Multi-resolution approach for interactively locating functionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device.

Olivier Delalande, Nicolas Ferey, Benoist Laurent, Marc Gueroult, Brigitte Hartmann, Marc Baaden

▶ **To cite this version:**

Olivier Delalande, Nicolas Ferey, Benoist Laurent, Marc Gueroult, Brigitte Hartmann, et al.. Multi-resolution approach for interactively locating functionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device.. Pacific Symposium on Biocomputing, 2010, pp.205-15. 10.1142/9789814295291_0023 . hal-00661430

HAL Id: hal-00661430
https://hal.science/hal-00661430

Submitted on 19 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MULTI-RESOLUTION APPROACH FOR INTERACTIVELY LOCATING FUNCTIONALLY LINKED ION BINDING SITES BY STEERING SMALL MOLECULES INTO ELECTROSTATIC POTENTIAL MAPS USING A HAPTIC DEVICE

OLIVIER DELALANDE, NICOLAS FEREY, BENOIST LAURENT, MARC GUEROULT[§]
Laboratoire de Biochimie Théorique, CNRS UPR9080/IBPC, 13 rue Pierre et Marie Curie, F-75005, Paris, France

BRIGITTE HARTMANN
[§] DSIMB team, INTS, 6 rue Alexandre Cabanel, F-75739 Paris Cedex 15, France

MARC BAADEN
Laboratoire de Biochimie Théorique, CNRS UPR9080/IBPC, 13 rue Pierre et Marie Curie, F-75005, Paris, France

Metal ions drive important parts of biology, yet it remains experimentally challenging to locate their binding sites. Here we present an innovative computational approach. We use interactive steering of charged ions or small molecules in an electrostatic potential map in order to identify potential binding sites. The user interacts with a haptic device and experiences tactile feedback related to the strength of binding at a given site. The potential field is the first level of resolution used in this model. Any type of potential field can be used, implicitly taking into account various conditions such as ionic strength, dielectric constants or the presence of a membrane. At a second level, we represent the accessibility of all binding sites by modelling the shape of the target macromolecule *via* non-bonded van der Waals interactions between its static atomic or coarse-grained structure and the probe molecule(s). The third independent level concerns the representation of the molecular probe itself. Ion selectivity can be assessed by using multiple interacting ions as probes. This method was successfully applied to the DNase I enzyme, where we recently identified two new cation binding sites by computationally expensive extended molecular dynamics simulations.

1. Introduction

Metal ions drive important parts of biology, yet it remains experimentally challenging to locate their binding sites. A particular example may concern the mechanisms underlying the formation of non-specific protein-DNA complexes. In this context, we have recently studied the DNase I/DNA system as a representative and rather simple model of a non-specific complex. DNase I is a glycoprotein hydrolyzing DNA phosphodiester linkages in the presence of divalent cations, Ca^{2+} and Mg^{2+}. We demonstrated that Ca^{2+} and Mg^{2+} are crucial for optimizing the electrostatic fit between DNA and enzyme. In particular, extended molecular dynamics simulations at atomic detail allowed us to identify two new cation binding sites that are functionally important [1]. Such high-resolution atomistic methods are computationally expensive and require long simulation times. Could these ion binding sites also have been detected with simpler methods? Maybe this could be achieved at lower resolution and requiring less computational power? In the present manuscript we set out to describe a new method aimed at these goals, using the molecular dynamics results as a reference for assessing our findings. The purpose of this method is to provide a fast user-guided search in order to locate potential binding sites, prior to detailed investigations of these sites using more accurate but computationally expensive approaches.

Recently, we have described the exciting possibilities of interactive molecular simulations for studying biological macromolecules [2]. Here, we demonstrate that this approach can be used for locating ion binding sites. The concept of our method is to interactively explore electrostatic potential fields while being guided towards the binding sites by force feedback. This interactive haptic approach may be traced back to the work of Nagata *et al.* [3] who explored the concept in the context of protein-ligand docking. Their work was somewhat ahead of its time, as may be illustrated by the following observations: "Certain limitations remain; for example, only twenty protein atoms can be used to generate the electrostatic field. Furthermore, the system can only use globular probes, preventing drug molecules or small chemical groups from being simulated. These limitations are the result of our insufficient computer resources". Today, this situation has changed and we did not observe any significant computational limitations. Our approach has been tested on systems as big as a pentameric ligand-gated ion channel comprising 1 500 amino acid residues [4]. Of course, the raw increase in computer speed is not the only reason for such an improvement. Adapted software solutions and improved algorithms further render current approaches more efficient.

Using such an interactive approach for locating ion binding sites is a new idea. It will be illustrated with the DNase I example system. Previously, electrostatic steering has been described for a variety of cases such as

enzyme-ligand binding [5], tRNA binding [6], antibody-antigen association [7] and Cdc42 recognition [8]. Computational methods have been used in several of these studies, but interactive virtual reality (VR) approaches only marginally. Why are such VR approaches useful? VR efficiently introduces a human element in the process and benefits from the user's experience and insight. Enabling the user to sense the electrostatic potential field via tactile feedback is a major advantage. The electrostatic potentials of biomolecules are often complex volume and multi-level data, rendering the visual perception of individual binding sites difficult if not impossible. Adding haptic feedback to their interactive exploration significantly simplifies this task. Here, we use such an approach on a biomolecular system in order to explore potential variations at the DNase surface, to locate favourable binding sites, and to discover far-reaching electrostatic pathways guiding ions to these sites.

The representation used for the simulation system is innovative. It could be summarized as a multi-level simulation combining multiple physics models and is intrinsically multi-resolution. The potential field is the first level of resolution used in this model. At a second level, we represent the accessibility of all binding sites by modelling the shape of the target macromolecule at atomic or coarse-grained resolution. The third independent level concerns the representation of the molecular probe itself. By using multiple scales and resolutions it is possible to focus on the essential degrees of freedom of a complex biological system. This makes the user's interaction with the system more efficient.

The interactive VR-based approach renders the scientific task nice and enjoyable, hence encouraging the user to pursue his investigation. Another example of this kind is the ''Fold It!'' project (http://fold.it) that recently brought interactive simulations to the fore. ''Fold It!'' is a 3D-puzzle desktop game, in which the user's task is to fold proteins interactively and without any knowledge prerequisites. The puzzle is essentially based on an interactive-intuitive learning process. The similarities such applications bear with video games should not delude scientists to underestimate the scientific value of such approaches.

2. Materials and Methods

2.1. *Structural data and model preparation for the DNase I system*

Several experimental structures of the DNase I enzyme from different biological contexts were considered as reference in this study: the enzyme crystallized in its *apo* form (PDB code 3DNI [9]), in a DNase I:Actin complex (PDB code 1ATN [10]) or bound to an oligonucleotide duplex (PDB code 1DNK [11]). Calcium ions are found in some of these structures: *apo* DNase I (3DNI) bears two Ca^{2+} binding sites (described below as sites 1 and 2) and the DNase I:Actin complex reveals three Ca^{2+} binding sites (sites 1, 2 and 3).

In the interactive explorations, a model derived from previous molecular dynamics (MD) simulations is used to represent the enzyme. This model corresponds to the most representative conformer of the whole trajectory as determined by an exhaustive cluster analysis [1]. The detection of ionic binding pockets and their localization in the crystallographic structures have been based on this model. In the MD simulations, the *AMBER* parm99 force field has been used [12]. Consistently, we generated the parameters for the electrostatic potential calculation (*pqr* files) with the *pdb2pqr* program [13] using the *--ff=amber* option, including for ionic probes. The potential was determined with the *APBS* software [14] and standard input was obtained with the *--apbs-input* option. Ionic strength was introduced using a 0.15 M concentration of both +1 and -1 charged ions with a 2.0 Å radius value (*APBS* parameters).

The shape of the DNase I molecule was represented *via* van der Waals parameters from the *AMBER* parm99 force field (atomic resolution) or from the coarse-grained model by Zacharias [15] (low resolution).

Every interactive experiment consisted in at least four independent trials to detect binding pockets for each probe. We concluded that a binding site was found when the probe got stuck in a location where it was impossible to escape without increasing the user force range. Attractive pockets at the enzyme surface where the ion could easily be retrieved were not considered as binding sites.

2.2. *MyPal: an interactive simulation approach combining an electrostatic potential field and pairwise non-bonded interactions*

MyPal stands for **M**olecular scrutin**Y** of **P**otenti**AL**s. The application has a corresponding French name with the same meaning: *MonPote*, "**MO**lécules **N**aviguant sur un **POTE**ntiel". In this section we describe the underlying methodology developed in order to study the behaviour of ions or small molecular probes interacting with a static macromolecular target molecule. In this approach, based on a classical Newtonian simulation, ionic probes are immersed into and guided by an electrostatic potential field induced by the target molecule. In addition, the 3D shape of the target molecule is taken into account by representing its static 3D structure interacting with the probe molecules. The simulation deals with both electrostatic properties and steric constraints induced by the 3D shape of the target molecule. Moreover, in order to study the behaviour of several interacting ions or molecules,

non-bonded Coulomb and van der Waals interactions between all the probes are calculated. These inter-probe interactions are merged with the electrostatic potential from the map and the van der Waals interactions with the target molecule. The user can remain passive and observe the movements of the ionic probes from their initial positions during a simulation in progress. Alternatively, the user can act on the probes by selecting and applying external forces taken into account by the simulation, in order to explore the potential and identify binding sites.

A *MyPal* simulation requires two datasets as schematically shown in Figure 1. The first dataset corresponds to the structure and potential of the target macromolecule, including all static particles in the simulation (left). The second dataset describes a separate dynamic molecular structure including the ionic probes (right).

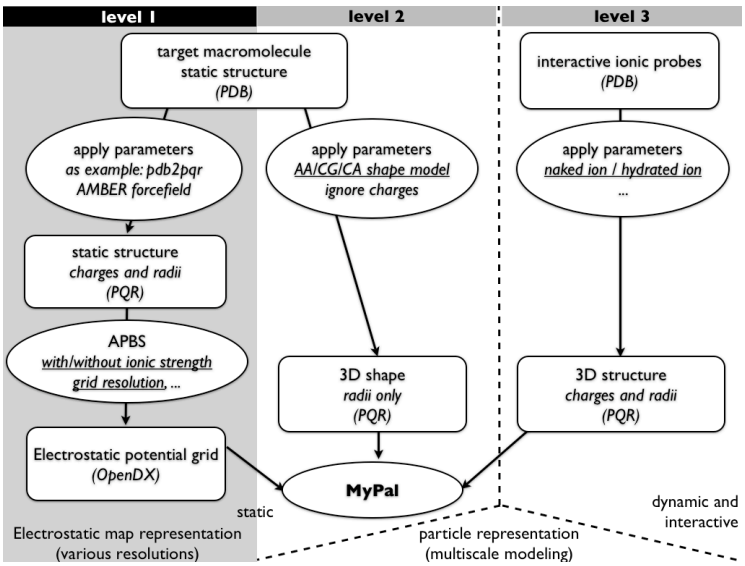

Figure 1: Schematic flow diagram for setting up a *MyPal* simulation. **Level 1** deals with the calculation of the electrostatic potential grid. Here we use the *pdb2pqr* tool in order to set up the radii and charges of the target macromolecule using the atomistic *AMBER* parm99 force field. Then we compute its electrostatic potential map using the *APBS* software [14]. The potential is saved as an *OpenDX* file and used as input for *MyPal*. At **level 2**, the 3D shape of the target molecule is modelled as a set of static spherical particles at the desired resolution (all-atom: AA, multi-bead coarse grained: CG or single-bead carbon alpha: CA). The ionic probes are modelled at **level 3**, using a set of radii and charges. The latter parameters may represent different resolutions such as naked *vs.* hydrated ions. Levels 1 and 2 are static, whereas the particles in level 3 are dynamic and can be guided by the user.

During the simulation, forces are computed and applied on the dynamic particle set ($P_{dynamic}$). We explicitly consider potential, van der Waals, Coulomb and external forces.

The potential forces F_{elect} act on the ionic probes and originate from the electrostatic potential map. They are defined by computing the gradient of the electrostatic potential. In Equation (1), we consider particle p belonging to the spatial cell $C_{i,j,k}$ of the electrostatic potential grid, and $E_{i,j,k}$ the value of the potential in this cell. We define the gradient as the mean of the difference between the $E_{i,j,k}$ potential and the potentials of the six adjacent cells, two for each axis. This method of computing the gradient reduces the bias related to the discretization of the grid.

$$\vec{F}_{elec}(p \in C_{i,j,k}) = \begin{bmatrix} \dfrac{\left(E_{i,j,k} - E_{i-1,j,k}\right) + \left(E_{i+1,j,k} - E_{i,j,k}\right)}{2.delta_x} q_p \\ \dfrac{\left(E_{i,j,k} - E_{i,j-1,k}\right) + \left(E_{i,j+1,k} - E_{i,j,k}\right)}{2.delta_y} q_p \\ \dfrac{\left(E_{i,j,k} - E_{i,j,k-1}\right) + \left(E_{i,j,k+1} - E_{i,j,k}\right)}{2.delta_z} q_p \end{bmatrix} \quad (1)$$

The van der Waals forces F_{vdw} applied to the ionic probes ($P_{dynamic}$) are computed taking into account all interactions between the ionic probes and the static particles defining the 3D shape of the target molecule, and between ionic probes themselves (P_{all}). As shown in Equation (2), these forces are computed using the following approximation for the van der Waals interactions.

$$\vec{F}_{vdw}(p \in P_{dynamic}) = \sum_{p' \in P_{all}} \vec{u}_{pp'} 4\varepsilon_{pp'} \left[\left(\frac{\sigma_{pp'}}{9d_{pp'}} \right)^9 - \left(\frac{\sigma_{pp'}}{7d_{pp'}} \right)^7 \right] \quad (2)$$

The Coulomb forces $F_{coulomb}$ applied to the ionic probes ($P_{dynamic}$) take into account all electrostatic interactions between the ionic probes according to Equation (3).

$$\vec{F}_{coulomb}(p \in P_{dynamic}) = \sum_{p' \in P_{dynamic}} -\frac{q_p q_{p'} \vec{u}_{pp'}}{4\Pi\varepsilon_0 d_{pp'}^2} \qquad (3)$$

Finally, these forces are summed with an external force provided by the user through the graphical interface during the simulation.

$$\vec{F}(p \in P_{dynamic}) = \vec{F}_{elec}(p) + \vec{F}_{vdw}(p) + \vec{F}_{coulomb}(p) + \vec{F}_{user}(p) \qquad (4)$$

This multi-resolution approach consists of three independent levels of modelling. The potential field is the first level of resolution used in this model (level 1 in Figure 1). In principle, any type of potential field could be used, not only electrostatic ones. Most of the scrutinized potentials in the present work are Poisson-Boltzmann type electrostatic potentials computed using the *APBS* software [14]. Such potentials are subject to several parameters (resolution, ionic strength, solvent and solute dielectric constants) and may thus implicitly take into account an averaging over selected solvent or solute degrees of freedom. More complex scenarios, for instance the presence of a biological membrane environment for a membrane protein target, could also be taken into account with tools such as AquaSol [16]. Deriving the potential field from an average over a molecular dynamics simulation could help to mimic the flexibility of the target molecule. These are only a few examples of how the resolution and underlying physics of the potential field can be varied. At a second level (level 2 in Figure 1), we model the accessibility of a binding site by representing the shape of the target macromolecule *via* non bonded van der Waals interactions between its static structure and the probe molecule(s). For this representation, a whole range of parameterizations commonly used in molecular mechanics is available. These comprise the detailed atomistic scale (AA), medium-resolution multi-bead coarse graining (CG) and single-bead carbon alpha models (CA) [17]. Level 3 in Figure 1 is the third, independent level of modelling and concerns the molecular probe itself. At this level, van der Waals and coulombic interactions are treated. Radii and charges can be chosen in order to fine-tune the desired properties of the ionic probes by considering effects such as solvation (increased radii) or charge screening (decreased charges).

It should be stressed that the *MyPal* approach is computationally cheap. The electrostatic map is computed offline prior to the interactive simulation and there are only a small number of dynamic ionic probes on which pairwise interactions are computed. Given the reduced number of degrees of freedom, namely the positions of the ionic probes, thorough sampling of phase space is easily achieved.

2.3. *Visualizing a simulation in progress*

During the design of the *MyPal* application, initial tests have been carried out with the VMD software [18]. This approach couples visualization and simulation with the *MDDriver* library [19], using the IMD protocol [20]. All screenshots in the present manuscript were obtained using VMD. VMD is however not optimally adapted to some of the specific tasks occurring during an interactive simulation. As an example, there is no visual feedback during the particle selection task, making it difficult to select the particle of interest with confidence. Multiple selections are not available using direct interactions with the mouse or a haptic device, but only using the Graphical User Interface (GUI). In order to surmount these shortcomings, we are now developing our own simple visualization tools specifically designed for interactive visualization. We use the visualization toolkit VTK [21] for visual rendering. VTK provides high-level features, such as isosurface rendering of electrostatic potential maps. In an enhanced version, our tool allows us to visualize up to several hundred thousand particles in interactive time using a GPU shader implementation of spherical representations. VTK was encapsulated into a Cocoa application, allowing us to quickly develop a GUI using the XCode and Interface Builder tools. As soon as this application will be finalized, it will be made freely available in order to simplify interactive potential exploration as described in this work [22].

2.4. *Interacting with the ionic probes during a simulation*

In order to interact with a *MyPal* simulation in progress, it is possible to use a mouse for adding force constraints on the probe particles. This approach provides two degrees of freedom, *e.g.* the x- and y-axes, for the interaction. However, using a haptic device is even better adapted to this task, in particular for selecting and moving particles in 3D space. Such a device with three instead of two degrees of freedom is more intuitive and efficient for interacting with a complex three-dimensional object. Furthermore, the immediate haptic feedback when a

particle is actually picked significantly improves the user experience and greatly helps to immerse the user in the molecular scene. When using visual feedback only, the user often asks for additional explanations before getting started. With force feedback, this barrier is lifted, as the interactive simulation becomes more intuitive and is comparable to dextrous manipulations such as those carried out in daily life. Hardware requirements are modest (details in section 2.6). In our experience, this approach is viable using a small and affordable haptic device, providing 3D positions and handling 3D directional force feedback. Such an entry-level solution designed for a desktop use is targeted at a large user community and is very easy to set up. The availability of convenient drivers that are compatible with software libraries such as *VRPN* [23] was a major criterion for our choice.

The interaction with the simulation was implemented as in VMD, allowing the user to impose forces on probe particles and experience a tactile feedback. The haptic device is used in order to control the direction of the forces applied to selected particles and to adjust the amplitude of these forces. This interaction method contains two stages. The first stage comprises the selection of a single probe particle or a set of particles that we will name $P_{selection}$, using a 3D tool attached to a haptic device and its buttons. In a second stage, the model described in Equation (5) is used in order to compute the forces $F_{simulation}$ applied to the selected particles and sent to the *MyPal* simulation as external force (see section 2.2). $F_{simulation}$ is proportional to the distance between the geometrical centre of the particle set and the tracker position P_T.

$$\vec{F}_{simulation}(p \in P_{selection}) = k_{simulation}(\vec{P}_T - \frac{1}{|P_{selection}|}\sum_{p' \in P_{selection}} \vec{P}_{p'}) \qquad (5)$$

The main idea of this approach is to link the selected atoms and the 3D haptic tool with a spring. Instead of providing direct haptic rendering of forces computed in the simulation, the force feedback $F_{feedback}$ only depends on the spring length according to Equation (6), which in turn is influenced by the way the simulation reacts to the applied force.

$$\vec{F}_{feedback} = -k_{feedback}(\vec{P}_T - \frac{1}{|P_{selection}|}\sum_{p \in P_{selection}} \vec{P}_p) \qquad (6)$$

We emphasize that the haptic loop computation frequency must be between at least 300 to 1000 Hz in order to provide a haptic rendering of good quality. A strong point of the approach described above is that a low physical simulation framerate does not cause instabilities and does not affect the quality of the haptic feedback. With this decoupled spring model, force feedback can be computed at a very high frequency required by the haptic device.

2.5. *Coupling simulation, visualization and interaction*

In this section, we briefly explain how simulation, interaction and visualisation modules exchange data. Figure 2 illustrates a schematic view of the data flow in the application presented in this work. We use the *MDDriver* library as central element in order to couple a visualization, interaction and *MyPal* simulation module. Previously, this library was successfully used to couple the *GROMACS* molecular dynamics engine [24] with *VMD* in order to carry out interactive studies of several biomolecular systems comprising enzymes and membrane proteins [2]. Here we use a similar approach in order to couple the *MyPal* program with our own visualization code based on *VTK* or with the visualization module *VMD*.

The *MDDriver* library provides a simplified and high-level API [19] for exchanging particle positions computed in a simulation module and custom user-generated forces from the visualization and interaction modules. The adapter layers shown in Figure 2 are specific for a given simulation or visualization software. They are mainly in charge of transforming the physical values into adequate units, but can also operate more complex transformations, such as changing data from a given discretization model to another one. The adapters are in charge of redistributing and collecting the data in a parallel application. For implementing *MDDriver* in a given software package, the major part of the work consists of locating the adequate routines in the simulation algorithm where the data exchange needs to take place. The exchange with the visualization tool includes sending atom positions, receiving forces and managing control events for the simulation.

Further coupling is needed by our interactive approach. It consists of coupling the peripherals management library, *e.g.* for working with a haptic device, and the visualization module. We use the *VRPN* library [23], which offers a device independent implementation, in order to read out the position and orientation of a tool used to select particles in the visualization module. *VRPN* is further used in order to generate force feedback along the lines described in section 2.4.

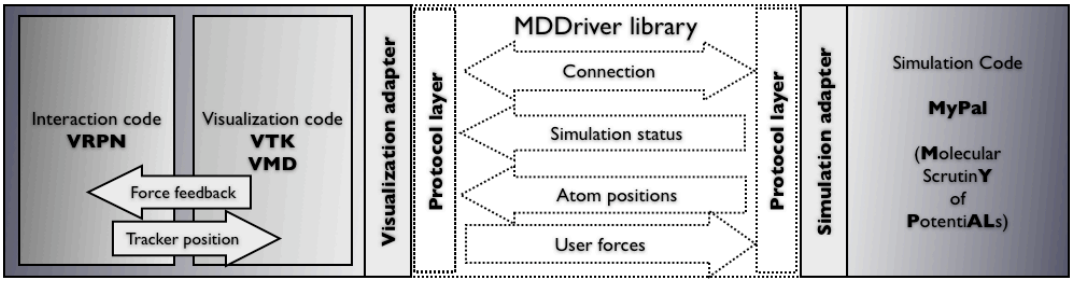

Figure 2: The *MDDriver* library couples a simulation code with a visualization code. *MDDriver* provides an API for exchanging particle positions computed in a simulation module and custom forces from the visualization and interaction modules [19]. Haptic device management is provided by the *VRPN* library [23] and force feedback is computed as described in section 2.4

2.6. *Hardware setup and typical configuration*

All interactive *MyPal* simulations were performed on a 3GHz dual quad-core MacPro Apple computer. Although the *MyPal* application itself is not parallelized, a multi-core workstation makes it possible to run all parts of the application on the same machine without performance issues. This includes visualization, haptic device server and calculation tasks. The molecular scene display was rendered using an NVIDIA Quadro FX 5600 graphics card. Stereo rendering was achieved via a Crystal Eyes stereovision device. Two devices were tested for user interaction and tactile feedback: a Phantom Omni haptic device and a higher precision Phantom Premium 1.5A, both by Sensable Technologies. Scene navigation was achieved with a Spaceball device providing six degrees-of-freedom.

The *MyPal* application can easily be run in a desktop context, with minimal spatial requirements, and does not require any sophisticated hardware setup. In our experiments, the typical configuration was to mount the haptic device, Spaceball, mouse and keyboard in front of a big screen. Figure 3 illustrates such a typical setup.

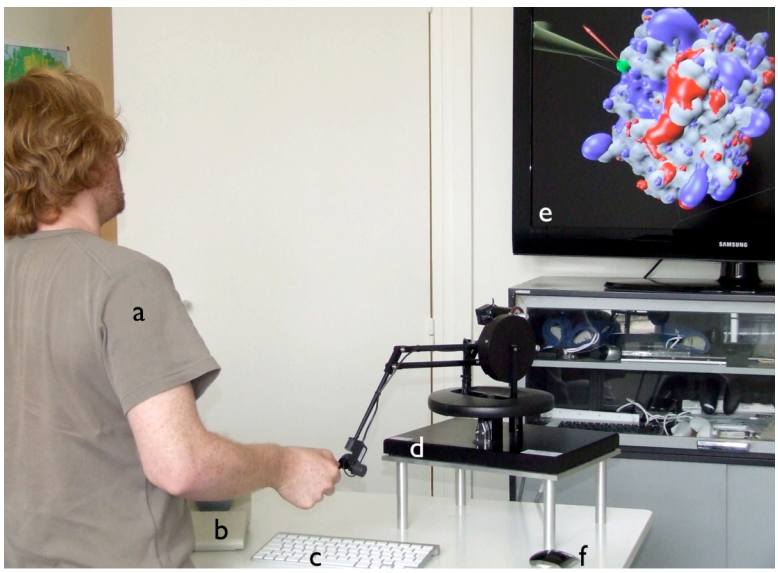

Figure 3: Typical hardware setup for an interactive exploration experiment based on the *MyPal* simulation tool. The user (a) can navigate the scene with a Spaceball device (b). Further control can be achieved *via* a keyboard (c) and a mouse (f). Interaction with the calculation and force feedback from *MyPal* occur *via* a haptic device (d). The scene is shown on a large screen (e) with an avatar representing the user-driven tool (cone in top-left corner of the screen), and a visual rendering indicating the user force (red arrow) acting on the ionic probe (green sphere).

3. Biological application – results and discussion

3.1. *DNase I cation binding sites revealed by atomistic molecular dynamics simulations*

We will use our previous results on cation binding sites in the DNase I enzyme as reference to compare our interactive simulations to. It has been shown elsewhere that molecular dynamics (MD) simulations are powerful – yet computationally expensive – tools for locating such sites [25]. Specific and non-specific protein-DNA interactions imply the formation of intermolecular interfaces requiring electrostatic and structural complementarity between the related partners. A precise functional role for metal ions in this process is certainly

an interesting hypothesis worthwhile exploring. The DNase I/DNA complex is a representative and rather simple model to study these non-specific interactions. We have recently carried out four 25 ns molecular dynamics simulations of the *apo* enzyme where the cation composition was varied between Na^{+}, Ca^{2+} and Mg^{2+} [1]. Detailed analysis of cation coordination in these simulations revealed four distinct sites. Two sites show a preference for calcium and two are selective for magnesium (Figure 4, left panel). The calcium binding sites were previously known [9], whereas the two magnesium sites were unexpected. We have shown that cations are essential for the biological function of DNase I, allowing DNA to bind to the enzyme. Indeed, the concurrent occupation of all four sites with the proper cation is required for an electrostatic fit between DNA, negatively charged, and the enzyme interface, also negatively charged in the absence of cations. This may imply that the information about the binding sites is encoded in the electrostatic potential map of the enzyme (Figure 4, right panel). In the following paragraphs, we reproduce key results of the MD study using the interactive *MyPal* application.

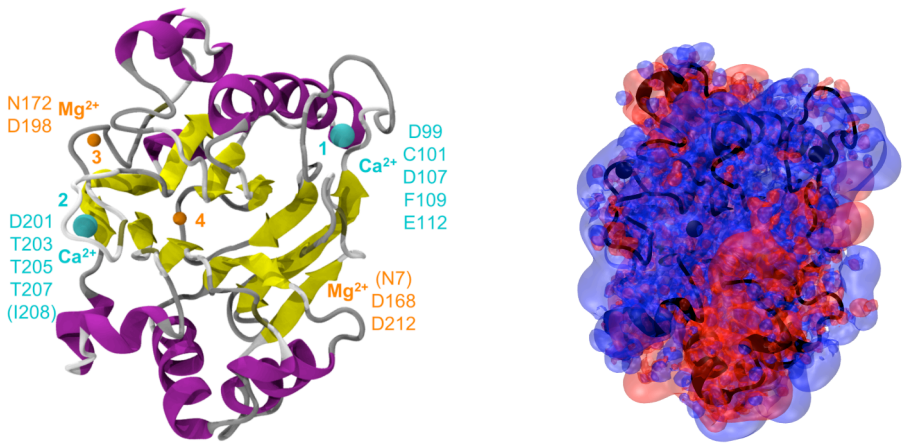

Figure 4: On the left, DNase I cation binding sites numbered from 1 to 4. Calcium binding sites are shown in cyan, magnesium binding sites in orange. The main amino acid residues constituting each site are indicated. On the right, electrostatic potential isosurfaces for potential values of -15.0(blue)/+15.0(red) kT/e are superposed onto the DNase I backbone structure and ions (in black).

3.2. *Locating the four previously identified ion binding sites with MyPal using interactive exploration at atomic resolution*

Figure 4 (right panel) illustrates the difficulty of visually identifying ion binding pockets due to the complexity of typical macromolecular electrostatic potentials. Even knowing their location (Figure 4, left panel), the binding sites cannot be distinguished. By using tactile feedback with our interactive *MyPal* application, it was however possible to locate all four binding sites with confidence and high precision (e.g. 0.9 Å positional deviation for the first site, mainly limited by the spacing of the potential grid). The interactive approach proved quick and intuitive.

Starting from a relaxed reference structure extracted from the MD trajectories, we interactively scanned the entire electrostatic potential surface of DNase I. An initial potential was calculated considering standard physiological ionic strength (0.15M NaCl). Na^{+}, Ca^{2+} and Mg^{2+} are potential ionic probes. For the identification of the binding sites we primarily used a hard sphere representation of the small divalent Mg^{2+} cation as it offers several advantages as a probe. Its double charge facilitates long-range electrostatic steering towards the binding pockets and its small size (0.79 Å) increases the accuracy for sensing the rough and detailed molecular surface at atomic resolution. The structural model of DNase I was extracted from MD simulations with binding sites occupied by ions that are mainly coordinated by amino acids and not water molecules. Consequently, it is important to consider a “naked” ion and not a hydrated one as probe. Otherwise the ion would not fit into the preformed coordination sphere at each site. More generally, considering hydrated ions using an implicit hydration shell remains feasible by increasing the probe radius and possibly varying the charge in order to take into account solvent shielding.

The parameters used for calculating the electrostatic potential affect the interactive exploration, in particular ionic strength. Taking into account ionic strength leads to locally more accurate potential maps. The downside of this for interactive experiments is that the detection of potential wells becomes harder as the range of electrostatic steering shortens. Without ionic strength, we achieved comparable precision, but did more easily detect several of the binding sites. The results for both series of experiments, with and without ionic strength, are presented in Figure 5. Despite the simplicity of the representations of protein surface and ions, all four ion binding sites identified by MD are retrieved by our approach. No other strong binding sites (false positive

locations) were detected, which can be a problem with other approaches [26]. We did observe several attractive interactions located at pockets on the enzyme surface. These locations correspond to shallow minima requiring little force to extract the ions and were not considered as binding sites.

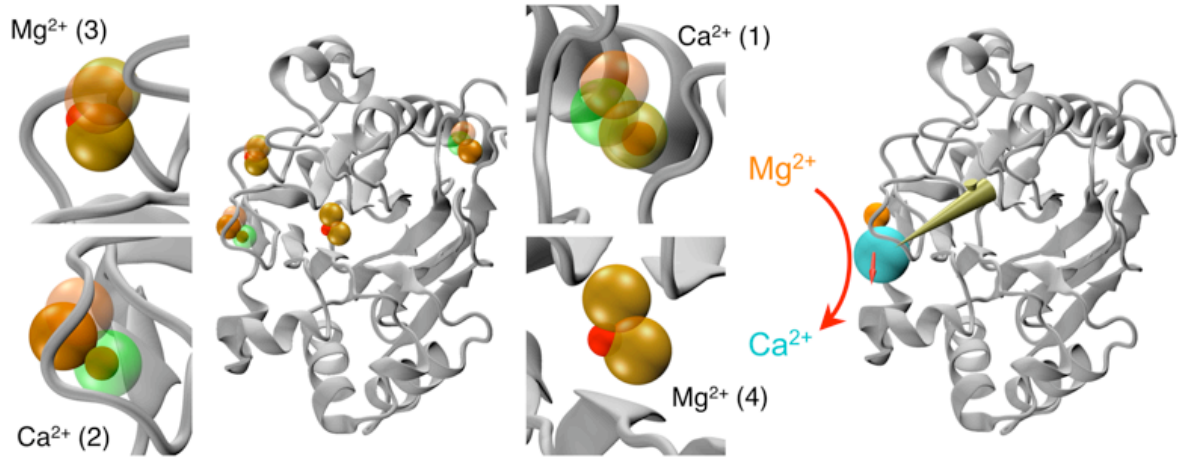

Figure 5: Visual summary of the interactive experiments. On the left, the results of a series of experiments for detecting *a priori* unknown ion binding sites are shown. The reference position of each binding pocket as determined by MD simulation is shown as a red sphere. Transparent spheres display *MyPal* predictions obtained for a potential map with (orange) or without (green) ionic strength. On the right, an ion substitution experiment ("molecular-billiard") at site 2 is depicted. Such an experiment probes the selectivity of a given ionic pocket for different ions.

3.3. *"Molecular-billiard" for probing site selectivity by ion substitution*

In the previous section, we successfully described how to locate the four DNase I cation binding sites. We did however not assess their selectivity. For this purpose, we have carried out a series of additional experiments, each starting with a different ionic probe at a given site. We then tried to interactively substitute the probe by another ion. We considered magnesium, calcium, sodium and chloride ions as probes. Figure 5 and Table 1 illustrate and summarize the results for these ion substitution "molecular-billiard" simulations.

As might be expected, chloride as an anion cannot be stabilized within any of the four cation binding pockets, nor can it displace a bound cation. Sites 1 and 2 are calcium selective, which is generally verified. It is however surprising that magnesium is able to substitute for calcium at site 1. This may be related to the simplicity of our model in which selectivity depends on the shape of the pocket itself and the pathway for accessing it. Generally speaking, buried and narrow sites are unreachable for large ions, whereas sites localized at the enzyme surface are readily subject to ion exchange. In the latter case, haptic feedback helps the user to distinguish between favourable and unfavourable substitutions. The favourable substitution of Na^+ with Ca^{2+} requires little user forces, whereas Na^+ can only displace Ca^{2+} using excessive force. Generally, Mg^{2+} can displace Ca^{2+} but the opposite remains impossible, except for the exposed calcium selective site 2. Sites 3 and 4 are selective towards the small divalent Mg^{2+} cation (r_{Mg2+}=0.79 Å). Site 4 is deeply buried and does not permit binding of the two bigger positive ions (r_{Ca2+}=1.71 Å ; r_{Na+}=1.87 Å). Site 3 can accommodate sodium and calcium in agreement with the 1ATN crystal structure [10]. Those ions are however easily displaced by magnesium. On the other hand, magnesium cannot be displaced by any of the other two cations, corroborating the magnesium selectivity of this site.

Table 1. Ion substitution simulation results. The table indicates whether exchange from X to Y is possible (→) or impossible (↛). For instance, Ca ↛ Cl means that Ca^{2+} cannot be displaced by Cl^-. A minus sign indicates that initial positioning of the chosen probe ion at the given binding pocket was not possible *via* our approach.

Probe (Site)	Ca^{2+} (1)	Ca^{2+} (2)	Mg^{2+} or Ca^{2+}(3)	Mg^{2+} (4)
Mg^{2+}/Ca^{2+}	Ca → Mg	Ca → Mg	Mg ↛ Ca	Mg ↛ Ca
	Mg ↛ Ca	Mg → Ca	Ca → Mg	-
Na^+	Ca ↛ Na	Ca → Na	Mg ↛ Na	Mg ↛ Na
	Na → Ca	Na → Ca	Na → Mg	-
Cl^-	Ca ↛ Cl	Ca ↛ Cl	Mg ↛ Cl	Mg ↛ Cl
	-	-	-	-

The selectivity assessment remains a qualitative one, in particular due to the simplicity of the metal model. Recent work seems to indicate that induced polarization effects in the coordination sphere might be required to fully capture metal ion binding [27]. Such an improved potential function would presumably require further refinements, such as a completely polarizable forcefield, flexible sidechains and explicit water molecules [28, 29]. In our approach, we intentionally use approximations in order to speed up the search. More refined representations should be used *a posteriori* in order to refine the results of this initial haptic search.

The current implementation of *MyPal* was not designed in order to provide precise quantitative binding affinity estimates, but to be capable of distinguishing in real time between non-existing, weak and strong ion binding sites and assess the relative selectivity of significantly different ionic probes. The approximations made in the choice of the model representation limit the precision that could be obtained with the current implementation. Despite these limitations, it remains in principle possible to quantify the strength of binding. This aspect will be addressed in future work. One possible approach would be to calculate the work required by the user to extract an ion from its binding site.

These future extensions would also be beneficial for applications in docking and drug discovery. *MyPal* is already capable of handling rigid multi-atomic drug-like molecules. It is possible to displace a probe and assess whether it interacts favourably with cavities and pockets of a macromolecule. Currently the application does however lack tools to assess the ranking of different binding poses and the comparison between different probes. The implementation of such tools using the non-bonded interactions between target and probe molecules would be straightforward.

3.4. *Predicting the location of ionic binding pockets: multi-resolution exploration using a coarse grained protein shape*

Up to this point we have described experiments on a model structure with preformed ion binding sites. In this sub-section we validate our interactive approach using DNase I extracted from the DNase I/DNA 1DNK structure [11], crystallized without any divalent cation (ion-free DNase I). The expected locations of the binding sites were defined by superimposing the MD model onto the crystallographic structure.

We start with the same representation as previously, using a high-resolution atomistic model in order to describe the protein shape. All four binding sites can be located, albeit with reduced accuracy. The static conformation captured by the ion-free crystal structure biases the exploration by imposing an ill-suited local shape for the putative binding sites. For instance, in the case of the Mg^{2+} binding sites, the most solvent accessible and thus the least affected site deviates 2.5 Å from the reference location. Site 4, the most buried – *e.g.* potentially most biased – pocket can only be detected with significantly less precision and a deviation up to 3.9 Å from the expected site.

By lowering the resolution of the molecular shape representation, it should be possible to reduce these artefacts observed at atomic resolution. We used a coarse-grain representation with several beads per residue. In these simulations, the offset of the most buried magnesium site largely decreased from 3.9 to 2.2 Å. This decrease represents an important gain in precision. Localization of the more solvent-accessible pockets benefits to a lesser extent from coarse graining, with improvements ranging from 0.2 Å (site 3) to 0.4 Å (site 2).

Introducing flexibility in the model is an alternative to lowering the resolution of the representation. The protein or the complex could then be deformed in order to fit a binding interaction. The haptic device would help to render deformations fast and easy. Flexibility requires the introduction of bonded interactions in the model. Such a development is currently under way in our laboratory.

A positive conclusion can be drawn on the predictivity of our method. The multi-resolution approach combining an electrostatic potential grid computed from an all-atom structure and a coarse-grained molecular surface representation for steric repulsion is efficient for predicting ion binding sites in a quick and interactive manner. As an explanation, we may venture that the electrostatic contribution is a cooperative effect of several coordinating charged groups. A single misplaced sidechain is not critical and atomic resolution is appropriate even for distorted binding sites. A single atom sterically occluding the binding site is however a severe problem, calling for a low resolution representation of the molecular shape or alternatively for the introduction of flexibility in the model.

4. Conclusion and outlook

MyPal is a method for interactively locating ion binding sites by steering ionic probes into electrostatic potential maps using a haptic device. This multi-resolution approach, coupled with interactive devices is an important improvement of existing methods. This new strategy already emerges as potentially useful for other applications such as docking small ligands on proteins. Here we show how it facilitates the discovery of new relevant ion

binding sites by successfully retrieving the location of cation binding sites in DNase I and assessing their selectivity, combining atomic and coarse-grained resolutions.

The interactive experiments described in this article are particularly well adapted in order to acquire an intuitive understanding of the possible role of counterions in a biomolecular system. They can be carried out without any specific equipment, just using a mouse and on-screen visual feedback. However, we show that tactile feedback is essential for an efficient exploration of complex electrostatic potentials. Generally speaking, immediate feedback from the simulation adds a new dimension of user perception. Haptic feedback in particular is very intuitive and further increases the understanding of a given system. The force resistance allows the user to experience physical effects such as electrostatic steering or steric repulsion while extracting or inserting an ion in a binding pocket. A movie illustrating some interactive *MyPal* experiments on DNase I is available [22].

MyPal is complementary to existing methods such as deterministic or random computer generated searches for ionic binding sites. The human element based on the user's expertise is an essential feature of our approach. At the same time this human element makes it difficult to exhaustively test *MyPal* and compare it to other methods. Different users will use *MyPal* differently and a user's attention will wear off with time. Locating four ion binding sites such as in the case of DNase I remains feasible in a single work session. Such an exploration guided by the expertise of the user should be used prior to more computationally expensive approaches. For example, the 4th ion binding site in DNase I was quickly identified with *MyPal*, whereas it only appeared after extensive equilibration in an MD simulation. The MD simulation does however enable a very detailed and quantitative characterization of ion binding with extensive statistics.

The *MDDriver* library developed in order to couple molecular simulation engines with a visualization tool was used here in the context of a multi-resolution and multi-physics application. *MDDriver* is readily available on http://mddriver.sourceforge.net. The *MyPal* software implementing the simulation methodology presented in this manuscript, as well as the graphical tool designed for interactive simulations and simple visualization tasks based on Cocoa and VTK, will both be made available very soon. Important future improvements may concern visualization, *e.g.* the implementation of volume rendering, and ways to quantify, record and analyze interactive experiments more systematically.

Many future applications for *MyPal* can be imagined. Ions play a particularly central role inside ion channels and their selectivity, gating mechanisms and preferred ionic transport pathways could be investigated *via* interactive simulations. Closer to the example of the DNase I enzyme, the nucleosome still remains uncharted territory when it comes to locating its ion binding sites and assessing their functional role. Ion channels in membrane proteins and ions bound to the nucleosome are good examples of very large systems for which *MyPal* could provide a simple and straightforward method to capture molecular properties. Substantial time savings regarding simulation time and analysis are expected compared to classical approaches using MD. Furthermore, even at the present stage, *MyPal* is not limited to the docking of ions or small ligands, and should therefore be generally applicable in many biological domains. For instance, it opens the prospect to investigate larger molecular assemblies and their interactions. As an example, one could imagine moving the DNase I enzyme described in this manuscript along its DNA substrate to probe sequence specificity. This would be particularly challenging for a nucleosomal DNA substrate.

Acknowledgments

This interdisciplinary work was supported by the DEISA Consortium (co-funded by the EU, FP6 projects 508830/031513) and IDRIS (CNRS's National Supercomputer Center in Orsay). The project was funded by the French Agency for Research (grants ANR-06-PCVI-0025 and ANR-07-CIS7-003-01). We thank Christopher Amourda, Anthony Bocahut, Chantal Prévost and Sophie Sacquin-Mora for providing useful feedback and testing the method presented in this manuscript on a selection of biological systems. Their critical comments enabled us to continually improve the *MyPal* application.

References

1. M. Guéroult, J. Abi Ghanem, B. Heddi, C. Prévost, P. Poulain, M. Baaden and B. Hartmann, *in Albany 2009: Conversation 16* (2009).
2. O. Delalande, N. Férey, G. Grasseau and M. Baaden, *J. Comput. Chem*, in press, published online (2009).
3. H. Nagata, H. Mizushima and H. Tanaka, *Bioinformatics* **18**, 140 (2002).
4. N. Bocquet, H. Nury, M. Baaden, C. Le Poupon, J.P. Changeux, M. Delarue and P. J. Corringer, *Nature* **457**, 111 (2009).
5. R. C. Wade, R. R. Gabdoulline, S. K. Lüdemann and V. Lounnas, *Proc. Natl. Acad. Sci. USA* **95**, 5942 (1998).

6. D. Tworowski, A. V. Feldman and M. G. Safro, *J. Mol. Biol.* **350**, 866 (2005).
7. R.E. Kozack, M.J. d'Mello and S. Subramaniam, *Biophys. J,* **68**, 807 (1995).
8. L. Hemsath, R. Dvorsky, D. Fiegen, M.-F. F. Carlier and M.R. Ahmadian, *Mol Cell* **20**, 313 (2005).
9. C. Oefner and D. Suck, *J. Mol. Biol.* **192**, 605 (1986).
10. W. Kabsch, H. G. Mannherz, D. Suck, E. F. Pai and K. C. Holmes, *Nature* **347**, 37 (1990).
11. S. A. Weston, A. Lahm and D. Suck, *J. Mol. Biol.* **226**, 1237 (1992).
12. D. A. Case *et al.* (2002) AMBER 7 (University of California, San Francisco).
13. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucl. Acids Res.* **32**, W665 (2004).
14. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **98**, 10037 (2001).
15. M. Zacharias, *Protein Sci.* **12**, 1271 (2003).
16. C. Azuara, H. Orland, M. Bon, P. Koehl and M. Delarue, *Biophys. J.* **95**, 5587 (2008).
17. M. Baaden and R. Lavery, *in Recent Adv. in Protein Engineering*, 173 (2007).
18. W. Humphrey, A. Dalke and K. Schulten, *J. Molec. Graphics* **14**, 33 (1996).
19. N. Férey, O. Delalande, G. Grasseau and M. Baaden, *in Proceedings of the 15th ACM Symposium on Virtual Reality Software and Technology*, 91 (2008).
20. J. E. Stone, J. Gullingsrud, K. Schulten and P. Grayson, *in Proceedings of the 2001 ACM Symposium on Interactive 3D Graphics*, 191 (2001).
21. W. Schroeder, K. Martin and B. Lorensen, The *Visualization Toolkit An Object-Oriented Approach To 3D Graphics, 4th Edition.*
22. http://mddriver.sourceforge.net/
23. R. M. Taylor II, T.C. Hudson, A. Seeger, H. Weber, J. Juliano and A.T. Helser, *in Proceedings of the ACM Symposium on Virtual Reality Software & Technology,* 55 (2001).
24. http://www.gromacs.org
25. D. S. Glazer, R. J. Radmer and R. B. Altman, *Structure* **17**, 919 (2009).
26. C. Claperon, R. Rozenfeld, X. Iturrioz, N. Inguimbert, M. Okada, B. Roques, B. Maigret and C. Llorens-Cortes, *Biochem. J.* **416**, 37 (2008).
27. G. Kupparaj, M. Dudev and C. Lim, *J. Phys. Chem. B* **113**, 2952 (2009).
28. D. Jiao, C. King, A. Grossfield, T. A. Darden and P. Ren, *J. Phys. Chem. B* **110**, 18553 (2006).
29. F. Jalilehvand, D. Spångberg, P. Lindqvist-Reis, K. Hermansson, I. Persson and M. Sandström, *J. Am. Chem. Soc.* **123**, 431 (2001).