

Haptic-Assisted Interactive Molecular Docking

Incorporating Receptor Flexibility

Nick Matthews,[†] Akio Kitao,[‡] Stephen Laycock,^{*,†} and Steven Hayward^{*,†}

*[†]School of Computing Sciences, University of East Anglia, Norwich Research Park,
Norwich NR4 7TJ, UK*

*[‡]School of Life Science and Technology, Tokyo Institute of Technology, 2-12-1 Ookayama,
M6-13, Meguro, Tokyo 152-8550, Japan*

E-mail: s.laycock@uea.ac.uk; steven.hayward@uea.ac.uk

Abstract

Haptic-assisted interactive docking tools immerse the user in an environment where intuition and knowledge can be used to help guide the docking process. Here we present such a tool where the user “holds” a rigid ligand via a haptic device through which they feel interaction forces with a flexible receptor biomolecule. To ensure forces transmitted through the haptic device are smooth and stable, they must be updated at a rate greater than 500 Hz. Due to this time constraint, the majority of haptic docking tools do not attempt to model the conformational changes that would occur when molecules interact during binding. Our haptic-assisted docking tool, “Haptimol FlexiDock”, models a receptor’s conformational response to forces of interaction with a ligand whilst maintaining the required haptic refresh rate. In order to model receptor flexibility we use the method of linear response for which we determine the variance-covariance matrix of atomic fluctuations from the trajectory of an explicit-solvent Molecular Dynamics simulation of the ligand-free receptor molecule. Key to satisfying the time constraint is an eigenvector decomposition of the variance-covariance matrix which enables a good approximation to the conformational response of the receptor to be calculated rapidly. This exploits a feature of protein dynamics whereby most fluctuation occurs within a relatively small subspace. The method is demonstrated on Glutamine Binding Protein in interaction with glutamine, and Maltose Binding Protein in interaction with maltose. For both proteins the movement that occurs when the ligand is docked near to its binding site matches the experimentally determined movement well. It is thought that this tool will be particularly useful for structure-based drug design.

Introduction

Molecular docking refers to computational methods used to predict how two molecular structures join together to form a complex. It is an important tool in the fields of protein-protein interaction¹, protein-ligand interaction² and structure-based drug design (SBDD)³.

One can classify molecular docking into two types: automated and interactive. For automated docking, pose selection algorithms are employed to search for possible binding poses⁴⁻⁷. This can result in a large number of conformations, which are then scored according to their binding energy. The “correct” pose should result in a high score. Interactive systems put the user in charge of the docking process, allowing them to use their knowledge and intuition to find a docked pose^{8,9}.

One way of enhancing interactive docking is to use a haptic device. Haptic-assisted docking systems allow 3D manipulation and interaction forces to be felt, enabling the user to guide the process through their sense of touch, naturally avoiding high-energy interactions. Haptics brings profound advantages over the use of a simple mouse. As Aspuru-Guzik et al.¹⁰ state, haptic devices add “a new level of intuition to the virtual experience of the molecular world that goes far beyond its archaic and fractured perception through computer mouse and keyboard.” It is thought that such state-of-the-art visualisation technologies will have considerable impact on SBDD¹¹.

Besides offering an environment to rapidly test new ideas and hypotheses, haptic-assisted interactive docking systems can be used in conjunction with automated systems to allow experts to test high scoring poses and either improve them or reject them^{12,13}. Interactive docking tools have also been shown to improve the users’ understanding of the process of molecular binding¹⁴⁻¹⁶.

During docking, the binding of a ligand to a receptor often gives rise to specific forces that induce conformational change¹⁷⁻¹⁹. Modelling this flexibility presents a challenge, in both automated and interactive docking systems, due to the large number of degrees of freedom in biomolecules making it a computationally expensive task.

Receptor Flexibility in Automated Docking

Below we refer to the large biomolecule to which another “ligand” molecule binds as the “receptor”. Incorporating receptor flexibility into automated docking has been an active area of research for many years. Recent review papers^{5,6} have covered the topic in detail.

Within automated docking, receptor flexibility can be modelled both explicitly by modelling a protein’s change in conformation during docking, or implicitly, that is without actively deforming the structure. Implicit approaches tend to be less computationally expensive than the explicit methods⁶.

The least computationally expensive way to model flexibility implicitly is to use the soft-docking approach²⁰. In soft docking, small overlaps are allowed between receptor and ligand atoms by softening the van der Waals’ repulsive term between atoms as a way of simulating minor conformational plasticity and resulting in a larger binding site⁶. The main advantage that soft docking has over other approaches to modelling molecular flexibility is that it adds no extra computational cost compared to rigid docking. However, soft docking can only model small, local conformational changes rather than large global ones that are often related to function. Sometimes, soft docking is used in conjunction with more advanced modelling methods^{21,22}.

Ensemble docking attempts to dock a ligand to multiple conformations of the same receptor instead of a single one²³. Flexibility is accounted for implicitly, in the sense that the protein will not deform in response to the ligand during the docking process. Studies have shown that ensemble docking provides significant improvement over single-conformation docking^{24,25}. However, it has been found that the additional poses can increase the number of false positives generated during docking²⁶.

In other docking approaches, receptor flexibility is modelled explicitly. However, due to the computational cost which results from the large number of degrees of freedom involved, methods have been developed that reduce the complexity of the problem.

The earliest approaches only modelled partial flexibility within the receptor, in particular,

only modelling side-chain flexibility²⁷. For example, rotamer libraries have been used to reduce the computational cost of calculating new side-chain positions, although the number of poses within the library⁶ is necessarily limited. Full molecular flexibility approaches have also been reported. One approach, presented by Sherman et al.²¹, is to dock the ligand into the receptor using soft docking and then to explore various conformations of the receptor using rotamer libraries. Another approach also starts with soft docking and after a potential pose has been found, the ligand and receptor poses are optimised using a molecular dynamics simulation (MD)²⁸, Monte-Carlo²⁹ or energy minimisation³⁰.

Lower-dimensional representations of a molecule’s fluctuation space have also been used to model its response to binding. These approaches use the dominant modes of the biomolecule’s motion. Both Normal Mode Analysis (NMA)^{29,31,32}, and Principle Component Analysis (PCA)^{33,34} have been used as dimensionality reduction techniques. These techniques have been shown to be an effective way of reducing the computational expense of incorporating flexibility. Zacharias³⁴ incorporated “soft” modes calculated using PCA into a docking application, and found that whilst rigid docking failed to identify a docking site close to the experimentally derived site, including them resulted in a docking pose local to the experimentally derived energy minimum. The protein used was the FK506-binding protein. Tatsumi et al.³³ notes that the system developed by Zacharias³⁴ ignores local flexibility, and so demonstrated a hybrid algorithm which uses a similar approach to Zacharias³⁴ for global motions and conventional MD for local motions. They found that their hybrid method reproduced global fluctuations that were not present in ordinary MD docking simulations.

Interactive Docking

Most interactive docking tools employ a haptic device. Despite the various ways automated docking has incorporated flexibility, few haptic-assisted interactive docking methods that have been reported model flexibility. The primary reason for this is due to the time constraints that come with working at an interactive refresh rate. Modern haptic technology

requires the haptic device to be updated at a refresh rate higher than 500Hz, ideally 1kHz, because of the sensitivity of the human haptic system^{35,36}. If a lower refresh rate is used, device vibrations caused by force discontinuities will occur. As a result of this the majority of haptic-assisted interactive molecular docking systems limit themselves to rigid docking.

Most existing haptic-assisted molecular docking systems also make use of precomputed force grids³⁷ to reduce the number of computations required. Grid based applications, first proposed by Brooks et al.³⁸ treat either one, or both of the docking molecules as rigid, and then pre-compute the desired interaction forces (usually van der Waals and electrostatic), around the receptor³⁸⁻⁴³, or part of the receptor⁴⁴. The fundamental limitation of this approach is that owing to the expensive pre-computation step, receptor flexibility cannot be accommodated, as the grid would need to be recomputed after every deformation step. Further limitations are that rough force transitions can be felt at grid cell boundaries⁴⁴, and the precomputed grid consumes large amounts of memory.

For very small molecules, it is possible to interactively solve a docking scenario as demonstrated by Nagata et al.⁴⁵, Férey et al.⁴⁶ and Hou and Sourina⁸, where interaction forces between all of the atoms within the ligand and the receptor are computed in real time. On a modern computer processor, this approach can accommodate molecules up to a few hundred atoms in size⁴⁷. Implementations with molecules much larger than this are impractical on the CPU. Modern GPUs (Graphics Processing Units) are, however, suited to this problem as a large number of computations can be performed in parallel. Haptimol_RD, a docking system presented by Iakovou et al.⁴⁸, made use of the GPU to model rigid docking of two large molecules. Importantly for our docking approach, in which receptor flexibility is modelled, the method by Iakovou et al.⁴⁸ computes the interaction forces in real time, rather than relying on any pre-computation and, as a result, it can be applied to a flexible docking problem. The effectiveness of this approach was demonstrated in Iakovou et al.⁴⁹.

Other approaches to docking include the work by Daunay et al.⁵⁰, Zonta et al.⁵¹ and Anthopoulos et al.⁵². Daunay et al.⁵⁰ developed a system that modelled flexibility by using

a molecular dynamics engine to compute the relevant forces. This approach proved too costly to compute within the haptic time constraint, so wave transformations were used to bridge the gap between rendering and simulation. Zonta et al.⁵¹ presented a system that modelled ligand flexibility by using a third party library to accelerate force computations. Anthopoulos et al.⁵² incorporated a GPU-accelerated force calculation approach within their molecular modelling system⁵³. Although methods presented by Anthopoulos et al.⁵³ model flexibility to some degree, the forces on the haptic device were updated at the 33Hz refresh rate of the display, well below that required for a smooth haptic experience.

Currently, haptic-assisted interactive molecular docking systems that achieve a haptic refresh rate of at least 500Hz have either limited themselves to docking rigid molecules, (comprising up to 184k atoms⁴⁸) and rigid receptor with a flexible ligand where the ligand comprises a few atoms⁵¹.

Stocks et al.⁵⁴ used an Elastic Network Model to study how large molecules fluctuate when forces are applied to individual atoms. In order to allow large proteins to be studied they utilised the “important subspace” to reduce the size of the matrices used to calculate the proteins’ response. By doing this, they reduced both the memory consumption and computational cost.

In this paper, we build on the work of Ikeguchi et al.¹⁹ who used the theory of linear response to model conformational change in a protein in interaction with a ligand. Linear response theory states that the response of a system to an external perturbation can be approximated using equilibrium fluctuations of the system in its unperturbed state. Ikeguchi et al.¹⁹ showed that equilibrium fluctuations of the ligand-free protein could indeed be used to accurately predict conformational change in response to the binding of a ligand which provides the external perturbation. We build on these findings and also those of Stocks et al.⁵⁴ to produce a system (see Figure 1) to calculate the conformational response of a receptor molecule to the binding of a rigid ligand molecule. The application also builds upon our previous work for rigid molecular docking implemented in Hap-

timol_RD⁹ to create Haptimol FlexiDock, which maintains an interactive haptic refresh rate on a GPU equipped desktop PC when performing interactive docking between a large flexible receptor and a rigid ligand. Haptimol FlexiDock is available for download here: <http://www.haptimol.co.uk/flexidock/haptimolflexidockinstaller.zip>.

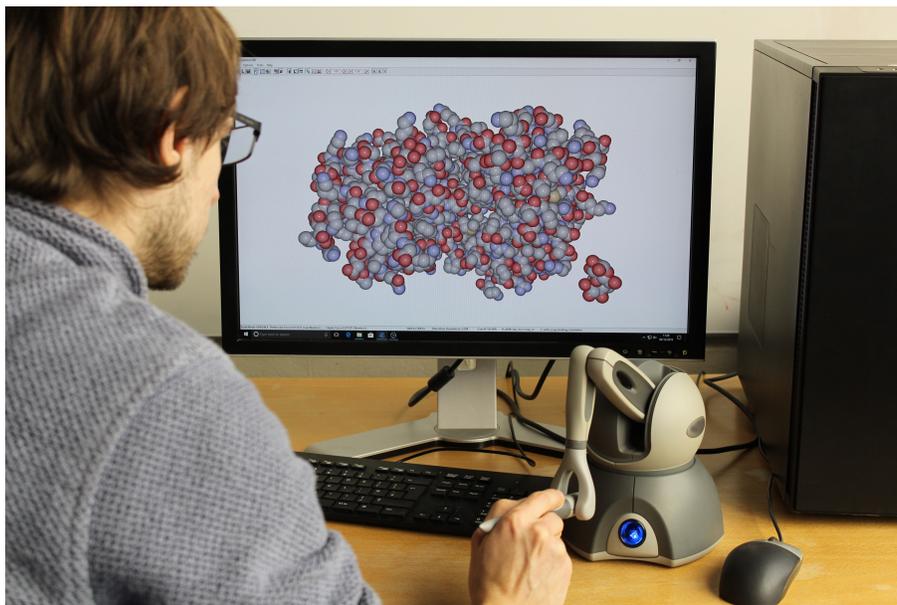


Figure 1: A user docking maltose to maltose binding protein using Haptimol FlexiDock

Methods

Force Calculation

Like other haptic-assisted interactive molecular docking approaches, we only model van der Waals and electrostatic interactions between the receptor and ligand molecule. Iakovou et al.⁵⁵ presented a GPU accelerated approach to calculate both of these forces between two molecular structures. Equation (1) gives the force, \vec{f}_i on receptor atom i from all of the ligand atoms, labelled j .

The haptic device “holds” the ligand and transmits the forces upon it from the receptor to the user. The force on the ligand from receptor is given by $-\sum_{i=1}^N \vec{f}_i$. The resulting

force is then scaled before being transmitted through the haptic device. A range of force scaling profiles, as described by Iakovou et al.⁹ are available for use. These profiles are used to ensure a good range of forces can be felt by the user. Within Haptimol FlexiDock the ligand is currently modelled as rigid.

$$\vec{\mathbf{f}}_i = - \sum_{j=1}^M \left(\left(24\varepsilon_{ij} \left[\frac{2\sigma_{ij}^{12}}{r_{ij}^{13}} - \frac{\sigma_{ij}^6}{r_{ij}^7} \right] + \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}^2} \right) \vec{\mathbf{r}}_{ij} \right) \quad (1)$$

In Equation (1) ε_{ij} and σ_{ij} are Lennard-Jones parameters that depend on the characteristics of the interacting atoms, r_{ij} is the distance between the two interacting atoms, q_i and q_j are the atomic charges of the two atoms, ϵ_0 is the permittivity of free space, ϵ is the relative permittivity dependent on the dielectric properties of the solvent and $\vec{\mathbf{r}}_{ij}$ is the unit vector in the direction of atom i to atom j . Note the negation, required in order to get the interaction force acting on receptor atom i from all ligand atoms, rather than the force on the whole ligand from receptor atom i . We write in this way to be consistent with notation in a previous paper⁵⁵. Torques are omitted, as it is not possible to render them on low cost haptic devices.

The parameters for ε_{ij} , σ_{ij} , q_i and q_j were taken from the AMBER ff03 forcefield⁵⁶. The Lorentz-Berthelot rules were used to compute ε_{ij} and σ_{ij} . These rules state that $\varepsilon_{ij} = (\varepsilon_i \varepsilon_j)^{\frac{1}{2}}$ and $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$, where σ_i , σ_j , ε_i and ε_j are the Lennard-Jones parameters of atoms i and j .

There is the option to use a distance-dependent ϵ as described by⁵⁷ in order to reproduce the effect of dielectric screening of electrostatic interactions by the water solvent.

The forces on each of the receptor's atoms are calculated in parallel, then the individual force vectors $\vec{\mathbf{f}}_i$ are combined to form a $3N \times 1$ column vector, $\mathbf{F} = (f_{x1} f_{y1} f_{z1} \dots f_{xN} f_{yN} f_{zN})^t$, where t denotes the transpose. We use linear response to calculate the response of the receptor to forces from the ligand. In order to do this equilibrium fluctuations of the receptor in its ligand free state need to be found. This is achieved using explicit-solvent Molecular Dynamics (MD) simulation.

Molecular Dynamics

MD simulations of ligand-free maltodextrin binding protein (MBP) and glutamine binding protein (GlnBP) were started from the structures deposited in the Protein Data Bank (MBP: PDB ID 1OMP⁵⁸; GlnBP: Chain A of 1GGG⁵⁹). AMBER ff14SB⁶⁰ was used for the proteins. Each protein was initially solvated in a cubic box with SPC/Eb water molecules⁶¹, and Na⁺ ions⁶² to neutralize the systems. The simulation boxes were constructed with a margin of at least 10Å from the proteins to the periodic box boundaries. The total number of atoms included the simulated systems for MBP and GlnBP including solvent molecules were 80,019 and 56,741, respectively. The simulation was conducted with the pmemd.cuda module⁶³ of AMBER16⁶⁴. The electrostatic interactions were treated with particle mesh Ewald method⁶⁵ and the real space cutoff distance was 10Å. In both cases, after a 200 step energy minimization with positional restraints imposed on experimentally-determined heavy atoms with a force constant of 1 kcal/molÅ⁶⁶, the systems were equilibrated at 300 K and 1 atm with weaker positional restraints with a force constant of 0.1 kcal/molÅ⁶⁶. MD simulations without the restraints were performed for 120 ns and the last 100 ns trajectories used for the linear response calculation.

Linear Response

Linear response theory can be used to calculate the structural change in a receptor, induced by interaction forces from the ligand. To do this, Equation (2), (equivalent to Equation 3 within Ikeguchi et al.¹⁹) can be used requiring calculation of the variance-covariance matrix of atomic fluctuations, i.e. the approximation is at the quasi-harmonic level. Taking the coordinate trajectory of the protein from the MD simulation, mass-weighted least-squares best fitting is used to superimpose each frame onto the reference structure which was the starting structure. The average structure is then calculated, and the $3N \times 3N$ variance-covariance matrix of atomic fluctuations, \mathbf{A} .

$$\Delta \mathbf{r} = \frac{1}{k_b T} \mathbf{A} \mathbf{F} \quad (2)$$

In Equation (2), k_b is Boltzmann’s constant, T , the absolute temperature, and $\Delta \mathbf{r} = (\Delta x_1 \ \Delta y_1 \ \Delta z_1 \dots \ \Delta x_N \ \Delta y_N \ \Delta z_N)^t$, a column vector of atomic coordinate displacements. Within the quasi-harmonic approximation displacements are from the average but as this is not a viable structure we use the structure from the trajectory with the smallest RMSD to the average. This structure is the undeformed structure rendered on the screen during a haptic session.

Given a receptor and ligand, we calculate the force on each receptor atom by using Equation (1), and then we use Equation (2), to calculate the atomic displacements in the receptor. If we allow the user to control the position of the ligand, via keyboard or haptic device, and perform the above calculations continually updating the receptor’s deformation in real time, we will have an interactive molecular docking system that supports realistic receptor flexibility. Note that due to the fitting procedure used to calculate \mathbf{A} whereby global translational and rotational movements are removed, forces applied to the receptor do not result in its global translation or rotation.

The main problem to overcome is the evaluation of $\Delta \mathbf{r}$ within the 2 ms time constraint. When using Equation (2), $9N^2$ multiplications are required every frame, in addition to the calculations made to evaluate the force. Even though these calculations are performed on the GPU, it was not possible to complete them within the 2 ms time limit even for the modestly sized proteins used here. A further problem may arise for larger proteins. GPU memory is a scarce resource in comparison to regular RAM, and it may not be sufficient to store the covariance matrix. For example, the covariance matrix of a protein comprising 15,000 atoms would consume 8.1 GB of GPU memory, which is greater than the available memory on the majority of consumer graphics cards.

We can overcome these issues by making use of the fact that a large amount of the total fluctuation of a protein occurs within a small subspace (sometimes this space is referred to

as the “important subspace” or the modes spanning it as “essential modes”^{67–69}), to reduce both the number of calculations required per frame, and the memory consumption of the algorithms. This is achieved with Equation (3), which is equivalent to Equation 4 in Ikeguchi et al.¹⁹.

Fast evaluation of response using high-fluctuation modes

$$\Delta \mathbf{r} \approx \Delta \mathbf{r}_m = \frac{1}{k_b T} (\mathbf{V}_m (\boldsymbol{\Lambda}_m (\mathbf{V}_m^t \mathbf{F}))) \quad (3)$$

An eigenvector decomposition of the covariance matrix \mathbf{A} leads to a set of eigenvectors and eigenvalues. These are sorted in descending order of the eigenvalues, each of which gives the mean square fluctuation of the collective coordinate described by its corresponding eigenvector. In Equation (3), $\boldsymbol{\Lambda}_m$ is the $m \times m$ diagonal matrix of the first m eigenvalues and \mathbf{V}_m is the $3N \times m$ matrix of corresponding eigenvectors. The fact that the $\boldsymbol{\Lambda}_m$ is a diagonal matrix further reduces the number of multiplications required to calculate the deformation from Equation (3). The matrix multiplications can be performed in any order, however, the parentheses are included to show the most efficient way to perform the calculation, i.e. from right to left.

By using Equation (3) instead of Equation (2) to calculate the receptor’s response, and by performing the calculation in the order shown by the parentheses, the number of multiplications is $m(6N + 1)$. Furthermore, when multiplying the matrices together in this order, the largest transient matrix will be $3N \times 1$ in size (although the largest matrix is the matrix of eigenvectors, \mathbf{V}_m , which is $3N \times m$ in size). This means for appropriate choice of m , use of Equation (3) can produce significant savings in memory and execution time over use of Equation (2).

We perform these multiplications in parallel on the GPU. Owing to the fact that $\boldsymbol{\Lambda}_m$ is a diagonal matrix, the entire equation can be solved with two kernels and a single global synchronisation point, ideal for a parallel environment. With an appropriate choice of m

it will be possible to evaluate the receptor’s response according to Equation (3) within the 2 ms time limit. A pertinent question is whether we are able to use a sufficient number of eigenvectors to get a realistic response whilst maintaining a smooth haptic refresh rate. $\rho(m)$ in Equation (4) measures the percentage of the protein’s total fluctuation contained within the first m eigenvectors.

$$\rho(m) = 100 \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^{3N-6} \lambda_i} \quad (4)$$

Where λ_i is the i th eigenvalue of $\mathbf{\Lambda}_m$, $3N - 6$ is the total number of non-zero eigenvalues and $\rho(m)$ is the percentage fluctuation contained within the first m eigenvalues. This formula will be used to determine if a sufficient proportion of the total fluctuation is represented by using m eigenvalues within a docking session. When $m = 3N - 6$, $\rho(m) = 100\%$.

Finding a stable state

In certain binding positions, the interaction forces between the receptor and ligand become unstable. In these positions, the initial interaction force between the two molecules is such that a conformational change occurs that is large enough to noticeably change the interaction force, which in turn alters the receptor’s conformation. This is illustrated in Figure 2.

In Figure 2, the initial force, labelled F_s , causes the atoms in the receptor to move to Δr_i . Because of this, the interaction force changes to F_i , causing the atoms to change position again. In the illustrated scenario depicted in the figure, the forces and atomic displacements get larger and larger, never reaching equilibrium. When using the software in combination with a haptic device, this scenario creates severe vibration.

In order to prevent this from occurring, we used an iterative approach. Rather than apply the entire receptor deformation in one step, a small amount of the total calculated deformation is applied and then the force is recomputed, which in turn will cause a change in deformation. In this fashion, the force and receptor response will eventually reach an

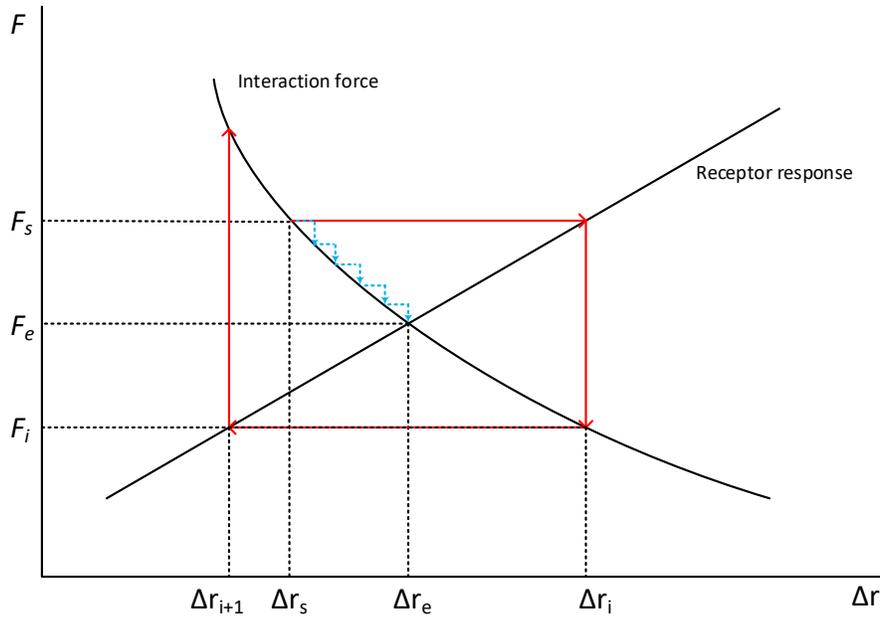


Figure 2: Diagram depicting how using an iterative approach when applying deformation to the receptor improves stability. The initial ligand receptor interaction occurs at Δr_s . The broken arrows describe the subsequent sequence of events with the iterative approach, the solid arrows without it. The solid arrows show that the initial interaction force, F_s , causes the receptor to deform to Δr_i , which in turn changes the interaction force to force F_i , prompting a different response from the receptor. This cycle can continue ad infinitum. The iterative approach breaks the deformation application up into smaller steps, re-computing the force after each step. In this way, the interaction force and deformation response will come into equilibrium. The global positions of the receptor and ligand are assumed constant throughout the process.

equilibrium. This is depicted as the stepped line from Δr_s to Δr_e in Figure 2.

Deforming the receptor in an iterative manner has further usability advantages. Firstly, as the deformation happens more slowly, the receptor’s movement becomes visible to the user, providing opportunity to learn about the deformation process. Furthermore, the reduced deformation speed gives the user opportunity to respond to haptic and visual cues generated as the protein deforms.

Haptic and Visual Rendering

Figure 3 shows each of the steps that need to be completed at least once every 2 ms, in order to keep the haptic refresh rate above the 500Hz target. As a result of the time constraint, all of the steps are implemented in OpenCL and performed on the GPU. The CPU is only responsible for selecting which path to take. During a docking session, communication between the CPU and GPU is kept to a minimum; overlap information is passed back to the CPU at the decision stages and a small amount of intercommunication happens during the force calculation stage, required in order to render the calculated force on the haptic device.

Stage 1, the force calculation stage, is discussed in detail by Iakovou et al.⁴⁸. The implementation differences within this application are limited to supporting implicit water by screening the Coulomb potential, and using a much larger cut-off distance than that used by Iakovou et al.⁴⁸. With a short cut-off distance it was found that vibrations occurred when atoms moved across the abrupt cut-off boundary. This was more apparent when modelling the interactions in vacuo rather than with electrostatic screening, but the effect is noticeable with either. To overcome this a larger cut-off distance is necessary and indeed with the small ligands used here it was possible to forgo the cut-off completely; however, with larger ligands it is likely to be a useful feature in order to maintain the haptic refresh rate.

The deformation response (Equation (3)), Stage 2, is calculated using two kernels. The first kernel performs the multiplication ($\Lambda_m(\mathbf{V}_m^t \mathbf{F})$), making use of the fact that Λ_m is a diagonal matrix. The second kernel performs the remaining multiplication.

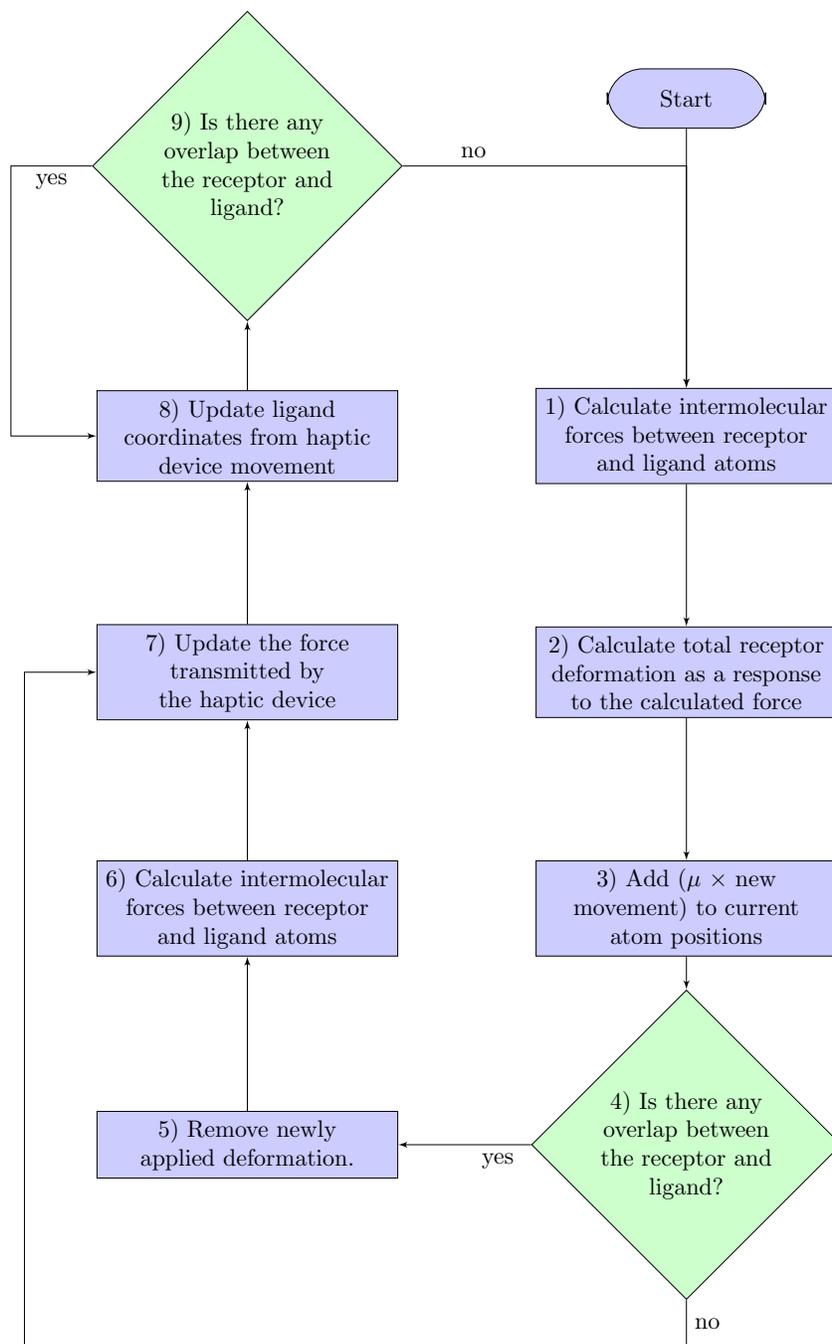


Figure 3: Flowchart showing Haptimol FlexiDock’s “Haptic Loop.” Each of the steps within the loop must be completed every time the haptic device is updated. μ specifies the size of the step used within the iterative approach.

During Stage 3, the iteration stage the difference between the new deformation and the current deformation is calculated, multiplied by μ , and then added to the existing deformation. In this way, μ controls the speed of the deformation; we leave it to the user to tune μ for their purposes. Within our application, we set the default value of μ to be 0.01, a value found to be a nice medium between the speed of the deformation, and smoothness of the haptic feedback.

If atoms within the ligand significantly overlap with those within the receptor, a large unrealistic force will be generated. A collision detection algorithm is used at Stage 4 to detect when large atom overlaps have occurred, so the docking process can be halted and returned to its most recent good pose. The maximum permissible overlap is set by the user, with a default value of 0.5 Å. Allowing a small amount of overlap allows the receptor to continue to deform, when atoms are in contact. This allows the receptor to deform away from the contact if necessary, or around the ligand if the ligand is positioned in a viable binding site. However, allowing a small amount of overlap can, in some conformations, cause a large force to occur. We therefore also provide the facility to halt the deformation when a large force occurs. The size of this force threshold is again user configurable.

The collision detection is performed on the GPU, where each atom in the receptor performs an intersection test with each of the ligand's atoms in parallel. If a receptor atom overlaps a ligand atom by more than the defined threshold, a counter is incremented. These counters are then summed in parallel; if the result is greater than zero, one of the atoms is in collision and the previous pose is held. Otherwise, the response movement is accepted and the loop is continued. A second overlap test is performed at Stage 9 of the haptic loop.

Visual Rendering

Haptimol FlexiDock renders proteins in space-filling mode, using CPK colouring⁷⁰ by default. The colour of individual residues can be changed by the user. Although our application is designed for use with a haptic device, we acknowledge that this may not always be feasible,

so mouse and keyboard control is supported.

In order to improve the usability of the software when used without a haptic device, we provide a colour system that varies the brightness of individual receptor atoms in order to highlight those which are contributing the most to the total interaction force. The effect can be seen in Figure 4.

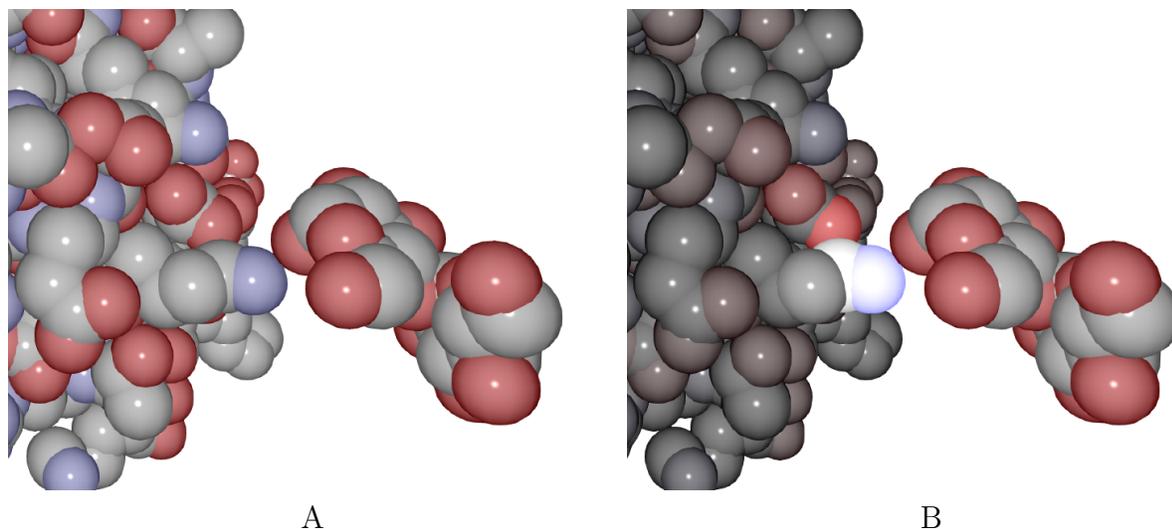


Figure 4: A receptor-ligand pose rendered in Haptimol FlexiDock with (A) ordinary colour, (B) force-contribution colouring. The size of each atom is given by its van der Waals radius.

Besides the force-interaction colouring, the interaction energy and forces can be plotted on a graph in real time, allowing the user some understanding of the force interactions that are prompting the rendered response.

Methods for evaluating performance and accuracy

To understand how well the described approach works as a haptic-assisted interactive molecular docking system that incorporates flexibility, two main aspects of our approach need to be analysed: the performance of the application, and how well the system models the receptor's functional movement.

Performance

As stated in the introduction, for continuous, smooth and stable force feedback, haptic devices need updating at a refresh rate greater than 500Hz. To ascertain whether our system meets this criteria, whilst using a sufficient number of eigenvalues to incorporate the important subspace of any tested molecule, we simulated ligand movement around the protein and timed how long the haptic loop (Figure 3) took to complete. The ligand movement was simulated by randomly positioning the ligand in proximity to the receptor, and then moving the ligand slowly toward the protein. When the ligand touched the receptor, its movement was halted, and it was held in that position for a short time. Whilst this was happening, the execution time of the haptic loop, which runs in a separate thread, was recorded. As processor exclusivity cannot be guaranteed it is difficult to accurately time how long methods take to run. For this reason a large number of samples need to be taken.

Methods to compare to experimentally determined movement

To assess how well the system models the receptor’s functional movement, we used the experimentally derived ligand-bound and unbound structures and compare them with poses generated within our application.

Before performing these comparisons it is important to know how much of the experimentally derived movement is contained within the high fluctuation eigenvectors. This is measured in Equation (5).

$$\tau(m) = 100 \sum_{i=1}^m \left(\mathbf{v}_i^t \left(\frac{\Delta \mathbf{r}_{exp}}{|\Delta \mathbf{r}_{exp}|} \right) \right)^2 \quad (5)$$

In Equation (5), \mathbf{v}_i is the i th column of \mathbf{V}_m , and $\Delta \mathbf{r}_{exp}$ is a column vector of the experimentally derived individual atomic coordinate displacements - the movement required to displace each atom from its position in the ligand free structure to its position in the ligand bound structure, after global superposition. $\tau(m)$ indicates the percentage of the

experimental movement that is represented in the space of the first m eigenvectors. When $m = 3N - 6$, $\tau(m) = 100\%$.

There are various methods that can be used to compare a docked pose generated within Haptimol FlexiDock to the experimentally derived ligand-bound structure. The most straightforward of these is the root-mean-square deviation of atomic positions (RMSD).

The atomic displacements that occur upon ligand docking during the haptic session can be compared with the experimentally derived displacements.

$$p = \frac{\Delta \mathbf{r}_{hap_bb}^t \Delta \mathbf{r}_{exp_bb}}{|\Delta \mathbf{r}_{hap_bb}| |\Delta \mathbf{r}_{exp_bb}|} \quad (6)$$

In Equation (6) $\Delta \mathbf{r}_{hap_bb}$ is a column vector of atomic coordinate displacements from the starting to bound conformations from a haptic docking session, $\Delta \mathbf{r}_{exp_bb}$ the experimentally derived individual backbone atom displacements. When using this equation, we only consider the backbone atoms. When calculating $\Delta \mathbf{r}_{hap_bb}$ the start structure and haptic docked structure must be globally aligned. The closer p is to 1.0, the more the direction of experimentally derived and haptic movement align. Equation (6) does not consider the relative magnitudes of the displacements, which is quantified in Equation 7. Again, we only use this on the molecule’s backbone. Combined, Equations (6) and (7) can be used to determine how close a docked ‘haptic’ structure is to the experimentally derived ligand-bound structure; the closer both values are to 1.0, the closer the pose is.

$$q = \frac{\Delta \mathbf{r}_{hap_bb}^t \Delta \mathbf{r}_{exp_bb}}{|\Delta \mathbf{r}_{exp_bb}|^2} \quad (7)$$

We also compare the distributions of displacements of C^α atoms as was previously done by Ikeguchi et al.¹⁹. The final metric used is ΔL which is the distance between the ligand in its position in the haptic docked pose and the experimental pose. As we keep the orientation of the haptic-controlled ligand relative to the protein the same as in the experimental structure of the protein-ligand complex, this is simply evaluated by the distance between any atom of

the ligand in its two positions.

Together these metrics will give a good picture of how well the haptic-derived movement compares to the experimentally derived movement.

Results

The performance testing of our application was performed using two different receptor ligand pairs, as shown in Table 1. The topology for MBP was generated using PDB2GMX⁷¹, The structure of maltose was taken from the closed structure of MBP (PDB: 1ANF), and its topology, along with the topologies of GlnBP and glutamine were generated using the Antechamber and leap programs within AmberTools16⁶⁴, and converted into GROMACS top files using ParmED⁷². The all-atom RMSD between the closest to average and average structure was 1.05 Å for MBP and 0.99 Å for GlnBP. The closest to average structures were the undeformed structures rendered on the screen during the haptic session. All of the testing was performed on a desktop computer with an Intel i7 processor and an Nvidia GTX 1080. As the GPU was equipped with 8 GB of memory and the test proteins are modest in size, the entire covariance matrix could fit in the GPU memory, however the 2 ms time constraint could not be achieved whilst using Equation 2.

Table 1: Table detailing the receptor/ligand pairs used for testing.

Receptor Name	Receptor PDB Code (N ^o of atoms)	Trajectory Length	Ligand Name (N ^o of atoms)	Liganded-Receptor PDB
Maltodextrin Binding Protein	1OMP (5737)	100 ns	Maltose (45)	1ANF
Glutamine Binding Protein	1GGG (3431)	100 ns	Glutamine (20)	1WDN

Determination of subspace size for haptic time constraint

For smooth operation with the haptic device, the haptic loop needs to complete in less than 2 ms. Testing was performed to determine how many eigenvectors we can use whilst achieving

this goal. To do this, the execution time of the haptic loop was recorded when used with different numbers of eigenvectors. For each different eigenvector count tested, approximately 400,000 loops were timed, and the average calculated. The results are shown in Figure 5.

The results show a fairly large standard deviation (approximately 0.3 ms). This can largely be explained by the fact that during the haptic loop, two paths are taken; if there is a collision, an extra two steps are performed, adding to the runtime, however as processor exclusivity during testing cannot be guaranteed other processes maybe claiming processor time.

To ensure smooth use of the haptic device, it is important that the majority of haptic loops complete in under 2 ms. By looking at the second standard deviation of the run time data, the broken line in the lower graphs contained in Figure 5, we see that for MBP, using fewer than 550 eigenvectors (3.2% of the 17205 eigenvectors), and for GlnBP, fewer than 1220 eigenvectors (21.4% of the 10287 eigenvectors), 97.7% of the calculations are completed in less than 2ms, meeting this goal.

To determine whether these values are sufficient to model the protein’s movements, Equation (4) can be used. The results for different numbers of eigenvectors can be seen in the upper plots of Figure 5 which shows the plot of $\rho(m)$ for the two test proteins: MBP and GlnBP.

Figure 5 shows that for MBP, $\rho(550) = 87\%$, and GlnBP $\rho(1220) = 97\%$. Thus despite only being able to use a relatively small number of eigenvectors to calculate the deformation during an interactive session, they are sufficient to span the important subspace.

For MBP, whilst using 550 eigenvalues, the deformation calculation is the most computationally expensive part of the haptic loop, making up 51% of the total loop execution time on average. The force calculation time is the second most time consuming part of the loop, taking 16% of the total run time. The remaining time is used updating the positions of the atoms and performing collision detection between the receptor and ligand.

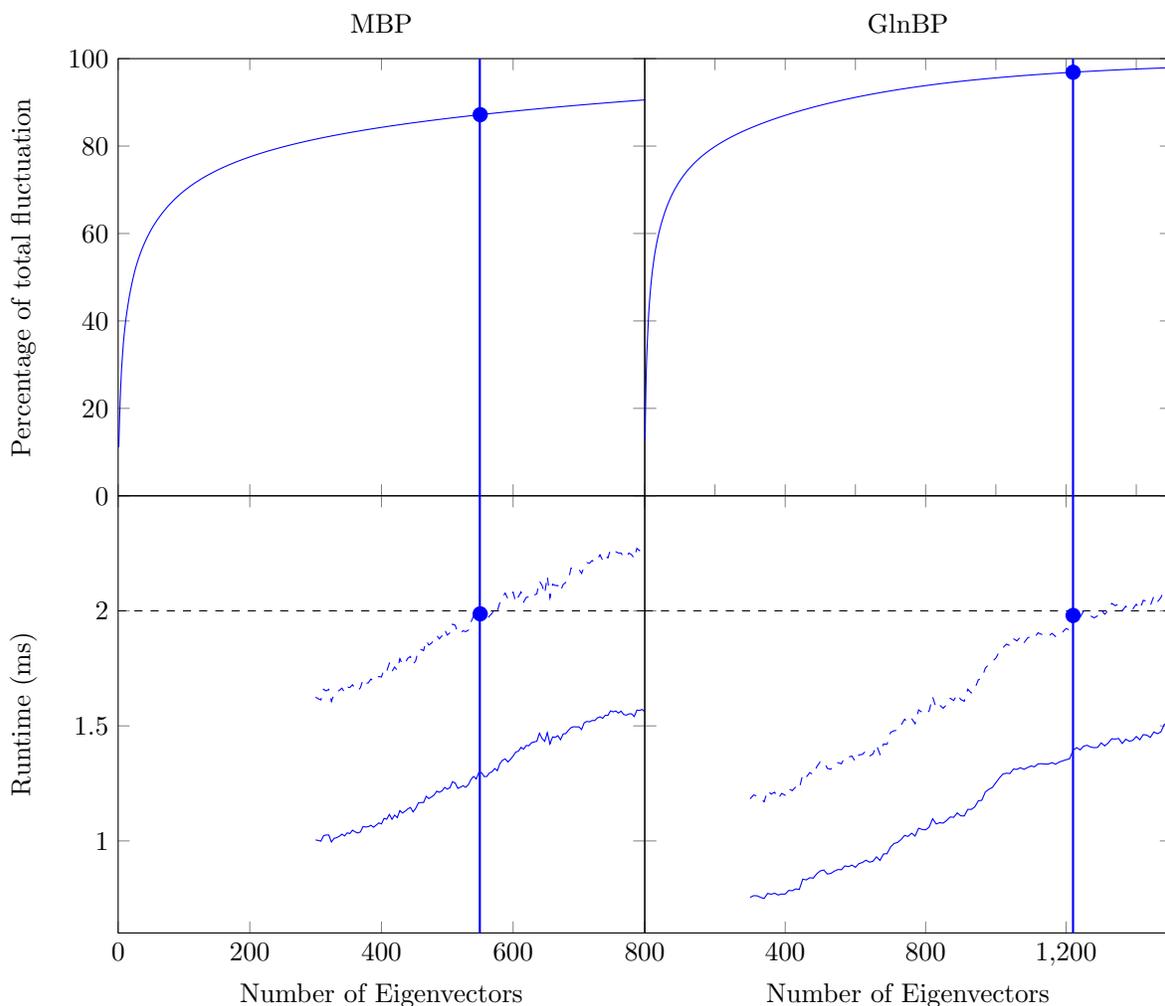


Figure 5: shows the time taken to complete the haptic loop (lower row) and the percentage of the total fluctuation incorporated within the corresponding number of eigenvectors (upper row) for the test proteins MBP (left column) and GlnBP (right column). The vertical line indicates the number of eigenvectors for which the 2ms constraint can be satisfied 98% of the time. Within the lower row, the mean runtime is shown by a solid line. The broken line is at two standard deviations from the mean which gives the duration required for 97.7% of calculations to be completed. The 2ms cut-off is drawn with a horizontal dashed line across the plots.

Functional movement and space explored in MD trajectory

In the previous section we have seen that using the first 550 eigenvalues and eigenvectors for MBP the first 1220 eigenvalues and eigenvectors for GlnBP, is sufficient to get below the 2 ms haptic time constraint. We now want to find out whether the functional movement as derived from the crystallographic ligand-free and ligand-bound structures is contained within the high-fluctuation modes from the MD trajectories. Equation 5 was used to determine how much of the proteins' experimentally derived movement is incorporated within the first 550 eigenvectors for MBP, and 1220 for GlnBP. The results show that for MBP $\tau(550) = 90.3\%$, and for GlnBP $\tau(1220) = 94.7\%$. Therefore, for both proteins, over 90% of the experimental motion is incorporated in the subspace for which a haptic refresh rate of over 500Hz is achievable. Consequently, a movement that closely matches the experimental functional movement would be possible even when comparatively few eigenvectors are used.

Haptic vs Experimental Deformation

Using a 3D Systems Touch haptic device (formally known as the SensAble Phantom Omni), we performed docking experiments, attempting to dock maltose into MBP and glutamine into GlnBP. The starting pose, or haptic start structure, is the conformation from the trajectory with the lowest RMSD to the trajectory average structure. In these experiments, 500 eigenvectors were used to calculate the flexibility for MBP and 1200 for GlnBP. The ligand was placed into its experimentally determined bound orientation by superimposing the experimental ligand-bound structure onto the haptic start structure, reorienting the ligand accordingly. Then, using the metrics outlined within Methods, we tested each of the poses to determine if the calculated movement aligned with that derived experimentally. The results are shown in Table 2.

The results show that for both proteins tested the calculated receptor deformation, when the ligand is docked into the approximate binding area (as determined from the experimental structure), is aligned with the experimentally derived motion. This is shown both with the

Table 2: Table detailing the quality of the ‘‘Haptic’’ docking poses for MBP and GlnBP. Start RMSD is the RMSD between the haptic docked pose of the receptor and the haptic starting pose. Docked RMSD is the RMSD between the haptic docked pose of the receptor and the experimentally derived docked structure. Both are calculated between C α atoms only. p is the result of using Equation (6) and q the result of Equation (7). ΔL is the distance between maltose in the haptic pose and the experimental pose.

MBP					
PoseID	ΔL (Å)	Start RMSD (Å)	Docked RMSD (Å)	p	q
H-MBP1	2.9	2.39	2.23	0.85	0.58
H-MBP2	3.0	2.91	1.99	0.85	0.67
H-MBP3	2.6	2.68	2.11	0.83	0.63
H-MBP4	3.3	2.30	2.47	0.83	0.53
H-MBP5	3.2	2.64	2.25	0.84	0.63

GlnBP					
PoseID	ΔL (Å)	Start RMSD (Å)	Docked RMSD (Å)	p	q
H-GlnBP1	1.5	3.36	3.31	0.85	0.50
H-GlnBP2	3.3	3.43	3.40	0.83	0.49
H-GlnBP3	1.3	4.41	3.09	0.83	0.63
H-GlnBP4	1.6	3.32	3.29	0.86	0.50
H-GlnBP5	1.8	4.34	3.08	0.83	0.62

RMSD values and the values of p and q .

Initially looking at the RMSD values, we can see that for most of the test poses, the haptic docked structure is ‘closer’ to the experimentally derived bound structure than it is to its starting structure, showing that the deformation modelled has deformed the structure toward its ligand-bound pose.

The given values for p and q further validate that the calculated movement is largely aligned with the experimental movement. For all of the reported poses at least 0.8 of the backbone movement is in the direction expected, in order for the starting pose to deform into the experimental ligand-bound pose. Furthermore, all of the poses exhibit at least half of the intra-molecular backbone movement required to fully translate the atoms from an open position to a closed position. Together, these metrics show that the described system calculates and renders a deformation that is broadly aligned with the experimentally derived movement. This can be seen in Figure 6 which shows the haptic docked pose (A and D) and the experimentally determined closed pose (B and E) superimposed onto the starting pose for both of the test proteins.

Reducing the subspace size in Haptimol FlexiDock

Figure 7 shows $\rho(m)$, (Equation 4) and $\tau(m)$ (Equation 5) for the first 20 eigenvectors of both of the test proteins, MBP and GlnBP. Figure 7 shows that for MBP, despite only using the first 20 eigenvectors, which is 0.1% of the total number, nearly 50% of the total fluctuation occurs within the space spanned by their corresponding eigenvectors. What’s more 75% of the experimentally determined movement is contained within this space. For GlnBP, the space spanned by the first 20 eigenvectors contains 78% of the experimentally determined movement and approximately 50% of the total fluctuation contained within the trajectory.

As a consequence of this, using a small number of eigenvectors will still reproduce the majority of the movement seen upon ligand binding from the crystal structures. This will enable

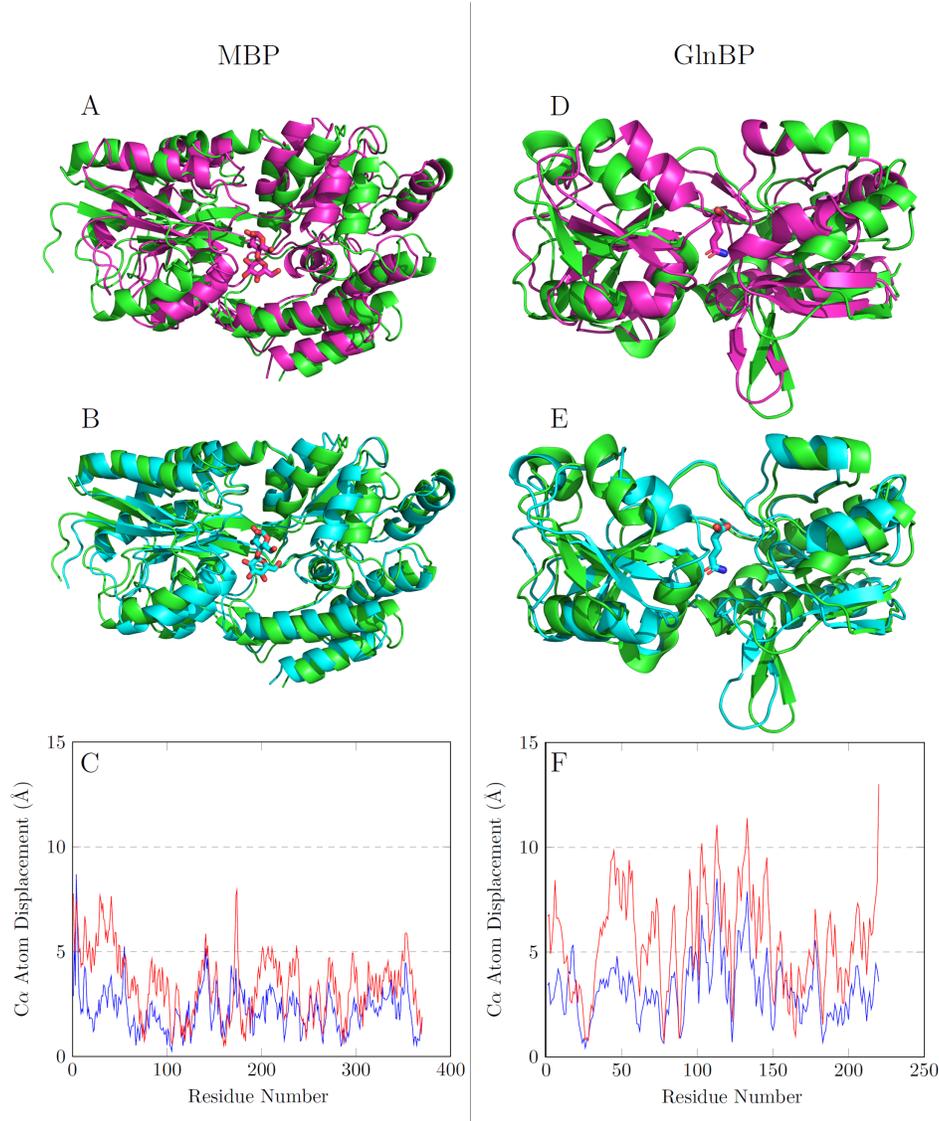


Figure 6: (A) Structure of MBP from an interactive session where maltose is bound close to its binding site (magenta - pose: H-MBP2 - see Table 2) superimposed upon the haptic starting structure of MBP (green). (B) Maltose bound X-ray structure (1ANF - cyan) superimposed on the haptic starting structure (green). (C) Displacements of C^α atoms for the Haptic (A - blue) and Experimental (B - red) cases (correlation coefficient is 0.628) for MBP. (D) Structure of GlnBP from an interactive session where glutamine is bound close to its binding site (magenta - pose: H-GlnBP5) superimposed upon the haptic starting structure of GlnBP (green). (E) Glutamine bound X-ray structure (1WDN - cyan) superimposed on the haptic starting structure (green). (F) Displacements of C^α atoms for the Haptic (A - blue) and Experimental (B - red) cases (correlation coefficient is 0.761) for GlnBP. (A),(B),(D) and (E) were generated using PyMOL⁷³.

our software to work well on less powerful computers, as reducing the number of eigenvectors used will reduce the computational cost of calculating the conformational response.

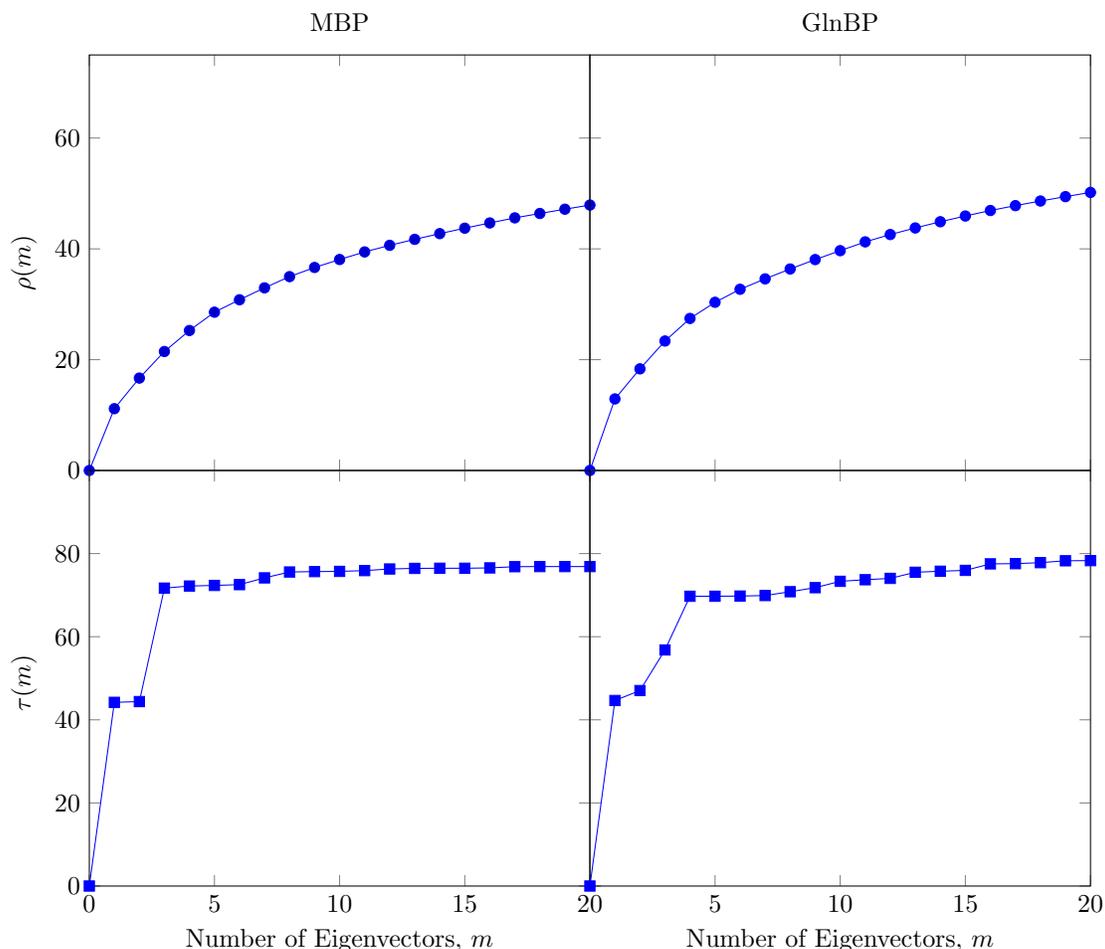


Figure 7: Upper graph shows $\rho(m)$, the percentage of the total fluctuation contained within the space of the first m eigenvectors and the lower graph shows $\tau(m)$, the percentage of the experimental movement contained within this space for MBP (left) and GlnBP (right).

The lower graphs in Figure 7 show that for GlnBP 56.8%, and for MBP, 71.6%, of the movement that occurs upon ligand binding is incorporated within the first 3 eigenvectors.

In order to allow users to study the resultant effect of adding and removing eigenvectors from consideration, the facility to modify m during a docking session is provided.

Haptimol FlexiDock also incorporates other features, including the option to select and remove side chains from the interaction force calculations whilst still rendering them graphically (“ghostify”) and to switch different force components (van der Waals or electrostatic)

on and off. See Iakovou et al.⁹ for more details on these features. Video S1 in the Supporting Information shows Haptimol FlexiDock in use.

Discussion and Conclusion

We have described a tool for interactive docking which incorporates receptor flexibility. The user “holds” a rigid ligand molecule using a haptic device being able to control its position and orientation as well as feel forces on the ligand from the receptor. As the ligand moves, the receptor responds to changes in the forces exerted on it from the ligand by changing conformation. Conformational change is modelled by application of linear response theory. Linear response has been shown to work very well on proteins by Ikeguchi et al.¹⁹, where it accurately reproduced patterns of displacements of C^α atoms derived from experiment.

Our approach is an implicit approach in that the effect of water is handled implicitly separating out the MD simulation from the interactive docking session. In contrast, interactive MD⁷⁴ allows the user to apply forces via the haptic feedback device during the MD simulation. If the MD simulation involves an explicit solvent model, then one would feel the forces of interaction between individual water molecules and the molecule controlled with the haptic device. This would produce a great deal of noise on the haptic device making it difficult to control. There is a further, perhaps more fundamental argument against using this approach. Protein functional movements, which are in response to interaction forces from the binding of ligands, have relaxation times of several nanoseconds at least, and are properly handled in a thermodynamic sense by taking long-timescale averages of solvent mediated interactions. The linear response approach taken here separates the MD simulation from the interactive session allowing some of the average solvent effects from a long MD simulation to be included in the interactive session implicitly. This approach suggests that an implicit solvent model forcefield should also be used for non-bonded interactions during the interactive session. Although we do include the possibility of screening electrostatic inter-

actions using a distance-dependent dielectric constant, we are not currently able to include forces that are a consequence of hydrophobic interactions.

Other groups^{39,53,75} have reported using energy minimisation to model the conformational response to interaction forces. This kind of approach may be appropriate for small ligands but for larger molecules such as proteins, using linear response would seem to be a better approach as large molecules are able to move between a large number of local minima on their potential energy surface at physiological temperature. This jumping amongst minima (JAM) is implicitly included in our approach⁷⁶.

We demonstrated in our previous application, Haptimol_RD⁹, for which only rigid molecules can be docked, that with modern GPUs, interaction forces between the ligand and receptor can be calculated within the 2ms time constraint without the need of a precomputed force grid. This was a crucial step as previous applications had employed precomputed force grids on the receptor in order to circumvent this time constraint, an approach which would preclude the modelling of the conformational change in the receptor. In Haptimol FlexiDock the force calculation followed by calculation of the conformational response of the receptor, must together be calculated within 2ms. Within the linear response approach calculation of the response of the receptor to forces from the ligand involves a matrix multiplication. Despite the considerable speed offered by modern GPUs in the execution of matrix multiplications, the full matrix multiplication cannot be executed sufficiently quickly even for modestly sized biomolecules. This is where the concept of the important subspace can be applied. By using only the eigenvectors spanning this relatively small subspace, we have shown that we can reduce the number of matrix multiplications considerably without an appreciable loss of accuracy. For MBP, even though only 3% of the total number of the eigenvectors (the total number of degrees of freedom of MBP), could be used in order to satisfy the 2ms time constraint, the subspace they defined contained nearly 90% of the fluctuation that occurred in the MD simulation. Furthermore, 90% of the functional movement that occurs upon binding maltose is within this subspace. Reducing the subspace still further, as necessitated

by a slower processor, can still give reasonable results as demonstrated in Figure 7 where it is shown that 70% of functional movement in MBP is within the space defined by the first three eigenvectors.

We also tried our approach on the cancer target, B-Raf. The structure used for the MD simulation was not a ligand-free structure but one bound to the drug sorafenib, which we removed from the binding site. In performing this simulation we expected the protein to open sufficiently to enable us to dock sorafenib during the interactive session. This approach was prompted by an MD simulation on the enzyme citrate synthase started from a ligand-bound X-ray structure, but with the ligands removed⁷⁷. This resulted in spontaneous opening to give a structure close to the open ligand-free X-ray structure. This, however, did not appear to happen in B-Raf indicating that it may require a much longer simulation to relax to a fully open conformation, a flexible sorafenib to dock to the binding site, or both of these.

Interactive methods should not be regarded as competing with automated methods but rather as complementary. Automated methods are used to search for the correct binding pose by exploring a vast number of different poses, something that clearly cannot be done with an interactive tool. The purpose of an interactive tool is to provide an environment that enables knowledge and intuition to find a solution. One could envisage interactive docking being used to refine the results of automated docking, or even, a hybrid tool which incorporates features of automated docking into an interactive docking environment. One application where interactive docking may be particularly useful is in SBDD, where the aim is not to find the correct binding site but to trial different lead molecules at a known binding site. Our tool would appear to be particularly suited to this application as it has been increasingly recognised amongst the drug development community that target flexibility needs to be properly modelled in ligand binding⁷⁸. Theoretically, the linear response approach means that for a single response matrix (the variance-covariance matrix), derived from equilibrium fluctuations of the ligand-free target molecule, the calculated response is dependent on the perturbing force only, irrespective of origin, i.e. irrespective of the ligand being used. This

means that in the context of SBDD, after investment through long, explicit solvent MD simulations on the target receptor molecule, any lead molecule could be used for that target. The tool could therefore not only be used for comparative studies on existing lead compounds, but in expert hands, help foster ideas as to how such leads might be adapted to, say, enhance binding. In other words it would provide an environment where hypotheses are nurtured and tested.

We expect Haptimol FlexiDock to be used for student engagement and education⁷⁹. Feeling forces whilst seeing high-quality renderings of molecules on the screen is highly engaging for students, especially when, as in Haptimol FlexiDock, there is an interplay between the forces felt through the haptic device and the shape of objects rendered visually. From an educational perspective, Haptimol FlexiDock could be used to demonstrate how interaction forces between molecules lead to molecular binding and conformational change.

There are a number of improvements that can be made to Haptimol FlexiDock. It is our intention to improve both the biophysical accuracy of the tool as well as to introduce more realistic graphical rendering. If the forces are large, the molecular geometry is distorted beyond reasonable bounds. It is a direct consequence of the linear approach taken and occurs particularly with electrostatic screening switched off or when there is atomic overlap. It would be desirable therefore to find a method that is able to restrain the molecular geometry within reasonable bounds. Another obvious improvement would be the inclusion of ligand flexibility. As stated above energy minimisation is often used and although it is not appropriate for a large molecule it could be implemented for small ligands. With regard to graphical rendering, real-time shadows and ambient occlusion as in our Protein Trajectory Viewer⁸⁰ will be incorporated, as well as the ability to use head mounted displays for an immersive 3D experience.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council [Grant Number EP/M508020/1]

References

- (1) Smith, G.; Sternberg, M. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **2002**, *12*, 28–35.
- (2) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
- (3) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395.
- (4) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- (5) Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophys. Rev.* **2017**, *9*, 91–102.
- (6) Antunes, D. A.; Devaurs, D.; Kaviraki, L. E. Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discovery* **2015**, *10*, 1301–1313.
- (7) Yuriev, E.; Agostino, M.; Ramsland, P. A. Challenges and advances in computational docking: 2009 in review. *J. Mol. Recognit.* **2011**, *24*, 149–164.
- (8) Hou, X.; Sourina, O. Haptic Rendering Algorithm for Biomolecular Docking with Torque Force. 2010 International Conference on Cyberworlds. 2010; pp 25–31.

- (9) Iakovou, G.; Hayward, S.; Laycock, S. D. Virtual Environment for Studying the Docking Interactions of Rigid Biomolecules with Haptics. *J. Chem. Inf. Model.* **2017**, *57*, 1142–1152.
- (10) Aspuru-Guzik, A.; Lindh, R.; Reiher, M. The Matter Simulation (R)evolution. *ACS Cent. Sci.* **2018**, *4*, 144–152.
- (11) Dalkas, G. A.; Vlachakis, D.; Tsagkrasoulis, D.; Kastania, A.; Kossida, S. State-of-the-art technology in modern computer-aided drug design. *Briefings Bioinf.* **2013**, *14*, 745–752.
- (12) Computing power revolution and new algorithms: GP-GPUs, clouds and more: general discussion. *Faraday Discuss.* **2014**, *169*, 379–401.
- (13) Ricci, A.; Anthopoulos, A.; Massarotti, A.; Grimstead, I.; Brancale, A. Haptic-driven applications to molecular modeling: state-of-the-art and perspectives. *Future Med. Chem.* **2012**, *4*, 1219–1228.
- (14) Bivall, P.; Ainsworth, S.; Tibell, L. A. E. Do haptic representations help complex molecular learning? *Sci. Educ.* **2011**, *95*, 700–719.
- (15) Persson, P. B.; Cooper, M. D.; Tibell, L. A. E.; Ainsworth, S.; Ynnerman, A.; Jonsson, B. Designing and Evaluating a Haptic System for Biomolecular Education. 2007 IEEE Virtual Reality Conference. 2007; pp 171–178.
- (16) Subasi, E.; Basdogan, C. A New Haptic Interaction and Visualization Approach for Rigid Molecular Docking in Virtual Environments. *Presence* **2008**, *17*, 73–90.
- (17) Hayward, S. Identification of Specific Interactions that Drive Ligand-induced Closure in Five Enzymes with Classic Domain Movements. *J. Mol. Biol.* **2004**, *339*, 1001–1021.
- (18) Hayward, S.; Kitao, A. Molecular Dynamics Simulations of NAD⁺-Induced Domain Closure in Horse Liver Alcohol Dehydrogenase. *Biophys. J.* **2006**, *91*, 1823–1831.

- (19) Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. Protein Structural Change Upon Ligand Binding: Linear Response Theory. *Phys. Rev. Lett.* **2005**, *94*, 078102.
- (20) Jiang, F.; Kim, S. H. "Soft docking": matching of molecular surface cubes. *J. Mol. Biol.* **1991**, *219*, 79–102.
- (21) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–553.
- (22) Venkatraman, V.; Ritchie, D. W. Flexible protein docking refinement using pose-dependent normal mode analysis. *Proteins* **2012**, *80*, 2262–2274.
- (23) Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424–440.
- (24) Sinko, W.; Lindert, S.; McCammon, J. A. Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. *Chem. Biol. Drug Des.* **2013**, *81*, 41–49.
- (25) Brooijmans, N.; Humblet, C. Chemical space sampling by different scoring functions and crystal structures. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 433–447.
- (26) Barril, X.; Morley, S. D. Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures. *J. Med. Chem.* **2005**, *48*, 4432–4443.
- (27) Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **1994**, *235*, 345–356.
- (28) Tsfadia, Y.; Friedman, R.; Kadmon, J.; Selzer, A.; Nachliel, E.; Gutman, M. Molecular dynamics simulations of palmitate entry into the hydrophobic pocket of the fatty acid binding protein. *FEBS lett.* **2007**, *581*, 1243–1247.

- (29) Borrelli, K. W.; Cossins, B.; Guallar, V. Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem.* **2010**, *31*, 1224–1235.
- (30) Apostolakis, J.; Plückthun, A.; Caffisch, A. Docking small ligands in flexible binding sites. *J. Comput. Chem.* **1998**, *19*, 21–37.
- (31) Zacharias, M.; Sklenar, H. Harmonic modes as variables to approximately account for receptor flexibility in ligand–receptor docking simulations: Application to DNA minor groove ligand complex. *J. Comput. Chem.* **1999**, *20*, 287–300.
- (32) Keserû, G. M.; Kolossváry, I. Fully Flexible Low-Mode Docking: Application to Induced Fit in HIV Integrase. *J. Am. Chem. Soc.* **2001**, *123*, 12708–12709.
- (33) Tatsumi, R.; Fukunishi, Y.; Nakamura, H. A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor. *J. Comput. Chem.* **2004**, *25*, 1995–2005.
- (34) Zacharias, M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: Binding of FK506 to FKBP. *Proteins: Struct., Funct., Bioinf.* **2004**, *54*, 759–767.
- (35) Minsky, M.; Ming, O.-y.; Steele, O.; Brooks, F. P., Jr.; Behensky, M. Feeling and Seeing: Issues in Force Display. Proceedings of the 1990 Symposium on Interactive 3D Graphics. New York, NY, USA, 1990; pp 235–241.
- (36) Ellis, R. E.; Ismaeil, O. M.; Lipsett, M. G. Design and evaluation of a high-performance haptic interface. *Robotica* **1996**, *14*, 321–327.
- (37) Pattabiraman, N.; Levitt, M.; Ferrin, T. E.; Langridge, R. Computer graphics in real-time docking with energy calculation and minimization. *J. Comput. Chem.* **1985**, *6*, 432–436.

- (38) Brooks, F. P., Jr.; Ouh-Young, M.; Batter, J. J.; Jerome Kilpatrick, P. Project GROPE-Haptic Displays for Scientific Visualization. Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques. New York, NY, USA, 1990; pp 177–185.
- (39) Bayazit, O. B.; Song, G.; Amato, N. M. Ligand binding with OBPRM and user input. Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164). 2001; pp 954–959 vol.1.
- (40) Lee, Y.-G.; Lyons, K. W. Smoothing haptic interaction using molecular force calculations. *Comput Aided Des* **2004**, *36*, 75–90.
- (41) Bivall, P. Touching the Essence of Life : Haptic Virtual Proteins for Learning. 2010.
- (42) Lai-Yuen, S. K.; Lee, Y.-S. Computer-aided molecular design (CAMD) with force-torque feedback. Ninth International Conference on Computer Aided Design and Computer Graphics (CAD-CG'05). 2005; pp 6 pp.–.
- (43) Subaai, E. Rigid Molecular Docking in Virtual Environments with Haptic Feedback. 2006.
- (44) Wollacott, A. M.; Merz, K. M. Haptic applications for molecular structure manipulation. *J. Mol. Graphics Modell.* **2007**, *25*, 801–805.
- (45) Nagata, H.; Mizushima, H.; Tanaka, H. Concept and prototype of protein–ligand docking simulator with force feedback technology. *Bioinformatics* **2002**, *18*, 140–146.
- (46) Férey, N.; Nelson, J.; Martin, C.; Picinali, L.; Bouyer, G.; Tek, A.; Bourdot, P.; Burkhardt, J. M.; Katz, B. F. G.; Ammi, M.; Etchebest, C.; Autin, L. Multisensory VR interaction for protein-docking in the CoRSAIRE project. *Virtual Real.* **2009**, *13*, 273.

- (47) Sourina, O.; Torres, J.; Wang, J. In *Transactions on Edutainment II*; Pan, Z., Cheok, A. D., Muller, W., Rhalibi, A. E., Eds.; Lecture Notes in Computer Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp 105–118.
- (48) Iakovou, G.; Hayward, S.; Laycock, S. D. Adaptive GPU-accelerated force calculation for interactive rigid molecular docking using haptics. *J. Mol. Graphics Modell.* **2015**, *61*, 1–12.
- (49) Iakovou, G.; Laycock, S.; Hayward, S. Determination of locked interfaces in biomolecular complexes using Haptimol.RD. *Biophys. Physicobiol.* **2016**, *13*, 97–103.
- (50) Daunay, B.; Micaelli, A.; Regnier, S. 6 DOF haptic feedback for molecular docking using wave variables. Proceedings 2007 IEEE International Conference on Robotics and Automation. 2007; pp 840–845.
- (51) Zonta, N.; Grimstead, I. J.; Avis, N. J.; Brancale, A. Accessible haptic technology for drug design applications. *J. Mol. Model.* **2009**, *15*, 193–196.
- (52) Anthopoulos, A.; Grimstead, I.; Brancale, A. GPU-accelerated molecular mechanics computations. *J. Comput. Chem.* **2013**, *34*, 2249–2260.
- (53) Anthopoulos, A.; Pasqualetto, G.; Grimstead, I.; Brancale, A. Haptic-driven, interactive drug design: implementing a GPU-based approach to evaluate the induced fit effect. *Faraday Discuss.* **2014**, *169*, 323–342.
- (54) Stocks, M. B.; Hayward, S.; Laycock, S. D. Interacting with the biomolecular solvent accessible surface via a haptic feedback device. *BMC Struct. Biol.* **2009**, *9*, 69.
- (55) Iakovou, G.; Hayward, S.; Laycock, S. A real-time proximity querying algorithm for haptic-based molecular docking. *Faraday Discuss.* **2014**, *169*, 359–377.
- (56) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge

- force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (57) Mehler, E. L.; Solmajer, T. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein Eng., Des. Sel.* **1991**, *4*, 903–910.
- (58) Sharff, A. J.; Rodseth, L. E.; Spurlino, J. C.; Quioco, F. A. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **1992**, *31*, 10657–10663.
- (59) Hsiao, C.-D.; Sun, Y.-J.; Rose, J.; Wang, B.-C. The Crystal Structure of Glutamine-binding Protein from *Escherichia coli*. *J. Mol. Biol.* **1996**, *262*, 225–242.
- (60) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**,
- (61) Takemura, K.; Kitao, A. Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density. *J. Phys. Chem.* **2012**,
- (62) Joung, I. S.; Cheatham III, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem.* **2008**,
- (63) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.
- (64) Case, D., A; Betz, R., M; Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.;

- Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollan, P. Amber 2016. **2016**,
- (65) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (66) Wan, P. T. C.; Garnett, M. J.; Roe, S. M.; Lee, S.; Niculescu-Duvaz, D.; Good, V. M.; Project, C. G.; Jones, C. M.; Marshall, C. J.; Springer, C. J.; Barford, D.; Marais, R. Mechanism of Activation of the RAF-ERK Signaling Pathway by Oncogenic Mutations of B-RAF. *Cell* **2004**, *116*, 855–867.
- (67) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–425.
- (68) Hayward, S.; Go, N. Collective Variable Description of Native Protein Dynamics. *Annu. Rev. Phys. Chem.* **1995**, *46*, 223–250.
- (69) Kitao, A.; Go, N. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164–169.
- (70) Koltun, W. L. Precision space-filling atomic models. *Biopolymers* **1965**, *3*, 665–679.
- (71) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (72) Swails, J.; Hernandez, C.; Mobley, D.; Nguyen, H.; Wang, L.; Janowski, P. ParmEd. 2018.
- (73) Schrödinger, LLC,

- (74) Stone, J. E.; Gullingsrud, J.; Schulten, K. A system for interactive molecular dynamics simulation. Proceedings of the 2001 symposium on Interactive 3D graphics. 2001; pp 191–194.
- (75) Daunay, B.; Micaelli, A.; Regnier, S. Energy-field reconstruction for haptic-based molecular docking using energy minimization processes. 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2007; pp 2704–2709.
- (76) Kitao, A.; Hayward, S.; Go, N. Energy landscape of a native protein: jumping-among-minima model. *Proteins* **1998**, *33*, 496–517.
- (77) Roccatano, D.; Mark, A. E.; Hayward, S. Investigation of the mechanism of domain closure in citrate synthase by molecular dynamics simulation¹¹Edited by R. Huber. *J. Mol. Biol.* **2001**, *310*, 1039–1053.
- (78) Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* **2003**, *2*, 527–541.
- (79) Lancaster, S. J. Immersed in virtual molecules. *J. Phys. Chem.* **2018**, *2*, 253–254.
- (80) Matthews, N.; Easdon, R.; Kitao, A.; Hayward, S.; Laycock, S. High quality rendering of protein dynamics in space filling mode. *J. Mol. Graphics Modell.* **2017**, *78*, 158–167.

Graphical TOC Entry

