JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

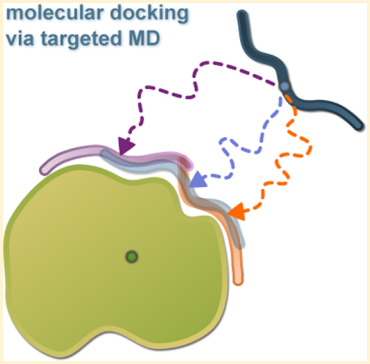
Flexibility and Explicit Solvent in Molecular-Dynamics-Based Docking of Protein−Glycosaminoglycan Systems

Sergey A. Samsonov,[†] Jan-Philip Gehrcke,[†] and M. Teresa Pisabarro*

Structural Bioinformatics, BIOTEC, TU Dresden, Tatzberg 47-51, 01307 Dresden, Germany

Supporting Information

ABSTRACT: We present Dynamic Molecular Docking (DMD), a novel targeted molecular dynamics-based protocol developed to address ligand and receptor flexibility as well as the inclusion of explicit solvent in local molecular docking. A class of ligands for which docking performance especially benefits from overcoming these challenges is the glycosaminoglycans (GAGs). GAGs are periodic, highly flexible, and negatively charged polysaccharides playing an important role in the extracellular matrix via interaction with proteins such as growth factors and chemokines. The goal of our work has been to develop a proof of concept for an MD-based docking approach and to analyze its applicability for protein−GAG systems. DMD exploits the electrostatics-driven attraction of a ligand to its receptor, treats both as entirely flexible, and considers solvent explicitly. We show that DMD has high predictive significance for systems dominated by electrostatic attraction and demonstrate its capability to reliably identify the receptor residues contributing most to binding.



INTRODUCTION

Molecular docking has been extensively applied for drug discovery in recent decades.[1,2] Whereas molecular docking methodologies demonstrate to be very successful and continue to develop rapidly, they still suffer from a number of limitations.[3−5] What holds true for practically all docking methods is that the reproduction of an experimentally observed docking pose is less challenging than proper scoring of this pose by energy.[6−8] In the majority of docking approaches, only small ligands can be treated as entirely flexible in order to limit the size of the conformational search space. Although flexible treatment of the ligand is crucial for obtaining meaningful results, ligand flexibility often can only be approximated, especially in protein−protein docking.[9] Another severe approximation applied in most docking methods is static treatment of the receptor even though protein side chain flexibility is crucial for ligand binding, and substantial conformational changes can occur in the receptor structure induced by interaction with a ligand.[10,11] Finally, most of the established docking approaches do not explicitly account for solvent molecules, whereas many studies point out their importance in docking calculations.[12−15]

A particularly challenging class of ligands for which docking performance is especially limited by the above-mentioned issues is the glycosaminoglycans (GAGs), which are periodic negatively charged linear polysaccharides mainly located in the extracellular matrix. Through interaction with their protein targets, they participate in a number of key processes involved in cell regeneration and proliferation, angiogenesis, metastasis, and lipid metabolism.[16−18] Due to the occurrence of numerous sulfate and carboxyl groups, GAGs have a high charge density, rendering long-range Coulomb interactions crucial for binding to proteins.[19] The significance of electrostatic interactions underlines the importance of taking explicit solvent molecules as binding mediators into account when GAGs are used as ligands in molecular docking.[20] Moreover, the orientation and conformation of long side chains of charged protein residues may be greatly influenced by interaction with a GAG ligand. As a consequence, flexible treatment of the receptor during GAG docking is of special relevance. In addition, a similar spatial distribution of functional groups in GAGs independent of the reducing/nonreducing end orientation[21] as well as their high flexibility[22] substantially contribute to the challenges in protein−GAG docking. Overall, the prediction of protein−GAG complexes via molecular docking comprises a good example of some of the current limitations in the field of classical molecular docking.

Classical docking approaches are generally optimized for having relatively modest computational requirements and therefore enable the quick investigation of single complexes as well as the execution of large-scale studies involving a large number of different complexes. Taking into account ligand and receptor flexibility as well as treating the solvent explicitly clearly increases the computational complexity of docking approaches. However, in view of the ever-increasing computing power, a new generation of docking methods should evolve, which, though being computationally more demanding, aims to deal with the above-mentioned challenges.

Molecular dynamics (MD) techniques are established for rigorous studies of intermolecular interactions.[23] Beyond that, MD methods have already been used in docking approaches for

Received: October 21, 2013
Published: January 31, 2014

dx.doi.org/10.1021/ci4006047 | J. Chem. Inf. Model. 2014, 54, 582−592

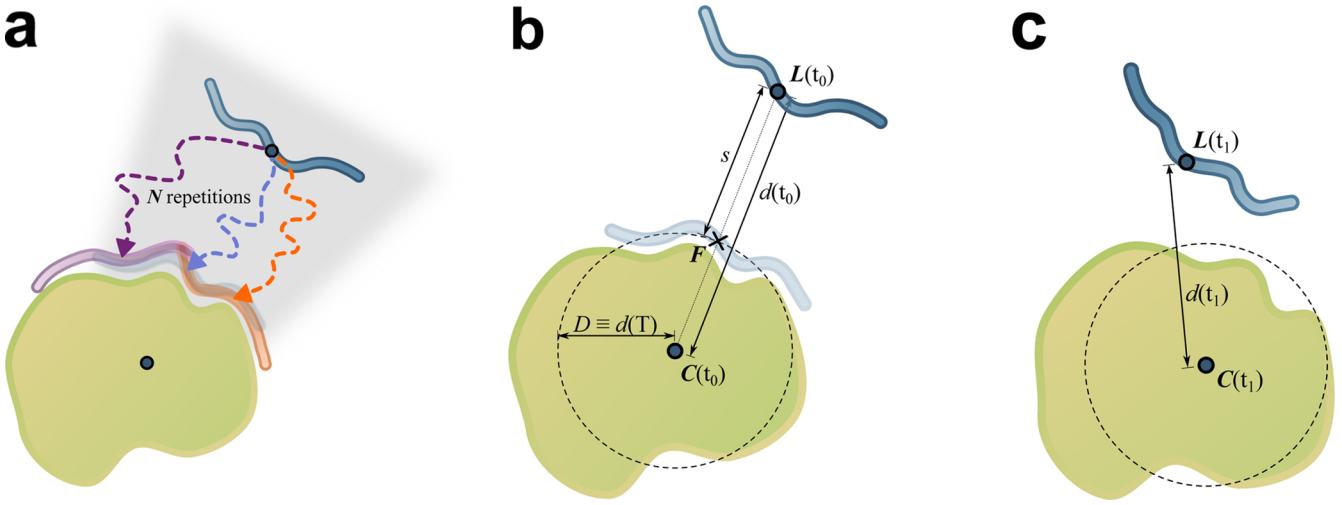


Figure 1. (a) Schematic representation of the DMD pulling process. Starting from a distal position, the ligand (blue) is pulled toward its receptor (green). During this process, the ligand moves along a random path. The pulling process is repeated N times in independent simulations. Most of the resulting ligand trajectories lie within a certain "entry lane", as depicted here in gray shade. Each pulling process yields an individual final ligand state (purple, blue, orange) near the receptor surface. (b) System configuration at time t_0 right before the pulling process, referred to as the ligand-reoriented model (LROM, with the displaced ligand in dark blue). The central atom of the experimentally determined ligand position (light blue) has been placed at a distal position $L(t_0)$ along the axis given by a receptor core atom at point $C(t_0)$ and the focus point F. s is the displacement length. The ligand has been randomly rotated around its central atom. The distance between $C(t_0)$ and F defines the final distance D for the ligand pulling process. The initial distance $d(t_0)$ between C and L is $D + s$. (c) Arbitrary state within the pulling process, during which the distance $d(t)$ between central ligand atom at point $L(t)$ and protein core atom at point $C(t)$ is decreased over time t. Right after the pulling process, all final ligand states have their central atom placed on the sphere that is indicated here with a dashed line.

overcoming the challenges of both ligand and receptor flexibility.[24,25] Furthermore, standard MD methods allow for the inclusion of explicit solvent via well-established water models. The application of MD techniques is limited by high computational cost, which previously has hindered their usage in high-throughput approaches for drug discovery. However, nowadays, advanced computational resources are available, and specialized hardware (such as graphics processing units, GPUs) can be used to dramatically increase the performance of MD simulations, making the establishment of MD in the field of docking gradually feasible.

We propose Dynamic Molecular Docking (DMD), a combination of established MD-based methods specifically designed for tackling the above-mentioned challenges in protein−GAG docking. DMD is a targeted MD-based approach where the ligand, which is initially placed at a distance from the receptor so that their interaction is negligible, is slowly pulled toward a receptor target region by applying a time-dependent distance restraint. During this process, both receptor and ligand are treated as entirely flexible in explicit solvent. The time-dependent distance restraint applied in DMD is usually used in steered MD simulations, which are performed for studying the energetics of processes that generally happen on time scales too large for being treated by classical MD simulations,[26] such as protein folding, ligand unbinding, and large-scale conformational alteration. In DMD, once the ligand reaches the receptor, the distance restraint is switched off and a long free MD simulation is carried out, allowing for the mutual adjustment of receptor residues and GAG as well as for extensive GAG-internal degree of freedom sampling. The obtained trajectory data are then used for extracting a docking solution (the coordinates) and for its characterization (binding energy estimate and other quantities). Binding pose search as well as binding pose scoring are consistent in terms of using the same potential as given by the MD force field and the parametrization of the molecular system.

In this validation study, we applied DMD to a set of reference systems and compared its performance to a classical docking approach, AutoDock 3 (AD3), which has successfully been applied to protein−GAG systems before.[20,27,28] The results obtained for our test data set comprised of five protein−GAG complexes, one protein−peptide, and one protein−small molecule complex show that DMD has high predictive significance while performing best when complex formation is driven by strong electrostatic interaction. Detailed analysis of the electrostatic potential of the corresponding binding sites, of the spatial distribution of docking solutions, and per-residue binding free energy decomposition support the promising potential of the DMD approach when used for docking in highly flexible and electrostatics-driven systems.

MATERIALS AND METHODS

Test Data Set. Seven protein−ligand complexes with experimentally determined 3D structures were used as test data set (TDS). Five of them are protein−GAG systems: basic fibroblast growth factor (FGF2) in complex with a heparin (HP) tetrasaccharide, PDB ID 1BFB, 1.9 Å resolution; cathepsin K (CathK) in complex with a chondroitin-4-sulfate hexasaccharide (CS4), PDB ID 3C9E, 1.8 Å; a CathK mutant (referred to as CathKmut) in complex with a CS4 hexasaccharide, PDB ID 3H7D, 2.2 Å; CD44 in complex with a hyaluronan heptasaccharide (HA), PDB ID 2JCQ, 1.3 Å; stromal cell-derived factor-1 (SDF-1) in complex with a HP disaccharide, PDB ID 2NWG, 2.1 Å. In our studies, we also included two complexes of different nature: Abl-SH3 domain complexed with a decapeptide (p41), PDB ID 1BBZ, 1.7 Å; trypsin in complex with the inhibitor benzo[b]thiophene-3-

dx.doi.org/10.1021/ci4006047 | J. Chem. Inf. Model. 2014, 54, 582−592

methanamine (referred to as Ihb.), taken from the DINGO data set.[29]

Dynamic Molecular Docking protocol. For each complex in the TDS, DMD was carried out in the following way. Starting with randomly oriented ligand molecules placed at a distance from the receptor (ligand reoriented models, LROMs), 100 independent DMD runs were performed. The first step of a DMD run is a targeted molecular dynamics (tMD) simulation in which the ligand is pulled toward a predefined target region on the receptor via a time-dependent decrease of the distance $d(t)$ between one central atom in the ligand and one core atom in the protein receptor. Among DMD run repetitions, most of the ligand trajectories lie within a certain "entry lane" which is focused on a point near the receptor surface, the focus point F, and therefore defining the target region (Figure 1). All final tMD states have the central ligand atom positioned on the surface of a sphere defined by the protein core atom (the center of the sphere) and the final distance D of the tMD pulling process (the radius of the sphere). On the basis of the final state of each tMD simulation, the second step of a DMD run relaxes the system via a free MD simulation. Geometrical definitions, system preparation, and details about tMD and free MD parametrization as well as subsequent trajectory data analysis methods are provided in the following sections.

Preparation of Ligand-Reoriented Models (LROMs). An LROM contains the receptor as well as the ligand placed in a distal, reoriented position. Per TDS complex, 10 LROMs have been created, differing only in ligand orientation around its central atom. For a given complex, LROM creation requires the selection of a *central ligand atom*, definition of a *focus point* $\boldsymbol{F}$ near the surface of the receptor within the anticipated binding region, definition of a *ligand displacement length* s, and selection of a *core atom* at point $\boldsymbol{C}$ within the receptor. The distance between focus point and core atom defines the final distance D of the tMD pulling process, i.e. $d(T) \equiv D \equiv \|\boldsymbol{F} - \boldsymbol{C}\|$. Initially, at time t_0, the ligand is placed distal from the receptor with its central atom lying on the axis defined by $\boldsymbol{F} - \boldsymbol{C}$ (Figure 1b). The starting coordinate for the central ligand atom is defined as

$$\boldsymbol{L}(t_0) = \boldsymbol{F} + s\frac{\boldsymbol{F} - \boldsymbol{C}}{D} \tag{1}$$

For each TDS complex, 10 LROMs were prepared as follows. First, the structure of the biological unit of the protein receptor was taken from the corresponding experimental data source. The coordinates of a central atom in the ligand as found in the experimentally determined structure were used as a focus point. The core atom in the receptor was selected fulfilling three criteria: (*i*) it is a backbone atom within a helix or β sheet in the protein core, (*ii*) the line connecting the core atom and $\boldsymbol{F}$ (defining the orientation of the "entry lane") is roughly perpendicular to the surface comprising the anticipated binding region, and (*iii*) the surface of the sphere around the core atom with radius D has significant overlap with the molecular receptor surface in the receptor target region. Distal ligand placement was followed by 10 random ligand rotations uniformly distributed in 3D space with $\boldsymbol{L}(t_0)$ being the rotation center. From each rotational ligand state, one LROM was built.

Molecular Dynamics Protocol. All MD simulations were set up, performed, and analyzed using Amber 11 and AmberTools 12.[30] The FF99SB force field was used for parametrization of the peptidic parts in the TDS complexes. Hyaluronan, chondroitin sulfate, and heparin monosaccharide force field parameters were created based on GLYCAM 06 version g[31] and sulfate partial charges obtained by RESP fitting calculations at the level of 6-31(d)G for methylsulfate. The trypsin inhibitor was parametrized using the GAFF force field via AmberTools' antechamber program. All systems were solvated in a box of TIP3P water with a minimum of 8 Å distance between solute and box boundaries. The systems were neutralized by adding Na^+ or Cl^- counterions. During MD, lengths of bonds including hydrogen atoms were constrained by SHAKE. The time integration step was set to 2 fs. Nonbonded interactions were cut off for distances larger than 8 Å. The Particle Mesh Ewald method was used for treating long-range electrostatic interactions.

Each simulated system went through minimization, heat up, equilibration, and production steps. During the first stage of minimization, only the solvent was relaxed. During the second stage, the entire system was minimized without restraints. System heat up to 300 K was performed within 20 ps in the canonical ensemble (NVT) using the Langevin thermostate and periodic boundary conditions. Subsequently, 500 ps of MD in the isothermal−isobaric ensemble (NPT) with Langevin thermostate and Berendsen barostat under periodic boundary conditions were carried out for system equilibration. The following production stage of duration T was performed in the NVT ensemble with the Berendsen thermostat and periodic boundary conditions.

For the complexes involving FGF2 and heparin, weak torsional restraints were applied in order to keep the pyranose rings of IdoA(2S) in the $^1C_4$ conformation. This conformation has been shown to be one of the two predominantly populated ones[32] and was observed experimentally in the structure of the FGF2-HP complex in our test data set (PDB ID 1BFB). The applied restraints enable definition and control of the specific conformation of each IdoA(2S) ring throughout the entire DMD study, since its natural ring conformer population is not properly reproduced by GLYCAM 06.[33]

Targeted and Free Molecular Dynamics. The LROMs were prepared for MD and time-evolved following the general MD protocol as described above. During the tMD production stage, the core atom and ligand center atom were exposed to an additional time-dependent harmonic potential

$$U(t) = \frac{1}{2}k(d(t) - d(t_0) + vt)^2 \tag{2}$$

with force constant k = 200 kcal mol^{-1} A^{-2}, pulling velocity $v = s/T$ and

$$d(t) = \|\boldsymbol{L}(t) - \boldsymbol{C}(t)\| \tag{3}$$

This potential enforces the distance between the selected ligand center atom and the protein core atom to linearly decrease with time t from $D + s$ to D (Figure 1c).

For all tMD simulations, we used a pulling velocity of s = 30 Å per T = 4 ns. The tMD pulling process was repeated 10 times with different random seeds for each of the 10 LROMs per complex, yielding 100 independent tMD simulations for each TDS complex. On the basis of the final state of the receptor and ligand in each tMD trajectory, an MD simulation with T = 10 ns without restraints was performed according to the protocol described above, hereafter referred to as free MD.

Energetic Evaluation of DMD Docking Results. From the last 200 ps of each free MD trajectory, 100 equidistantly distributed frames were extracted for energy analysis. The MM-

dx.doi.org/10.1021/ci4006047 | J. Chem. Inf. Model. 2014, 54, 582−592

PBSA[34] approach was applied for calculating the time-averaged Coulomb interaction energy ΔE between receptor and ligand as well as an estimate for the free energy of binding ΔG. MM-GBSA[34] single residue energy decomposition (SRED) was applied to estimate the energy contribution of single receptor residues to the bound state.

In order to identify the receptor residues mostly contributing to the binding from all 100 DMD runs corresponding to one TDS complex, the SRED data of all free MD trajectories were filtered and merged: we excluded DMD runs resulting in weakly bound docking solutions (MM-PBSA $\Delta G > -1$ kcal mol^{-1}) and averaged the SRED-energy ΔG_R for each receptor residue over the remaining independent DMD runs. We discarded all receptor residues with an average SRED-energy $\langle \Delta G_R \rangle \geq 0$ kcal mol^{-1} and ranked the remaining ones by $\langle \Delta G_R \rangle$. For each TDS complex, we extracted the 10 top-ranked residues, referred to as the set of receptor *anchoring residues*. For reference, we set up an identical MD simulation (T = 10 ns) of each experimentally determined TDS complex and used SRED to obtain a set of reference anchoring residues for comparison.

MM-PBSA free energy calculations as well as MM-GBSA single residue energy decompositions were performed with default parameters using the Python version of the MM-PBSA application provided with AmberTools 12. We did not consider an entropic contribution to binding since we aimed to compare very similar systems where taking into account entropy may increase the overall uncertainty in the calculated binding energies.[35,36]

Classical Molecular Docking. For comparison with DMD, classical docking based on AutoDock 3[37] (AD3) was applied to all TDS complexes. AD3 has been shown to produce reasonable results for GAG-protein systems, especially compared to other docking methods.[20,27,38,39]

With AD3, a semiflexible ligand is docked to a static receptor without the presence of explicit solvent. In order not to be biased toward a ligand-induced receptor conformation, prior to docking with AD3, we relaxed all TDS receptor structures via energy minimization in the Amber 99 force field as implemented in MOE[40] using a convergence criterion of 0.01 Å heavy atom RMSD.

For each TDS complex, the potential grid of the receptor was calculated with a grid step of 0.38 Å. Grid size, position, and orientation were chosen so that the spatial sampling volume is comparable to the one in DMD. We set ligand torsional degrees of freedom for glycosidic linkages as well as sulfate and carboxyl groups to be flexible during the search. A total of 10^3 independent docking runs were performed for each complex. Each docking run was performed using a genetic algorithm (initial population size, 300; abortion condition, 10^5 generations) followed by a local search. If not stated otherwise, default parameters were used. Spatial clustering was performed with the top 100 solutions according to the AD3 score.

Spatial Clustering of Docking Results. In order to evaluate the spatial distribution of docking solutions generated by either docking method, we used the DBSCAN clustering algorithm.[41] Based on the parameters ε (neighborhood search radius) and m (the minimal neighborhood size), DBSCAN assigns data points to clusters or classifies them as noise.

Since spatial clustering evaluates distances between data points rather than data points themselves, the distance metric used for calculation of the similarity between two structures must be properly adjusted to the scientific aim. Thus, we define a distance metric δ as the root-mean-square of pairwise atomic distances while pairing up the spatially closest atoms of the same type. This distance metric accounts for periodicity of functional groups in GAGs and considers two GAGs shifted by n periodic units as structurally similar, while classical RMSD with identity-based matching ignores this characteristic.

Prior to clustering, we transformed all docking solutions into the same coordinate system via structural alignment of corresponding protein receptors. The parameters ε and m were determined individually for each set of docking solutions with the goal to produce one cluster with at least four members and a minimal ε value.

Electrostatic Potential Analysis. For each TDS complex, we analyzed the electrostatic potential of the receptor calculated numerically with AmberTools' PBSA program (choosing the modified ICCG solver applied to the linearized Poisson−Boltzmann equation). The receptor atoms were parametrized according to the FF99SB force field. The grid spacing was set to 1 Å. Isosurface visualization of the potential was done in VMD.[42]

Computational Resources. Most of the simulations were performed on a compute cluster comprised of AMD Opteron 6274 CPUs. On these machines, a DMD study as described above requires about 100.000 CPU hours per TDS complex. For testing purposes, some MD simulations were executed on Nvidia Tesla C2070 as well as Nvidia GTX 580 GPUs, taking advantage of Amber's recent optimizations for such hardware.[43]

RESULTS AND DISCUSSION

Sampling of the Degrees of Freedom of the Ligand during DMD. A molecular docking method should be able to extensively sample the internal degrees of freedom (DOFs) of the ligand as well as translational and rotational DOFs of the entire ligand within a certain volume including part of the receptor surface (the anticipated binding region). It is important to note that ligand pulling via tMD does not occur along a predefined axis: the ligand follows a random trajectory and samples its conformational space (Figure 1a). While doing so, the pulling velocity has a crucial impact on the sampling extent of especially the rotational and translational DOFs of the ligand: the slower the ligand is pulled, the more time it has to respond to the potential of the receptor and—with that—to deviate from the shortest path toward the receptor (an extreme scenario with translation along the initially defined displacement axis only).

For the FGF2-HP complex, we compared tMD runs with different pulling velocities with the goal to find a moderate velocity which still allows significant translational and rotational ligand movement. In such a case, entirely different trajectories among tMD repetitions are produced so that Cartesian space sampling as well as ligand orientation sampling can be enhanced by increasing the number of tMD repetitions and leaving all other parameters constant. In tMD test runs with varying duration T but constant ligand displacement length s, we measured the structural difference (RMSD) between the current and initial ligand configuration for each point in time during tMD. Ligand translation, rotation, and internal conformational changes contribute to this time-dependent structural difference denoted as $\mathrm{RMSD}_{\mathrm{rot,trans,int}}(t)$. The difference $\mathrm{RMSD}_{\mathrm{rot,trans,int}}(t) - ts/T$ describes the extent of ligand translation and rotation in space compared to the shortest-path scenario.

dx.doi.org/10.1021/ci4006047 | J. Chem. Inf. Model. 2014, 54, 582−592

For single test cases with T = 4 ns, we observed a quickly fluctuating $\mathrm{RMSD}_{\mathrm{rot,trans,int}}(t)$ curve with ligand displacements from the shortest path between 5 and 40 Å RMSD. In contrast, when using T = 0.5 ns, the displacement varied in the range between 0 and 10 Å RMSD. From these data, we could already conclude that when using T = 4 ns, the limiting factor for translational and rotational ligand sampling is the number of tMD repetitions rather than the pulling velocity. To further quantify the translational and rotational freedom of the ligand during tMD with s = 30 Å and T = 4 ns in detail, we analyzed the deviation of the ligand from the shortest path for an ensemble of 100 independent tMD runs. To this end, the structural difference between current and initial ligand configuration was calculated as described above and then averaged over all 100 independent pulling processes for each point in time during tMD, leading to $\langle\mathrm{RMSD}_{\mathrm{rot,trans,int}}\rangle(t)$ (Figure 2). The average extent of ligand translation and rotation in space turned out to have its maximum at an RMSD value of about 20 Å. The path variation among ligand trajectories (the standard deviation of the averaged data) is about ± 10 Å RMSD. The data suggest that with s = 30 Å and T = 4 ns, enough translational and rotational freedom are provided to the ligand during tMD in order to be applicable in a local docking method. This also verifies that the extent of sampling of all ligand DOFs is determined and can be adjusted by the number of tMD repetitions N.

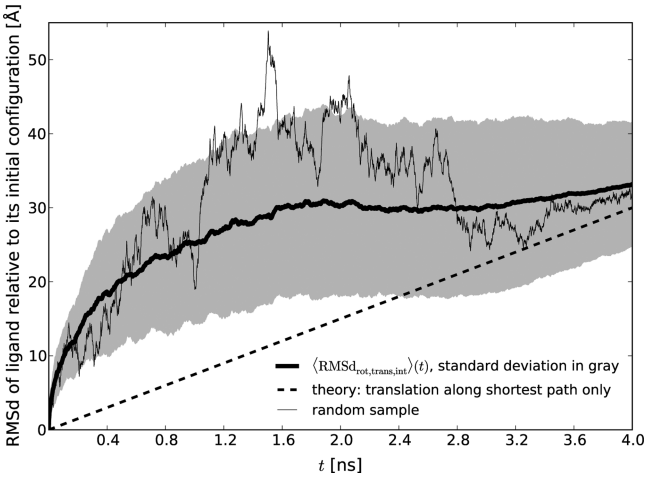


Figure 2. Translational and rotational freedom of a heparin tetramer while approaching FGF2 during the tMD pulling process with s = 30 Å and T = 4 ns. For each point in time during tMD, the structural difference between the current and initial ligand configuration was calculated via classical RMSD and averaged over 100 independent pulling processes. The variation in ligand movement among different ligand trajectories is visualized in terms of the standard deviation (gray background).

$\langle\mathrm{RMSD}_{\mathrm{rot,trans,int}}\rangle$ increases slower after about t = 2 ns (Figure 2), which is probably caused by the electrostatic potential of the receptor prevailing against thermally driven ligand movement. Having this transition included in the $\langle\mathrm{RMSD}_{\mathrm{rot,trans,int}}\rangle(t)$ curve is a good indication for a properly (long enough) selected initial displacement length s. As a general rule, s should be chosen in a way that no atoms from the initially placed ligand lie within the nonbonded interaction cutoff range of the receptor.

Due to the type of distance restraint applied, all final tMD states have the central ligand atom positioned on the surface of the sphere given by the protein core atom and the final restraint distance D. A more robust realization of the basic DMD concept would not require assigning specific roles to certain points (core atom and focus point) but implement a dynamic distance restraint between receptor surface and ligand during the pulling process.

In the current setup, we have compensated for the spherical—and therefore limited—distribution of tMD final ligand states via careful selection of the core atom, of D, and via a long free MD simulation stage with T = 10 ns. This allows for significant ligand refinement according to the characteristics of the receptor surface as well as for an exhaustive sampling of the internal DOFs of the ligand. We have quantified the extent of glycosidic linkage sampling by analyzing the distribution of glycosidic linkage dihedral values within all free MD simulations of a DMD study with FGF2 in complex with heparin. Since GLYCAM 06 is known for its proper description of the glycosidic linkage torsional potential,[31] we extracted the distribution of equivalent torsion angles from an independently performed 760 ns MD simulation of the same heparin molecule free in solution for reference. We compared both distributions and observed high similarity between them (Supporting Information Figure 3). In particular, the distribution as observed for the GAG-only simulation is entirely included in the distribution as observed from the DMD MD simulations where the GAG is in contact with its protein receptor. In the latter case, some additional conformers seem to be accessible to the ligand when compared to the GAG-only simulation, which is to be expected due to its interaction with the receptor. For further characterization of the sampling, we extracted glycosidic linkage angle values from 12 PDB entries containing free heparin as well as heparin−protein complexes (1AXM, 1BFB, 1BFC, 1E0O, 1FQ9, 1G5N, 1GMN, 1HPN, 1QQP, 2AXM, 2HYU, 2HYV) and observed that this range of experimentally determined angles is included within the glycosidic torsion angle distribution as sampled by our MD simulations (also shown in Supporting Information Figure 3).

Most important, we have observed that due to the long free MD stage and a large number of DMD repetitions, the overall result of a DMD study becomes insensitive to moderate changes in focus point coordinates as well as in the final tMD restraint distance D.

When visualizing trajectories of tMD and free MD, we repeatedly observed conceptually different scenarios. In one type of scenario, the ligand ended up in its native binding pose in the final state of tMD. The conformation of the ligand then did not undergo significant changes during free MD. In other cases, the ligand arrived near its anchoring residues on the surface of the receptor during tMD and then adopted the native binding mode within the free MD simulation. We also observed the ligand to end up in a false-positive bound state as well as in an unbound state after tMD or to unbind during free MD.

From these observations, we can deduce that if after DMD no agglomeration of docking solutions stands out, the chosen receptor target region most likely does not contain a real binding site for the ligand. Furthermore, even if the focus point of the method is not centered on but only in the vicinity of a real binding site, the latter most likely stands out during data analysis.

Electrostatic Potential Analysis in Protein−GAG Systems. The electrostatic potential often dominates protein−GAG interaction.[44] In this section, we discuss the electrostatic properties of the receptors in the TDS with the goal to determine how these properties could assist defining a

dx.doi.org/10.1021/ci4006047 | J. Chem. Inf. Model. 2014, 54, 582−592

receptor target region for DMD and also to be able to relate docking performance to electrostatic characteristics of the receptor.

Poisson−Boltzmann-based numeric calculations of the electrostatic potential of molecules are most error-prone near the dielectric boundary, i.e., on the molecular surface. We therefore did not simply map the potential on the molecular surface but analyzed the topology of the potential with an isosurface representation while varying the isosurface value. This allows for an understanding of the distribution of the potential in space and how strongly it would affect a ligand. Figure 3 shows an isosurface of the Coulomb potential of SDF-1. Supporting Information Figure 1 shows analogous isosurface representations of the electrostatic potential for the other TDS complexes. For this type of visualization, the isovalue selection was done individually for each receptor in the TDS such that only a small part of the isosurface is protruding into space further than the molecular surface of the receptor.

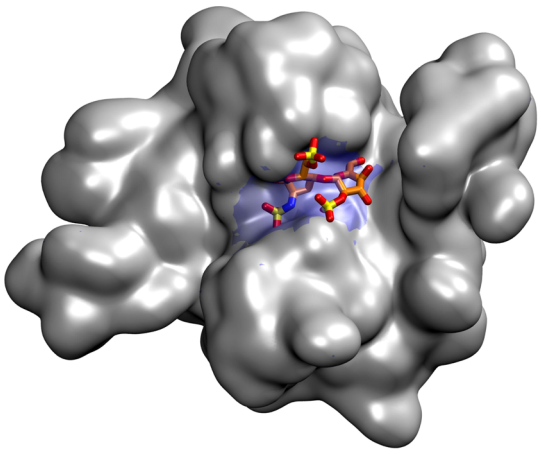


Figure 3. Isosurface representation of SDF-1's electrostatic potential with the isovalue Φ = 8.5 kcal mol^{-1} e^{-1} (blue). The molecular surface of the SDF-1 dimer is shown in gray; the heparin ligand as determined experimentally (PDB ID 2NWG) is shown as sticks with C atoms in orange.

Regarding the SDF-1-HP complex, our electrostatic potential evaluation procedure unambiguously identifies the GAG binding site as determined experimentally. With 8.5 kcal mol^{-1} e^{-1}, the isovalue chosen is the largest among the TDS complexes, indicating that SDF-1 has the strongest electrostatic attraction to its ligand. In case of FGF2-HP (Supporting Information Figure 1), the binding site is also unambiguously defined by the electrostatic potential. For CD44-HA, the net electrostatic interaction between both molecules is slightly repulsive. There is no obvious relation between the electrostatic properties of the receptor and the binding site location. CathK and CathKmut exhibit electrostatic attraction for negatively charged molecules in the experimentally observed ligand binding site as well as in an adjacent region. We observe that electrostatic potential analysis provides a clear idea whether GAG binding to a given receptor is mainly driven by Coulomb interaction. If a protein is known to bind GAGs, and the electrostatic potential topology is as unambiguous as in case of SDF-1 or FGF2, a GAG binding site prediction based on the presented procedure is reliable. Visualization of the electrostatic potential can also be helpful to *a priori* exclude regions of the receptor surface when repulsive to negatively charged ligands. Furthermore, this analysis shows that knowledge about the electrostatic potential distribution in space can be used to choose a reasonable ligand "entry lane" orientation for the tMD pulling process.

The SH3-p41 complex is dominated by nonpolar interactions, rendering the Coulomb potential analysis inconclusive. Recognition of the trypsin inhibitor in its binding pocket is affected by polar interactions. The electrostatic potential, however, does not display clear characteristics to predict a binding region.

DMD Performance. *Spatial Distribution of Docking Solutions.* We compared the performance of DMD with the performance of the well-established AD3 docking approach, which was used for other molecular systems including GAGs. We applied DMD and AD3 to all TDS complexes and analyzed the spatial distribution of docking results via clustering (clustering parameters listed in Supporting Information Table 3). For each complex, we compared the mean structural distance $\langle\delta\rangle$ between the experimentally determined ligand position and all members of the most populated cluster (Figure 4). In terms of the capability to produce solutions structurally

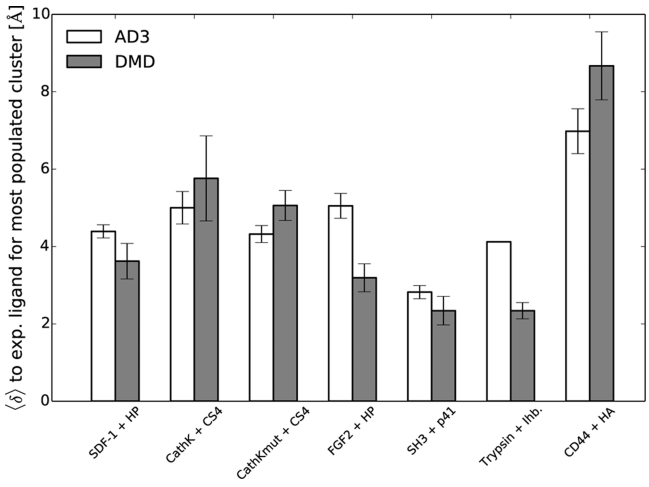


Figure 4. Mean structural distance $\langle\delta\rangle$ between all members of the most populated cluster and the experimentally determined ligand for all investigated complexes with both docking methods. The error bars show the standard deviation of the mean.

close to the experimentally determined ones, both docking methods perform comparably well. However, interesting differences are observable. AD3 generally yields smaller spatial scattering of solutions as given by the standard deviation of data points in Figure 4. This can be attributed to the static treatment of the receptor by AD3. Notably, we can generally conclude that DMD performed better than AD3 for the complexes with strongest electrostatic attraction, namely SDF-1-HP and FGF2-HP. For the FGF2-HP complex, the most populated cluster from DMD reproduces the experimentally determined HP binding pose, while the most populated cluster from AD3 docking only partially overlaps with this pose (Figure 5A). In fact, DMD was able to identify the "higher affinity" binding site for heparin (as denoted in the original publication of the FGF2-HP crystal structure), while AD3 identified the "lower affinity" site, which becomes occupied upon heparin hexasaccharide binding.[45] In addition, DMD was able to properly predict the positioning of the two sulfate groups making specific high-affinity contact with FGF2 (Figure 5B). These two sulfate

dx.doi.org/10.1021/ci4006047 | J. Chem. Inf. Model. 2014, 54, 582−592

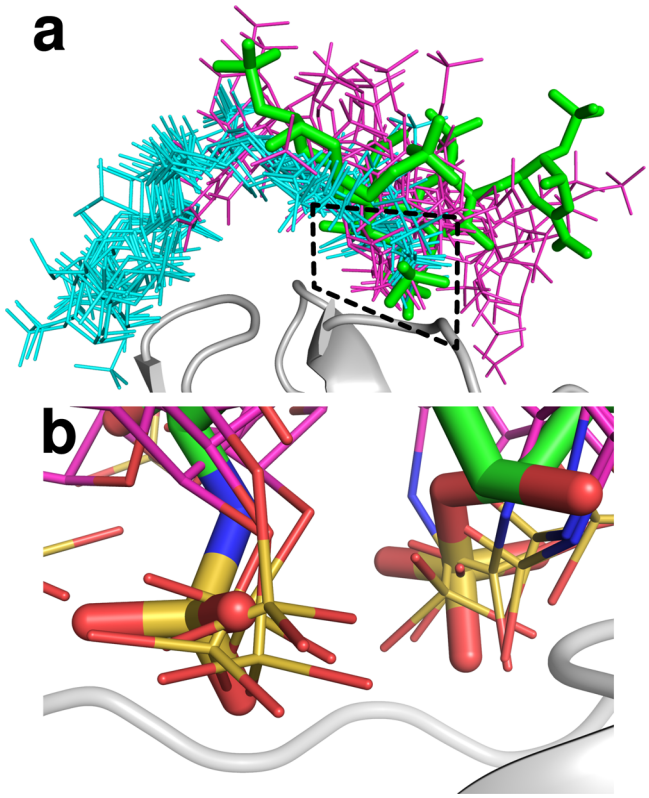


Figure 5. Docking results from DMD and AD3 for FGF2 (gray cartoon representation) and HP. (a) Ligand from crystal structure in green sticks. Most populated cluster of DMD solutions in magenta. Most populated cluster of AD3 solutions in cyan. (b) Zoom on two sulfate groups making specific high-affinity contact to FGF2[45] (as marked in (a) via dashed line). Ligand from crystal structure with C atoms in green (thick sticks). Most populated DMD cluster with C atoms in magenta (thin sticks).

groups form seven of the nine polar contacts between FGF2 and HP and therefore are important anchoring groups for the molecular recognition.[45] The fact that DMD was able to reproduce this key feature as opposed to AD3 most probably reflects the crucial role of receptor flexibility considered in the docking simulation.

In order to assess the ability of DMD to predict consistent binding poses for GAGs differing in length, we carried out additional DMD studies for SDF-1 as well as FGF2. Eventually, both systems were investigated with independent DMD studies for each of di-, tetra-, and hexasaccharide of heparin. For FGF2-HP, the two sulfate groups making specific contact with FGF2 were identified when docking both an HP tetrasaccharide and an HP hexasaccharide. These observations support the assumption that GAG fragments differing in length share key interactions with the protein. When docking an HP disaccharide, those key interactions were not identified, suggesting that the disaccharide is not long enough to maintain specificity. Nevertheless, for the SDF-1-HP and FGF2-HP systems, we observe that all obtained poses overlap regardless of GAG length (Supporting Information Figures 4 and 5). A direct visual comparison between the crystal structure of FGF2 with a heparin hexasaccharide[45] and the corresponding DMD result can be seen in Supporting Information Figure 5.

As has been conceptually described for the FGF2-HP complex, also regarding the SH3-p41 and trypsin-inhibitor complexes, DMD provided solutions closer to the experimentally determined ones while displaying a more diverse scattering of docking solutions than AD3.

For CD44 in complex with the HA heptamer, both docking methods were only partially able to reproduce the experimentally determined ligand pose. Weak net electrostatic repulsion between receptor and ligand seemingly imposes a major challenge for both docking methods. AD3 and DMD solutions, however, are spatially consistent (data not shown).

Both docking methods had difficulties reproducing the experimentally determined binding poses of CathK and CathKmut in complex with CS4. In the case of CathK-CS4, the crystal structure contains a single high molecular weight polymeric CS4 molecule interacting with multiple copies of the same protein.[46] Within the CathKmut-CS4 crystal, each GAG hexamer tightly interacts with two CathKmut proteins.[47] Under these conditions, it is to be expected that considering only a short GAG and a single protein in the docking experiment is not sufficient to allow for an accurate reproduction of the experimentally determined structures.

The presented structural difference comparison between DMD and AD3-based docking solutions and the corresponding experimentally determined ligand coordinates is subject to a systematic error. Structural relaxation prior to docking with AD3 changed the receptor side chains in the binding region by a heavy atom RMSD of 1.0 ± 0.4 Å (mean value and standard deviation derived from all TDS complexes) when compared to the experimentally determined receptor structure. The solutions from DMD display a mean binding region alteration of 2.3 ± 0.6 Å (corresponding raw data are shown in Supporting Information Table 4). The larger the binding site alteration, the more error-prone the comparison between docked solution and experimentally determined ligand becomes. Under the assumption that this error systematically increases the structural difference, DMD might have performed better compared to AD3 than shown in Figure 4.

Energetic Evaluation of Docking Results. *Scoring.* Assigning an energy-related score to each obtained docking solution allowing for an energetic ranking of such is well-established. While AD3 provides its own scoring scheme, we built two different score values per DMD docking solution: the Coulomb energy ΔE between receptor and ligand as well as an MM-PBSA estimate for the free energy of binding, ΔG (as described in Materials and Methods).

Although debatable, in order to validate a scoring method, the rank by score is often expected to decrease with increasing distance between the ideal solution and the scored docking solutions.[7] We calculated Spearman's rank correlation between δ and AD3 score, DMD ΔG, and DMD ΔE for all TDS complexes. For each scoring method, two Spearman coefficients were determined based on two different docking solution subsets: the first subset included all solutions with $\delta \leq 20$ Å; the solutions in the second subset fulfill $\delta \leq 6$ Å. Coefficients have been calculated only when at least eight value pairs were available. Between the two distance filter cases, the resulting correlation coefficients varied significantly: for both DMD scoring methods, we generally observed higher coefficients for $\delta \leq 20$ Å than for $\delta \leq 6$ Å in the case of the four protein−GAG complexes dominated by electrostatic attraction (FGF2-HP, SDF-1-HP, CathK-CS4, CathKmut-CS4, see Supporting Information Figure 2). This is an effect of long-range electrostatic potential, which slowly decays with increasing distance from the GAG binding site. Within the vicinity of the experimentally determined ligand position ($\delta \leq 6$

dx.doi.org/10.1021/ci4006047 | J. Chem. Inf. Model. 2014, 54, 582−592

Å), we observed that only DMD ΔE yielded a correlation coefficient larger than 0.2 for most of the complexes. From the data, we cannot derive a clear relation between δ and any of the scoring metrics.

In order to analyze the influence of GAG length on the DMD ΔG score, we investigated the MM-PBSA results of our DMD studies with SDF-1 and FGF2 in complex with heparin di-, tetra-, and hexasaccharides. We observe that the free energy of binding estimates becomes more favorable with increasing GAG length: the average ΔG values in kcal/mol for the most populated clusters of SDF-1-HP are −59.2, −96.7, and −144.4 and of FGF2-HP are −37.0, −73.8, and −80.4 for di-, tetra-, and hexasaccharides, respectively. For FGF2, the difference between the values for hexa- and tetrasaccharide is rather small compared to the difference between values for tetra- and disaccharide. This is in agreement with experimental data showing that a heparin tetrasaccharide is sufficient to occupy the high affinity part of the FGF2 heparin binding site.[45]

Identification of Anchoring Residues in the Ligand Binding Region. The prediction of receptor residues important for ligand binding is of particular practical value. While AD3 does not provide energetic data on the single-residue level, DMD allows for a single-residue energy decomposition (SRED) based on time-averaged free MD interaction energy data.

For each TDS complex, we merged SRED data from the entire ensemble of DMD runs and extracted the set of anchoring residues, which are at most 10 of the receptor residues contributing most (positively) to ligand binding. We also determined a reference set of anchoring residues from MD simulations of the experimentally determined structures. For each TDS complex, we then identified the intersection of both sets (Table 1).

Table 1. Predictive Power of DMD for the Identification of Receptor Anchoring Residues

complex	number of correctly predicted residues[a]	number of charged residues	number of polar uncharged residues	r^b
SDF-1 − HP	7 of 10	7	0	0.52
CathK − CS4	6 of 10	4	2	−0.21
CathKmut − CS4	6 of 10	4	1	−0.12
FGF2 − HP	9 of 10	6	2	0.84
SH3 − p41	2 of 5	0	0	
Trypsin − Ihb.	5 of 10	2	2	0.73
CD44 − HA	1 of 7	1	0	

[a]Number of predicted receptor anchoring residues having a match in the reference set (as obtained from SRED via MD of the experimentally determined structures). [b]Spearman's rank correlation r between the SRED energies of both residue sets (calculated only if five or more residues match).

For trypsin in complex with its inhibitor, half of the anchoring residues in the crystal structure were correctly found. For SDF-1-HP, CathK-CS4, and CathKmut-CS4, more than half of the anchoring residues were identified by SRED. Furthermore, for the FGF2-HP complex, nine of 10 anchoring residues were predicted properly.

In order to evaluate the capability of DMD to rank the receptor anchoring residues according to their importance for ligand binding, we calculated Spearman's rank correlation between the SRED energies of residues in both the reference and comparison sets (Table 1). For three of those five complexes with at least five predicted anchoring residues, the DMD ranking turned out to reproduce the reference ranking quite well (SDF-1-HP, FGF2-HP, Trypsin-Ihb). For the CathK-CS4 and CathKmut-CS4 complexes, the anchoring residue ranking was not reproduced.

On the basis of the obtained results, we can conclude that for systems dominated by electrostatic attraction such as protein−GAG systems, DMD is capable of correctly identifying those amino acid residues responsible for forming key interactions with the ligand. Regarding the FGF2-HP system, we found that both heparin tetra- and hexasaccharide docking via DMD enabled us to consistently identify the same chemical groups in the GAG molecule being key for binding. Moreover, using SRED data, we were able to demonstrate that the identified key amino acid residues as well as their energy ranking are identical when comparing heparin tetra- and hexasaccharide docking performed in independent DMD studies (see Figure 6). This

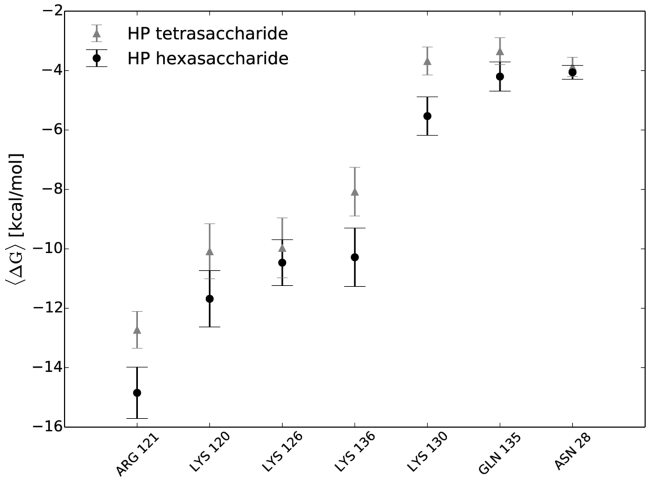


Figure 6. FGF2 amino acid residues identified by DMD as having the greatest impact on FGF2-HP binding. Results are shown for both heparin tetra- and hexasaccharide ligands as retrieved from two independent DMD studies. For each residue, an average energy (±standard error of the mean) for the interaction with the ligand is shown as obtained from 40 independent free MD simulations.

observation supports the idea that, regardless of their length, GAG fragments preserve their key interactions with certain protein amino acid residues. The set of FGF2 key amino acid residues as identified by SRED contains all residues that were crystallographically characterized as most important for binding (R121, K126, N28, Q135).[45] On the basis of our data, we can additionally state that also residue K120 is of major importance, since it acts as a hydrogen bond donor in a similar way to the way K126 does. The latter observation is in agreement with previous calorimetry studies pointing out the importance of K120 in the FGF2-HP system.[48] These findings motivate the application of DMD to systems in which a detailed and appropriate description of GAG recognition is required.

In summary, our data suggest that DMD is able to yield docking results of high significance, especially in the case of receptor−ligand systems dominated by attractive electrostatic interaction. This success is well-grounded by the concept of the ligand conformational sampling while responding to the long-range Coulomb potential of the receptor when slowly approaching its surface. During this step, the Coulomb

potential significantly contributes to steering the entirely flexible and charged ligand toward its binding site and to let it finally adopt its binding pose in an explicit solvent environment. The subsequent free MD step allows for an unbiased mutual adjustment of receptor residues and ligand.

DMD is a local docking method focused toward a certain region on the receptor surface. If the confidence in the binding region assumption is to be increased, one could repeat the DMD study while varying the focus point for covering a larger search space. There is no general dependence of the DMD protocol on the confidence in the binding region assumption (however, the conclusiveness of the DMD results clearly depends on the validity of this assumption). When selecting the focus point, core atom, and target distance *D*, one needs to ensure that after the pulling process, the ligand is able to establish short-range interactions within the receptor binding region but is not forced into clashes with the receptor. In our study, we selected the focus point based on the experimentally determined ligand position (Supporting Information Table 1). In practice, when the structure of the bound ligand is unknown, the focus point has to be defined on the basis of the residues comprising the putative binding region. Via analysis of the atomic coordinates of 10 protein−GAG complexes (PDB IDs 1BFB, 1G5N, 1GMO, 1RID, 1T8U, 2BRS, 2HYU, 2JQR, 3C9E, and 1XMN, including GAGs comprised of two to seven monosaccharide units), we found that the distance as well as the direction from the center of mass of the receptor residues comprising the putative binding site to the ligand center is consistent within these complexes (Supporting Information Table 2). This relationship can be used to derive a focus point from the residues constituting the receptor binding region.

One of the main concepts of DMD is that pulling process and subsequent free MD are repeated many times in independent simulations. This allows for the creation and evaluation of an ensemble of docking solutions rather than the interpretation of single trajectories. In this ensemble, the energetically more favorable states are the more likely ones. Spatial clustering of this docking solution ensemble identifies those docking solutions that appeared with highest probability and, therefore, lowest energy. With respect to the identification of anchoring residues in the receptor and their energetical ranking, meaningful results can be obtained by merging the properties of multiple docking solutions.

Sampling performance is one of the main limitations of any docking method. Regarding our DMD protocol applied to the TDS complexes, we have shown that with 100 independent DMD runs the sampling performance was sufficient for generating meaningful docking solution ensembles. Although the predictive significance of DMD in this study has been good enough, it can still be largely improved by a sampling enhancement. An obvious way for achieving this would be to increase the number of independent DMD runs, which would result in an even clearer picture upon clustering of a docking solution ensemble. Furthermore, the current DMD protocol leaves room for optimizing the sampling performance per compute time via careful reduction of the ligand starting distance and an increase of the pulling velocity.

Beyond the proof of concept provided by this study for including receptor flexibility and explicit solvent in molecular dynamics-based docking of protein−GAG systems, the analysis of MD data collected in the course of DMD can be largely extended to further increase the atomic resolution of results, improve the overall predictive significance, and gain new insights into molecular recognition mechanisms. The behavior and role of single water molecules in ligand binding, for instance, could be investigated thoroughly. Furthermore, the dynamic nature of DMD data enables and motivates the creation of specialized scoring schemes, e.g. the evaluation of ligand position fluctuations during free MD as an indicator for the stability of a docking solution. Also, the single-residue energy decomposition analysis can be accompanied by an evaluation of hydrogen bond formation and occupancy via simple distance and angle criteria.

An important challenge in GAG−protein docking is to properly deal with the conformational flexibility of IdoA(2S) in heparin.[49,50] Unfortunately, GLYCAM 06 is known to not properly reproduce the natural ring conformer population of IdoA(2S),[33] and its ring conformation interconversion time scale is beyond simulation times accessible in DMD anyway.[32] Nevertheless, this issue can be addressed by applying ring-internal restraints to explicitly impose a certain conformation on individual rings (as outlined in the Materials and Methods section) and then perform a number of independent DMD studies differing only in the combination of ring conformations—an approach which has recently been proposed by Muñoz-García and co-workers.[51] As we have shown previously, the MM-PBSA approach is capable of distinguishing GAG poses differing only in the conformation of one single ring.[52]

Although the computational demands for DMD are higher than for conventional docking methods such as AD3, investigations of single systems via DMD are entirely feasible when having access to reasonable computing resources, especially using MD software optimized for GPU hardware such as Amber: a DMD study for the FGF2-HP complex as presented here can be performed within one week incorporating six modern GPU cards (GTX 580 in this case).

CONCLUSIONS

In this study, we have established and tested DMD, a targeted molecular dynamics-based protocol for local docking developed for specifically tackling the challenges imposed by highly flexible systems dominated by electrostatic interaction, such as protein−GAG systems.

In particular, flexible treatment of the receptor copes with the fact that GAGs usually bind to charged surface patches on the receptor comprised of highly flexible side chains. The identification of proper binding poses in these surface patches requires the flexible adjustments of protein residues to the GAG ligand. Furthermore, the inclusion of explicit solvent in DMD considers the solvent-accessibility of such binding sites and—most important—lives up to the supremacy of charge−charge interactions in most protein−GAG complexes. The dominating long-range Coulomb potential is explicitly exploited as a driving force in DMD while slowly approaching the ligand toward the receptor. A long free MD step and the flexible treatment of the ligand allow for an extensive sampling of the GAG-internal degrees of freedom.

We show that DMD has high predictive significance for systems dominated by electrostatic attraction. Our data implicate that via proper selection of DMD parameters such as pulling velocity, ligand displacement distance, and the number of independent repetitions, sufficient sampling is achievable. We demonstrate that in some cases the simple analysis of the spatial distribution of the electrostatic potential of the receptor can lead to a reliable prediction of the GAG binding region and therefore offers itself as a useful tool for

defining the target region on the receptor surface as required by DMD.

Regarding the spatial distribution of docking solutions, DMD generally yields results comparable to AD3. Nevertheless, DMD performs better in terms of achieving agreement with atomic details inferred from experimental data. The time-dependent data obtained via DMD allow for a reliable prediction of receptor anchoring residues via MM-PB(GB)SA single-residue energy decomposition, yielding consistent results when docking GAGs of different lengths. Furthermore, obtained MD trajectory data pave the way for the evaluation of various dynamics-based measures, the development of specialized scoring schemes, and the investigation of the dynamic nature of molecular interaction mechanisms within a particular system.

■ ASSOCIATED CONTENT

Supporting Information

Supporting Tables 1−4 and Supporting Figures 1−5. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mayte@biotec.tu-dresden.de.

Author Contributions

†These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the German Research Council (SFB-TRR67, project A7). J.-P.G. is funded by a fellowship of the Studienstiftung des deutschen Volkes. The authors thank the ZIH TU Dresden for providing high-performance computational resources, Ralf Gey for technical support, and Malte Lichtner for helpful discussions.

■ REFERENCES

(1) Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med.* **2000**, *78*, 269−281.

(2) Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S. Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J.* **2012**, *14*, 133−141.

(3) Moreira, I.; Fernandes, P.; Ramos, M. Protein-protein docking dealing with the unknown. *J. Comput. Chem.* **2010**, *31*, 317−342.

(4) Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. Principles of flexible protein-protein docking. *Proteins* **2008**, *73*, 271−289.

(5) Lensink, M.; Wodak, S. Docking and scoring protein interactions: CAPRI 2009. *Proteins* **2010**, *78*, 3073−3084.

(6) Kim, R.; Skolnick, J. Assessment of programs for ligand binding affinity prediction. *J. Comput. Chem.* **2008**, *29*, 1316−1331.

(7) Plewczynski, D.; Lazniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **2011**, *32*, 742−755.

(8) Smith, R.; Dunbar, J.; Ung, P.; Esposito, E.; Yang, C.-Y.; Wang, S.; Carlson, H. CSAR Benchmark Exercise of 2010: Combined Evaluation Across All Submitted Scoring Functions. *J. Chem. Inf. Model.* **2011**, *51*, 2115−2131.

(9) Ritchie, D. Recent progress and future directions in protein-protein docking. *Curr. Protein Pept. Sci.* **2008**, *9*, 1−15.

(10) Gunasekaran, K.; Nussinov, R. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J. Mol. Biol.* **2007**, *365*, 257−273.

(11) Gutteridge, A.; Thornton, J. Conformational changes observed in enzyme crystal structures upon substrate binding. *J. Mol. Biol.* **2005**, *346*, 21−28.

(12) van Dijk, A.; Bonvin, A. Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics* **2006**,

(13) Baron, R.; Setny, P.; Andrew McCammon, J. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.* **2010**, *132*, 12091−12097.

(14) Roberts, B.; Mancera, R. Ligand-Protein Docking with Water Molecules. *J. Chem. Inf. Model.* **2008**, *48*, 397−408.

(15) Thilagavathi, R.; Mancera, R. Ligand-protein cross-docking with water molecules. *J. Chem. Inf. Model.* **2010**, *50*, 415−421.

(16) Hynes, R. The Extracellular Matrix: Not Just Pretty Fibrils. *Science* **2009**, *326*, 1216−1219.

(17) MacRi, L.; Silverstein, D.; Clark, R. Growth factor binding to the pericellular matrix and its importance in tissue engineering. *Adv. Drug Delivery Rev.* **2007**, *59*, 1366−1381.

(18) Nieto, L.; Canales, A.; Fernández, I. S.; Santillana, E.; González-Corrochano, R.; Redondo-Horcajo, M.; Cañada, F. J.; Nieto, P.; Martín-Lomas, M.; Giménez-Gallego, G.; Jiménez-Barbero, J. Heparin Modulates the Mitogenic Activity of Fibroblast Growth Factor by Inducing Dimerization of its Receptor. A 3D View by Using NMR. *ChemBioChem.* **2013**, *14*, 1732−1744.

(19) Mulloy, B. The specificity of interactions between proteins and sulfated polysaccharides. *An. Acad. Bras. Cienc.* **2005**, *77*, 651−664.

(20) Samsonov, S.; Teyra, J.; Pisabarro, T. Docking glycosaminoglycans to proteins: analysis of solvent inclusion. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 477−489.

(21) Forster, M.; Mulloy, B. Computational approaches to the identification of heparin-binding sites on the surfaces of proteins. *Biochem. Soc. Trans.* **2006**, *34*, 431−434.

(22) Bitomsky, W.; Wade, R. Docking of Glycosaminoglycans to Heparin-Binding Proteins: Validation for aFGF, bFGF, and Antithrombin and Application to IL-8. *J. Am. Chem. Soc.* **1999**, *121*, 3004−3013.

(23) Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6679−6685.

(24) Chaudhuri, R.; Carrillo, O.; Laughton, C.; Orozco, M. Application of Drug-Perturbed Essential Dynamics/Molecular Dynamics (ED/MD) to Virtual Screening and Rational Drug Design. *J. Chem. Theory Comput.* **2012**, *8*, 2204−2214.

(25) Antes, I. DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins* **2010**, *78*, 1084−1104.

(26) Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. E. Free Energy Calculations with Non-Equilibrium Methods: Applications of the Jarzynski Relationship. *Theor. Chem. Acc.* **2006**, *116*, 338−346.

(27) Takaoka, T.; Mori, K.; Okimoto, N.; Neya, S.; Hoshino, T. Prediction of the Structure of Complexes Comprised of Proteins and Glycosaminoglycans Using Docking Simulation and Cluster Analysis. *J. Chem. Theory Comput.* **2007**, *3*, 2347−2356.

(28) Pichert, A.; Samsonov, S.; Theisgen, S.; Thomas, L.; Baumann, L.; Schiller, J.; Beck-Sickinger, A.; Huster, D.; Pisabarro, T. Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. *Glycobiology* **2012**, *22*, 134−145.

(29) Newman, J.; Dolezal, O.; Fazio, V.; Caradoc-Davies, T.; Peat, T. The DINGO dataset: a comprehensive set of data for the SAMPL challenge. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 497−503.

(30) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; J. Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, CA, 2010.

(31) Kirschner, K.; Yongye, A.; Tschampel, S.; González-Outeirino, J.; Daniels, C.; Foley, L.; Woods, R. GLYCAM06: A generalizable

biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622−655.
(32) Sattelle, B. M.; Hansen, S. U.; Gardiner, J.; Almond, A. Free Energy Landscapes of Iduronic Acid and Related Monosaccharides. *J. Am. Chem. Soc.* **2010**, *132*, 13132−13134.
(33) Gandhi, N. S.; Mancera, R. L. Can current force fields reproduce ring puckering in 2-O-sulfo-α-l-iduronic acid? A molecular dynamics simulation study. *Carbohydr. Res.* **2010**, *345*, 689−695.
(34) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314−3321.
(35) Gandhi, N. S.; Mancera, R. L. Free energy calculations of glycosaminoglycan-protein interactions. *Glycobiology* **2009**, *19*, 1103−1115.
(36) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol. Inf.* **2012**, *31*, 114−122.
(37) Morris, G.; Goodsell, D.; Halliday, R.; Huey, R.; Hart, W.; Belew, R.; Olson, A. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1999**, *19*, 1662−1639.
(38) Imberty, A.; Perez, S. In *New Developments in Therapeutic Glycomics*; Delehedde, M., Lortat-Jacob, H., Eds.; Research Signpost: Kerala, India, 2006; pp 185−201.
(39) Sage, J.; Mallévre, F.; Barbarin-Costes, F.; Samsonov, S. A.; Gehrcke, J.-P.; Pisabarro, M. T.; Perrier, E.; Schnebert, S.; Roget, A.; Livache, T.; Nizard, C.; Lalmanach, G.; Lecaille, F. Binding of Chondroitin 4-Sulfate to Cathepsin S Regulates Its Enzymatic Activity. *Biochemistry* **2013**, *52*, 6487−6498.
(40) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2010.
(41) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*; American Association for Artificial Intelligence: Menlo Park, CA, 1996.
(42) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(43) 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.
(44) Gandhi, N.; Mancera, R. The structure of glycosaminoglycans and their interactions with proteins. *Chem. Biol. Drug Des.* **2008**, *72*, 455−482.
(45) Faham, S.; Hileman, R.; Fromm, J.; Linhardt, R.; Rees, D. Heparin structure and interactions with basic fibroblast growth factor. *Science* **1996**, *271*, 1116−1120.
(46) Li, Z.; Kienetz, M.; Cherney, M. M.; James, M. N.; Brömme, D. The Crystal and Molecular Structures of a Cathepsin K:Chondroitin Sulfate Complex. *J. Mol. Biol.* **2008**, *383*, 78−91.
(47) Cherney, M. M.; Lecaille, F.; Kienitz, M.; Nallaseth, F. S.; Li, Z.; James, M. N. G.; Brömme, D. Structure-Activity Analysis of Cathepsin K/Chondroitin 4-Sulfate Interactions. *J. Biol. Chem.* **2011**, *286*, 8988−8998.
(48) Thompson, L. D.; Pantoliano, M. W.; Springer, B. A. Energetic Characterization of the Basic Fibroblast Growth Factor-Heparin Interaction: Identification of the Heparin Binding Domain. *Biochemistry* **1994**, *33*, 3831−3840.
(49) Mulloy, B.; Forster, M. J. Conformation and dynamics of heparin and heparan sulfate. *Glycobiology* **2000**, *10*, 1147−1156.
(50) Canales, A.; Angulo, J.; Ojeda, R.; Bruix, M.; Fayos, R.; Lozano, R.; Giménez-Gallego, G.; Martín-Lomas, M.; Nieto, P. M.; Jiménez-Barbero, J. Conformational Flexibility of a Synthetic Glycosylaminoglycan Bound to a Fibroblast Growth Factor. FGF-1 Recognizes Both the 1C4 and 2SO Conformations of a Bioactive Heparin-like Hexasaccharide. *J. Am. Chem. Soc.* **2005**, *127*, 5778−5779.
(51) Muñoz-García, J. C.; Corzana, F.; de Paz, J. L.; Angulo, J.; Nieto, P. M. Conformations of the iduronate ring in short heparin fragments described by time-averaged distance restrained molecular dynamics. *Glycobiology* **2013**, *23*, 1220−1229.
(52) Samsonov, S. A.; Pisabarro, M. T. Importance of IdoA and IdoA(2S) ring conformations in computational studies of glycosaminoglycan-protein interactions. *Carbohydr. Res.* **2013**, *381*, 133−137.