbf{v}_{p_2} \right\| \tag{4.5}$$

gives the depth of penetration. The direction of the molecular forces, calculated using Eq.4.7, is along the line connecting the centers of the atoms. Based on the action-reaction principle, the magnitude of the forces acting on the ligand and protein are the same, but they are opposite in directions.

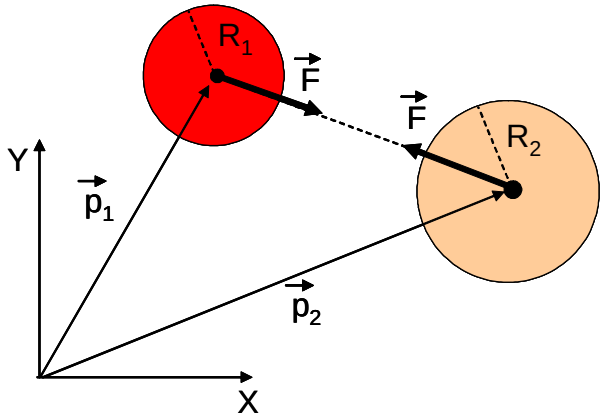


Figure 4.2: Haptic rendering of force interactions between two atoms of different characteristic radii.

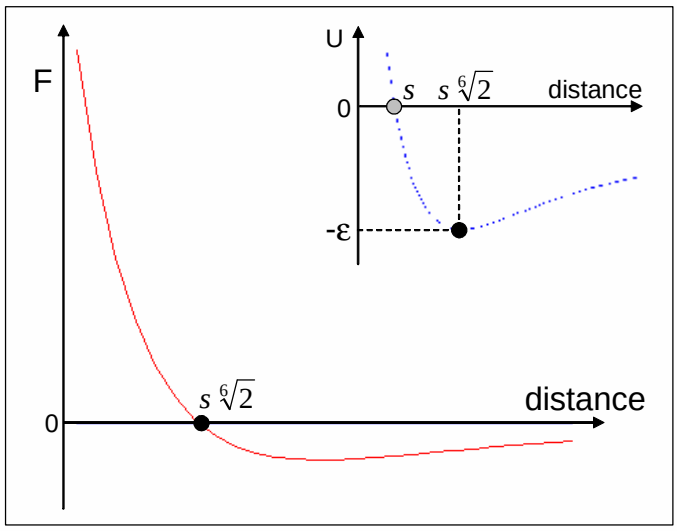


Figure 4.3: Long range molecular forces due to Lennard-Jones potential.

The total force acting on the atoms shown in Figure 4.3 is a combination of long range attraction and repulsion forces due to Lennard-Jones potential. The LJ potential is made of two terms: the first term $(1/r^{12})$ models the repulsion and dominates when the two atoms are close to each other and the second term $(1/r^{6})$ models the attraction and

dominates when the two atoms are far away from each other. The potential energy due to long range interactions alone is zero when $r = s$ and minimum when $r = s\sqrt[6]{2}$ (equal to twice the characteristic VDW radius). The atoms attract each other when they are far away from each and if the distance between the centers of the atoms is less than $s\sqrt[6]{2}$, they start to repel each other very strongly. This sudden change, first in force direction, then in force magnitude may cause instabilities in haptic devices. In particular, when two atoms penetrate into each other, the force magnitude reaches to extremely high levels that the current haptic devices cannot render at all. If all the forces are scaled down such that the forces during the penetration can be displayed to a user, then the forces before the penetration becomes too small to be sensed by the user. For this reason, Lee and Lyons [17] render van der Waals interaction forces until a contact occurs between two atoms and then add a spring-based force component (depth of penetration times a spring constant) to the total force if the atoms further penetrate into each other. In Figure 4.4 above, an atom of a ligand molecule (located at L_1) penetrates into three atoms of a protein molecule (located at P_1, P_2 and P_3). The contribution of each protein atom to the total spring force acting on the ligand atom

$$\left(\vec{F}_{total}^{\,spring} = \sum_{i}^{n} F_i^{spring} \right) \tag{4.6}$$

is

$$\vec{F}_{i}^{\,spring} = kd_i \frac{\vec{L}_i - \vec{P}_i}{\left| \vec{L}_i - \vec{P}_i \right|} \tag{4.7}$$

where d_i is the depth of penetration of the ligand atom into the i^{th} atom of the protein.

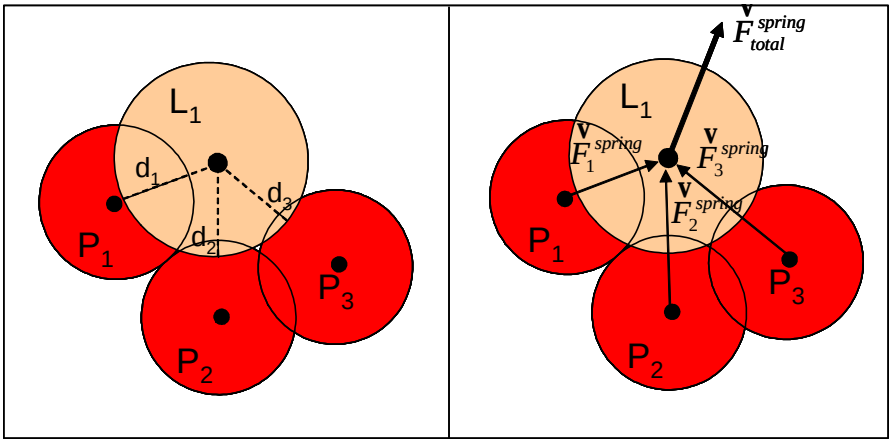


Figure 4.4: Haptic rendering of interaction forces during the penetration.

Now, the total force acting on the ligand atom is the summation of the total molecular force at the point of contact (due to LJ potential) and the total spring force due to the penetration of the atoms into each other.

4.3 Active Haptic Workspace Concept

The haptic device used in our simulations has a limited workspace (defined by the physical dimensions of the haptic device) and sensing resolution which prevents the user from exploring the surface of large protein molecules efficiently. Moreover, there may be a mismatch between the visual and haptic coordinates of a molecular surface due to the difference in the size of haptic and visual workspaces. To resolve this mismatch, the absolute coordinates of a 3D molecular surface can be transformed to the visual and haptic workspaces independently such that the visual representation is small enough to be enclosed by the viewing volume and the haptic representation is large enough to be explored with the haptic device, A mapping (i.e. transformation) between the two workspaces must be

considered for the implementation. The absolute 3D coordinates of the protein surface must be multiplied by the proper transformation matrices ${}^{V}_{A}T$ and ${}^{H}_{A}T$ such that the model fits into haptic and visual workspaces.

$$ {}^{H}PROTEIN = {}^{H}_{A}T \; {}^{A}PROTEIN \tag{4.8} $$

$$ {}^{V}PROTEIN = {}^{V}_{A}T \; {}^{A}PROTEIN \tag{4.9} $$

where, ${}^{A}PROTEIN$ represents the absolute coordinates of the protein molecule extracted from PDB and ${}^{H}PROTEIN$, and ${}^{V}PROTEIN$ represent the coordinates of the same molecule in haptic and visual workspaces. Then, the transformation between the visual and haptic coordinate frames, ${}^{V}_{H}T$ is defined as

$$ {}^{V}_{H}T = {}^{V}_{A}T \, ({}^{H}_{A}T)^{-1} \tag{4.10} $$

and used in displaying the movements of the ligand molecule in visual workspace. In our simulations, the ligand molecule is virtually coupled with the haptic stylus (probe). The position and orientation of the ligand molecule are updated in visual workspace every time the haptic stylus is moved in haptic workspace as

$$ {}^{H}LIGAND_{current} = {}^{H}t_{stylus} \; {}^{H}LIGAND \tag{4.11} $$

where, ${}^{H}t_{stylus}$ represents the transformation matrix of the haptic stylus, ${}^{H}LIGAND_{current}$ and ${}^{H}LIGAND$ are the recent and initial haptic coordinates of the ligand molecule. In each cycle of the haptic loop, the haptic coordinates of the ligand molecule at the current time frame is multiplied by the ${}^{V}_{H}T$ to display of the movements ligand molecule in visual workspace

$$ {}^{V}LIGAND_{current} = {}^{V}_{H}T \; {}^{H}LIGAND_{current} \tag{4.12} $$

To calculate the reaction forces that will be reflected to the user through the haptic device, we must return to the absolute coordinates since all the parameters for calculating the molecular interaction forces are defined in terms of the absolute coordinates of the ligand and protein atom. For this purpose, we first transform the haptic coordinates acquired at the current time frame to the corresponding absolute coordinates (${}^{A}LIGAND_{current} = ({}^{H}_{A}T)^{-1} \; {}^{H}LIGAND_{current}$) and then calculate the molecular interaction forces between the current configuration of the ligand molecule and the protein molecule (${}^{A}PROTEIN$). The calculated forces are scaled and transformed back to the haptic workspace using the rotation matrix as

$$ {}^{H}Force = Scale\,Factor \times ({}^{H}_{A}R \; {}^{A}Force) \tag{4.13} $$

and then displayed to the user through the haptic device as if the user is feeling the force in the visual workspace ${}^{V}({}^{H}Force) = {}^{V}_{H}R \; {}^{H}Force$.

This approach is effective in haptic exploration of 3D objects if the object surface is continuous, and does not contain small scale bumps and cavities or other fine surface details such as textures. For the case of very large objects, scaling the original object down to fit it into the haptic workspace may significantly compress those surface details such that the bumps and cavities may be perceived as textures and the existing textures may not be perceived at all. For the case of very small objects, scaling the original object up to fit it into the haptic workspace may not give us the satisfactory spatial resolution for the detailed haptic exploration of bumps and cavities since the scaling factor is limited by the size of the haptic

workspace. In fact, this is exactly the situation that we have observed during the haptic exploration of protein surfaces for the potential binding sites. The actual dimensions of a protein molecule is in the order of a few hundreds angstroms and the spatial resolution achieved by scaling the original atom coordinates such that it completely fits into the haptic workspace does not give sufficient results for the detailed haptic exploration of the potential binding sites. To resolve this problem, we propose the concept of A*ctive Haptic Workspace* (*AHW*). The AHW is a transparent haptic subspace that actively travels in the large 3D visual workspace. The part of the object that is inside the AHW can be explored in high resolution with the haptic device. AHW moves to a new location if the tip point of the haptic stylus (this point is actually the center of mass of the ligand molecule and it will be referred as Haptic Interface Point –HIP- now on) exceeds its pre-defined boundaries. This approach enables the user to manipulate the ligand molecule and explore all parts of the protein surface to search for potential binding sites in high spatial resolution (see Figure 4.2). For example, if the user translates the ligand molecule in x-direction to explore the parts of the object that are not currently accessible by AHW and the pre-defined boundary of the AHW is crossed during this process, the protein surface is automatically translated in the negative x-direction in visual workspace to make this exploration possible. In addition, if the switch on the haptic stylus is activated at the same time, the object in the scene is also rotated about the y-axis in counterclockwise direction to allow the exploration of invisible parts of the object such as its back surface. The relations between the ligand and the protein translations as well as the switch settings and protein rotations are given in Table 4.1.

Table 4.1: The relations between the ligand and protein translations and the switch settings on the haptic stylus and the protein rotations in the visual workspace.

Ligand Translation	**Protein Translation** (rate = 5mm/sec)	**Protein Rotation** (if the switch is ON) (rate = 36 degrees/sec)
Left (+x)	Right (+x)	CCW about y-axis
Right (-x)	Left (-x)	CC about y-axis
Up (+y)	Down (+y)	CW about x-axis
Down (-y)	Up (-y)	CCW about x-axis
Front (+z)	Back (+z)	CW about z-axis
Back (-z)	Fronts (-z)	CCW about z-axis

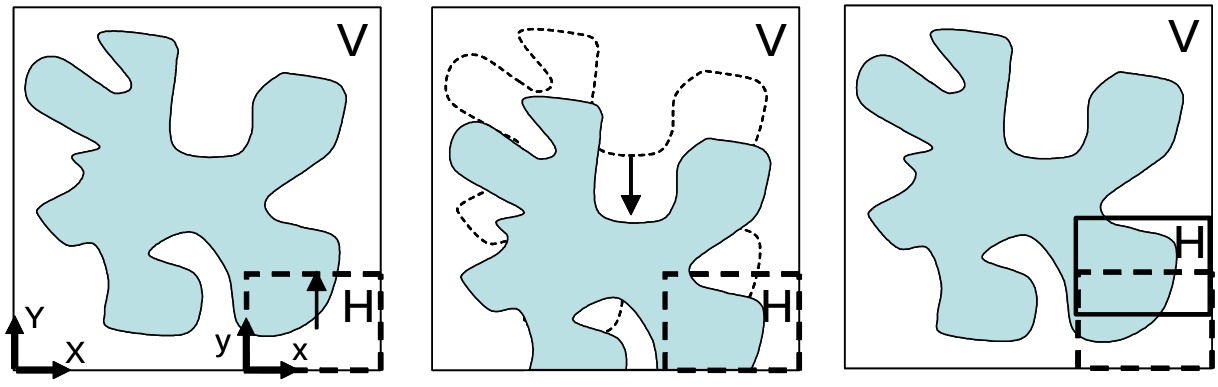


Figure 4.5: In this example, we show the concept of "moving" active haptic workspace for the haptic exploration of large molecular surfaces: **a)** The user reaches to the boundary of the AHW in Y direction. **b)** The object is moved in –Y direction in constant speed to let the exploration of the other parts of the molecular surface. **c)** The user perceives as if the haptic workspace moves up in the +Y direction.

In order to further explain the use of AHW in molecular visualization, lets define the transformations of AHW in haptic coordinates as ${}^{H}t_{AHW}$. If the user exceeds the boundaries of the AHW by translating the ligand molecule and if the switch on the stylus is pressed simultaneously to get a new orientation of the protein molecule for visualization, the transformation matrix ${}^{H}t_{AHW}$ is updated incrementally at a constant rate (5 mm/sec for translation and 36 degrees/sec for rotation) based on relations given in Table 4.1 until the center of mass of the ligand molecule is in the new boundaries of AHW.

$$ {}^{H}t_{AHW} = {}^{H}t_{AHW}\ {}^{H}t_{increment} \tag{4.14} $$

where, ${}^{H}t_{increment}$ represents the incremental translations and rotations. Now, the visual model of the protein surface that is accessible by the user for active haptic exploration is updated by

$$ {}^{V}PROTEIN_{visible} = ({}^{V}_{H}T\ {}^{H}t^{-1}_{AHW})\ {}^{V}PROTEIN \tag{4.15} $$

where, ${}^{V}PROTEIN = {}^{V}_{A}T\ {}^{A}PROTEIN$ represents the initial model of the protein molecule in visual workspace. During the real-time interactions, the transformation of the haptic stylus, ${}^{H}t_{stylus}$, is acquired at 1 kHz in haptic coordinates using the I/O library of the device and utilized to update the position and orientation of the ligand molecule at 30 Hz in the visual workspace (see Eq. 4.12). As the user manipulates the stylus of the haptic device, the virtual model of the ligand molecule is manipulated in the visual workspace and the interaction forces are calculated in the absolute coordinate frame. The van der Waals and electrostatic interaction forces between the ligand and protein molecules have to be calculated in absolute coordinate frame for the proper utilization of constants defined in Eq. 4.1 and 4.3. As the user

exceeds the boundaries of the AHW to explore the invisible parts of the large protein molecule, the protein molecule is transformed to a new position and orientation (see Table 4.1) to accommodate the needs of the user. Hence, the visual coordinates of the protein molecule that is currently visible to the user ($^{V}PROTEIN_{visible}$) must be projected to the absolute coordinate frame to detect collisions and calculate interaction forces. However, this projection can be computationally quite expensive if the number of atoms of the protein molecule is considered. An alternative solution is to apply the affect of $^{H}t_{AHW}$ to the ligand molecule to calculate its current position and orientation in absolute coordinate frame with respect to the fixed protein molecule. Since the ligand molecule typically contains much less number of atoms than the protein molecule, there are less number computations involved in the process. The absolute coordinates of the ligand molecule can be updated as

$$^{A}LIGAND_{current} = (^{H}_{A}T^{-1})\,(^{H}t_{stylus}\,^{H}t_{AHW})\;\;^{H}LIGAND \tag{4.16}$$

Now, the interaction forces between the absolute coordinates of the protein molecule ($^{A}PROTEIN$) and the current absolute coordinates of the ligand molecule ($^{A}LIGAND_{current}$) can be calculated in the absolute coordinate frame and transformed back to the haptic coordinate frame using the following transformation

$$^{H}Force = Scale\,Factor \times (^{H}t_{AHW})^{-1}(^{H}_{A}R\;\;^{A}Force) \tag{4.17}$$

If the ligand molecule collides with the protein molecule, we stop calculating the molecular interaction forces and add a spring-based component ($\vec{F}^{spring}_{total}$) to the existing force vector

$({}^{H}Force)$. Finally, the force vector is displayed with respect to the visual coordinate frame ${}^{V}({}^{H}Force) = {}^{V}_{H}R\ {}^{H}Force$, to provide the effect of what you see is what you feel.

.

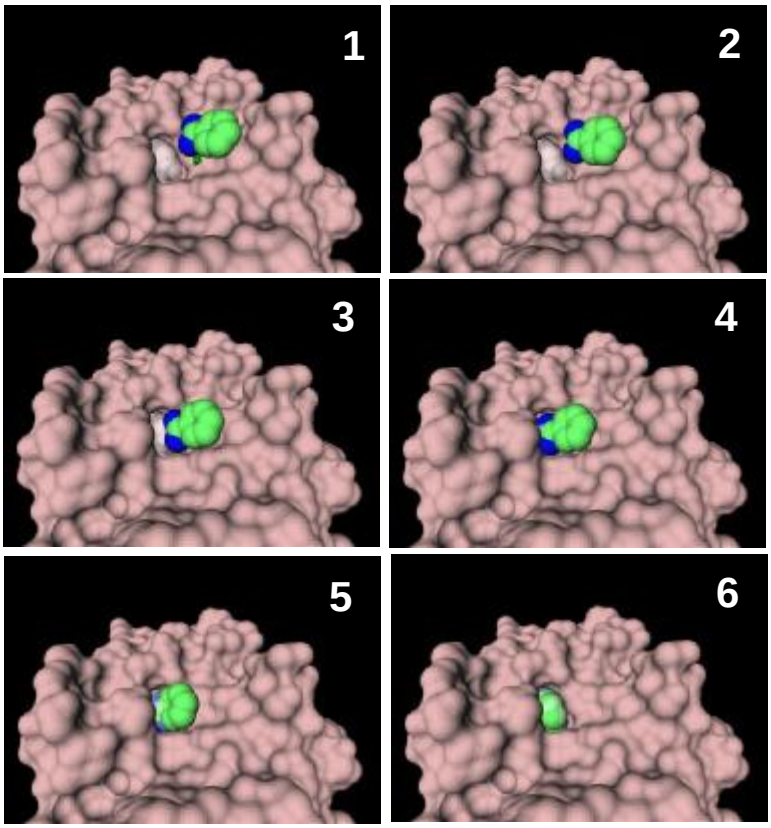


Figure 4.6: Snapshots showing the binding process of Benzamidine molecule (ligand) with 9 atoms to Beta-Trypsin molecule (protein) with 1701 atoms. The ligand and protein molecules were extracted from the 3PTB complex in Protein Data Bank (PDB). During the real-time interactions, the user manipulates the ligand molecule in virtual space by manipulating the end-effector of the haptic device in physical space. The molecular interaction forces are felt through the haptic device as the ligand molecule is inserted into the cavity (binding site) as shown in the figure.

Chapter 5

EXPERIMENTAL STUDY

5.1 Experimental Procedures Introduction

In order to investigate the role of the haptic feedback in molecular docking, we have designed two sets of experiments. To our knowledge, there is only a single experimental study conducted with human subjects in this area, which dates back to 1990 [15]. In that study, subjects were experienced biochemists and they were asked to dock four ligand molecules to the active side of a protein molecule to discover which ligand molecule is the best choice. Subjects were allowed to orient the ligand molecule and adjust up to six of its internal twistable bonds to achieve the lowest potential energy configuration. Two energy bars showing the total potential binding energy, which is to be minimized, and the internal potential energy of the ligand molecule in its current conformation were displayed. The word "GOAL" signaled the subjects that they have achievement the target energy tolerance. Experiments were repeated with and without force feedback, but with the same visual display and energy bars for both. The results showed that the haptic display improved the performance of subjects. For example, 6-D rigid body docking part of the task was 30 % faster and the trajectory path lengths were 41 % shorter with haptic feedback.

We have designed two experiments as stated before with two different purposes. In the first experiment user supposed to choose true binding pocket among 5 candidates, whereas second experiment targets to study the ability of how well the user using our virtual reality system finds binding conformation in the true binding pocket. The details and results are following.

5.2.1 Experiment I

5.2.1.1 Experiment I Architecture and Overview

The aim of our first experimental study is to test if the subjects can find the active binding site of a protein molecule for the given ligand molecule. Experiments were conducted with 6 naïve subjects. Subjects were college students with no formal training in molecular biology. They were displayed 3D models of molecular surfaces in pairs. Using the haptic device, they were asked to manipulate the ligand molecule and insert it into the cavities on the surface of the protein molecule to search for the true binding site. During this process, they felt the molecular interaction forces through the haptic device. There were 5 different cavities on the surface of each protein molecule for exploration, but only one of them was the true binding site. The task of the subjects was to find the true binding site by ranking the cavities from most likely to least likely. The cavities on the surface of protein molecules were selected using *Pocket[30]*, a public domain package developed by Edelsbrunner and Koehl (2005). Pocket uses the Alpha Shape theory to detect pockets (i.e. potential binding sites) in proteins and ranks them according to their volume and surface area. In all of the six ligand-protein (LP) couples used in our experiments, the true binding site was among the first 5 cavities returned by the Pocket program.

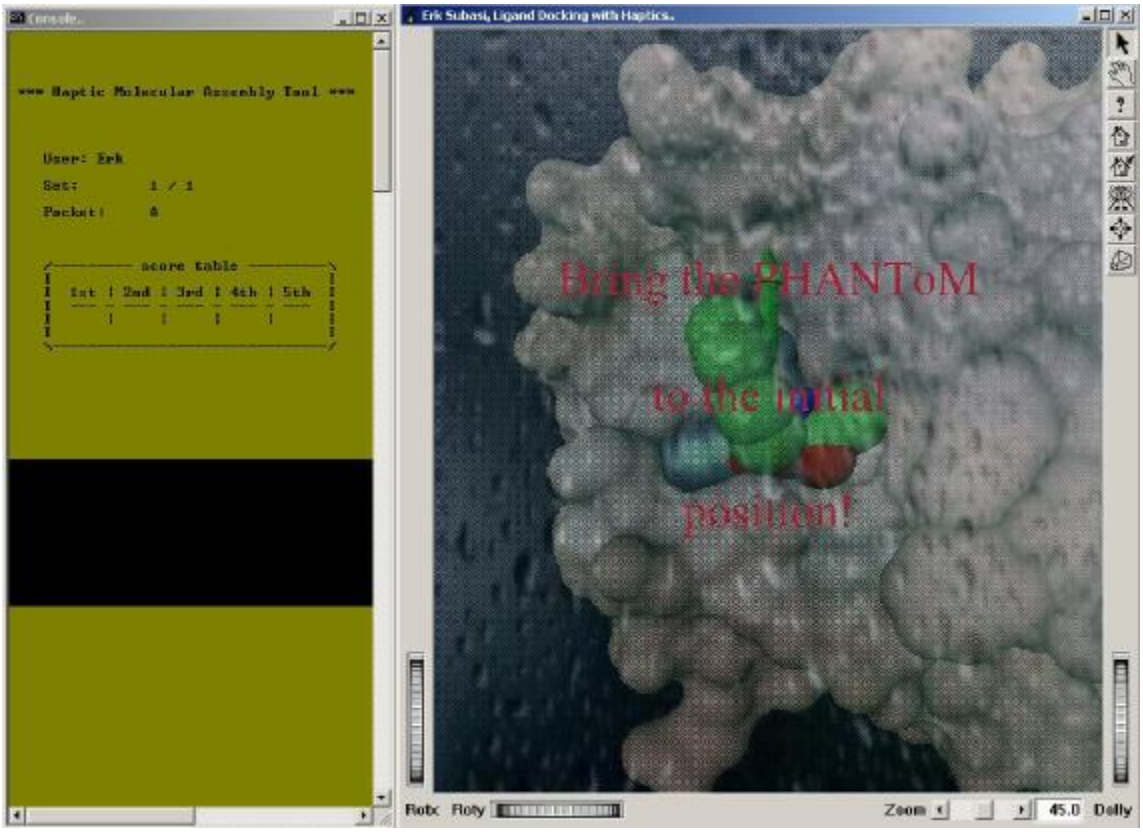


Figure 5.1: Screen capture from our program at initial stage.

Before the experiment, subjects were instructed about the haptic device and the molecular docking problem. Subjects were asked to read a document that describes how to find the true binding site using visual and haptic cues. They were also displayed a slide show summarizing what they read in the document. Subjects were instructed that

(1) the shape complimentary is an important factor in selecting the true binding site.
(2) the magnitude of the net force at the true binding site is close to zero.
(3) the true binding site generates a “tunneling effect” and pulls the ligand molecule towards the inside of the binding cavity.
(4) the true binding site traps the ligand molecule and does not easily let it go.

All subjects were also trained with the 3PTB complex (see Figure 5.2) before the actual experiments. With the help of an expert user, they were educated to find the true binding site using the clues given above. During the actual experiments, subjects were displayed 6 pairs of molecules. For each pair, they were asked to explore the 5 different cavities on the protein surface (only one of them was the true binding cavity) and ranked them from most likely true binding site (first place) to least likely true binding site (fifth place). A total of 30 cavities (6 pairs x 5 cavities) were displayed to the subjects in random order. Subjects repeated the experiment with one day rest period. The second time, the same cavities were displayed to the subjects in different order and different spatial orientations to reduce the visual bias.

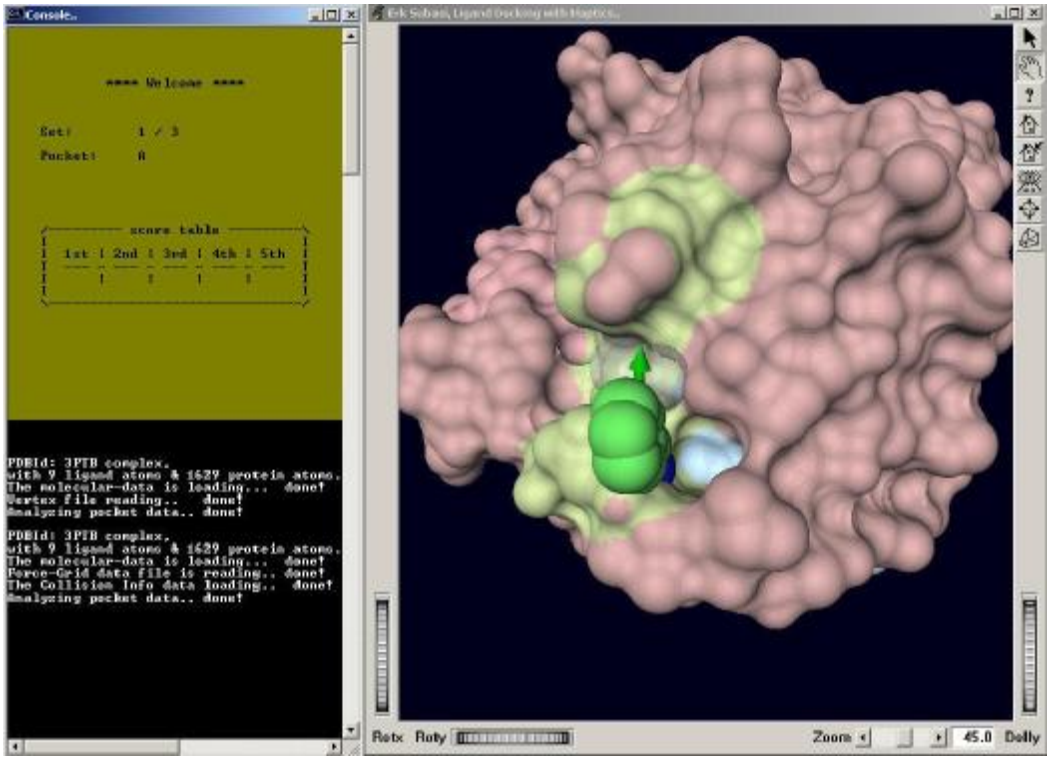


Figure 5.2: Screen capture of system on action, the user is trained at sample case with 3PTB complex.

During the experiments, subjects were allowed to navigate between the cavities labeled as A, B, C, D, and E freely by pressing the left and right arrows on the keyboard. We implemented a fly-through from one cavity to the other to better perception of the molecular surface. A fly-through is an effect created by moving the camera through three-dimensional space, giving the impression that you are flying along with the camera. The polygons around the entrance and inside of each cavity were highlighted with a color different than the color of the protein surface to help the users locate the cavities more easily. In addition, a different color was used for each cavity to help the users differentiate the cavities of a protein molecule. The colors of the cavities were also randomized in each trial. Subjects were asked to rank the cavities by pressing the "1" (most likely), "2", "3", "4", and "5" (least likely) keys on the keyboard. The ranking was displayed on the screen and subjects were allowed to update their ranking at anytime (i.e. they were allowed to test any of the 5 cavities at any time to change their ranking) before they start working on a new LP pair by pressing the "N" key on the keyboard.

5.2.1.2 Experiment I Results

The results of this experiment demonstrate that the subjects successfully discovered the true binding site. The average score of the subjects was 1.47 ± 0.36 in the first trial and 1.44 ± 0.27 in the second trial. The score was calculated using a conservative penalty-based method. The average of two trials is 1.46±0.30, which indicates that subjects have ranked the true binding sites as their first choice in most of the pairs and as their second choice in some pairs. After the experiments, a simple questionnaire, made of 10 questions (5 questions related to the role of visual information and 5 questions related to the role of haptic information) was also given to each subject to better understand his/her perception of the individual role of visual and haptic feedback in discovering the true binding cavity. In

particular, we wanted to know if the subjects had effectively used the haptic cues taught them during the training phase of the experiments. All questions have started with a phrase of "How much did … help you find the true insertion site?" (e.g. how much did the size and depth of a cavity help you find the true binding site?, how much did the effect of being pulled inside a cavity help you find the true binding site?) Each question rated on a scale that varies from 1 (very little)-to-5 (very much). The results of the questionnaire shows that subjects used the haptic cues (Ave = 4.33 ± 0.8) more effectively than the visual ones (Ave = 3.67 ± 0.8) in finding the true binding site ($p < 0.05$). More detailed experiment I results data are given in Appendix A.

Table 5.1: The results of the first experiment. The raw scores were obtained by simply counting the number of times the true binding site was discovered by the subjects. The penalty-based scores are more conservative and they were obtained by assigning a penalty score to the wrong choices made by the subjects. For example, a penalty score of "5" was used in the calculation of total score if the subject had selected the true binding site as his/her last choice.

LP pairs (PDB ID)	**Trial-First Trial (raw score)**	**Trial – Second Trial (raw score)**	**Total Combined (raw score)**	**Trial-I First Trial (penalty-based score)**	**Trial-II Second Trial (penalty-based score)**	**Total Combined (penalty-based score)**
1BU4	5/6	5/6	10/12	7/6	8/6	15/12
1STP	6/6	6/6	12/12	6/6	6/6	12/12
1MFA	3/6	3/6	6/12	15/6	10/6	25/12
3VGC	4/6	4/6	8/12	9/6	9/6	18/12
1XIG	5/6	3/6	8/12	7/6	9/6	16/12
1CSI	3/6	3/6	6/12	9/6	10/6	19/12
Total	26/36	24/36	50/72	53/36	52/36	105/72
Success Rate	72%	67%	69%	-	-	-
Average Score	-	-	-	1.47	1.44	1.46

5.2.2 Experiment II

5.2.2.1 Experiment II Architecture and Overview

In the second set of experiments, only the true binding site of each pair was displayed to the subjects. Subjects were asked to find the binding configuration (position and orientation) of the ligand molecule inside the true binding cavity using the visual and haptic cues. In addition, the total potential energy of interactions was displayed on the screen. Subjects were simply told that they should adjust the position and orientation of the ligand molecule inside the binding cavity until the level of net force and the energy magnitudes were close to a minimal value. As the subjects manipulated the ligand molecule, the total energy of the complex was calculated at the graphics update rate of 30 Hz. When the ligand molecule passed through a low energy configuration during the exploration of the binding cavity, a visual copy of the ligand molecule was left there as a decoy to draw the attention of the subject (See Figure 5.3). When the subject discovered a lower energy configuration than the previously stored one, the new configuration of the visual decoy was updated to inform the subject. Subjects could easily turn the visual decoy on and off by pressing the "Backspace" key on the keyboard. The same molecular complexes were used in the second experiment. The true binding cavity of each pair was displayed to the subjects in two different orientations. The order of the cavities was randomized, with the same order displayed to each subject. Subjects pressed the "Enter" key to stop the docking process. The binding configuration of the ligand molecule was saved to file in the form of a transformation matrix for further processing by the proposed potential approach as discussed in Chapter 3, Section 3.3.

Table 5.2: Protein – Ligand Pairs that we have used in our experiments.

LP pairs (PDB ID)	**Name of the complex**	**Ligand Molecule**	**Number of Protein Atoms**	**Number of Ligand Atoms**
1BU4	Endoribonuclease	2GP	782	24
1STP	Streptavidin Complex with Biotin	BTN	901	16
1MFA	Immunoglobulin	ABE	1712	9
3VGC	Serine Protease	SRB	1738	19
1XIG	Isomerase	XYL	3031	10
1CSI	Lyase(Oxo-Acid)	OAA	3391	9

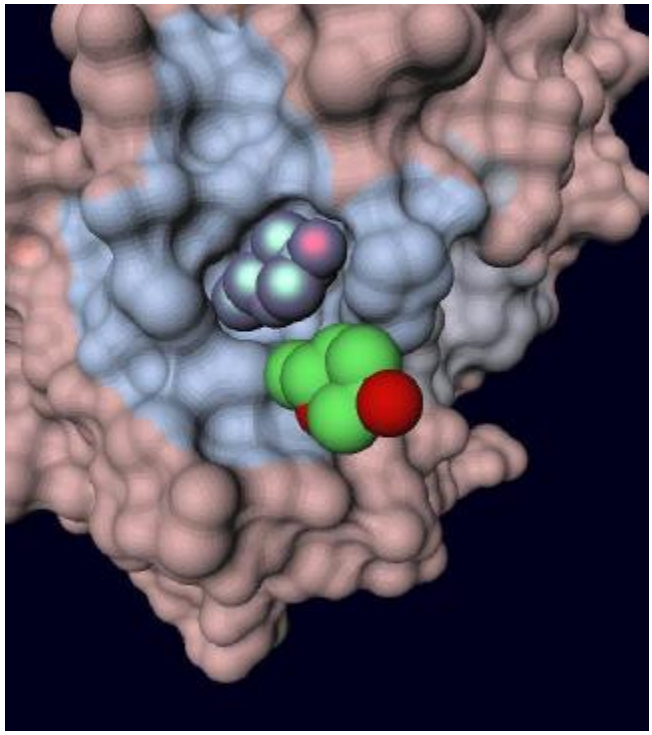


Figure 5.3: Screen capture of system on action, the user is examining around the binding site of 1MFA. For the best energy value obtained so far system keeps the decoy as seen.

5.2.2.2 Experiment II Results

In this experiment, the subjects estimated the position and orientation of a ligand molecule inside the binding cavity of a protein molecule. Following this initial alignment, we executed our distance-minimization algorithm discussed in Section 3.3, to calculate the final configuration of the ligand molecule. One of the drawbacks of the potential energy based approaches used for molecular docking is the possibility of trapping to a local minima.

Potential field created from some potential functions can have local minima and thus create a ‘trapped situation’ for the ligand molecule. Local minima are caused by the potential function and the geometry of the protein surface. Hence, there are more than one configuration on the surface of the protein in which the user feels minimal force (though the energy values might be different). For this reason, potential approaches are typically combined with stochastic approaches to escape from local minima. We used Monte Carlo approach which is based on exploring the energy surface by randomly changing the configuration of the ligand molecule to avoid local minima. If the random move causes an energy drop, we accept the random move. If not, a probabilistic number is generated based on the energy change and compared it with a random number and decision is made accordingly. Details of this approach are explained in Section 3.4. The results are shown in Table 5.3 and Figure 5.4. The results for LSM cases are obtained such that algorithms start to run from same initial conformations and run with a fixed amount of iterations which is equal for both algorithms. Throughout iterations energy levels of system is calculated and finally system is taken to minimum energy conformation faced among iteration history. More detailed results data for experiment II can be examined in Appendix B.

Table 5.3: Experiment II results.

LP pairs (PDB ID)	RMSD of Initial orientation set by user (A°)	RMSD after execution of LSM algorithm (A°)	RMSD after LSM algorithm with Metropolis1 (A°)	RMSD after LSM algorithm with Metropolis2 (A°)
1XIG	4,352583	4.24056	4.273939	4.213841
1MFA	2,187775	1.936589	2.024329	1.711405
1STP	6,15318	5.935669	5.938884	5.031299
3VGC	4,526952	4.173884	3.940335	3.917051
1BU4	3,955809	3.660165	3.426154	3.629912
1CSI	4,293905	4.103001	3.887026	3.749829

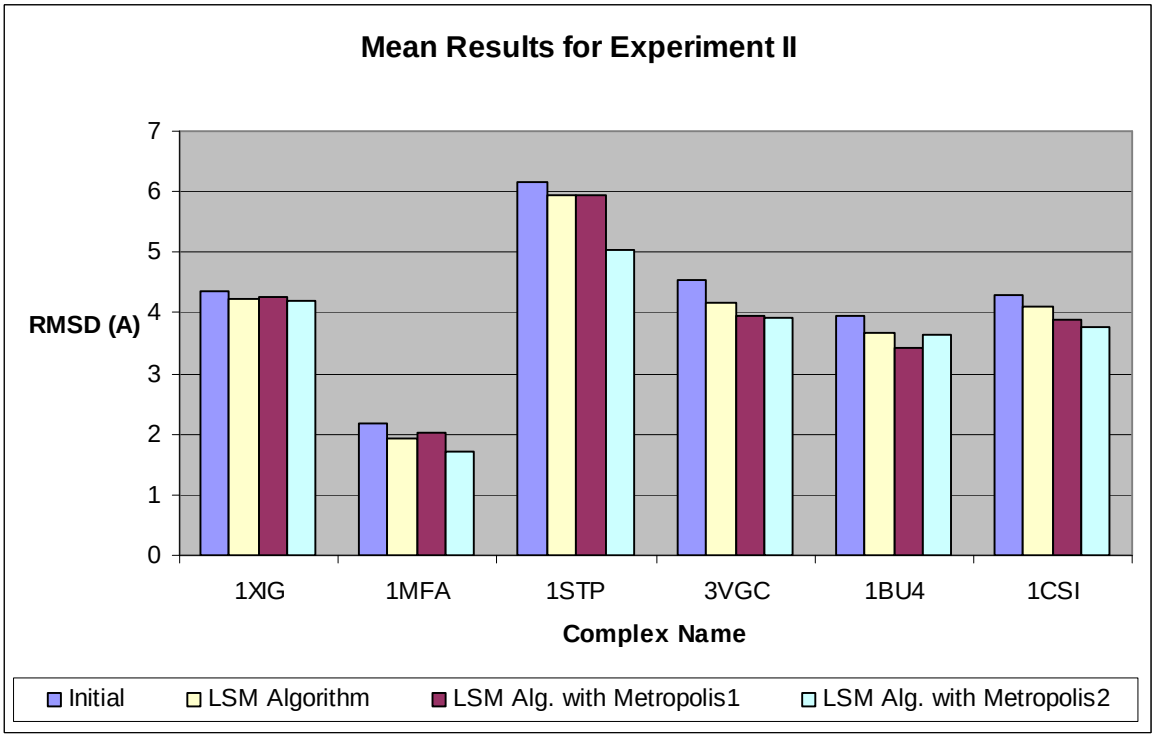


Figure 5.4: Final Mean Results for Experiment II

Chapter 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

It can be argued that human operators are better than computers in complex assembly and disassembly tasks involving insertion and removal of parts. On the other hand, computers are known to be better than human operators in fine-tuning and precision. The work presented in this thesis demonstrates an example of human-machine collaboration in virtual environments for achieving better results in molecular docking. It is expected that the recent advances in bioinformatics and high powered computing continue to stimulate interest in the area of molecular modeling and docking in virtual environments. In this thesis we have investigated the role of force feedback in molecular docking simulations and presented efficient and user friendly rendering methods for haptic visualization of molecular surfaces. The system is used interactively to search for potential binding sites in virtual environments. We also presented a computationally efficient algorithm for rigid molecular docking. After placement and rough alignment of the ligand molecule inside the binding cavity, we executed a new distance-minimizing algorithm to fine tune the position and orientation of the ligand molecule. The results of our experiments with 6 ligand-protein pairs suggest that initial alignment of the ligand molecule via haptic feedback can be used as a precursor to docking algorithms. The proposed two-stage docking approach can be tested with more advanced docking engines to obtain more accurate results of course.

The proposed rigid docking approach and the molecular interaction models have some shortcomings. We assumed that ligand and protein molecules are rigid structures in our

simulations. Although frequently used with success in small-ligand docking, rigidity has some disadvantages. A rigid model does not allow for energy exchange between body and environment. In addition, rigid models are unable to display the conformational changes that some proteins exhibit upon binding. However, it is argued that conformational changes in both protein and ligand are necessary for a successful docking process especially for some particular complexes. In this work we showed attention to select our sample protein-ligand pairs such that they do not exhibit such kind of behavior during binding process. As another missing element in our simulations, we have neglected the contribution of hydrogen bonds. Two interacting molecules can also make a hydrogen bond (if one has a hydrogen atom and interacts with the atoms of the other) during binding, which significantly strengthen the bonds between them. Recent work shows that hydrogen bonding and hydrophobic characteristics of molecules can be critically important in docking process [45]. Another important point regarding our system shows itself as local minima problem. There is a gradient descent method at the heart of our SVD algorithm in order to overcome the local minima problem which is pretty common for these algorithm family we have implemented a Monte Carlo variation of algorithm. MC approach yielded better results, but still due to our mechanism of randomization we do not observe total freedom around local minima. Briefly, randomizations for each atom sum up and as final result molecule does not exhibit big conformational changes.

Finally, the haptic device that we have used in our experiments has had 3 DOF output, i.e. we can feel point forces, whereas torque feedback was not exist. Around the true binding conformation we believe that an additional torque feedback will cause to a big improvement in the quality of haptic feeling. And especially, non-experienced users will come up with a lot better results in a torque feed-backed system.

6.2 Future Work

Speaking for the system resources efficiency, our real time force calculation mechanism's regular force grid based approach is not the optimal solution. Instead, an octree mechanism should yield more efficient memory usage and probably better access times. Throughout thesis work we have implemented an octree based grid structure and integrated it into our system. But the final results did not showed up as promising as we expected so we gave that effort away. A more careful implementation of the method should yield a more efficient and faster system.

The most important improvements for the system will be obviously related with molecular modeling part. Integration of hydrogen bonds and hydrophobicity effects will make the system a lot more reliable for molecular docking problem in a general sense. Also a definitely big improvement can be achieved through implementation of non-rigid molecular behavior. As a first step it is pretty realizable to add non-rigid ligand behavior while SVD fine-tuning algorithm is working. Only thing we need to do is to relax the rigidity of body during iterations. A similar work in computer graphics literature is achieved recently [46]. All the calculations are based upon a non-perfect energy model of van der Waals interactions and electrostatic forces. A better representation of these energies through more precise energy and force models will definitely improve the results.

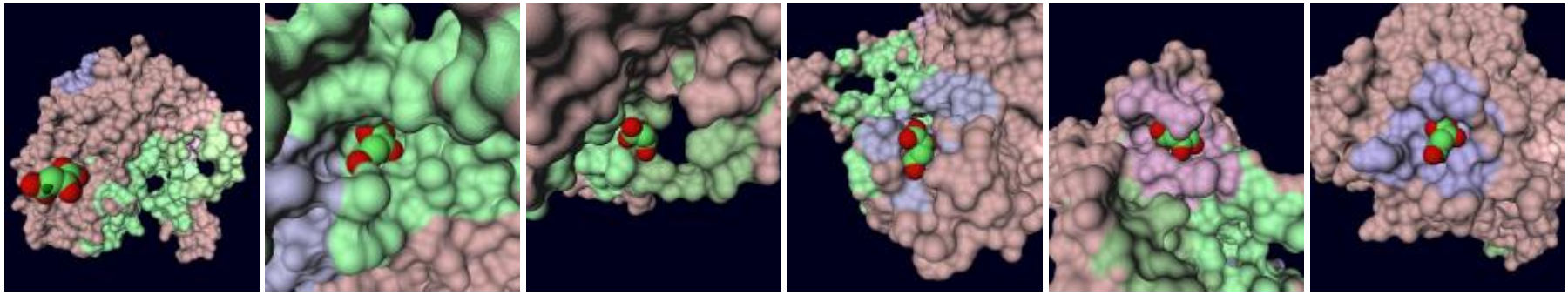
As we have discussed in the conclusion section the MC approach that we have implemented does not provide big conformational changes for ligand molecule. In that case for a totally false initial configuration it is pretty hard for our system to recover it. Instead of randomizing the forces on particles, another approach might be to give random conformational changes to ligand around its principal axes. For most cases molecular structures show symmetric behaviors, so we envision a symmetric but false initial position can be corrected using such an approach. Though, implementation of such an approach in our

system is pretty straight forward, since we are making use of SVD calculations we already have principal axes of ligand molecule.

APPENDIX A

In this appendix, the results of the first experiment (Experiment I) are given in detail. There are 6 pages: one for each molecular complex used in the experiment and tested by 6 different subjects. Each page contains a series of 6 different snapshots of the same molecular complex and a table of results related to that complex. The table presents the results of the subject. Note that the experiment was repeated after one day rest period (i.e. there are two numbers for each subject separated by a comma). The first snapshot contains surface models of ligand and protein molecules. The following 5 snapshots are the potential binding sites displayed to the subjects of the experiment. These sites are determined using a public domain package *Pocket*. This package returns the potential binding sites in the order of importance. We randomized this order to eliminate the bias. The little red star on the top of one of the 5 snapshots indicates the true binding site. It is interesting to note that while the *Pocket* ranked the true binding site for 1 BU4 as the 3rd cavity based on the surface area, the subjects were successful in discovering it using visual and haptic cues (10 out of 12 cases).

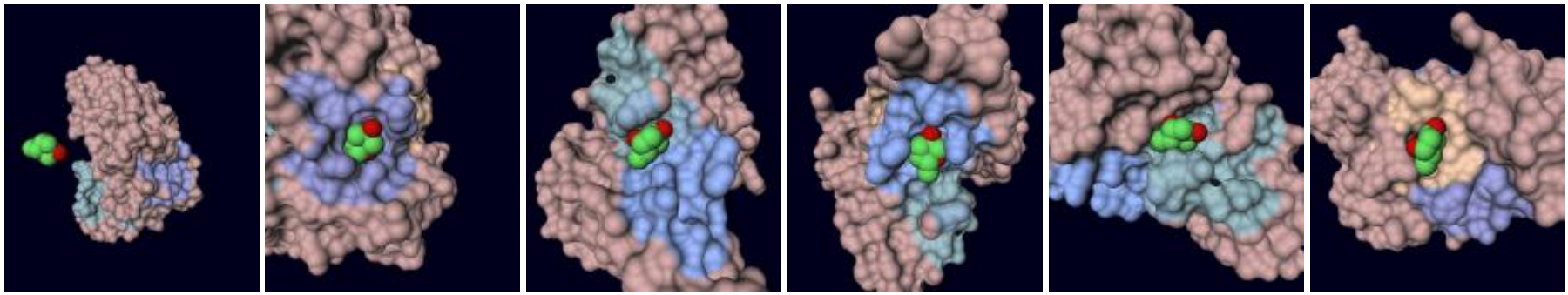
Complex Name: 1CSI



# of times selected as best candidate	6	1	-	5	-
Aydın	1,1	5,5	2,2	3,3	4,4
İbrahim	1,2	5,4	4,3	2,1	3,5
İhsan	2,3	1,4	5,5	3,1	4,2
Mert	2,2	5,5	3,3	1,1	4,4
Nesra	1,1	4,2	3,4	5,5	2,3
Yaşar	2,1	3,5	4,3	1,2	5,4
Mean :	1.58	4.00	3.41	2.33	3.66

Complex Name: 1MFA

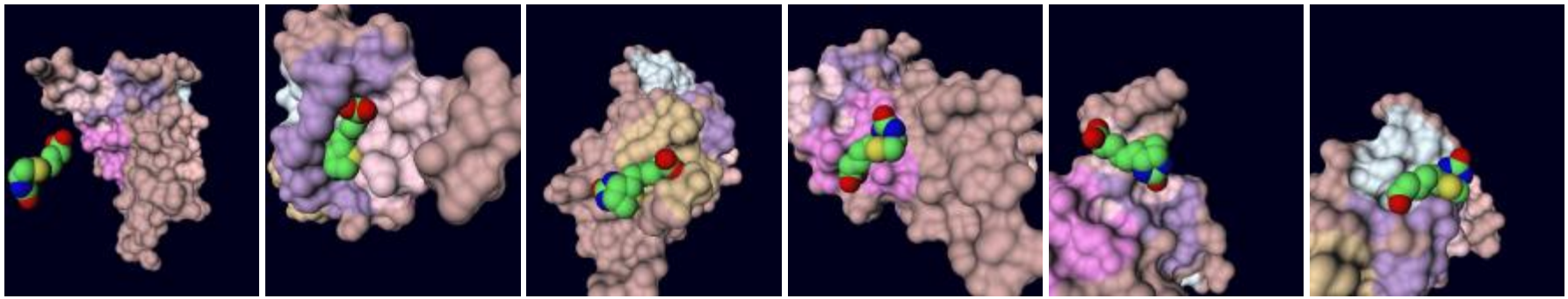
*



# of times selected As best candidate	6	1	1	1	3
Aydın	3,3	1,5	4,4	2,2	5,1
İbrahim	1,1	4,3	5,4	2,5	3,2
Ihsan	5,2	4,5	3,3	1,4	2,1
Mert	1,1	3,5	4,2	2,4	5,3
Nesra	1,1	3,4	4,5	2,3	5,2
Yaşar	4,2	3,4	1,3	2,5	5,1
Mean :	2.08	3.66	3.5	2.83	2.91

Complex Name: 1STP

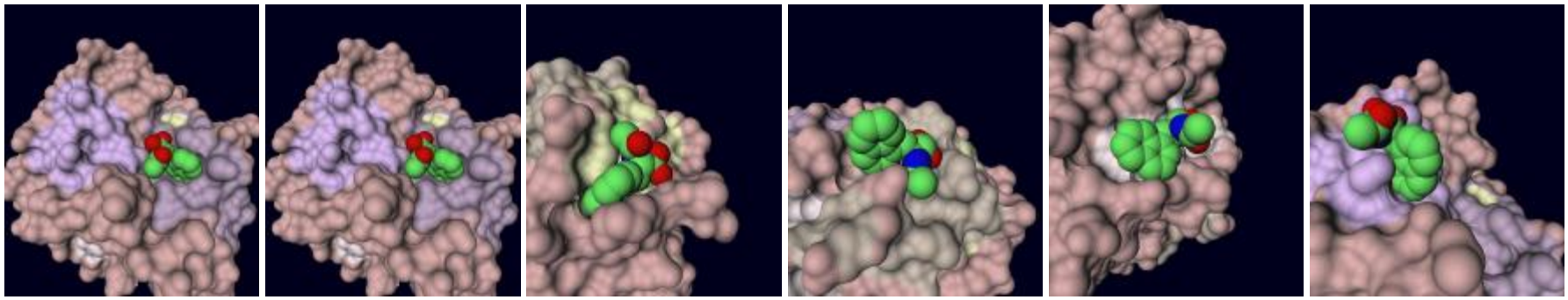
*



# of times selected as best candidate	12	-	-	-	-
Aydın	1,1	4,2	5,4	3,5	2,3
İbrahim	1,1	2,5	5,3	3,2	4,4
İhsan	1,1	2,2	5,3	3,5	4,4
Mert	1,1	3,2	2,3	5,5	4,4
Nesra	1,1	2,2	3,4	5,3	4,5
Yaşar	1,1	2,4	3,3	5,5	4,2
Mean :	1.0	2.66	3.58	4.08	3.66

Complex Name: 3VGC

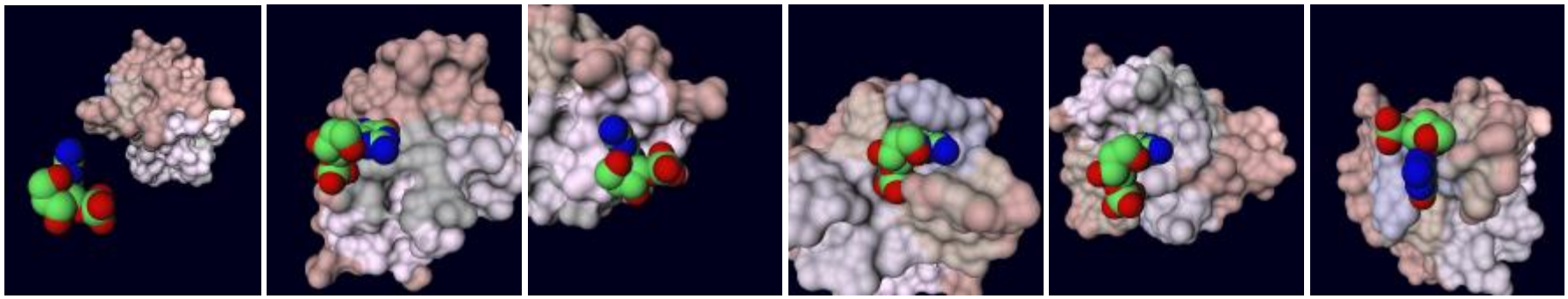
*



# of times selected as best candidate	8	2	1	1	-
Aydın	3,2	4,1	2,3	1,5	5,4
İbrahim	1,1	2,5	4,3	3,2	5,4
İhsan	1,3	3,2	2,1	4,4	5,5
Mert	2,1	1,2	4,4	3,5	5,3
Nesra	1,1	4,2	3,5	2,4	5,3
Yaşar	1,1	3,2	2,4	4,5	5,3
Mean :	1.50	2.58	3.08	3.5	4.33

Complex Name: 1BU4

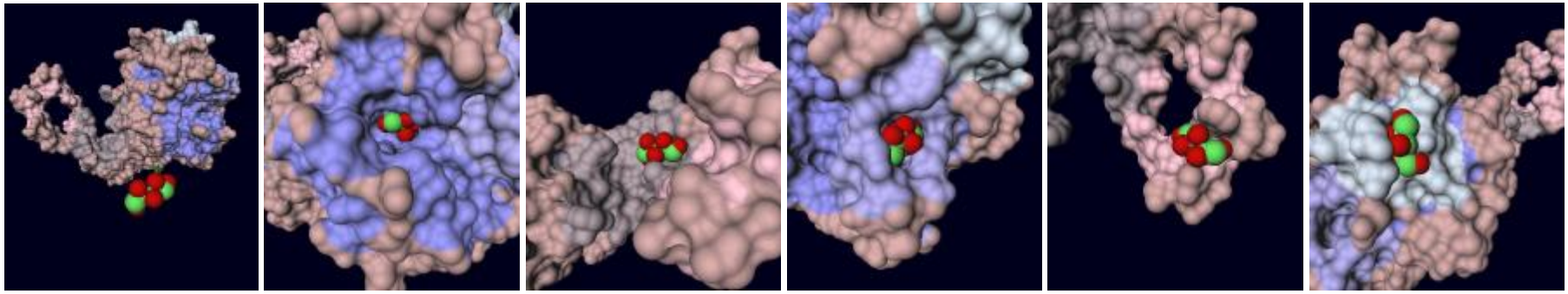
*



# of times selected as best candidate	2	-	10	-	-
Aydın	3,3	5,2	1,1	4,5	2,4
İbrahim	3,1	2,5	1,3	5,4	4,2
İhsan	4,3	2,2	1,1	3,5	5,4
Mert	4,5	2,2	1,1	3,3	5,4
Nesra	1,4	5,3	2,1	3,5	4,2
Yaşar	3,5	5,2	1,1	2,4	4,3
Mean :	3.25	3.08	1.25	3.83	3.58

Complex Name: 1XIG

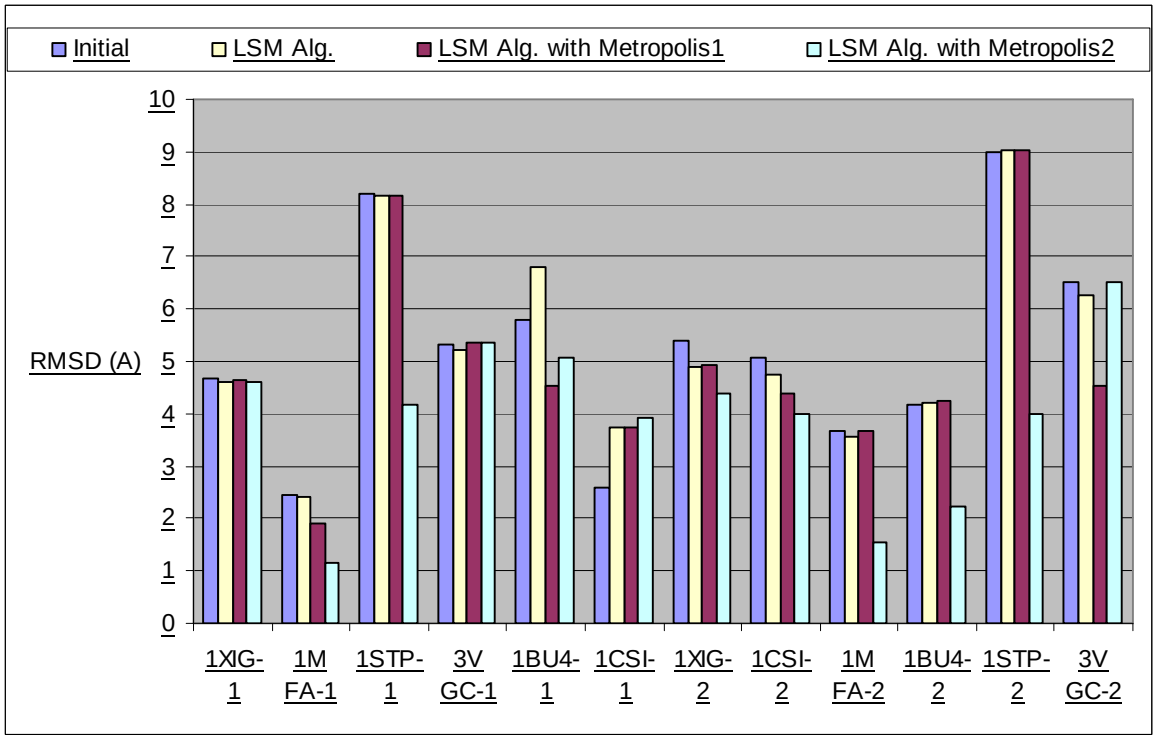
*



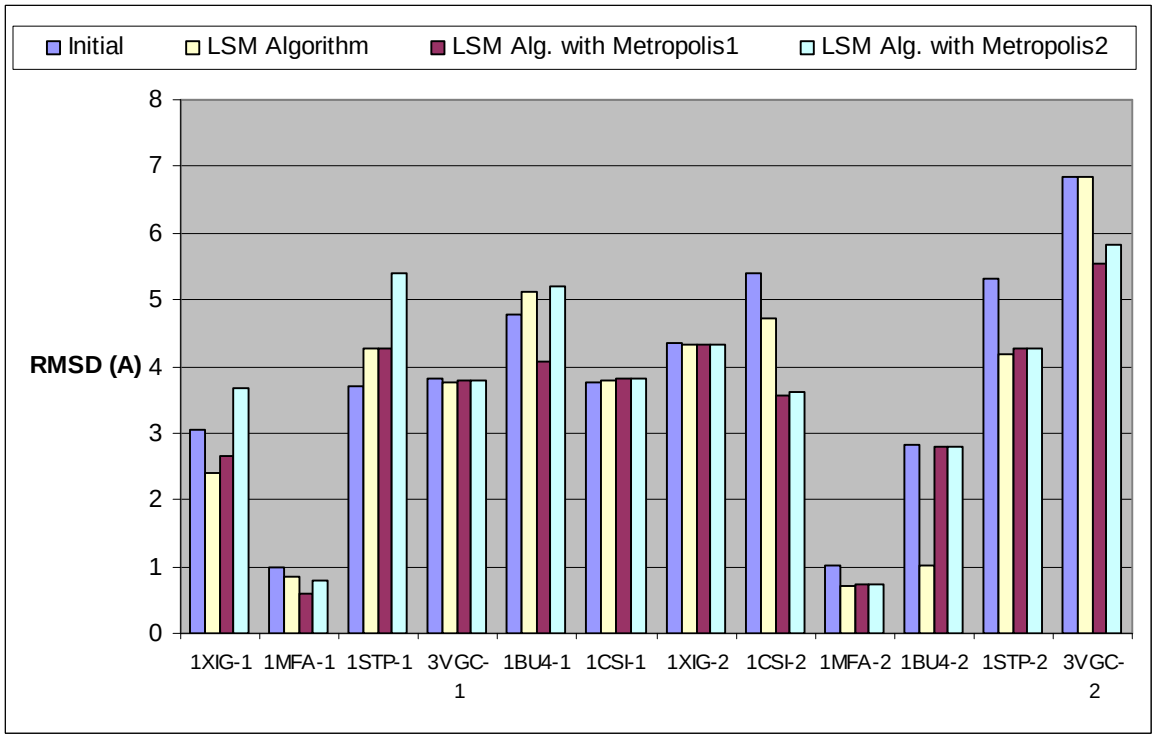
# of times selected as best candidate	8	-	3	-	1
Aydın	2,2	3,5	1,4	4,3	5,1
İbrahim	1,1	5,5	4,2	3,4	2,3
İhsan	1,1	5,5	4,2	3,4	2,3
Mert	1,2	5,4	3,1	4,5	2,3
Nesra	1,2	3,4	4,1	5,3	2,5
Yaşar	1,1	5,5	4,2	3,4	2,3
Mean :	1.33	4.5	2.66	3.75	2.75

APPENDIX B

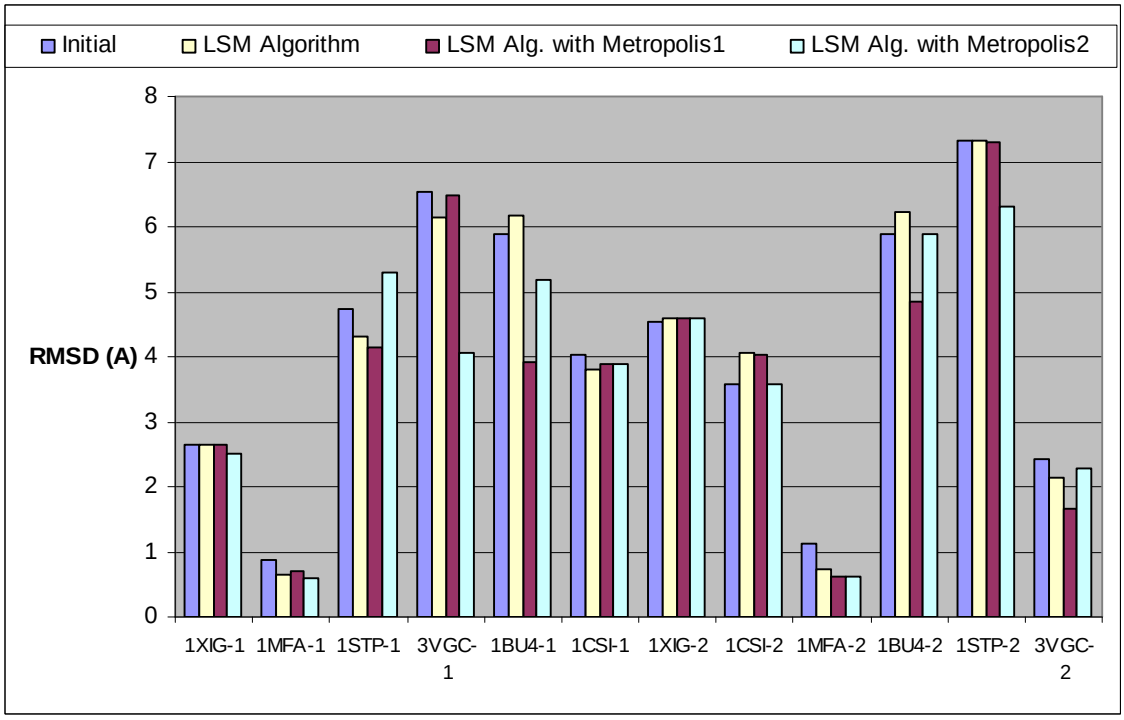
In this appendix, the results of the second experiment (Experiment II) are given in detail. There are 3 pages: each page contains the results of 2 subjects. The bars in each graph represents the initial alignment achieved by the subject, the results obtained by the distance-minimization algorithm (SVD algorithm is used to calculate the optimum transformation) after the initial alignment, and the results of the Monte Carlo approach after the initial alignment. Note that the experiment was repeated after one-day rest period with different molecular conformations (we have selected two of the three previously decided view angles of same molecule while doing this). Hence, for example, 1MFA1 and 1MFA2 represents the first and second trials on the x-axis of the plots.



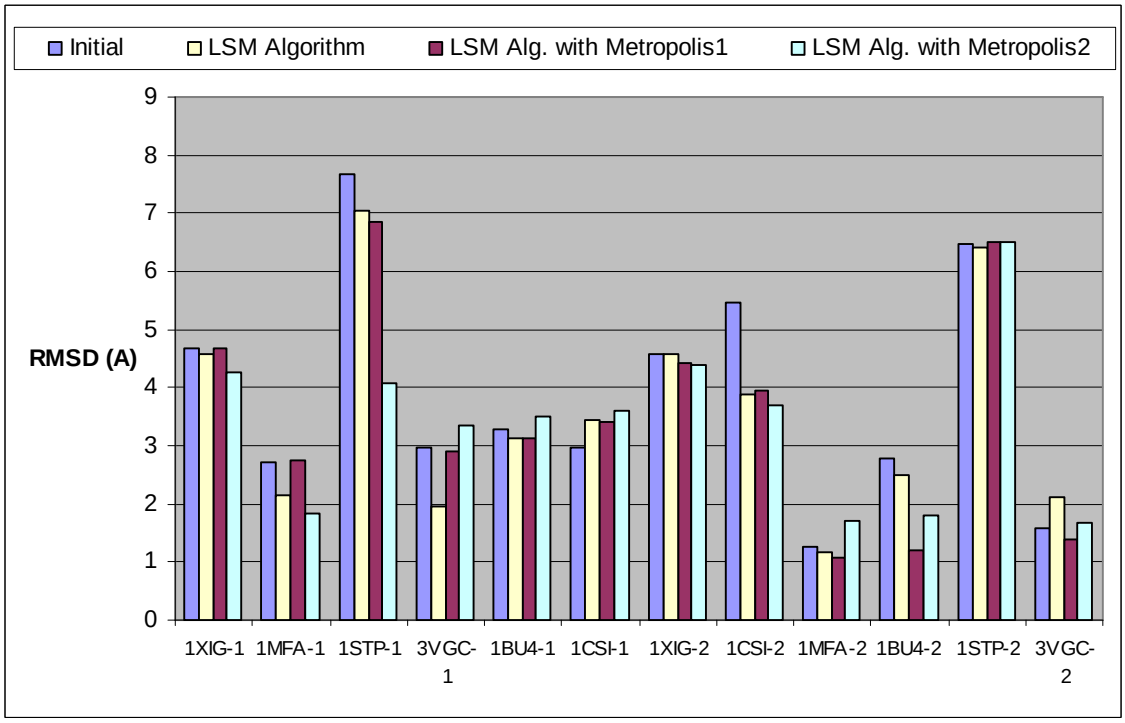
Results for Experiment II for User Aydın



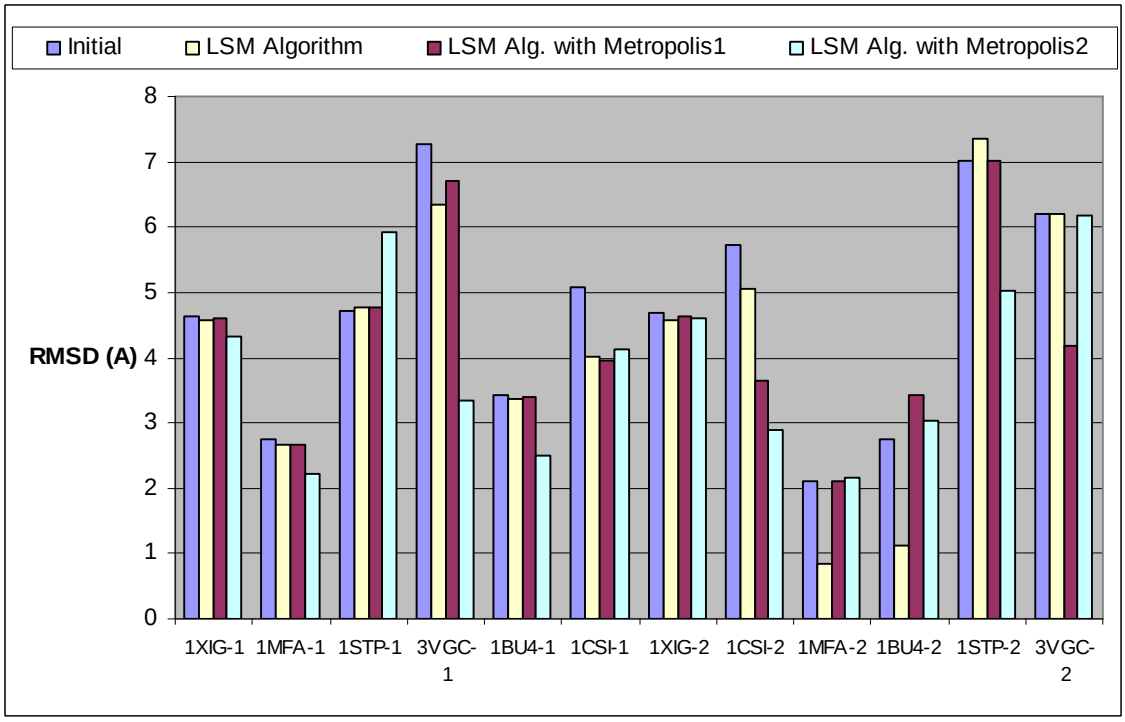
Results for Experiment II for User İbrahim



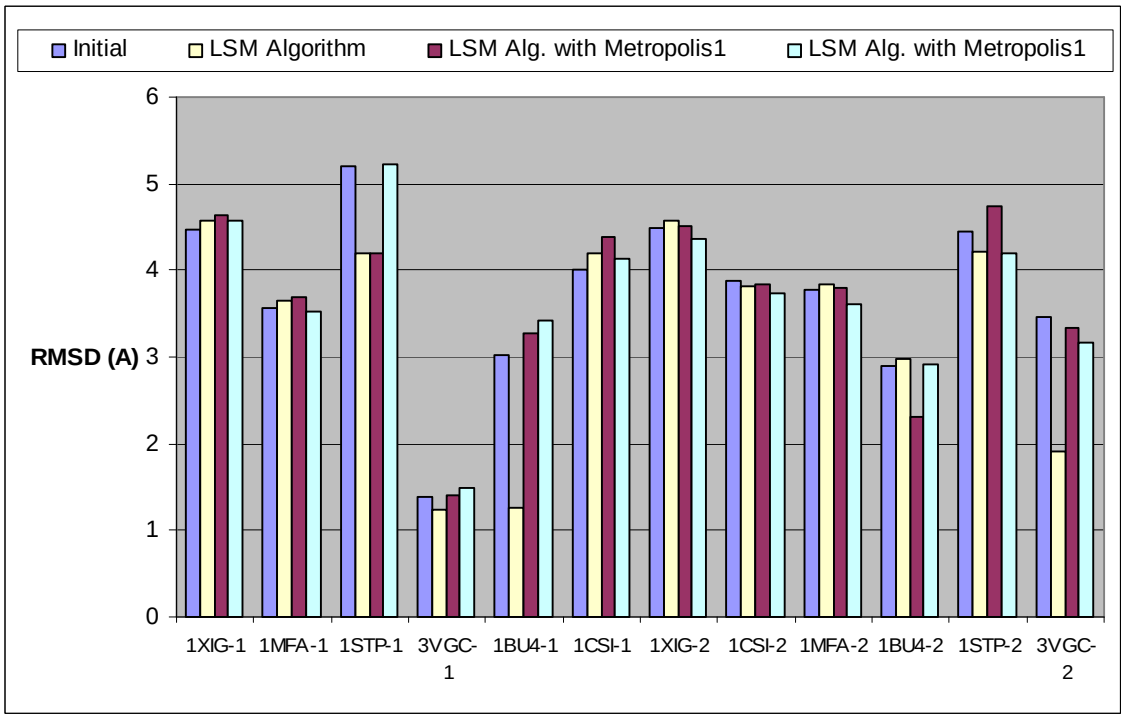
Results for Experiment II for User İhsan



Results for Experiment II for User Mert



Results for Experiment II for User Nesra



Results for Experiment II for User Yaşar

BIBLIOGRAPHY

[1] M. Teodoro, G.N. Phillips, Jr., L. Kavraki. “Molecular Docking: A Problem with hundreds of thousands of freedom”, IEEE International Conference on Robotics and Automation (ICRA), Seoul, Korea, pp.960-900, 2001.

[2] Halperin, Ma, Wolfson, Nussinov, "Principles of docking: An overview of search algorithms and a guide to scoring functions", Proteins, 2002.

[3] Ewing, Makino, Skillman, and Kuntz, "DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases," J. CAMD, 2001.

[4] G. Morris, D. Goodsell, R. Halliday, R. Huey, W. Hart, R. Belew, and A. Olson, “Automated docking using a lamarckian genetic algorithm and empirical binding free energy function,” J. Comp. Chem., vol. 19, pp. 1639 -1662, 1998.

[5] I. Kuntz, “Structure-based strategies for drug design and discovery,” Science, vol. 257, pp. 1078 - 1082, 1992.

[6] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, “A fast flexible docking method using an incremental construction algorithm”, J. Mol. Biol., vol. 261, pp. 470-489, 1996.

[7] M. Miller, S. Kearsley, D. Underwood, and R. Sheridan, “Flog: A system to select 'quasi flexible' ligands complementary to a receptor if known three dimensional structure” J. Comp.-Aid. Mol. Des., vol. 8, pp. 153-174, 1994.

[8] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. Friesem, and C. A. I. Vakser, “Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques”, in Natl. Acad. Sci. USA, vol. 89 pp. 2195-2199, 1992.

[9] Cole JC, Murray CW, Nissink JW, Taylor RD, Taylor R., “Comparing protein-ligand docking programs is difficult.” Proteins. , vol. 60(3), pp. 325-332, 2005.

[10] Anderson A, Weng Z., “VRDD: applying virtual reality visualization to protein docking and design.”, J Mol Graph Model. vol.17 (3-4), pp.180-186, 1999.

[11] Levine, D., Facello, M., Hallstrom, P., Reeder, G., Walenz, B., and Stevens, F. Stalk: An interactive system for virtual molecular docking. IEEE Comput. Sci. Eng., vol. 4, pp. 55–65, 1997.

[12] Akkiraju, N., Edelsbrunner, H., Fu, P., and Qian, J. Viewing geometric protein structures from inside a CAVE. IEEE Comput. Graphics Appl. ,vol. 16, pp. 58–61, 1996.

[13] Czernuszenko, M., Pape, D., Sandin, D.J., DeFanti, T.A., Dawe, G.L., and Brown. M. The ImmersaDesk and Infinity Wall projection-based virtual reality displays. Comput. Graphics, vol. 31, pp. 46–49, 1997.

[14] Stefan Birmanns and Willy Wriggers. Interactive Fitting augmented by force-feedback and virtual reality. J. Struct. Biol., vol. 144, pp. 123-131, 2003.

[15] Frederick P. Brooks, Jr. , Ming Ouh-Young , James J. Batter , P.Jerome Kilpatrick,"Project GROPE Haptic displays for scientific visualization", ACM SIGGRAPH Computer Graphics, vol24, pp.177-185, Aug. 1990.

[16] Ganesh Sankaranarayanan, Suzanne Weghorst, Michel Sanner, Alexandre Gillet, Arthur Olson. "Role of Haptics in Teaching Structural Molecular Biology," Haptics, vol. 1, pp. 363, Nov. 2003.

[17] Yong-Gu Lee and Kevin Lyons, "Smoothing haptic interaction using molecular force calculations", Computer-Aided Design, vol.36, pp. 75-90, 2004.

[18] Rapaport, D.C., The art of molecular dynamics simulation, Cambridge University Press, 2002.

[19] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE., "A Geometric approach to macromolecule-ligand interactions.", Journal of Molecular Biology ,vol. 161, pp. 269-288, 1982.

[20] Apaydin MS, Guestrin C, Varma C, Brutlag DL, Latombe J-C, "Stochastic roadmap simulation for the study of ligand-protein interactions.", Bioinformatics, vol.18, pp.18-S26, 2002.

[21] Singh,A.P., Latombe,J.C. and Brutlag,D.L., "A motion planning approach to flexible ligand binding. International Conference on Intelligent Systems for Molecular Biology." Heidelberg, pp. 252–261, 1999.

[22] O. Burchan Bayazit, Guang Song, Nancy Amato, “Ligand Binding with OBPRM and Haptic User Input.”, Proceedings of the 2001 IEEE International Conference on Robotics and Automation (ICRA'01), pp. 954—959, 2001.

[23] Gillilan and Lilien, “Optimization and Dynamics of Protein–Protein Complexes Using B-Splines”, Journal of Computational Chemistry, vol. 25, no. 13, 2004.

[24] Timothy R. Forester, William Smith, “SHAKE, Rattle, and Roll: Efficient Constraint Algorithms for Linked Rigid Bodies”, Journal of Computational Chemistry, vol. 19, no. 1, pp. 102-111, 1998.

[25] Pyne, S., “Geometric methods in Molecular Docking”, http://www.cs.sunysb.edu/~spyne/

[26] John E. Stone, Justin Gullingsrud, Klaus Schulten, Paul Grayson , “A System for Interactive Molecular Dynamics Simulation", 2001 ACM Symposium on Interactive 3D Graphics, ACM SIGGRAPH, pp. 191-194, 2001.

[27] Besl, P.J., McKay N.D. “A method for registration of 3d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence”, vol. 14, no. 2, pp. 239-256, 1992.

[28] Chen, Y., Medioni. “Object modeling by registration of multiple depth maps. Image and Vision Computing”, vol. 10, no. 3, pp. 145-155, 1992.

[29] Horn, B.K., Hilden, H., Negahdaripour, S. “Closed-form solution of absolute orientation using orthonormal matrices.”, Journal of the Optical Society of America A, vol. 5, no. 7, 1988.

[30] Available at http://hyper.stanford.edu/~koehl/ProShape/

[31] H. Edelsbrunner and P. Koehl., "The geometry of biomolecular solvation.” Combinatorial and Computational Geometry (MSRI Publications) vol. 52, pp. 243-275, 2005.

[32] Y. Koren and J. Borenstein, "Potential field methods and their inherent limitations for mobile robot navigation.", Proc. International Conference on Robotics and Automation, Sacramento, California, 1991.

[33] Taylor R., et al., “VRPN: a device-independent, network-transparent VR peripheral system”. Proceedings of the ACM symposium on Virtual reality Software and Technology, Baniff, Alberta, Canada, pp. 55 – 61, 2001.

[34] GHOST Software Developer's Toolkit Programmer's Guide Version 2. Cambridge, MA: Sensable Technologies, 1998.

[35] T. H. Massie and J. K. Salisbury, “The PHANToM haptic interface: A device for probing virtual objects,” Int. Mechanical Engineering Exposition and Congress (Chicago), DSC 55-1, pp. 295-302, 1994.

[36] Basdogan, C., Ho, C, and Srinivasan, M.A., “Virtual Environments for Medical Training: Graphical and Haptic Simulation of Common Bile Duct Exploration,” IEEE-ASME Trans. Mechatronics, vol. 6, no. 3, pp. 267-285, 2001.

[37] Basdogan C., Srinivasan M.A., “Haptic Rendering in Virtual Environments”, Handbook of Virtual Environments, Ed: K. Stanney, pp. 117-134, 2002.

[38] Available at http://oss.sgi.com/projects/inventor/

[39] Srinivasan, M.A., Basdogan, C., “Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges (PDF)”, *Computers and Graphics*, vol. 21, no. 4, pp. 393-404, 1997.

[40] ML Connolly. “Solvent-accessible surfaces of proteins and nucleic acids.” Science, vol. 221, pp. 709-713, 1983.

[41] Kollman, P.A. “AMBER 5.0”, University of California, San Francisco, 1996.

[42] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. J. Comput. Chem., vol 4, pp. 187–217, 1983.

[43] Refson K., Moldy: “A portable molecular dynamics simulation program for serial and parallel computers.”, Computer Physics Communications, vol126, no.3, pp. 309-328, 2000.

[44] Mark Nelson, William Humphrey, Attila Gursoy, Andrew Dalke, Laxmikant Kalé, Robert D. Skeel, and Klaus Schulten.. “NAMD - A parallel, object-oriented molecular dynamics program.”, *International Journal of Supercomputer Applications and High Performance Computing*, vol. 10, pp. 251-268, 1996.

[45] O. Keskin, B. Ma, R. Nussinov, "Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues". *J Mol Biol.* vol. 345, no. 5, pp. 1281-94, 2005.

[46] Muller M., Heidelberger B., Teschner M., Gross M., “Meshless deformations based on shape matching.” Siggraph 2005.

[47] Guthold, M., G. Matthews, A. Negishi, R. M. Taylor II, D. Erie, F. P. Brooks Jr. and R. Superfine. "Quantitative Manipulation of DNA and Viruses with the nanoManipulator Scanning Force Microscope." Surf. Interf. Analysis vol. 27, pp. 437-443, 1999.