JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Docking Server for the Identification of Heparin Binding Sites on Proteins

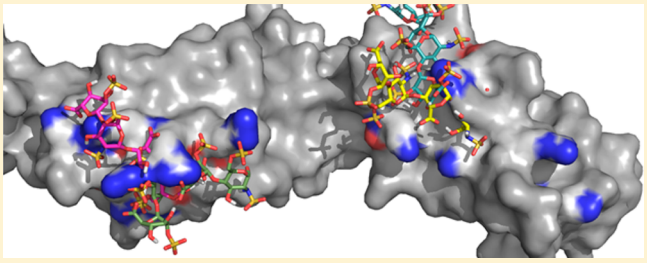
Scott E. Mottarella,[†] Dmitri Beglov,[‡] Natalia Beglova,[§] Matthew A. Nugent,[⊥] Dima Kozakov,*[,‡] and Sandor Vajda*[,‡]

[†]Program in Bioinformatics and [‡]Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, United States

[§]Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, CLS 941, Boston, Massachusetts 02215, United States

[⊥]Department of Biological Sciences, University of Massachusetts Lowell, Lowell, Massachusetts 01854, United States

ABSTRACT: Many proteins of widely differing functionality and structure are capable of binding heparin and heparan sulfate. Since crystallizing protein−heparin complexes for structure determination is generally difficult, computational docking can be a useful approach for understanding specific interactions. Previous studies used programs originally developed for docking small molecules to well-defined pockets, rather than for docking polysaccharides to highly charged shallow crevices that usually bind heparin. We have extended the program PIPER and the automated protein−protein docking server ClusPro to heparin docking. Using a molecular mechanics energy function for scoring and the fast Fourier transform correlation approach, the method generates and evaluates close to a billion poses of a heparin tetrasaccharide probe. The docked structures are clustered using pairwise root-mean-square deviations as the distance measure. It was shown that clustering of heparin molecules close to each other but having different orientations and selecting the clusters with the highest protein−ligand contacts reliably predicts the heparin binding site. In addition, the centers of the five most populated clusters include structures close to the native orientation of the heparin. These structures can provide starting points for further refinement by methods that account for flexibility such as molecular dynamics. The heparin docking method is available as an advanced option of the ClusPro server at http://cluspro.bu.edu/.



INTRODUCTION

Glycosaminoglycans (GAGs) are a diverse group of polysaccharides that participate in many biological processes through the regulation of their protein partners.[1−4] They are produced by almost every cell type and are most frequently found in the extracellular space and on cell surfaces where they play important roles in mediating cell−extracellular matrix interactions and cell−cell communication and in regulating extracellular matrix structure and function.[5−8] Heparan sulfate (HS) represents a structurally varied family of GAGs. HS is a polysaccharide of disaccharide units of alternating hexuronic acid and D-glucosamine. Variations in disaccharide sulfation and hexuronic acid structure are responsible for the differences between and even within HS chains.[1,3,9,10] Heparin, a particular member of the HS family, consists of highly sulfated disaccharides, and is frequently used as a model compound in experimental and theoretical studies of protein−HS interactions.[11] Heparin is very important on its own right, and has long been known for its capability as an anticoagulant, which has been used in humans for almost one hundred years.[8] The mechanism of heparin action involves catalyzing the inactivation of thrombin by antithrombin III by causing a conformational change in antithrombin III.[1,4] The ability of heparin to bind to proteins involved in regulating other cell processes such as cell proliferation suggests other medically relevant interactions.[3] Indeed, over the past two decades a growing number of biological activities have been discovered to be regulated by the interaction of heparin/HS with proteins that play major roles in cancer, wound healing, infectious diseases, and inflammatory processes. Thus, it is not surprising that a 2002 review on heparin-protein interactions has been cited over 1000 times.[8]

Given the importance of interactions between heparin/HS and a very large variety of proteins, considerable effort has been invested in the identification of protein regions that form such interactions. Structures have been determined by X-ray crystallography for a number of important protein−heparin/HS complexes, including complexes of antithrombin III, annexin V, and fibroblast growth factors.[2,12] However, the crystallization of protein−heparin/HS complexes is challenging, primarily due to inhomogeneity of GAG fragments and the nature of ionic interactions that may allow for multiple binding orientations.[13] Because of these problems, X-ray structures are

Received: February 22, 2014
Published: June 29, 2014

   dx.doi.org/10.1021/ci500115j | J. Chem. Inf. Model. 2014, 54, 2068−2078

available only for a small fraction of proteins that are known to interact with heparin/HS, and computational methods have been frequently used for predicting the structures of specific complexes. Early attempts were made to identify sequence determinants required for binding, and although some consensus sequences have been found, they are neither necessary nor sufficient.[14] A minimum requirement that appears to be common to all heparin binding sites is the concentration of basic residues in a particular area of the protein surface (not necessarily close to each other in sequence), oriented in a geometry that matches the pattern of sulfate groups along the heparin/HS chain.[11]

The combination of the charged residues and the requirements of a geometry that can accommodate an elongated heparin/HS molecule is expected to facilitate the identification of heparin binding sites and orientations using docking methods, and a large number of such calculations have been described in the literature. In most cases docking has been predictively applied to specific proteins without much validation, although in some cases the results were qualitatively confirmed by NMR titration or site-directed mutagenesis experiments.[11,15−22] The number of proper method validation studies, involving known protein−heparin/HS structures as a test set, appears to be relatively limited. In an early but still very influential study,[23] Bitomsky and Wade docked mono- and disaccharide probes using a number of programs, namely GRID,[24] AutoDock,[25] and DOCK.[26] The test set included three proteins: two structures taken from complexes of heparin with fibroblast growth factors, and the third the unbound structure of antithrombin III. Searches were performed both globally, i.e, considering the entire protein surface, and locally, restricting considerations to a box around the known heparin binding site. For each probe, results were given in terms of "interaction probabilities", defined as the normalized number of contacts between each residue and a set of low energy probe poses. While this measure is less informative than RMSD (root-mean-square deviation) from the native state used in most traditional docking tests, the relatively high interaction probabilities obtained for binding site residues indicate that the methods were able to correctly localize the heparin binding sites. However, it was also apparent that probes had comparably low energies in several different orientations. Rigid body docking of hexasaccharides was also performed. It was observed that in all docking runs, the crystal structure did not represent the most favorable conformation with respect to the force fields of the docking programs. These results were confirmed by more recent validation studies. Forster and Mulloy[11] used Autodock to dock an NMR structure of heparin to separately crystallized structures of the three proteins considered by Bitomsky and Wade.[23] While the methodology used by Forster and Mulloy[11] appears to be fairly rigorous, they did not present results beyond qualitative statements. More detailed results were given by Samsonov et al.,[27] who analyzed how solvent inclusion affects the results obtained by a number of docking programs. Although the main result of the paper was that inclusion of solvent improves results and that docking generally yields near-native structures, it was also reported that, on average, there were only 1.8 to 4.5 correct poses among the top 10 lowest energy structures. It was also shown that, in terms of the energy, the average rank of the best (lowest RMSD) pose was between 11 and 33, depending on the target protein.[27] Thus, there seems to be an agreement in the literature that docking generates near-native conformations, but selecting them among the ones generated is still a problem.

Essentially all methods used in the above studies have been developed for docking small molecules (with a few rotatable bonds) to traditional drug target proteins with well-defined binding pockets. In addition, the methods generally assume that the approximate location of the binding pocket is known, and hence restrict the search to a box around the putative binding site. In contrast, heparin/HS is substantially larger than the small ligands considered as potential drugs, and it also has more rotatable bonds. In addition, heparin/HS generally binds to shallow solvent exposed crevices or even slightly protruding regions on the protein surface rather than in deep pockets. Although the binding region must be highly charged, a number of sites on a protein may satisfy this condition, each allowing for several orientations of the ligand. Thus, it is not clear whether the small molecule docking methods offer the best tools for heparin/HS docking, and whether these methods can predict position and orientations with an accuracy that is useful in applications.

We have encountered the problem of heparin docking problem as a participant in the CAPRI (Critical Assessment of PRotein Interactions) worldwide protein docking experiment. While the CAPRI targets are generally protein−protein complexes, Target 57 in Round 27 of the experiment required docking a six-sugar heparin to the unbound structure of fragment 423−700 of the protein BT4661, a polysaccharide binding protein from the *Bacteroides thetaiotaomicron* heparin utilization locus. The predictor groups in CAPRI were given the atomic coordinates of the protein and a generic heparin conformation, and were expected to model the complex as accurately as possible. The BT4661-heparin complex is now deposited in the Protein Data Bank (PDB ID 4AK2), but at the time of the CAPRI assignment no information was available on the location of heparin binding, and the ligand-free structure of the protein is still unpublished. Since the 423−700 fragment of the protein BT4661 has two domains with a deep crevice between the domains, and many positive charges on the surface, locating the most likely binding pose with any confidence was far from trivial. We tried several approaches described in previous papers on heparin docking. First we mapped the protein using FTMap, a program developed for the identification and characterization of ligand binding sites,[28] which can be considered as a substantially improved version of the GRID program.[29] FTMap found several pockets that could favorably accommodate small ligands, but were not very likely to bind an elongated heparin molecule. Next, as in a number of previous studies, we docked mono and disaccharide fragments of heparin using Autodock, but the program returned too many potential binding sites, leaving us with a high level of uncertainty. This results has shown that using existing program is far from simple when there is no information on the site, as in the case of CAPRI targets.

After experimenting with FTMap and Autodock, we decided to use the docking program PIPER,[30] also implemented in our heavily used automated docking server ClusPro.[31,32] ClusPro was the first automated protein−protein server, and based on the results of the last three CAPRI evaluation meetings, it consistently has been the most accurate among such servers.[33−35] The server is heavily used: by December 2013 we registered over 8600 unique user IPs, and the server completed almost 57 000 docking jobs, currently about 2000 per month. Models built by ClusPro have been reported in over

250 publications. Since PIPER and ClusPro were developed for protein−protein docking, neither had the parameters needed for heparin docking, and thus had to be appropriately expanded. This approach was fairly successful, at least when compared to the results of the other 30 groups that participated in CAPRI and submitted predictions for Target 57. Based on the evaluator's report,[35] we obtained two medium quality models, with ligand RMSD (after superimposing the receptors) below 5 Å and interface RMSD (for residues within 10 Å of the interface) around 2 Å. There were two other groups that achieved comparable accuracy, but our model predicted the highest fraction of native contacts.[35]

On the basis of our successful prediction at CAPRI we decided to establish a general method and to add heparin docking as an advanced option to the ClusPro server. The goal of this paper is to describe this new development and the resulting capabilities. As shown by the CAPRI results, ClusPro is competitive with other approaches. However, the method also has clear limitations, and can be used only as the first step toward determining heparin−protein interactions. In fact, ClusPro performs rigid body docking of a generic tetrasaccharide heparin probe, allowing for some flexibility in a refinement step. In this step we perform off-grid energy minimization during which all atoms of the heparin chain and all protein side chain atoms within 5 Å of the heparin molecule are allowed to move freely. Although we explored the possibility of pregenerating and docking a number of conformers of the heparin tetramer, we decided to perform only local minimization for three reasons. First, the backbone in short heparin fragments is relatively conserved. Second, although the orientation of the sulfate groups is highly variable, due to the strong ionic interactions these are heavily affected by the positions of charges on the protein, and hence precalculating conformations generally does not improve the accuracy of predicting the binding pose. Third, generating and docking possible conformers would increase our algorithm's runtime to unreasonable lengths, making the calculations too expensive for a web-based server. Therefore, our main objective is to predict the location of the heparin binding site, defined in terms of amino acid residues that are in contact with the ligand, and possibly some of the most likely orientations of the heparin chain, and we do not attempt to predict detailed interactions at the atom−atom level. As will be shown, the binding sites identified by the server are reliable. In addition, some of the top poses are generally not very far from the native orientation of the heparin. Thus, we hope that adding heparin docking to ClusPro will be useful to the research community as the first step toward constructing more accurate interaction models. The models generated by ClusPro can be confirmed or rejected using data from low resolution experiments, and can be refined by molecular dynamics (MD) or Monte Carlo simulations that use the specific heparin/HS molecule rather than the generic tetrasaccharide, allow for flexibility, and may explicitly account for the effects of the solvent.

METHODS

ClusPro Server. Heparin docking was added as an advanced option to our protein−protein docking server ClusPro. The server has been in operation since 2004,[31] and it was substantially upgraded to Version 2.0 in 2007.[36] The first step of the current version is a global rigid body docking using PIPER, which as based on the fast Fourier transform (FFT) correlation approach.[30] In PIPER the smaller molecule (i.e., in our current application the heparin), is considered the ligand, which is moved about the protein receptor on a 3D grid with spacing 1.0 Å, using 70 000 rotations at each grid point. For a protein of average size this means close to a billion function evaluations. The advantage of the FFT approach is that a scoring function, written as a sum of correlation functions, can be evaluated extremely efficiently, and thus sampling the very large number of conformations is computationally feasible.[30] For heparin docking, the interaction energy type scoring function is calculated using three terms: van der Waals attractive, van der Waals repulsive, and electrostatic energy. Weights for each of these terms were selected to maximize accuracy in a test set of heparin binding proteins (see Results). As will be described, in the heparin docking mode we perform sampling using two different weight sets, and from each sample we retain the 900 lowest energy structures.

The second step performed by ClusPro is clustering the retained structures using pairwise RMSD as the distance measure.[37] The biophysical meaning of clustering is isolating highly populated low-energy basins of the energy landscape.[38] Several studies, including ours,[31] have demonstrated that clustering algorithms generally perform better for isolating near native structures as compared with selecting low-energy conformations, if used in conjunction with exhaustive energy-based sampling such as in PIPER. The clustering of the 1800 poses starts with the lowest energy pose and grouping all poses within 9 Å. From the remaining poses not already grouped, the lowest energy pose is selected and the process repeats until no poses remain. Clusters of 10 or fewer poses are not considered. The clusters are then ranked by population and represented by the lowest energy structure. Unless requested otherwise by the user, ClusPro returns the representatives of the 10 largest clusters along with cluster populations. Finally, the structures are refined by minimizing the Charmm energy of the complexes. All heparin atoms and all side chain atoms within 5 Å of the heparin are allowed off-grid flexibility during minimization, then the protein atoms are reset and a final minimization resolves collisions. While the minimization generally removes potential steric clashes, it does not substantially change the conformation of the complexes, and thus the RMSD of our ClusPro submissions from the native complexes is fully determined by the rigid-body docking and clustering steps.[38]

Test Set Selection. To select a test set of heparin-binding proteins, the RCSB Protein Data Bank (www.pdb.org)[39] was searched for structures containing both 2-*O*-sulfo-alpha-L-idopyranuronic acid and *N*,*O*6-disulfo-glucosamine; the two most common sugars making up heparin chains. Several structures contained a single 4-deoxy-2-*O*-sulfo-alpha-L-threo-hex-4-enopyranuronic acid in place of 2-*O*-sulfo-alpha-L-idopyranuronic acid when that sugar was the terminal sugar in the heparin chain. From these structures, we selected only proteins for which an unbound form with sequence similarity of 95% or better was also available. Additional cases were removed from consideration for containing essential ions in the heparin binding site, for structures whose heparin coordinates contained errors that proved insurmountable in our preparation steps, and for multiple conformations of the same protein target. These selection criteria resulted in a small test set of five proteins listed in Table 1.

Heparin Parametrization. The heparin molecule was extracted from each bound structure and parametrized for use in the ClusPro program. The molecules were given unique

Table 1. List of Heparin Bound Protein Structures and their Unbound Forms[a]

protein	unbound PDB	bound PDB	heparin length
human 3-*O*-Sulfotransferase-3	1T8T[40]	1T8U[40]	2
E2 domain of amyloid precursor-like protein 1	3Q7L[41]	3QMK[42]	2
NK1 fragment of human hepatocyte growth factor/scatter factor (HGF/SF)	1NK1[43]	1GMN[44]	2.5
plasma serine protease inhibitor	1LQ8[45]	3DY0[46]	2.5
basic fibroblast growth factor	1BFG[47]	1BFC[48]	3

[a]Length of heparin chain present in bound structures given in terms of disaccharides.

atom naming and each atom was assigned a value for partial charge, radius, mass, and hydrogen bond acceptor/donor status. Each atom was also assigned a mol2 atom type and bond lengths, bond angles, and bond torsions were calculated for each unique bond. The quantum chemistry system GAMESS[49] was used to compute Austin Model 1 (AM1) atomic charges; the information was then piped through molecule manipulation scripts (http://charles.karney.info/b2d-scripts/) using the molecular mechanics suite ANTECHAMBER[50] to perform bond charge corrections (BCCs) and to generate the final AM1-BCC atomic charges. ANTECHAMBER generates GAFF-based[51] topology files and parameter files in the Charmm format.[52] GAFF is a complete force field, i.e., parameters are either available for all atom types or can be computed using empirical rules.[51] The molecular parameters were then copied to the ClusPro server to allow for each heparin molecule to be used as a ligand in interaction predictions.

Selecting a Quality Measure. Since we will cross-dock heparin/HS ligands that may have different chemical structures, we need a quality measure that allows for superimposing such molecules. We will therefore use RMSD defined in terms of a subset of atoms from each heparin chain. Each subset contains a single carbon atom from each sugar residue. This carbon atom is a member of the six-member ring in each sugar bonded to the oxygen in the ring as well as a second oxygen not in the ring. It is typically named C_1 in all heparin sugars and is the only carbon in the ring bound to two unique oxygen atoms. In addition, this atom is along the heparin backbone if we consider the backbone to be a string that passes through the connectors between sugar residues and around which the residues may rotate, and hence the RMSD defined in terms of these atoms will describe the overall position and orientation of the docked heparin molecule relative to the native ligand. We note that the RMSD measure involving a single carbon atom from each heparin subunit is similar to a backbone RMSD frequently used for describing the accuracy of results in protein−protein docking, and allows for heparin chains of varying length to be compared using a method similar to a sliding window. It is also more forgiving to differences in heparin conformation while remaining a valid indicator of heparin position and orientation.

Binding Site Prediction. Using the representative poses from ClusPro, an additional round of clustering was added to form binding sites. In this step heparin poses are clustered by proximity as before, but this time each pose is represented as a single point at its center of mass rather than the positions of all C_1 carbon atoms. Clusters that fail to meet a minimum population threshold are not considered. Another difference is that the clusters are ranked by the highest number of protein−heparin contacts per contacting protein atoms rather than by cluster population. Thus, the cluster making most contacts with the protein at any single atom is ranked highest.

RESULTS

Selecting Scoring Function Coefficients. In a standard run of the ClusPro server, a ligand (usually the smaller molecule) is systematically positioned about a receptor and the receptor−ligand interaction energy is calculated by several parameters in our rigid docking program PIPER. The best poses are then clustered to select the most likely positions and orientations of the ligand to the receptor. This functional approach was originally optimized for protein−protein docking, including minor adjustments for specific interactions such as for enzyme−inhibitor or antibody−antigen pairs. Our objective here is to optimize ClusPro for predicting protein−heparin interactions, involving the selection of scoring function coefficients and the selection of an appropriate heparin probe.

The molecular mechanics scoring function of PIPER we use for heparin docking has the general form $E = w_{attr}E_{attr} + w_{rep}E_{rep} + w_{elec}E_{elec}$, where E_{attr} and E_{rep} denote the attractive and repulsive contributions to the van der Waals interaction energy and E_{elec} is an electrostatic energy term.[30] The coefficients w_{attr}, w_{rep}, and w_{elec} specify the weights of the corresponding terms, and are optimally selected for different types of docking problems.[38] In order to select appropriate weight coefficients for heparin docking we considered the test set of heparin-bound protein structures listed in Table 1. The heparin structure was extracted from each complex, parametrized, loaded into ClusPro, and docked to the unbound (separately crystallized) protein structure. We attempted, but were unable, to discover a single set of weights capable of successfully finding the binding site on all test systems, so an approach that takes two weight sets and concatenates them prior to clustering was developed. In both sets $w_{rep} = 0.40$ and $w_{elec} = 300$, and only the weights of the attractive van der Waals contributions differ: $w_{attr} = -0.40$ in one set and $w_{attr} = -0.10$ in the other, the latter generating complexes with weaker shape complementarity. Since the FFT method evaluates the different energy terms separately, generating results using two different sets of weights does not increase the computing time. For each set, we retain the 900 lowest energy docked structures, and the resulting 1800 structures are clustered as described previously.

Selection of a Heparin Probe. In applying CluPro to heparin docking, the next step was to find a heparin molecule capable of representing all heparin chains in any protein−heparin interaction. Each of the heparin ligands was then docked to all five unbound protein structures in the test set. The results of this all-against-all comparison of the structures show that just one heparin molecule, from PDB structure 3QMK, is capable of predicting the bound pose of the five test cases within 10 Å. The results of docking this heparin structure against the five targets in the test set are shown in Table 2.

Docking Results for the Test Set. As shown in Table 2, docking of the selected heparin probe, the binding site for heparin can be predicted by the consensus of calculated heparin poses. In addition, ClusPro is capable of returning a high ranking, low RMSD result for heparin binding. These best poses are shown in the first column of Figure 1 for each test system. In these figures, the unbound protein is shown with the actual heparin pose (obtained by aligning the bound protein to the unbound protein) in green sticks, while the best predicted pose is shown in thin, cyan sticks. For each system, the

dx.doi.org/10.1021/ci500115j | J. Chem. Inf. Model. 2014, 54, 2068−2078

Table 2. List of Bound and Unbound Protein Conformations Used for Docking Prediction as well as Rank and RMSD of Best Prediction and the Rank of the Binding Site

bound PDB ID	unbound PDB ID	chain	best probe rank	RMSD (Å)	binding site rank
1T8U	1T8T	B	5th	7.995	1st
3QMK	3Q7L	A	5th	5.961	1st
1GMN	1NK1	A	3rd	8.959	1st
3DY0	1LQ8	A	1st	7.510	1st
1BFC	1BFG	A	2nd	3.499	1st

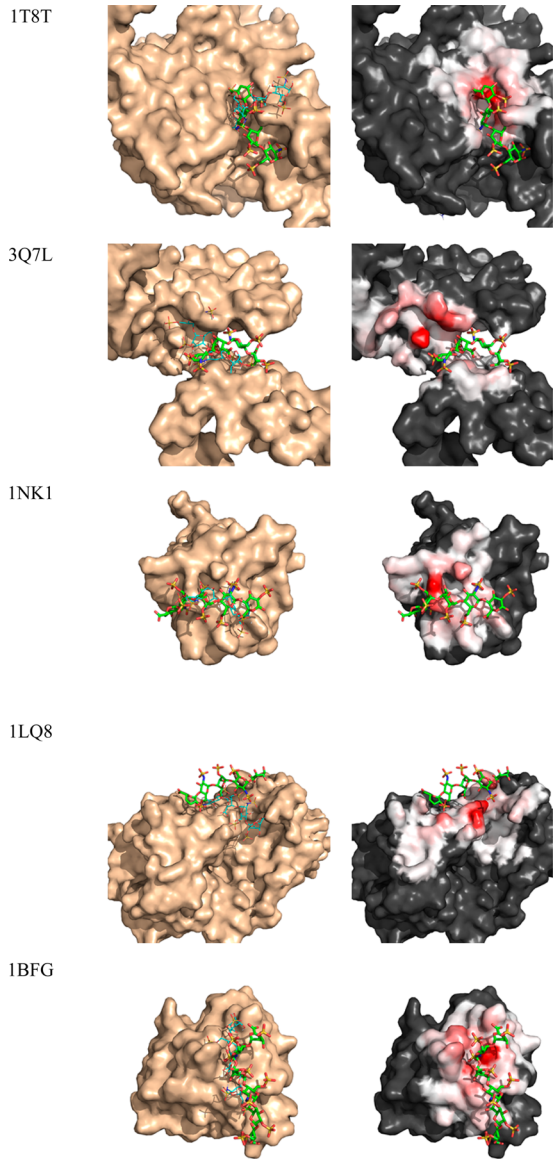


Figure 1. Best results and contact maps of the five test systems. (left) Unbound protein shown with the actual heparin pose (in green sticks). The best predicted pose is shown in thin, cyan sticks. (right) Contact maps. Red (hot) areas have large number of contacts with the docked heparin poses, white areas have fewer, and dark areas have none.

predicted pose forms more atom−atom interactions than the actual heparin pose. This is a consequence of the ClusPro algorithm's energy calculation, which results in several predictions in the correct relative position but closer to the target protein. However, in the case of 1LQ8 the actual heparin pose bridges a gap between two projecting amino acid residues (Arg26 and Asn230), while the predicted pose runs perpendicular down the valley between those residues, entirely beneath the actual heparin bridge.

As shown in Table 2, the five largest clusters generally include heparin poses that have the right orientation (with less than 10 Å RMSD from the native), but not necessarily the best rank. An atom−atom contact map is provided to show useful features even in such cases if no a priori information is available. For this map, each interaction of 4 Å or less from all returned representative poses is counted, and each protein atom is colored by its number of contacts. The second column in Figure 1 shows the contact map for the five proteins in the test set. Clearly shown along with the correct pose (obtained by aligning the bound protein to the unbound protein) are the contact maps. In each case, the hottest (reddest) area of the contact map correctly predicts the location of the bound heparin. Whiter areas have fewer contacts while dark areas have none. Only in the case of 1NK1 was there a secondary contact site (not shown), though it was clearly less important than the primary contact site shown here.

Analysis of Predicted Contacts. To observe the accuracy of predicting heparin binding sites on unbound protein structures, the contacts per protein residue were charted. These charts are given below in Figure 2.

For these charts, the number of contacts made (atom−atom distance $\leq$ 4.0 Å; y-axis) were counted per residue of the protein (x-axis). Counts of the contacts in the actual bound case (orange) were normalized to the same scale as the predicted unbound case (blue). These graphs show a general matching trend that any contacting residues are correctly predicted to have atom−atom contacts with the heparin chain. They also show many cases where a residue is predicted to form contacts with the heparin, though no true contacts are formed. In fact, in some cases ClusPro, which tends to optimize contacts, generates models that have better shape complementarity than the real protein−heparin complex. This problem will be discussed further in the paper.

Applications to Further Proteins. The extended ClusPro server was further tested against additional proteins that did not meet our original criteria for inclusion into the test set. In one such system, Annexin V, calcium ions present in the binding site create a favorable pocket for binding that would not otherwise exist.[53] These ions were also found in a PDB entry of the unbound protein.[54] This unbound conformation with the included ions was run with the new heparin docking approach. The best result, shown in Figure 3, had an RMSD of 6.66 Å and was ranked second.

We also docked heparin to segment FN12−FN14 of human fibronectin for which the structure of the bound conformation was not available, but one heparin binding site and one potential heparin binding site were described in the literature.[55] Results for this fibronectin model are shown in Figure 4 (PDB ID 1FNH). To run this protein model, the structure was divided into separate domains FN13 and FN14 as described.[55] The FN13 domain contains the heparin binding site confirmed by mutation studies, and consisting of residues R98, R99, R101, R115, and K117 (residue numbering is based on the X-ray structure of the FN12−FN14 fragment).[55] R146 may also be part of the binding site, but it is slightly removed from the other residues, in line with the mutagenesis data where change of R146 to S reduces heparin binding only 3-fold, whereas mutation of any of the other five residues reduces it at least 10-fold.[55] For this domain, the correct binding site was the only

dx.doi.org/10.1021/ci500115j | J. Chem. Inf. Model. 2014, 54, 2068−2078

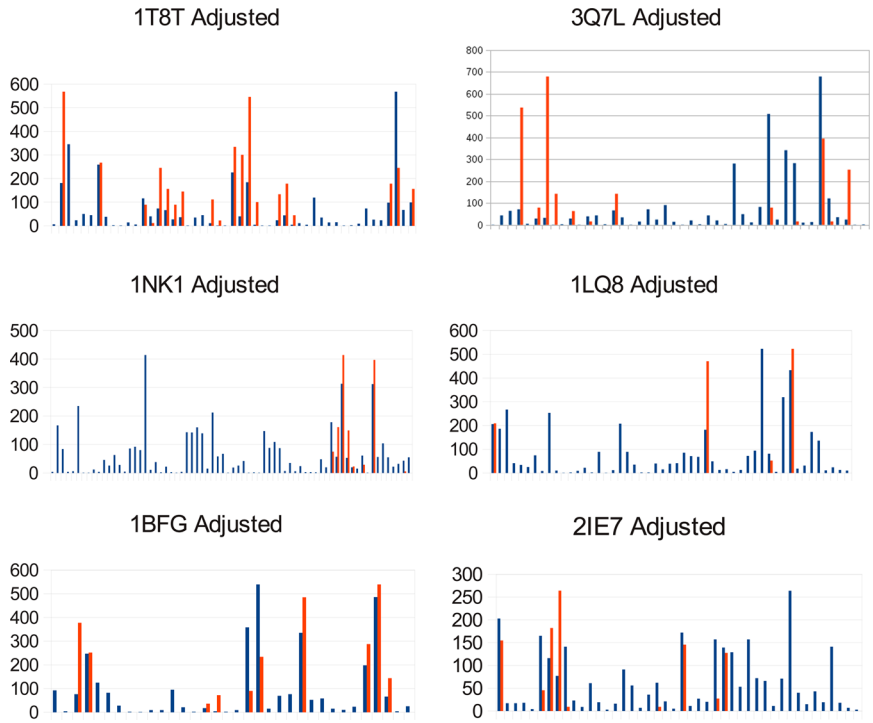


Figure 2. Atom−atom contacts between heparin chains and protein residues. Predicted heparin chains are shown in blue; actual chains (normalized) are in orange. The *x*-axis is protein residue sorted by residue number where only residues that had at least one atom−atom contact are included for clarity.

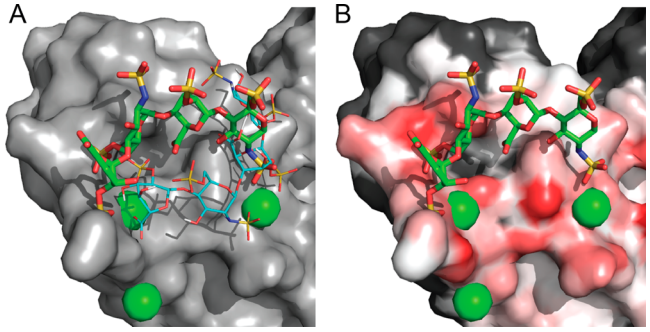


Figure 3. Results for Annexin V (PDB IDs 1G5N for the bound structure, 2IE7 for the unbound). Calcium ions are shown as green spheres. (A) Best predicted structure shown as thin cyan sticks. The native binding mode is shown in green. Notice that as in many other cases, the predicted structure is closer to the surface than the native one. (B) Heat map based on heparin docking. Red (hot) areas have large number of contacts with the docked heparin poses, white areas have fewer, and dark areas have none.

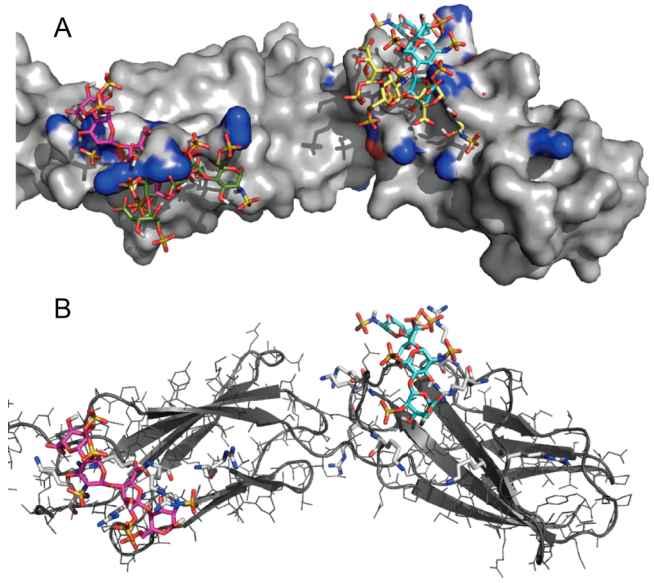


Figure 4. Results for human fibronectin (Unbound PDB 1FNH). Subunit FN13 is on the left and FN14 is on the right. (A) Predicted heparin binding residues are shown in blue. For FN13 we show the predicted heparin poses ranked 1 (magenta) and 8 (green). For FN14 we show the poses ranked 1 (cyan) and 2 (yellow). (B) Same as A with the protein shown as a cartoon. The R and K residues of the heparin binding site are shown as sticks. Only the top ranked heparin poses are shown.

ClusPro predicted binding site and it accurately centers on the implicated bound residues. The top three clusters predict an orientation very similar to the cluster ranked 1 (shown in magenta in Figure 4), and they all interact with the side chains R98 and R99, considered key to heparin binding. The FN14 domain contains a predicted binding site with several positively charged residues.[55] These residues are more scattered than the heparin binding residues on FN13. Indeed, peptides from FN14 spanning residues 204−210, 217−235, and 257−271 have been implicated independently, and thus there is a level of uncertainty concerning the binding site on FN14. Accordingly, ClusPro predicts more diverse heparin poses than for FN13, and based on cluster sizes the binding to FN14 is substantially weaker. In fact, when docking to the entire FN12−FN14 segment, the top 7 clusters are at the heparin binding site of FN13, and only cluster 8 is located on FN14. Mapping of the latter segment separately, the top three poses are found at the site represented by poses 1 and 2 in Figure 4A. Poses 4 and 5 (not shown) are shifted as the continuation of pose 1 (cyan)

dx.doi.org/10.1021/ci500115j | J. Chem. Inf. Model. 2014, 54, 2068−2078

toward the intersegment region on the back of the protein. Based on these five clusters, we predict that the residues most likely involved in heparin binding are K216, R230, R232, and K261. While R225 (shown on the right from the other binding residues) seems to be separated, its long side chain can also interact with the bound heparin. In addition, R207, R209, and K257 are oriented toward the intersegment region, and most likely can interact with longer heparin chains. Based on the repeated appearance of poses similar to the top ranked ones (magenta and cyan in Figure 4) among the docked structures, we consider it likely that such orientations occur.

It is generally assumed that a heparin chain of 12−16 saccharide units bridges the two FN13 and FN14 segments.[55] We show the poses ranked 8 for FN13 (green) and ranked 2 for FN14 (yellow) in Figure 4A, because their orientation extends the top ranked poses toward the other FN segment. Indeed, the two poses have the right directionality and distance (19.2 Å) to be connected by adding a tetrasaccharide unit bridging the intersegment region, without any major interaction with the protein. The resulting model is more curved than a recent model obtained by rigid docking of a standard heparin dodecasaccharide to the FN12−FN14 structure, where some interactions also occur with the FN12 segment.[21] In contrast, predicting only the tetrasaccharides binding to FN13 and FN14, our model does not reach FN12. Most of our top poses on FN13 are placed between the side chains of R98, R99, R101, and R115. According to an NMR study, only these residues are required for the binding of an octasaccharide to the FN13−FN14 units.[55] Of course, the predicted importance of the four residues does not prove that the model is correct, and further experimental studies, particularly exploring the heparin binding residues on FN14, are required for a better understanding of the binding mode.

Comparison to Other Computational Approaches. The method presented in this paper has been developed from our protein−protein docking program ClusPro specifically for heparin docking. Thus, for comparisons we restrict considerations to published heparin docking studies, as we have no intention to apply this option to any other problem. Although predictive heparin docking is performed in a substantial number of papers, very few studies evaluate the algorithm on test problems involving known protein−heparin structures. As mentioned in the introduction, systematic evaluations can be found in the papers by Bitomsky and Wade[23] and Samsonov et al.[27] However, comparing their results to ours is far from straightforward, since in each test problem both studies considered the known structure of a protein−heparin complex, separated the two molecules, and docked them back. Although some of the methods allowed for heparin flexibility, the protein was always kept fixed in its heparin-bound conformation. Furthermore, Samsonov et al.[27] restricted the search to a very tight box centered on the true bound position of the ligand. In contrast, our method has been developed to solve realistic heparin docking problems, and thus in each test case we consider the unbound (i.e., separately crystallized) protein and the generic tetrasaccharide built into ClusPro, and perform a global search over the entire protein surface without any assumption on the location of the binding site. We were able to find unbound protein structures for most complexes considered by Bitomsky and Wade[23] and Samsonov et al.,[27] and used these structures in our calculations. Based on the known complex structures we were able to determine that our method correctly identified the binding site in all cases. However, since we solved a very different problem using much less information on the targets, assessing the accuracy of docking in terms of RMSD from the native complex structures does not provide a very informative comparison to earlier results.

Bitomsky and Wade studied two test cases using known heparin-bound structures, namely the complexes of fibroblast growth factor proteins bFGF (PDB ID 1BFC[48]) and aFGF (PDB ID 2AXM[56]). In both cases the bound protein structures served as the receptors. The heparins were first docked as rigid body chains, then a single disaccharide was used as a probe, and lastly the flexible sugar chains were docked to the proteins. The initial rigid body approach was a simple test of the algorithms. The disaccharide probe was used to identify the heparin biding site. Finally, the flexible docking performed a complete test of these algorithms. Docking our generic tetrasaccharide to the separately crystallized fibroblast growth factor structures, we were able to identify the correct binding sites in both cases. For bFGF (unbound PDB ID 1BFG),[47] our best pose was ranked first with an RMSD of 3.51 Å using the previously defined backbone RMSD measure. This result was similar to the one obtained by Bitomsky and Wade,[23] who docked the bound heparin structure extracted from the complex to the heparin-bound structure of bFGF, and reported 4.1 Å as the best RMSD from the native complex, both using AutoDock and DOCK. For aFGF (unbound PDB ID 3K1X[57]), our best pose was ranked third with an RMSD of 6.56 Å. Again, Bitomsky and Wade[23] used the original bound heparin conformation as the ligand, and docking to the heparin-bound structure of the protein they were able to achieve their best result at 4.1 Å with AutoDock and 14.2 Å with DOCK. In view of this broad range, our result seems to be completely acceptable, particularly considering that we globally dock to the unbound protein structure.

Samsonov et al.[27] examined ten proteins that were cocrystallized with various glycoaminoglycans (GAGs) rather than heparin, and hence were not selected for our original test set. Of these ten systems, nine were found to have unbound structures available (see Table 3). The authors docked the minimized GAG structures to the bound protein structures using a box around the bound ligand position, thus performing

Table 3. List of Bound and Unbound Protein Conformations Used for Docking Prediction by Samsonov et al. as well as Rank and RMSD of Best Prediction and the RMSD of the Top Ranked Pose

bound PDB ID	unbound PDB ID	best probe rank	best probe RMSD (Å)	top rank RMSD (Å)	binding site rank
1OJN[58]	1N7O[58]	5th	4.35	6.84	1st
1RWH[59]	1RWA[59]	2nd	4.32	9.71	1st
1G5N[60]	2IE7[54]	2nd	6.66	8.60	1st
1T8U[40]	1T8T[60]	5th	8.00	8.36	1st
3E7J[61]	2FUQ[62]	7th	2.55	12.20	1st
2HYU[63]	2HYW[63]	9th	8.05	17.23	1st
2BRS[64]	1H8U[65]	12th	4.21	13.38	1st
1BFB[48]	1BFG[47]	2nd	3.50	8.54	1st
3IN9[66]	3IMN[66]	4th	2.50	6.07	1st[a]
		2nd	5.13	8.68	1st[a]

[a]System 3IN9/3IMN contained two GAG binding sites, both are shown. The binding site for 3IN9 spanned the two bound disaccharides, resulting in both bound poses in the top ranked binding site.

local rather than global docking. In contrast, we performed global docking of the generic tetrasaccharide to unbound proteins. As shown in Table 3, ClusPro ranked first the correct binding site in all cases, and for most proteins it also generated poses with fairly low RMSD from the bound structure. However, in some cases the best-ranked orientations had substantially higher RMSD values. As we will further discuss, this result emphasizes that lower ranked clusters generated by ClusPro should also be considered for refinement.

DISCUSSION

Heparin and other glycosaminoglycans are garnering attention as important regulators of function of many proteins, as well as potential drug molecules. Heparin has been used as an anticoagulant for almost a century and its mechanism of action suggests that many other uses may be possible.[3,4] Since crystallizing protein−heparin complexes for structure determination is generally difficult, computational docking can be a useful approach for understanding specific interactions. The size of heparin chains has provided a challenge for methods originally developed for docking small molecules such as AutoDock or DOCK, as well as methods that specifically target the prediction of protein−protein interactions such as our PIPER program and the ClusPro server. Our representative heparin probe is much larger than the small molecules handled by AutoDock, but significantly smaller than the typical protein uploaded to ClusPro. This required new parameters adjusted for molecules of this size. With appropriate updates we have extended the capabilities of the ClusPro server to include heparin molecules. To our knowledge, there does not already exist an interaction prediction software specific to heparin chain binding site prediction.

A single heparin tetrasaccharide was selected to serve as the default representative for heparin docking using ClusPro, and the approach has proven to be successful against all test targets in predicting the heparin binding location. This observation is in good agreement with previous reports. As noted by Mulloy and Linhradt,[12] a relatively short heparin is a well-behaved ligand for rigid docking and the site on a protein surface where heparin binds can be identified with good reliability. Indeed, the conformation of heparin appears to be unusually well-defined compared with other polysaccharides: in both solution studies of unbound heparin and crystal structures of complexes, heparin is a ribbon-like molecule with sulfo groups arranged in clusters along opposite sides of the polysaccharide chain.[12] Forster and Mulloy used several heparin structures for docking, but noted that in no case has the difference between the two conformations given rise to substantially different predictions.[11] The number of rotatable bonds, 22 even in this short chain, could generate over one billion conformations. These two facts precluded using multiple conformations of the heparin chain. However, it was also noted that the exact details of the interactions with proteins cannot be accurately predicted by docking calculations, whether or not flexibility is taken into account.[12] Accordingly, our objective is the reliable identification of the heparin binding site and possibly determining the most likely orientations of the tetrasaccharide. Using ClusPro we generate and evaluate a very large number of docked structures and determine a consensus binding area with good reliability. As expected, the method does not provide a simulation of the interaction or give any details of interactions between specific atoms in either structure. However, we expect that the results provide information for further exploration of the interactions between the protein and the specific heparin/HS molecule using more detailed simulation tools, primarily molecular dynamics. In fact, without the initial docking and clustering for the identification of the binding site it may be premature to focus on details.

Of the heparin chains examined, the selected tetrasaccharide probe was the shortest. With the longer chains, it was typical to see the tails peel away from the protein surface and move out into space toward the position of the neighboring protein in the crystal lattice. This caused heparin conformations with only half of the chain bound, followed by an unusual bend away from the protein surface. This was not seen in the shorter heparin molecules that were bound to the protein surface for the entire length of the chain. Therefore, we selected a heparin tetrasaccharide probe without the bend for predicting the binding site on proteins. In addition, this chain length is representative of a majority of available heparin bound structures in the Protein Data Bank, as longer structures are not free of the artifacts just described. In view of this limitation on the available structures, we do not generate three-dimensional models of chains longer than four saccharide units, as we have no basis for comparison.

In spite of correctly predicting the binding sites, we encountered difficulties that required some attention. First, several target proteins contained multiple chains and ClusPro favored the crevices found in the region between chains. Our solution was to run only a single chain at a time, though masking the region between chains (an advanced ClusPro option) would also have worked. Second, several structures contained near identical chains, and ClusPro found the same site on an identical chain, even if there was no heparin molecule present there in the bound protein structure. This caused very large RMSD values from the native complex. Lastly, on the HGF target protein (PDB ID 1NK1), a second location attracted many heparin poses and caused a predicted secondary binding site. This secondary site, however, occurs between two distinct protein domains and looked very similar to the inaccurate results found between multiple protein chains. It would be easy to exclude this fault by masking such a site or by cutting the chains into distinct domains.

More generally, we already mentioned that ClusPro frequently predicts heparin binding closer to the surface than the one observed in X-ray structures. This is particularly troubling for longer heparin chains because in many cases it is assumed that some part of the chain does not directly interact with the protein, but it is simply held in place by heparin segments that do. The assumed bridging of the fibronectin segments FN13 and FN14 is an example. Indeed, it is recognized that an important function of heparin/HS is to mediate and regulate protein−protein interactions, and this generally involves much longer heparin/HS molecules than the tetrasaccharide we routinely dock. Thus, it appears that it would be reasonable to modify the scoring function by further reducing the van der Waals attraction term, or penalizing the removal of solvent from charged residues, i.e., reducing direct interactions with the protein. However, a scoring function with such properties would not necessarily be able to place the short (tetrameric) heparin fragments. Thus, the optimal scoring function would depend on the length of the heparin chain. In principle, this would not be a problem, because we could optimize the interaction energy weight in a length-dependent matter. The real problem is that, apart from a few very special cases, there are no structures available for proteins with heparin

chains longer than eight saccharide units, and thus we have no data to parametrize and test the potential. Additional complications arise when one considers that most HS chains have considerably more structural variability than heparin, with long stretches of un- and undersulfated regions that might be involved in protein binding through mechanisms not captured by our model.

In spite of the above limitations, we hope that the new heparin binding site detection option in ClusPro will provide preliminary assistance in many areas where heparin is a focus of research. As shown, the server reliably identifies the heparin binding site, including the list of contact residues, on any protein. In combination with other methods, both computational and experimental, ClusPro may fill an imperative, though seemingly absent, role providing users with accurate information quickly and easily. The ClusPro server is freely available for academic and governmental research at http://cluspro.bu.edu/.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: (D.K.) midas@bu.edu.
*E-mail: (S.V.) vajda@bu.edu.

Notes

The authors declare the following competing financial interest(s): PIPER has been licensed to Acpharis Inc. for commercial applications. S.V., D. K., and D. B. own stock in Acpharis. PIPER and the use of the server ClusPro are free of charge for academic and governmental research.

ACKNOWLEDGMENTS

Research is supported in part by grants GM061867 and GM093147 from the National Institute of General Medical Sciences, grant HL096693 from the National Heart, Lung, and Blood Institute, grant DBI1147082 from the National Science Foundation, and grant M2012014 from the BrightFocus Foundation.

ABBREVIATIONS

GAG, glycosaminoglycan; HS, heparan sulfate; NMR, nuclear magnetic resonance; RMSD, root-mean-square deviation; CAPR, Critical Assessment of PRotein Interactions; MD, molecular dynamics; FFT, fast Fourier transform; AM1, Austin Model 1; BCC, bond charge correction

REFERENCES

(1) Bernfield, M.; Gotte, M.; Park, P. W.; Reizes, O.; Fitzgerald, M. L.; Lincecum, J.; Zako, M. Functions of cell surface heparan sulfate proteoglycans. *Annu. Rev. Biochem.* **1999**, *68*, 729−77.

(2) Gandhi, N. S.; Mancera, R. L. The structure of glycosaminoglycans and their interactions with proteins. *Chem. Biol. Drug Des* **2008**, *72*, 455−482.

(3) Lindahl, U. Heparan sulfate-protein interactions - a concept for drug design? *Thromb Haemostasis* **2007**, *98*, 109−115.

(4) Bourin, M. C.; Lindahl, U. Glycosaminoglycans and the regulation of blood-coagulation. *Biochem. J.* **1993**, *289*, 313−330.

(5) Spencer, J. L.; Bernanke, J. A.; Buczek-Thomas, J. A.; Nugent, M. A. A computational approach for deciphering the organization of glycosaminoglycans. *PLoS One* **2010**, 5.

(6) Raman, R.; Raguram, S.; Venkataraman, G.; Paulson, J. C.; Sasisekharan, R. Glycomics: An integrated systems approach to structure-function relationships of glycans. *Nat. Methods* **2005**, *2*, 817−24.

(7) Sasisekharan, R.; Raman, R.; Prabhakar, V. Glycomics approach to structure-function relationships of glycosaminoglycans. *Annu. Rev. Biomed Eng.* **2006**, *8*, 181−231.

(8) Capila, I.; Linhardt, R. J. Heparin-protein interactions. *Angew. Chem., Int. Ed. Engl.* **2002**, *41*, 391−412.

(9) Fugedi, P. The potential of the molecular diversity of heparin and heparan sulfate for drug development. *Mini Rev. Med. Chem.* **2003**, *3*, 659−67.

(10) Esko, J. D.; Lindahl, U. Molecular diversity of heparan sulfate. *J. Clin Invest* **2001**, *108*, 169−73.

(11) Forster, M.; Mulloy, B. Computational approaches to the identification of heparin-binding sites on the surfaces of proteins. *Biochem. Soc. Trans.* **2006**, *34*, 431−4.

(12) Mulloy, B.; Linhardt, R. J. Order out of complexity - protein structures that interact with heparin. *Curr. Opin Struct Biol.* **2001**, *11*, 623−628.

(13) Imberty, A.; Lortat-Jacob, H.; Perez, S. Structural view of glycosaminoglycan-protein interactions. *Carbohydr. Res.* **2007**, *342*, 430−439.

(14) Cardin, A. D.; Weintraub, H. J. R. Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **1989**, *9*, 21−32.

(15) Gandhi, N. S.; Coombe, D. R.; Mancera, R. L. Platelet endothelial cell adhesion molecule 1 (PECAM-1) and its interactions with glycosarninoglycans: 1. Molecular modeling studies. *Biochemistry* **2008**, *47*, 4851−4862.

(16) Ricard-Blum, S.; Feraud, O.; Lortat-Jacob, H.; Rencurosi, A.; Fukai, N.; Dkhissi, F.; Vittet, D.; Imberty, A.; Olsen, B. R.; van der Rest, M. Characterization of endostatin binding to heparin and heparan sulfate by surface plasmon resonance and molecular modeling - role of divalent cations. *J. Biol. Chem.* **2004**, *279*, 2927−2936.

(17) Mulloy, B.; Forster, M. J. Application of drug discovery software to the identification of heparin-binding sites on protein surfaces: A computational survey of the 4-helix cytokines. *Mol. Simulat* **2008**, *34*, 481−489.

(18) Pita, S. S. D.; Fernandes, T. V. A.; Caffarena, E. R.; Pascutti, P. G. Studies of molecular docking between fibroblast growth factor and heparin using generalized simulated annealing. *Int. J. Quantum Chem.* **2008**, *108*, 2608−2614.

(19) Kern, A.; Schmidt, K.; Leder, C.; Muller, O. J.; Wobus, C. E.; Bettinger, K.; Von der Lieth, C. W.; King, J. A.; Kleinschmidt, J. A. Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. *J. Virol* **2003**, *77*, 11072−11081.

(20) Ballut, L.; Sapay, N.; Chautard, E.; Imberty, A.; Ricard-Blum, S. Mapping of heparin/heparan sulfate binding sites on alpha v beta 3 integrin by molecular docking. *J. Mol. Recognit* **2013**, *26*, 76−85.

(21) Carpentier, M.; Denys, A.; Allain, F.; Vergoten, G. Molecular docking of heparin oligosaccharides with hep-ii heparin-binding domain of fibronectin reveals an interplay between the different positions of sulfate groups. *Glycoconj J.* **2013**, *31*, 161−169.

(22) Hung, T. J.; Tomiya, N.; Chang, T. H.; Cheng, W. C.; Kuo, P. H.; Ng, S. K.; Lien, P. C.; Lee, Y. C.; Chang, M. D. Functional characterization of ECP-heparin interaction: A novel molecular model. *PLoS One* **2013**, *8*, e82585.

(23) Bitomsky, W.; Wade, R. C. 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.

(24) Goodford, P. J. A computational-procedure for determining energetically favorable binding-sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(25) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(26) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aid Mol. Des* **2001**, *15*, 411−428.

(27) Samsonov, S. A.; Teyra, J.; Pisabarro, M. T. Docking glycosaminoglycans to proteins: Analysis of solvent inclusion. *J. Comput. Aid Mol. Des* **2011**, *25*, 477−489.
(28) Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621−7.
(29) Vajda, S.; Guarnieri, F. Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr. Opin Drug Discov Devel* **2006**, *9*, 354−62.
(30) Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins* **2006**, *65*, 392−406.
(31) Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: An automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **2004**, *20*, 45−50.
(32) Kozakov, D.; Hall, D. R.; Beglov, D.; Brenke, R.; Comeau, S. R.; Shen, Y.; Li, K.; Zheng, J.; Vakili, P.; Paschalidis, I.; Vajda, S. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13−19. *Proteins* **2010**, *78*, 3124−30.
(33) Lensink, M. F.; Mendez, R.; Wodak, S. J. Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins* **2007**, *69*, 704−18.
(34) Lensink, M. F.; Wodak, S. J. Docking and scoring protein interactions: CAPRI 2009. *Proteins* **2010**, *78*, 3073−84.
(35) Lensink, M. F.; Wodak, S. J. Docking, scoring, and affinity prediction in CAPRI. *Proteins* **2013**, *81*, 2082−95.
(36) Comeau, S. R.; Kozakov, D.; Brenke, R.; Shen, Y.; Beglov, D.; Vajda, S. ClusPro: Performance in CAPRI rounds 6−11 and the new server. *Proteins* **2007**, *69*, 781−5.
(37) Kozakov, D.; Clodfelter, K. H.; Vajda, S.; Camacho, C. J. Optimal clustering for detecting near-native conformations in protein docking. *Biophys. J.* **2005**, *89*, 867−75.
(38) Kozakov, D.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Vajda, S. How good is automated protein docking? *Proteins* **2013**, *81*, 2159−2166.
(39) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−42.
(40) Moon, A. F.; Edavettal, S. C.; Krahn, J. M.; Munoz, E. M.; Negishi, M.; Linhardt, R. J.; Liu, J.; Pedersen, L. C. Structural analysis of the sulfotransferase (3-o-sulfotransferase isoform 3) involved in the biosynthesis of an entry receptor for herpes simplex virus 1. *J. Biol. Chem.* **2004**, *279*, 45185−45193.
(41) Xue, Y.; Lee, S. W.; Wang, Y. C.; Ha, Y. Crystal structure of the E2 domain of amyloid precursor protein-like protein 1 in complex with sucrose octasulfate. *J. Biol. Chem.* **2011**, *286*, 29748−29757.
(42) Xue, Y.; Lee, S.; Ha, Y. Crystal structure of amyloid precursor-like protein 1 and heparin complex suggests a dual role of heparin in E2 dimerization. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16229−16234.
(43) Chirgadze, D. Y.; Hepple, J. P.; Zhou, H. J.; Byrd, R. A.; Blundell, T. L.; Gherardi, E. Crystal structure of the NK1 fragment of HGF/SF suggests a novel mode for growth factor dimerization and receptor binding. *Nat. Struct. Biol.* **1999**, *6*, 72−79.
(44) Lietha, D.; Chirgadze, D. Y.; Mulloy, B.; Blundell, T. L.; Gherardi, E. Crystal structures of NK1-heparin complexes reveal the basis for NK1 activity and enable engineering of potent agonists of the MET receptor. *EMBO J.* **2001**, *20*, 5543−5555.
(45) Huntington, J. A.; Kjellberg, M.; Stenflo, J. Crystal structure of protein C inhibitor provides insights into hormone binding and heparin activation. *Structure* **2003**, *11*, 205−215.
(46) Li, W.; Huntington, J. A. The heparin binding site of protein c inhibitor is protease-dependent. *J. Biol. Chem.* **2008**, *283*, 36039−36045.
(47) Ago, H.; Kitagawa, Y.; Fujishima, A.; Matsuura, Y.; Katsube, Y. Crystal-structure of basic fibroblast growth-factor at 1.6 a resolution. *J. Biochem-Tokyo* **1991**, *110*, 360−363.
(48) Faham, S.; Hileman, R. E.; Fromm, J. R.; Linhardt, R. J.; Rees, D. C. Heparin structure and interactions with basic fibroblast growth factor. *Science* **1996**, *271*, 1116−1120.
(49) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General atomic and molecular electronic-structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(50) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph Model* **2006**, *25*, 247−60.
(51) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−74.
(52) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(53) Capila, I.; Hernaiz, M. J.; Mo, Y. D.; Mealy, T. R.; Campos, B.; Dedman, J. R.; Linhardt, R. J.; Seaton, B. A. Annexin V–heparin oligosaccharide complex suggests heparan sulfate–mediated assembly on cell surfaces. *Structure* **2001**, *9*, 57−64.
(54) Colloc'h, N.; Sopkova-de Oliveira Santos, J.; Retailleau, P.; Vivares, D.; Bonnete, F.; Langlois d'Estainto, B.; Gallois, B.; Brisson, A.; Risso, J. J.; Lemaire, M.; Prange, T.; Abraini, J. H. Protein crystallography under xenon and nitrous oxide pressure: Comparison with in vivo pharmacology studies and implications for the mechanism of inhaled anesthetic action. *Biophys. J.* **2007**, *92*, 217−24.
(55) Sachchidanand; Lequin, O.; Staunton, D.; Mulloy, B.; Forster, M. J.; Yoshida, K.; Campbell, I. D. Mapping the heparin-binding site on the 13−14F3 fragment of fibronectin. *J. Biol. Chem.* **2002**, *277*, 50629−35.
(56) DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schlessinger, J.; Hendrickson, W. A. Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature* **1998**, *393*, 812−7.
(57) Fernandez, I. S.; Cuevas, P.; Angulo, J.; Lopez-Navajas, P.; Canales-Mayordomo, A.; Gonzalez-Corrochano, R.; Lozano, R. M.; Valverde, S.; Jimenez-Barbero, J.; Romero, A.; Gimenez-Gallego, G. Gentisic acid, a compound associated with plant defense and a metabolite of aspirin, heads a new class of in vivo fibroblast growth factor inhibitors. *J. Biol. Chem.* **2010**, *285*, 11714−11729.
(58) Rigden, D. J.; Jedrzejas, M. J. Structures of Streptococcus pneumoniae hyaluronate lyase in complex with chondroitin and chondroitin sulfate disaccharides - insights into specificity and mechanism of action. *J. Biol. Chem.* **2003**, *278*, 50596−50606.
(59) Lunin, V. V.; Li, Y. G.; Linhardt, R. J.; Miyazono, H.; Kyogashima, M.; Kaneko, T.; Bell, A. W.; Cygler, M. High-resolution crystal structure of Arthrobacter aurescens chondroitin ac lyase: An enzyme-substrate complex defines the catalytic mechanism. *J. Mol. Biol.* **2004**, *337*, 367−386.
(60) Capila, I.; Hernaiz, M. J.; Mo, Y. D.; Mealy, T. R.; Campos, B.; Dedman, J. R.; Linhardt, R. J.; Seaton, B. A. Annexin V-heparin oligosaccharide complex suggests heparan sulfate-mediated assembly on cell surfaces. *Structure* **2001**, *9*, 57−64.
(61) Shaya, D.; Zhao, W. J.; Garron, M. L.; Xiao, Z. P.; Cui, Q. Z.; Zhang, Z. Q.; Sulea, T.; Linhardt, R. J.; Cygler, M. Catalytic mechanism of heparinase ii investigated by site-directed mutagenesis and the crystal structure with its substrate. *J. Biol. Chem.* **2010**, *285*, 20051−20061.
(62) Shaya, D.; Tocilj, A.; Li, Y. G.; Myette, J.; Venkataraman, G.; Sasisekharan, R.; Cygler, M. Crystal structure of heparinase II from Pedobacter heparinus and its complex with a disaccharide product. *J. Biol. Chem.* **2006**, *281*, 15525−15535.
(63) Shao, C. H.; Zhang, F. M.; Kemp, M. M.; Linhardt, R. J.; Waisman, D. M.; Head, J. F.; Seaton, B. A. Crystallographic analysis of calcium-dependent heparin binding to annexin A2. *J. Biol. Chem.* **2006**, *281*, 31689−31695.

(64) Swaminathan, G. J.; Myszka, D. G.; Katsamba, P. S.; Ohnuki, L. E.; Gleich, G. J.; Acharya, K. R. Eosinophil-granule major basic protein, a C-type lectin, binds heparin. *Biochemistry* **2005**, *44*, 14152−14158.
(65) Swaminathan, G. J.; Weaver, A. J.; Loegering, D. A.; Checkel, J. L.; Leonidas, D. D.; Gleich, G. J.; Acharya, K. R. Crystal structure of the eosinophil major basic protein at 1.8 angstrom - an atypical lectin with a paradigm shift in specificity. *J. Biol. Chem.* **2001**, *276*, 26197−26203.
(66) Han, Y. H.; Garron, M. L.; Kim, H. Y.; Kim, W. S.; Zhang, Z. Q.; Ryu, K. S.; Shaya, D.; Xiao, Z. P.; Cheong, C.; Kim, Y. S.; Linhardt, R. J.; Jeon, Y. H.; Cygler, M. Structural snapshots of heparin depolymerization by heparin lyase i. *J. Biol. Chem.* **2009**, *284*, 34019−34027.