JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

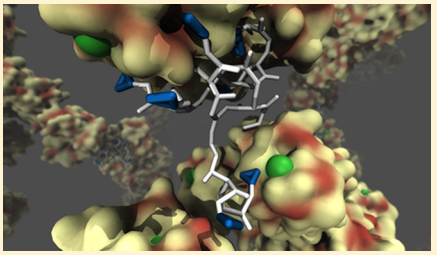
Optimizing the Multivalent Binding of the Bacterial Lectin LecA by Glycopeptide Dendrimers for Therapeutic Purposes

Benjamin Bouvier*

Laboratoire de Glycochimie, des Antimicrobiens et des Agroressources, CNRS UMR7378/Université de Picardie Jules Verne, 10 rue Baudelocque, 80039 Amiens Cedex, France

Supporting Information

ABSTRACT: Bacterial lectins are nonenzymatic sugar-binding proteins involved in the formation of biofilms and the onset of virulence. The weakness of individual sugar−lectin interactions is compensated by the potentially large number of simultaneous copies of such contacts, resulting in high overall sugar−lectin affinities and marked specificities. Therapeutic compounds functionalized with sugar residues can compete with the host glycans for binding to lectins only if they are able to take advantage of this multivalent binding mechanism. Glycopeptide dendrimers, featuring treelike topologies with sugar moieties at their leaves, have already shown great promise in this regard. However, optimizing the dendrimers' amino acid sequence is necessary to match the dynamics of the lectin active sites with that of the multivalent ligands. This work combines long-time-scale coarse-grained simulations of dendrimers and lectins with a reasoned exploration of the dendrimer sequence space in an attempt to suggest sequences that could maximize multivalent binding to the galactose-specific bacterial lectin LecA. These candidates are validated by simulations of mixed dendrimer/lectin solutions, and the effects of the dendrimers on lectin dynamics are discussed. This approach is an attractive first step in the conception of therapeutic compounds based on the dendrimer scaffold and contributes to the understanding of the various classes of multivalency that underpin the ubiquitous "sugar code".



INTRODUCTION

Carbohydrates play a prominent role in the mediation of biological processes, notably via the recognition of lectins by cellular glycans. This recognition process, dubbed the "sugar code",[1] remains the subject of active investigation, with fields of relevance ranging from oncology,[2] through biosensors[3] and synthetic supramolecular chemistry[4] to bionergy production[5] and infectiology.[6,7] Lectins feature a wide array of binding-site topologies that can bind sugar moieties, albeit with affinities that are generally weak. Despite this, a very strong specificity for certain cellular glycans can be achieved.[1] This apparent contradiction can be resolved by introducing the concept of multivalency, in which multiple simultaneous weak interactions between carbohydrates and their receptors contribute to an overall strong affinity that can be fine-tuned through its individual components.[8,9] Consequently, the conception of synthetic scaffolds bearing multiple covalently bound carbohydrates as a means of taking advantage of the multivalency effect in the recognition and binding of lectins has been an ongoing research goal over the past decade.[10−15]

Dendrimers are molecules that feature a treelike topology, with branches splitting off a trunk and then subdividing again following a regular pattern. The structural and dynamical properties of the branch termini, where functionalization usually takes place, depend on the chemical nature of the branches and the interactions between them, making dendrimers particulary versatile scaffolds in numerous molecular recognition applications, including lectin/glycan recognition, for which they have been used for several years.[14,15] Peptidic dendrimers, inspired by the antimicrobial peptides that are a part of the defense mechanisms of most multicellular organisms,[16] have the advantage of being very convenient to synthesize; as an added bonus, their compact yet flexible molten-globule structure makes them very resilient to proteolysis,[17] facilitating their delivery. As potential antibiotic agents, they have already shown promising effects against the human opportunistic bacterium *Pseudomonas aeruginosa*: non-functionalized, amphiphilic dendrimers have proven able to disrupt the bacterial membrane,[18,19] whereas glycoconjugated peptidic dendrimers targeting bacterial lectins have been shown to hinder bacterial adhesion to human cells and biofilm formation, limiting the onset of virulence and resistance to antibiotics.[15,20,21]

Bacterial lectins such as galactose-specific LecA and fucose-specific LecB play a role in specific recognition of host cells, attachment to target cells, and triggering of virulence.[22,23] As tetramers, they feature four sugar-binding sites. Designing peptidic dendrimer glycoconjugates in which the spacing between sugar moieties coincides with the intra- and/or intertetramer binding-site distances would in theory allow maximization of the impact of multivalency on the recognition and binding process and ensure optimal competition with the binding of host glycans, enhancing the therapeutic efficiency of

Received: March 11, 2016
Published: May 25, 2016

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

the compounds. A thorough and systematic optimization of the amino acid sequence of the dendrimer "arms" to achieve this goal is therefore highly desirable but also very difficult because of the unfavorable sequence combinatorics involved. In silico studies, which are in principle well-suited for the cost-effective preliminary screening of a large number of candidates, also stumble on this difficulty: flexible molecular constructs such as peptide dendrimers have extensive conformational spaces that are very difficult to fully characterize at an acceptable computational cost. Thus, theoretical studies on the matter have mostly been used either for a posteriori validation of experimental results[20] or to explore the effects of limited variations upon sequences gleaned from databases.[24−26]

Coarse-grained models (in which atoms making up stand-alone chemical functions are grouped into beads) can be used to simulate much longer time scales than their all-atom counterparts for the same computational cost; this is due not only to the straightforward reduction in the dimensionality of the problem but also to an effective kinetic speedup associated with the "smoothing out" of local features on the energy landscape brought about by coarse-graining.[27] This study applies the MARTINI[28,29] coarse-grained model (which has been successfully applied to unstructured antimicrobial peptides in the past[30−32]) and long-time-scale molecular dynamics (MD) simulations to the systematic exploration of the effects of the amino acid sequence on the dynamics of octavalent peptide dendrimers functionalized with galactose as well as their ability to bind the galactose-specific lectin LecA. The sequence-to-dynamics relationship thus inferred is then used to suggest potential candidates in which the recognition and binding of LecA would benefit the most from multivalency. Finally, the best and worst candidates are simulated in the presence of LecA tetramers to evaluate the validity of these predictions.

METHODS

Root-Mean-Square Deviation Calculations. Because of the topology of the peptide dendrimers considered (see Figure 1), amino acid chains branching from the same parent residue are chemically equivalent, and their numbering in the structure can be freely swapped. This has to be taken into account when computing the root-mean-square deviation[33] (RMSD) between two dendrimer structures: both the preliminary least-squares alignment procedure (used to remove the effects of rigid-body rotation and translation) and the actual deviation calculation require a mapping of atoms between the two structures, which for such dendrimers is not unique. The ability to swap equivalent branches results in two possible numberings per branching point. The dendrimers studied herein have seven branching points, for a total of $2^7 = 128$ possible equivalent numberings. The RMSD between two structures can thus be computed as the minimum of the set of RMSDs obtained between structure 1 and the 128 equivalent numberings of structure 2. A tool for the automatic generation of all possible mappings between structures and calculation of the corresponding RMSDs has been designed for this study.

Statistical Tools. The Kolmogorov−Smirnov (KS) distance[34] was used to quantify the similarity between distributions of structural values (RMSD, intersugar distances, etc.) obtained from the simulation of the different dendrimer compounds. Clustering of dendrimers was performed using the affinity propagation method.[35] The method iteratively computes the pairwise responsibility $r(i,j)$ and availability $a(i,j)$ between pairs of samples (i,j) taken from a data set.

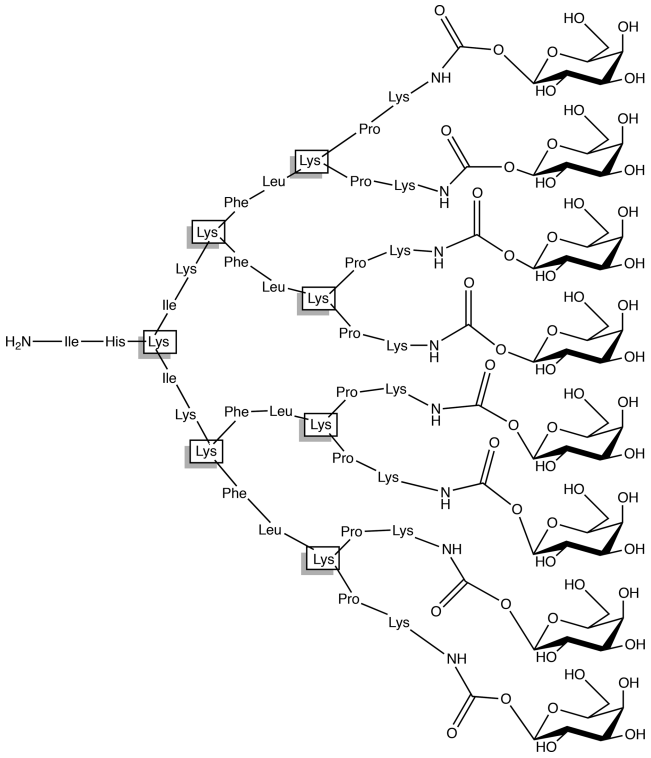


Figure 1. Two-dimensional sketch of the multivalent glycoconjugated peptide dendrimers under study. The lysine residues that implement branching via side-chain isopeptide bonds are framed.

Responsibility is the accumulated evidence that sample j should be an exemplar for sample i (in other words, that sample j should be the representative structure of a cluster containing both itself and sample i); availability is the accumulated evidence that sample i should choose sample j as its exemplar when compared with sample j's exemplarity for all other samples:

$$\begin{cases} r(i,j) = s(i,j) - \max[a(i,k) + s(i,k) \\ \quad \forall\, k \neq j] \\ a(i,j) = \min[0, r(j,j) + \sum_{k \notin \{i,j\}} r(k,j)] \end{cases} \quad (1)$$

where $s(i,j)$ is the similarity between samples i and j, which was taken as the opposite value of the KS distance between the corresponding distributions. The system of eqs 1 is iterated to convergence. The affinity propagation method was chosen because of its ability to take the similarities between samples, rather than the samples themselves, as input; this is very convenient when clustering complex objects (such as distributions) using nonstandard metrics. As an additional benefit, affinity propagation does not require the number of clusters to be specified by the user, a source of bias found in most other clustering techniques. However, on complex data sets, the method has a tendency to generate many clusters. To alleviate this behavior, affinity propagation was performed hierarchically on the data sets: the exemplars generated at generation n were used as the input set for generation $n - 1$; in this scheme, the generation-n exemplars are children of the generation-$(n - 1)$ exemplars to whose cluster they belong. The process was repeated until further clustering did not result in a simplification of the data set. This hierarchical approach

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

retains the advantage of not having to choose the number of clusters yet provides different levels of classification that facilitate the understanding of the structure of the underlying data set by the human mind.

Sequence logos are a widely used representation of multiple sequence alignments of proteins or nucleic acids.[36] A logo consists of a stack of symbols per position in the sequence; the height of each character in the stack relates to the frequency of the corresponding amino acid/nucleobase at the position, while the total height of the stack denotes sequence conservation. The latter is expressed as the difference between the maximum Shannon entropy, S_{max}, and the entropy in the observed character distribution at the considered position, S_{obs}:

$$R_{seq} = S_{max} - S_{obs} = \log_2 N - \left(-\sum_{n=1}^{N} p_n \log_2 p_n\right) \quad (2)$$

where p_n is the frequency of symbol n at the location and N is the number of possible distinct symbols.

Computational Details. All of the coarse-grained MD simulations were performed using version 2.2 of the MARTINI force field[28,29] inside GROMACS versions 4.6.5 and 5.1.[37,38] The parameters for the branching lysines were derived from those of the standard amino acid,[28] replacing the final Qd-type grain by a P1-type particle that was bound to the backbone grain of the next amino acid (bond equilibrium length 0.35 nm, bond force constant 1250 kJ mol^{-1} nm^{-2}; angle equilibrium value 124°, angle force constant 20 kJ mol^{-1} rad^{-2}). These parameters were chosen to provide the best match with the corresponding all-atom model on a set of distance distributions (see Figure S1 in the Supporting Information for details). The parameters for the terminal galactose residues were taken from López and co-workers' extension of MARTINI to carbohydrates.[39] The sugars were covalently bound to the backbone bead of the corresponding N-terminal residue (bond equilibrium length 0.35 nm, bond force constant 1250 kJ mol^{-1} nm^{-2}; angle to sugar bead 1: equilibrium value 127°, force constant 20 kJ mol^{-1} rad^{-2}; angle to sugar bead 2: equilibrium value 149°, force constant 25 kJ mol^{-1} rad^{-2}; angle to sugar bead 3: equilibrium value 123°, force constant 25 kJ mol^{-1} rad^{-2}). The dendrimers were treated as extended regions (coils) within the MARTINI framework and as such did not feature secondary-structure restraining potentials.

The dendrimers were minimized in vacuo and placed in a truncated octahedral box of MARTINI water beads (in which one bead replaces four actual water molecules[28]) extending at least 14 Å from the solute in all directions. Interestingly, while the simple MARTINI water model cannot account for the possible mediation of contacts by single water molecules or the directionality of hydrogen-bonding patterns, the bundling of water molecules inside a bead has been shown to have little overall effect on the physical and thermodynamic aspects of the interaction of water with most solutes.[40] The solvated system was then minimized to convergence and equilibrated at 300 K with fixed backbone beads for 5 ns. Production simulations without restraints were then performed for 5 μs. All of the MD simulations used a time step of 20 fs. Conditions of constant temperature (300 K) and pressure (1 bar) were maintained using velocity-rescaling (τ = 1.0 ps) and Parrinello−Rahman algorithms (τ = 12 ps, β = 1.8 × 10^{-5} bar^{-1}), respectively, applied separately to the solute and solvent. Coulomb interactions extended to 12 Å, with the forces decaying smoothly from 0 Å. The van der Waals interactions extended to 12 Å, with the forces decaying smoothly from 9 Å.

The structure of the LecA tetramer was taken from the work of Novoa and co-workers[20] (PDB entry 4CP9), converted to a coarse-grained representation using Martinize, and minimized in vacuo. As is usual for globular proteins in MARTINI, the secondary structure of each of the monomers was preserved by employing an additional elastic network model connecting the backbone beads of amino acids at distances of less than 9 Å and separated by at least three sequence positions (force constant 500 kJ mol^{-1} nm^{-2}). Spurious LecA tetramer dissociation events, although rarely observed, could have a negative impact on the efficiency of conformational sampling; to prevent them, the disposition of the LecA units inside a tetramer was constrained using harmonic potentials between the centers of masses of the units (500 kJ mol^{-1} nm^{-2}). Similarly, rare exchanges of the Ca^{2+} ion at each monomer's binding site were seen to occur over time scales compatible with the ion's experimental binding affinity;[41] to prevent such "blue moon" events from tainting the sampling of galactose recognition and binding events, the ions were restrained relative to the backbone beads of active-site amino acids Tyr36 and Thr104 with a 500 kJ mol^{-1} nm^{-2} harmonic potential. Five copies of the minimized LecA tetramer were inserted at random positions and orientations inside an octahedral box that was then filled with MARTINI water beads. Minimization and equilibration were performed as previously described, upon which 2 μs of unrestrained production simulation was begun. This was repeated for three distinct starting conditions (different random arrangements of the five tetramers), yielding 6 μs of total simulation time. In all cases, the equilibrated box had a typical volume of around 6000 nm^3, corresponding to an approximate LecA concentration of 1.4 mM, and comprised approximately 48000 beads.

The simulations of the LecA/dendrimer mixtures were carried out similarly by randomly introducing five copies of the relevant dendrimer and five LecA tetramers into the simulation cell. Three distinct starting conditions were thus generated, in which (to ensure generality) the absence of initial dendrimer−lectin contacts was checked. The protocols and simulation lengths were the same as for the $LecA_5$ system; the approximate volume of the equilibrated box, the number of particles, and the concentrations of LecA and dendrimers were also very similar to their values for $LecA_5$. For each dendrimer, an additional set of three 2 μs simulations from distinct starting conditions was generated to verify the convergence of the active-site distance distributions (see the Results for details).

Miscellaneous. All of the statistical analyses were performed using Scientific Python[42] and Scikit-learn.[43] All of the molecular graphics were produced using VMD.[44] Sequence logos were generated using the WebLogo Python API.[45] Trees were represented using the ETE toolkit.[46] All other plots and figures were generated using Matplotlib.[47]

RESULTS

The functionalized peptide dendrimers under study possess a branching, treelike topology with eight "leaf" nodes, each of which bears a galactose moiety (see Figure 1). Suitably functionalized branching peptide dendrimers of this nature have been shown to be highly effective in the recognition and binding of both the galactose-specific lectin LecA[21] and the fucose-specific lectin LecB.[48] The particular framework selected for this study was chosen to be a good representative of this

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

class: the octavalent scaffold can potentially maximize multivalent effects, and the use of a common-length spacer between any two branching points allows a consistent comparison of sequence effects as a function of residue depth. The framework consists of chains of homogeneous sequences: at any given depth in the tree, only one sort of amino acid is found. There are 11 such levels, three of which correspond to branching lysine residues that cannot be replaced by another amino acid because they effectively implement the tree topology. Consequently, eight levels remain for variation of the dendrimer amino acid sequence.

The number of possible sequences, on the order of 25 billion (20^8), precludes an exhaustive study even at the coarse-grained level. The number of amino acid types employed to generate the dendrimer sequences was thus reduced to five, representative of the major classes: small hydrophobic (alanine, A), large hydrophobic (methionine, M), polar neutral (serine, S), positively charged (arginine, R), and negatively charged (aspartate, D). Proline (P), whose unique conformational rigidity makes it stand out from the group of hydrophobic amino acids with which it is usually bundled, was verified to be reasonably represented by alanine for this study (see below and Figure S5 in the Supporting Information). Although volume and chemical character are not independent variables, this subset of amino acids was also chosen as reasonably representative of the entire set of amino acid volumes in solution[49] (75.1 ± 21.8 vs 85.1 ± 24.0 mL mol^{-1}).

To further simplify the combinatorics of the problem, it is possible to vary the sequence inside a sliding window rather than on a per-residue basis; the number of possible sequences is thus reduced from N^L to $(L - W + 1) \times N^W$, where N is the number of possible residue types per position, L is the number of positions at which the sequence is varied, and W is the width of the sliding window. Longer windows are more costly but offer a better description of the collective effects between neighboring residues. The dendrimers under study consist of branching lysines separating spans of two amino acids; the lysine residues act as buffers between the dipeptides on either side, interfering with the propagation of correlated information between them. Consequently, using a window length of two in these systems was deemed a good trade-off between computational cost and the inclusion of collective effects (a hypothesis that will be verified further down); it reduced the number of sequences to be considered from ∼391 000 to 175, allowing for adequately long simulation times on each. The sequence variations were performed upon a dendrimer of sequence H_2N-IH(K)IK(K)FL(K)PK-Gal (from trunk to leaves, with branching lysines shown in parentheses), a variant of which has been experimentally shown to be a good binder of the lectin LecB (25 nM binding affinity);[48] although the system-wide sequence exploration subsequently performed renders this choice somewhat irrelevant, it is justified by the similarities between LecA and LecB (charged active sites, comparable shortest intratetramer active-site distances). The solubility of dendrimers in water, which is essential to their druggability, is theoretically favored by the presence of polar and charged amino acids; however, the high number of galactose moieties per dendrimer and their location at the extremities of flexible chains, combined with the very high solubility of galactose (∼4 mol L^{-1}), allow for a hydrophilic external shell that can potentially shield hydrophobic amino acids from water, akin to the hydrophobic collapse observed in protein folding. Therefore, solubility issues at millimolar or lower concentrations are not expected regardless of sequence. Indeed, similar glycopeptide dendrimers have been shown to be water-soluble at such concentrations during isothermal titration calorimetry experiments.[21,26]

The 175 possible sequences were constructed and simulated in explicit solvent for 5 μs. Quantifying differences between peptide structures is traditionally performed by computing the RMSD of the backbone atoms after removal of rigid-body rotation and translation effects via least-squares fitting; this approach has already been applied to peptidic dendrimers of varying sequences.[25] The methodology was adapted to the symmetric topology of the current dendrimers (see Methods for details). The "central" structure of each dendrimer (whose RMSD to the average structure over the entire simulation is minimal) was extracted, and a square, symmetric, rank-175 similarity matrix was built from the RMSD values between the central structures of all possible dendrimer pairs. This matrix was used as input to the affinity propagation clustering method (see Methods for details), which grouped the 175 dendrimers into six unevenly populated clusters. To visualize the properties of the clusters thus obtained, statistics of hydropathy indices[50] (a well-adopted indication of an amino acid's polarity), volumes, and sequence positions were performed over each of them. The volume and hydropathy values for each dendrimer were taken as the sums of the corresponding values for the two amino acids forming the mutation window; the sequence position was taken to be the index of the amino acid within the window that is closest to the dendrimer trunk (Figure 2).

As can be seen, the clusters show no clear discrimination of dendrimers on the basis of position, hydropathy, or volume. All of the clusters contain dendrimers mutated on at least six of the seven possible positions with residues of all sizes and hydropathies; only a limited specificity for small and hydrophobic residues can be observed in cluster number 5, of population 18. This is in part due to the fact that the RMSD is a very degenerate measure of structural differences: two structures with equal RMSDs with respect to a third are not necessarily similar, especially if the common RMSD is large; the RMSD is more sensitive to minute differences between structures than to very large ones. However, clustering based on this measure still makes a degree of sense since pairwise RMSDs inside each cluster are expected to be small if the number of clusters is sufficiently large (in the present case, the mean intracluster RMSD was found to be equal to 2.1 Å, which is indeed small compared to the mean intercluster RMSD of 6.4 Å). Therefore, one has to conclude that mutations in the dendrimer sequence have little impact on the mean structure. The other structural measure usually employed in such cases, the macromolecular radius of gyration, was also verified to feature a relative insensitivity to amino acid nature (data not shown).

The average backbone structures of the six dendrimers representative of the six clusters, aligned along their inertial axes, are presented on Figure 3. The figure also provides for each cluster the moment along each of the three inertial axes. As can be seen, the average structures of the dendrimers are quite diverse, ranging from extended (purple structure, RR mutation on the last branch) to compact (orange structure, AA mutation on the middle branch) and from highly isotropic (red structure, AS mutation on the last middle-branch and first last-branch residues) to highly anisotropic (purple structure). The existence of this structural diversity shows that the sequence of

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

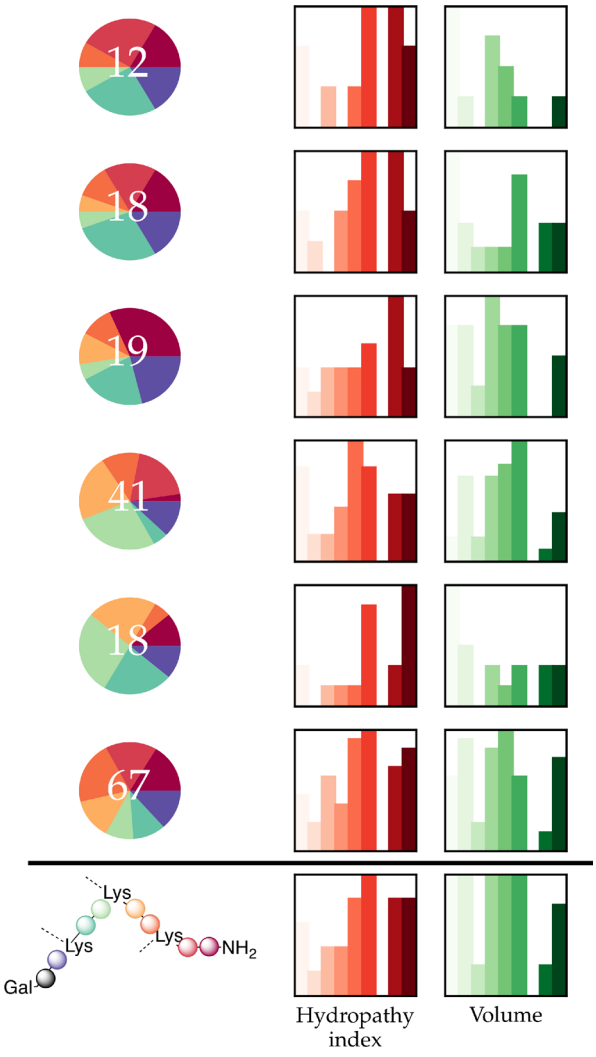


Figure 2. Properties of the clusters inferred from the RMSD matrix between the average structures of the dendrimers. Each row (except the last) represents a cluster, characterized by (i) its member count (white numbers on left panel) and statistics over (ii) mutation position (left panel, pie chart using the position of the mutated amino acid closest to the dendrimer trunk and the color code given on the last line of the left panel), (iii) hydropathy index (center panel, red bars, sum of values over the two mutated amino acids), and (iv) volume (right panel, green bars, sum of values over the two mutated amino acids). The hydropathy and volume bar plots on the last line correspond to the entire population of 175 dendrimers. Hydropathy values are binned from −9.0 (left bar, light red) to 3.8 (right bar, dark red) in steps of ∼1.4; volumes are binned from 106.4 mL mol^{-1} (left bar, light green) to 208.2 mL mol^{-1} (right bar, dark green) in steps of ∼11.4 mL mol^{-1}.

the dendrimer chains does have an impact on the system's behavior but that the static picture provided by the use of degenerate structural measures (RMSD, radius of gyration) on time-averaged structures is not suitable to its elucidation. There is a clear need to take into account the dynamics of the dendrimers in the classification mechanism (different dendrimer sequences could be characterized by different dynamics around similar average structures) and to optimize the dynamics−activity relationship using statistics over structural measures that relate as closely as possible to lectin recognition.

To this end, I have focused on the distribution of distances between sugar "leaves" over time, as revealed by the 5 μs of MD

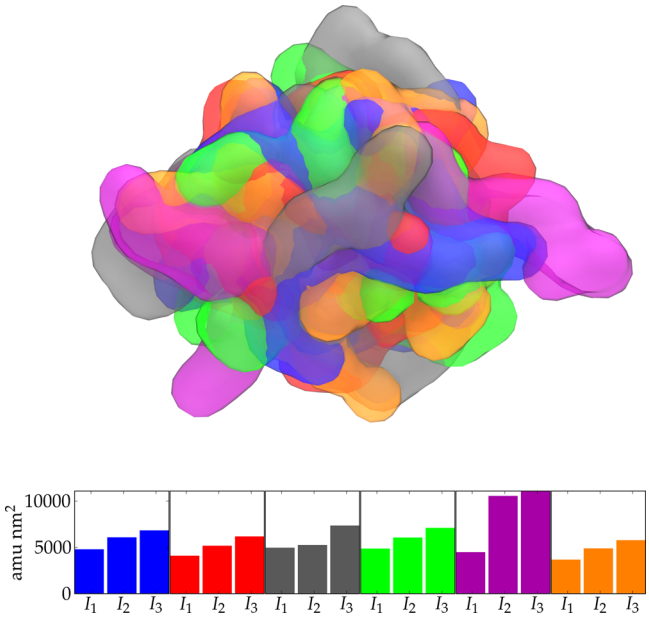


Figure 3. (top) Superposition of the representative structures of the clusters (as backbone bead surfaces, with one color per cluster). The structures are aligned along their inertial axes (first axis I_1, horizontal; second axis I_2, vertical; third axis I_3, perpendicular to the figure plane). (bottom) Moment along each inertial axis for each cluster representative (colors match those of the structures in the top panel).

simulation of each dendrimer sequence. This measure of the dendrimers' dynamical behavior does not suffer from the issues mentioned above and is critically related to the ability of the dendrimers to multivalently bind LecA (see Figure S2 in the Supporting Information for examples of such distributions). A similarity matrix between all possible pairs of distributions was calculated on the basis of the Kolmogorov−Smirnov distance (see Methods for details). As before, this matrix was subjected to clustering using affinity propagation, which yielded 28 distinct clusters—a testimony to the complexity of the underlying data set. To facilitate the understanding of these results, a hierarchical approach was employed to generate larger "parent" clusters. To visualize the nature of the clusters in the resulting "tree", statistics of hydropathy indices, volumes, and sequence positions were performed over each of them as previously described. Figure 4 represents the cluster tree, decorated with these per-cluster statistics.

Unlike before, there is now a clear influence of the position at which the sequence is varied on the constitution of the clusters; the impact of position on the overall dynamics of the systems appears to be much larger than that of the chemical nature or size of the amino acids introduced. Three of the six first-generation clusters are exclusively composed of dendrimers mutated on the last branch; the fourth cluster mixes in a limited contribution of the second amino acid of the middle branch. The two remaining clusters, which are much larger in population (28 and 124 members vs a total of 23 for the previously mentioned four clusters), are mostly composed of dendrimers mutated on the second and third branches or the first and second branches, respectively. The first-generation clusters composed of last-branch mutants show a limited, but apparent, discrimination between amino acid polarities and sizes: the RR mutation of the last branch is alone in its cluster; the remainder of the hydrophilic combinations are divided into two clusters (size 3, small amino acids, and size 6, larger ones);

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

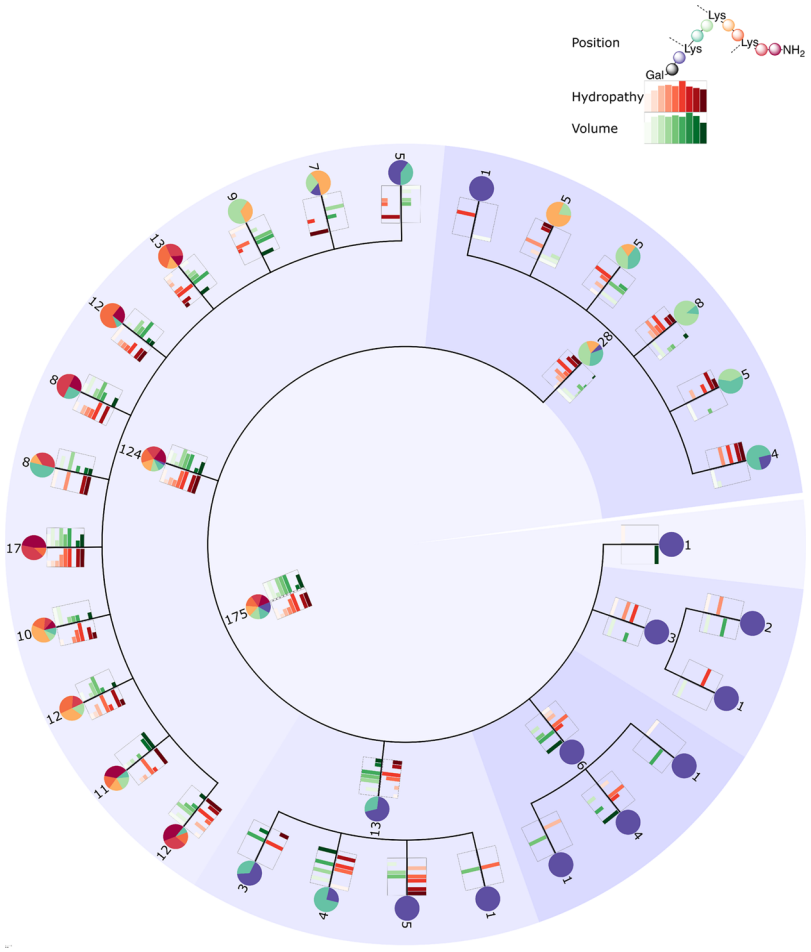


Figure 4. Hierarchical clutering of dendrimer sequences with respect to the distributions of intersugar distances over time. The filiation of clusters is represented as a tree from the center (parents) toward the periphery (children). Each cluster is described by its member count and statistics over mutation position (pie chart, using the position of the mutated amino acid closest to the dendrimer trunk), hydropathy index (red bars, sum of values over the two mutated amino acids), and volume (green bars, sum of values over the two mutated amino acids).

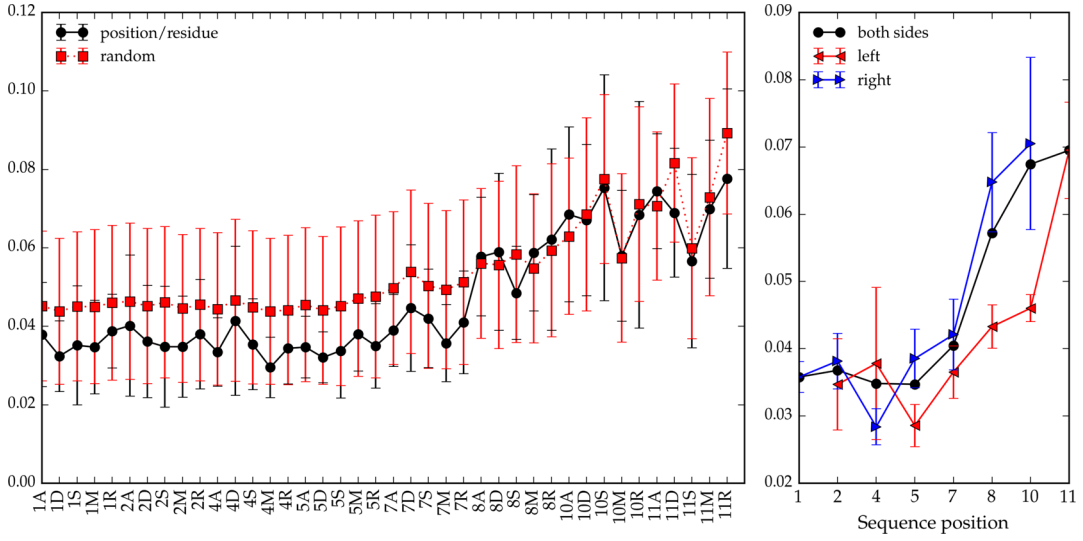


Figure 5. (left) Means and standard deviations of KS distances between sugar distance distributions for (i) pairs of dendrimers bearing a specific residue at a specific location (solid black line/circles) and (ii) a dendrimer bearing a specific residue at a specific location and any other dendrimer (dotted red line/squares). (right) Means and standard deviations of KS distances between sugar distance distributions for pairs of dendrimers mutated at a given position that have been simultaneously mutated (i) on the left (trunkward) side (red), (ii) on the right (leafward) side (blue), or (iii) on either side (black).

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

hydrophobic combinations of all sizes are bundled into the fourth cluster of population 13, for which discrimination over amino acid sizes occurs only at generation 2. On the other hand, the two large first-generation clusters consisting of first- to third-branch mutants do not give rise, at the second generation, to a marked separation based on size or hydropathy; the corresponding second-generation clusters are complex mixtures of dendrimer sequences, most of which are difficult to describe in simple terms. To summarize, the position of the mutation on the chains is the most important parameter to take into account when optimizing dendrimer sequences, followed by polarity and then size; in practice, the last two properties seem to matter only for last-branch mutants. Interestingly, the dependence of the structure and dynamics upon the polarity of the dendrimer-arm amino acids is strikingly more complex than the expected simple preference of polar (respectively hydrophobic) sequences for extended (respectively compact) conformations.

I now focus on the collective effects on the dendrimer dynamics induced by the mutation of neighboring residues. In view of the fact that the amino acid sequence was varied inside a sliding window of length 2, there are 10 instances of dendrimers having a given mutation at a given position (five possible neighbors on either side), except for the first and final positions, which, having only one neighbor, are represented by five instances only. To evaluate the importance of neighbor effects for a given mutation at a given position, the average and standard deviation of the set of dissimilarities between all possible pairs of instances (the "signal") were compared to the same statistical measures on the set of dissimilarities between each instance and all other sequences in the data set (the "background"). A signal that is lower than the background means that neighbor effects are not important; conversely, similar values of the signal and background hint at sizable neighbor effects. The left panel of Figure 5 shows that for the dendrimers under study, three domains can be distinguished. From positions 1 to 7, the average signal is clearly lower than the background and its spread is much reduced, meaning that it is probably safe to consider point mutations on an individual basis. On the other hand, at positions 8 and 10, neighbor effects seem to be important, and varying the sequence over two or more positions at a time is mandatory. Finally, neighbor effects abate again on average at position 11, although the spread in values remains high. The nature of the amino acid, on the other hand, does not seem to consistently affect neighbor effects: amino acid types for which neighbor mutations resulted in dissimilar sugar distance distributions were found to differ depending on their positions. Just like for the clustering discussed above, the position in the dendrimer sequence has the greatest impact on collective sequence effects. To check whether this effect is directional, the right panel of Figure 5 shows the means and standard deviations of the dissimilarities of sugar distance distributions generated when changing the left (trunk-side) or right (leaf-side) neighbor of an amino acid at a given position. The trunkward (left) part of the sequence, up to the end of the first branch (position 5), does not show significantly different right-side and left-side correlations. When moving toward the leaves, however, the average KS distance between distributions for right-side mutations becomes markedly higher than for left-side ones.

While it may indeed seem natural that the dynamics of the sugar residues is most influenced by the close-lying (leafward) amino acids, it is quite remarkable that the nature of "parent" branches in the dendrimer topology seems to have little effect on the behavior of their "children", even though the flow of correlation is expected to be directed from the dendrimer core to the branches.

The ability of a dendrimer of given sequence to multivalently bind LecA is very likely to depend on the extent of the overlap between its distribution of sugar distances and the distribution of LecA active-site distances. To obtain the latter, five LecA tetramers were simulated at the coarse-grained level in a solvent box corresponding to a concentration of 1.4 mM for a total of 6 μs, and all possible distances between active-site pairs were computed (a total of 190 per time step). As before, the overlaps between this distribution and each of the dendrimer sugar distance data sets were computed using the KS distance. The 20% of sequences with the best overlaps were used as input to the calculation of the per-position Shannon entropy; the resulting logo, presented in Figure 6, shows the amino acid

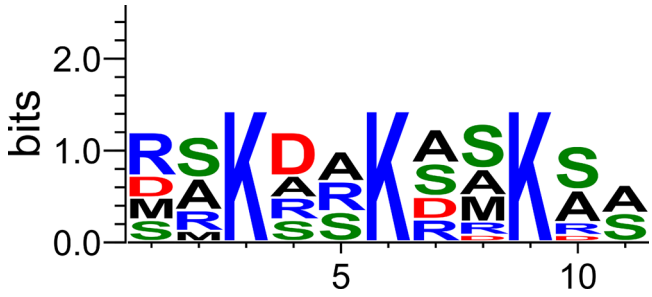


Figure 6. Putative dendrimer sequence preferences for the multivalent binding of LecA. The height of each letter is proportional to its statistical relevance in the set of sequences for which the overlap of ligand and receptor distance distributions is optimal. The sequence is ordered from trunk (left) to leaves (right).

preference at each position in the dendrimer sequence with respect to the multivalent binding of LecA (insofar as the hypothesis that the overlap of distance distributions between receptors and ligands correlates with multivalency is correct).

The sequence logo comforts the previous finding that the last-branch amino acids play a crucial role in the dendrimer dynamics; there is an almost equally high specificity for serine or alanine at the corresponding positions. However, the sequence of the first branch also appears to be important, with a marked preference for aspartate at position 4 and its absence at position 5 (along with methionine at both). The dendrimer trunk features a marked preference for arginine at position 1. The rest of the sequence is not as constrained but displays an overall preference for either serine or alanine. To check whether the bundling of proline within the group of hydrophobic amino acids despite its unique conformational rigidity is reasonable, proline was added to the previous set of five amino acids for the generation of all possible sequences of the last dendrimer branch (in which sequence effects have just been shown to be the strongest). The resulting sequence logo, presented in Figure S5 in the Supporting Information, shows that while the presence of a proline residue at the most leafward position is indeed favorable, it is not more so than alanine: for binding to LecA, the hydrophobic character of proline appears to prevail over its peculiar dynamical characteristics. As an aside, the fact that several experimentally validated dendrimers also feature leafward prolines[21,48,51] provides additional evidence of the predictive power of the model employed in this work.

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

To check whether the sugar distance properties inferred from simulations on a set of dendrimers mutated inside a two-amino acid sliding window (effectively neglecting correlations between second neighbors and beyond) could be used to predict the entire dendrimer sequence, the sequences with the best and worst overlaps with the distribution of LecA active sites (as predicted from the sequence logo in Figure 6) were simulated under the same conditions as previously. The best sequence was RSKDAKASKSA. Because there were multiple amino acids with zero occurrence at positions 5 and 11, several choices were possible for the worst sequence; ADKMMKMDKMM was chosen here. With a KS distance of 0.16 to the distribution of LecA binding-site distances, the best sequence was indeed found to be better than any other of the 175 previously simulated sequences, while the worst sequence (with a KS distance of 0.32) was found to be worse than 98.9% of the sequence set (the histogram of KS distances to the LecA active-site distance distribution for all dendrimer sequences under study can be found in Figure S3 in the Supporting Information). While this does not constitute proof that the best sequence is indeed the very best (which would require the entire set of 5^8 sequences to be simulated), it provides convincing evidence that potential collective effects do not reach far enough along the sequence, or are not important enough, to nullify the previously made hypothesis of the additivity of two-amino acid spans.

Finally, to check whether the overlap of distance distributions between receptors and ligands can be considered a good measure of multivalency and to evaluate the effect of the dendrimers on the dynamics of LecA tetramers, I have simulated the best and worst dendrimer sequences in the presence of LecA (five LecA tetramers and five dendrimer copies; see Methods for details)—possibly the most realistic simulation of multivalent lectin binding to date, albeit tempered by the simplicity of the coarse-grained model used and the relatively high LecA concentration chosen to reduce computational costs. First, the effect of the dendrimers on the distribution of distances between active sites on different LecA tetramers was monitored (Figure 7); because of the rigidity of the LecA tetramer assembly, the intratetramer active-site distances are the same in all cases and were disregarded). The distribution of distances observed in the absence of dendrimers is not very strongly affected by the addition of the best dendrimer candidate: apart from the appearance of small peaks from 40 to 70 Å, the overall distribution remains mostly devoid of salient features (narrow, well-defined peaks). The distributions are characteristic of systems with high mobility, where multiple interactions occur on comparatively short time scales (Figure 8 shows the structure of a typical multivalent complex encountered during the simulation). On the other hand, the addition of the worst dendrimer candidate to a solution of LecA tetramers profoundly affects the distribution of active-site distances, which now features well-defined peaks corresponding to long-lasting aggregates. The mobility of the LecA tetramers relative to one another is reduced by the addition of the dendrimer, which brings about a structuration of the system. This is not incompatible with multivalency—actually, long-lasting LecA−tetramer aggregates could conceivably facilitate the simultaneous binding of dendrimers to multiple tetramers. Conversely, high mobility and interactions over smaller time scales certainly promote multivalency via facilitated diffusion. The satisfactory convergence of the distributions with respect to simulation lengths was checked by running, for both dendrimers, an additional 6 μs of simulation from independent starting conditions and comparing the resulting distributions to the ones discussed here (see Figure S4 in the Supporting Information).

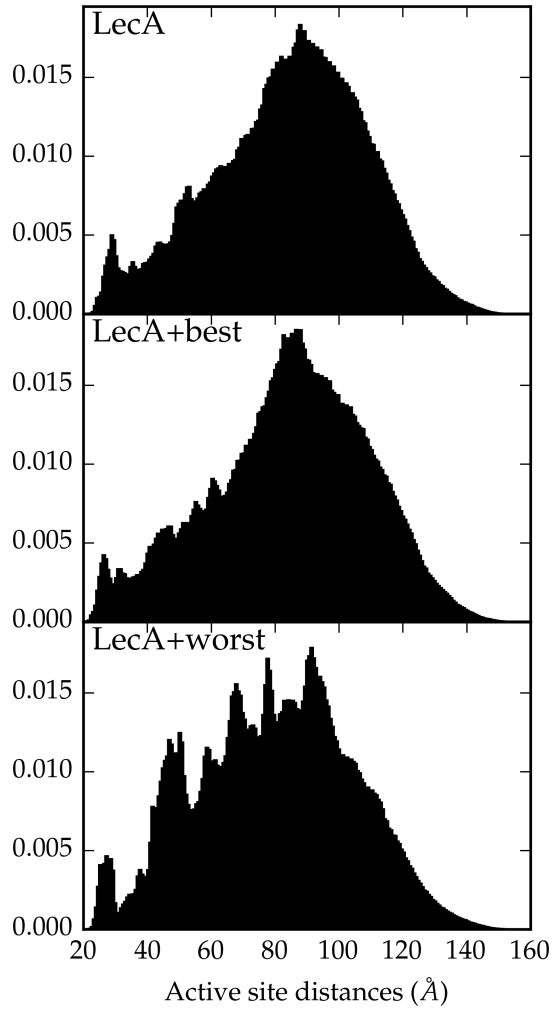


Figure 7. Distributions of distances between active sites on distinct LecA tetramers in the absence of dendrimers (top) and in the presence of the best (middle) and worst (bottom) dendrimer candidates.

To examine in more detail the occurrence of multivalent dendrimer−lectin interactions in both cases, the number of lectin active sites within a certain distance of all dendrimer sugar moieties was computed over the trajectories, and the results are plotted in Figure 9. For the best dendrimer candidate, a rapid onset of potential multivalency can be seen: there is a nonzero (if small) probability that all five simulated LecA tetramers are within distances as small as 7.5 to 10 Å of a galactose, and this probability increases linearly with distance. On the other hand, for the worst dendrimer a negligible probability of encountering four or five LecA tetramers is still observed even at distances as large as 22.5 to 25 Å; conversely, as a result of the greater ordering of LecA tetramers, the overall probability of encountering any number of LecA tetramers is consistently larger than for the best dendrimer candidate at distances larger than 5 Å.

DISCUSSION

Simulating multivalent lectin recognition and binding in silico is a difficult task. Dendrimers are flexible molecules with large

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

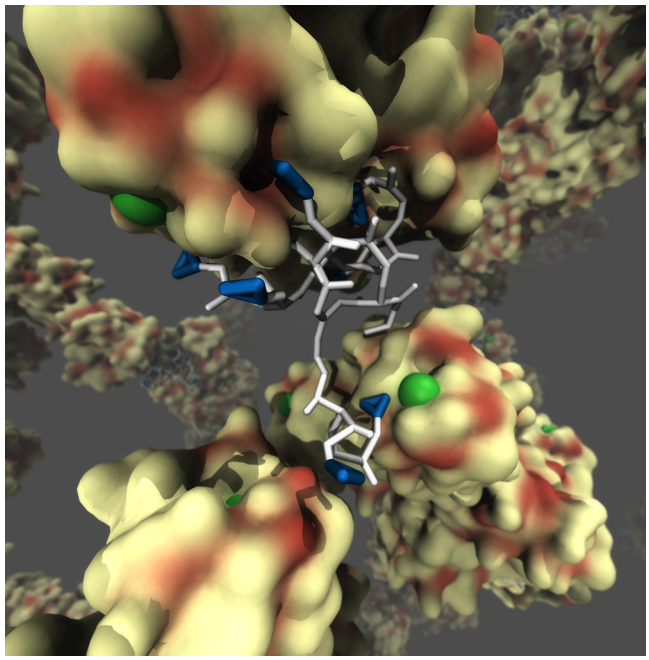


Figure 8. Coarse-grained structural representation of the "best" galactopeptide dendrimer (white sticks, peptidic arms; blue triangles, galactose) multivalently bound to three LecA tetramers (yellow/red surface), extracted from the corresponding molecular dynamics simulation. LecA active sites can be spotted by their Ca^{2+} ions (green spheres).

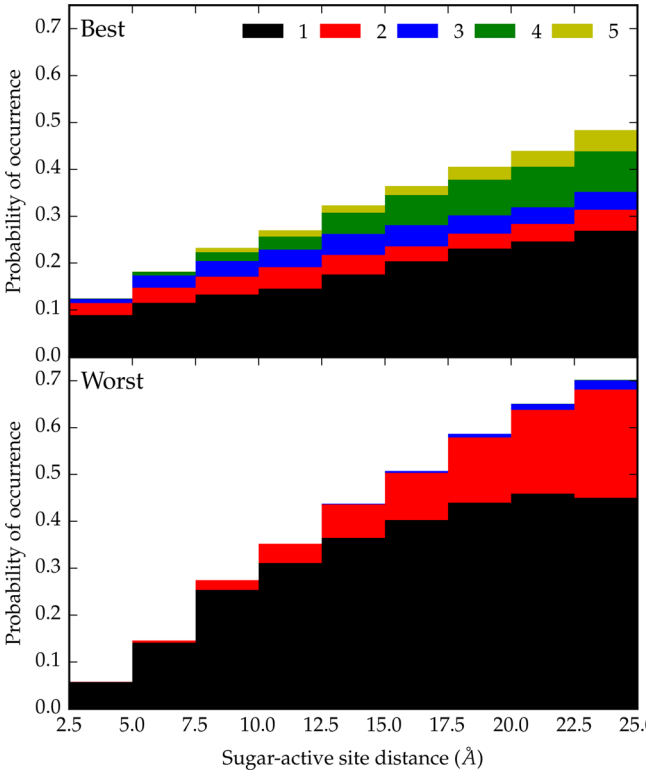


Figure 9. Probability of simultaneously finding one to five LecA tetramer active sites within a given distance of a dendrimer galactose for the best (top) and worst (bottom) dendrimer candidates in solution with LecA.

conformational spaces; encounters between dendrimers and lectins are diffusive in nature and as such occur over long time scales; the relative weakness of dendrimer−lectin interactions makes it difficult to distinguish relevant binding modes from random encounter complexes (which could also be yet-unidentified alternate binding sites, contributing to the complexity of the sugar code, which at the present time is still largely beyond our grasp[1]). Thus, previous studies have often simulated dendrimers and lectins on their own or focused on intratetramer (chelation) multivalency. Only very recently have all-atom simulations of inter-LecA multivalency been published,[51] although the time scales involved (on the order of 10 ns) preclude any major structural rearrangements or the possibility of observing multiple unbinding and rebinding events. By the use of a coarse-grained model that has been thoroughly validated both on proteins and carbohydrates, this study trades the precise, physics-based description of atomic interactions for the possibility of reaching meaningful time scales on large systems containing both partners and their aqueous environment (due in part to accelerated kinetics) and considering sequence effects over a relatively large set of dendrimer sequences. Despite this, numerous limitations still remain. For instance, the concentrations of lectins and dendrimers employed were deliberately overestimated compared with their typical experimental values in order to limit the dimension of the simulation cells and facilitate encounters; this particularly applies to the dendrimers, some of which have shown lectin-binding activity at stoichiometries as low as 1:20000 in experiments![52] Other potential issues are linked to the MARTINI framework: the relative overattractiveness of protein−protein interactions in MARTINI[53] could overemphasize the statistical relevance of long-lived aggregates observed in, e.g., the "worst" dendrimer candidate, and the conformational space of unstructured peptides may not be as faithfully rendered in MARTINI as in an all-atom model, despite a growing corpus of successful applications of MARTINI to unstructured antimicrobial peptides.[30−32] Since these issues should equally affect all dendrimer sequences, which can still be compared on a relative (if not absolute) basis, the trade-off was deemed worthy of adoption.

The first interesting point raised by this work is the nontrivial effect of the dendrimer amino acid sequence on the dynamical properties of the dendrimers, whether isolated or in solution with LecA. In agreement with previous atomistic simulations,[24] isolated dendrimers adopt different structure types (globular, extended, etc.), but with little apparent correlation between sequence and structure. Sequence matters somewhat more for the dynamics of the dendrimers; however, the choice of residues seems more important close to the galactose substituents than close to the dendrimer stem, where different sequences can yield indistinguishable dynamics. As a consequence, the sequences providing comparable distance distributions of galactose residues and LecA active sites show only a rather weak specificity for apolar amino acids toward the end of the dendrimer arms. This is reminiscent of experimental observations about the recognition of LecB by comparable peptide dendrimers, in which the degree of multivalency of the dendrimers was found to have more effect than the dendrimer sequence: despite an apparent preference for positively charged amino acids, hydrophobic and anionic sequences with comparable binding avidities were identified.[48] In view of these points, it is all the more interesting that the multivalent binding of dendrimer ligand functions to LecA active sites in a mixed dendrimer/lectin solution shows a marked sequence effect in the two cases studied here. In any case, the traditional

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

structure−activity relationship paradigm that has been a staple of in silico drug design for many years[54] should be replaced, for the dendrimers under study but also for other conformationally flexible and/or multivalent drugs, by the concept of a dynamics−activity relationship, in which molecular dynamics methods will naturally play a prominent role.

Multivalent binding comes in several flavors.[10] The most straightforward, chelation, is achieved by simultaneous binding of two or more binding sites on the same receptor with as many groups on the same ligand. A LecA tetramer features four binding sites, roughly arranged at the vertices of a rectangle. The small side of this rectangle measures 26 Å, which is close to the upper limit of distances between dendrimer galactose moieties observed in the simulations performed herein (see Figure S2 in the Supporting Information). Consequently, while this mode of multivalency cannot be ruled out in these particular systems, it is entropically disfavored. Although the sequence has an impact on the extension of the galactose distance distribution to larger values (e.g., arginine residues at the end of the arms promote extended conformations), this impact is too limited to provide room for improvement without modifying the dendrimer topology. This conclusion is in line with experimental findings on the binding of LecA and LecB by similar multivalent scaffolds;[15] however, it has very recently been experimentally demonstrated that four-generation peptidic dendrimer scaffolds (featuring four levels of branching and 16 sugar moieties), which could in theory span larger extents of space than the three-generation dendrimers discussed here, in fact showed reduced multivalent potency.[55] Although steric clashes between the lectin tetramers around each dendrimer, which the authors hold responsible for this phenomenon, are not due to chelation multivalency only, there appears to be little point in seeking to lengthen the dendrimer "arms".

Another multivalent mode, dubbed clustering, involves several receptors coming into close vicinity and being subsequently bound by a single multivalent ligand. This is likely what happens with the worst dendrimer candidate: it is able to modify the dynamics of the LecA tetramers, bringing them together to form long-lasting structural aggregates that it could then, in theory, multivalently bind. From the results presented herein, this approach seems only moderately effective: despite promising contact probabilities, the number of simultaneously contacted LecA tetramers appears to be limited to two at relevant distances. However, it has to be remembered that the so-called "worst" dendrimer was not optimized to achieve clustering multivalency: it was chosen within the pool of possible dendrimer sequences to maximize the difference between the distribution of galactose distances and the distribution of active sites in a solution of LecA tetramers. The modifications that the dendrimer triggered in the latter were not expected; consequently, other dendrimer sequences might have an even more pronounced effect, which could result in more effective clustering-type multivalency. The effect of the dendrimer on the dynamics of LecA is in line with the experimental finding that even a moderate dendrimer stoichiometry can significantly impact the global organization of lectins, while higher concentrations have an even more pronounced effect.[52] The dendrimer acts as a tensioactive molecule, decreasing lectin−lectin interactions and creating new structures by intercalating between lectins. It has recently been suggested that for the inhibition of *P. aeruginosa* biofilms, the capacity of a dendrimer to aggregate lectins might be more important than its capacity to bind individual lectins strongly.[51]

Considering the size of the $(LecA_4)_5$−$(dendrimer)_5$ system and the time scales involved, optimizing the dendrimer sequence to this effect using MD simulations will be very costly, even at the coarse-grained level.

The last instance of multivalent binding mechanisms is called statistical rebinding: a high concentration and/or high mobility of receptors and ligands allows multiple, short-lived interactions between the two species, in which contacts that break are quickly replaced by other equivalent ones. From the results presented above, the best dendrimer candidate optimizes this multivalent mechanism by mimicking the distribution of LecA active-site distances without affecting it. Dendrimers with high multivalency are particularly suited to this purpose; indeed, a sharp drop in efficiency was experimentally observed in going from octavalent to tetravalent or trivalent dendrimers of the same sequence in the case of LecB binding (with the latter not being significantly more active than fucose itself).[48] The importance of multivalency was similarly demonstrated on LecA.[56] Moreover, the statistical-rebinding mechanism explains the broad distributions of the number of effective partners in LecA−dendrimer complexes that have been observed both experimentally[57] and in the simulations presented here. As for clustering multivalency, there is no guarantee that the "best" dendrimer suggested in this work has the optimal sequence to maximize statistical rebinding; however, it does seem to result in a good trade-off between high LecA mobility and relevant interaction lifetimes.

In fact, much as the "worst" dendrimer candidate profoundly modifies the dynamics of a solution of LecA tetramers, transient interactions between the two species (at the active sites or otherwise) have a good chance of affecting the distribution of dendrimer galactose distances. Thus, searching for the optimal sequence on the basis of the distance distributions of the isolated dendrimers will only reveal part of the global picture. Since the "best" dendrimer candidate was shown not to affect the dynamics of the lectin active sites, a model akin to the kinetic model of protein−protein association can be considered:[58] the formation of the precomplex (a structure in which the ligand and receptor are aligned and primed for binding but still do not form strong interactions) is the obligatory prerequisite to the binding process, while noncognate contacts from random encounters between partners quickly dissociate. In this model, the dynamic features of the precomplex can reasonably be approximated from the dynamics of the isolated partners. This has been verified for the "best" dendrimer candidate, in which 92% of the noncognate contacts (defined as distances smaller than 10 Å between any atoms of the partners belonging neither to the galactose residues nor to the lectin active-site amino acids) last less than 200 ps, versus more than 5 ns for 73% of cognate contacts. Additionally, this model is even more likely to hold true at lower dendrimer concentrations; however, it might not apply to dendrimers that have a strong impact on the lectin dynamics. Therefore, while the dynamics of the isolated dendrimer is definitely a relevant indicator of multivalency (as the prediction of leafward prolines by the model, in agreement with experiments, indicates), it is only one part of a more complex equation. Unfortunately, the simulation of lectin/dendrimer mixtures over relevant time scales is much too costly to allow the screening of a large number of possible dendrimer sequences; in fact, the nature, occurrence, and impact of dendrimer−LecA interactions are likely to depend on the absolute and relative concentrations of the two species, which, as previously mentioned, have to be

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204

overestimated to bring down computational costs to acceptable levels.

Lectin-based design, which focuses on the structural and dynamical properties of lectins as the basis for the conception of potent ligands, has been the driving force behind the vast majority of studies in the field (including the present one).[12] However, the profound impact of the dendrimer sequence on the dynamics of LecA tetramers revealed in this work calls for a re-evaluation of this traditional assumption toward a more "ligand-centric" approach in which the ligands and lectins, rather than the lectins alone, are considered simultaneously from the start. This approach involves very large systems over extensive timeframes, pushing the boundaries of coarse-graining techniques and providing incentive for the development of robust methods combining a simplified representation of the systems with enhanced conformational sampling techniques.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.6b00146.

Validation of the branching lysine coarse-grained model against all-atom data; distributions of sugar distances for cluster exemplars; histogram of KS distances between sugar distance and LecA active-site distance distributions for all dendrimer sequences; convergence assessment of LecA active-site distance distributions; and validation of the inclusion of proline inside the group of hydrophobic amino acids (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: benjamin.bouvier@u-picardie.fr.

Funding

The author is employed by Centre National de la Recherche Scientifique.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

I thank Mohammed Benazza for fruitful discussions leading to this work. The calculations presented herein were performed using HPC resources from the MeCS computing platform of Université de Picardie - Jules Verne, Amiens, France.

■ REFERENCES

(1) Gabius, H.-J. J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D. From Lectin Structure to Functional Glycomics: Principles of the Sugar Code. *Trends Biochem. Sci.* **2011**, *36*, 298−313.

(2) Pinho, S. S.; Reis, C. A. Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nat. Rev. Cancer* **2015**, *15*, 540−555.

(3) Paleček, E.; Tkáč, J.; Bartošík, M.; Bertók, T. s.; Ostatná, V.; Paleček, J. Electrochemistry of Nonconjugated Proteins and Glycoproteins. Toward Sensors for Biomedicine and Glycomics. *Chem. Rev.* **2015**, *115*, 2045−2108.

(4) André, S.; Kaltner, H.; Manning, J. C.; Murphy, P. V.; Gabius, H.-J. J. Lectins: Getting Familiar with Translators of the Sugar Code. *Molecules* **2015**, *20*, 1788−1823.

(5) Buckeridge, M. S.; de Souza, A. P. Breaking the Glycomic Code of Cell Wall Polysaccharides May Improve Second-Generation Bioenergy Production from Biomass. *BioEnergy Res.* **2014**, *7*, 1065−1073.

(6) Mason, C. P.; Tarr, A. W. Human Lectins and their Roles in Viral Infections. *Molecules* **2015**, *20*, 2229−2271.

(7) Jo, E.-K. K. Mycobacterial Interaction with Innate Receptors: TLRs, C-Type Lectins, and NLRs. *Curr. Opin. Infect. Dis.* **2008**, *21*, 279−286.

(8) Jiménez Blanco, J. L.; Ortiz Mellet, C.; García Fernández, J. M. Multivalency in Heterogeneous Glycoenvironments: Hetero-Glycoclusters, -Glycopolymers and -Glycoassemblies. *Chem. Soc. Rev.* **2013**, *42*, 4518−4531.

(9) Lundquist, J. J.; Toone, E. J. The Cluster Glycoside Effect. *Chem. Rev.* **2002**, *102*, 555−578.

(10) Levine, P. M.; Carberry, T. P.; Holub, J. M.; Kirshenbaum, K. Crafting Precise Multivalent Architectures. *MedChemComm* **2013**, *4*, 493−509.

(11) Compain, P.; Bodlenner, A. The Multivalent Effect in Glycosidase Inhibition: a New, Rapidly Emerging Topic in Glycoscience. *ChemBioChem* **2014**, *15*, 1239−1251.

(12) Cecioni, S.; Imberty, A.; Vidal, S. Glycomimetics Versus Multivalent Glycoconjugates for the Design of High Affinity Lectin Ligands. *Chem. Rev.* **2015**, *115*, 525−561.

(13) Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E. Synthetic Multivalent Ligands in the Exploration of Cell-Surface Interactions. *Curr. Opin. Chem. Biol.* **2000**, *4*, 696−703.

(14) Wittmann, V.; Pieters, R. J. Bridging Lectin Binding Sites by Multivalent Carbohydrates. *Chem. Soc. Rev.* **2013**, *42*, 4492−4503.

(15) Bernardi, A.; et al. Multivalent Glycoconjugates as Anti-Pathogenic Agents. *Chem. Soc. Rev.* **2013**, *42*, 4709−4727.

(16) Giuliani, A.; Rinaldi, A. C. Beyond Natural Antimicrobial Peptides: Multimeric Peptides and Other Peptidomimetic Approaches. *Cell. Mol. Life Sci.* **2011**, *68*, 2255−2266.

(17) Sommer, P.; Fluxa, V. S.; Darbre, T.; Reymond, J.-L. L. Proteolysis of Peptide Dendrimers. *ChemBioChem* **2009**, *10*, 1527−1536.

(18) Stach, M.; Siriwardena, T. N.; Köhler, T.; van Delden, C.; Darbre, T.; Reymond, J.-L. L. Combining Topology and Sequence Design for the Discovery of Potent Antimicrobial Peptide Dendrimers Against Multidrug-Resistant Pseudomonas Aeruginosa. *Angew. Chem., Int. Ed.* **2014**, *53*, 12827−12831.

(19) Ravi, H. K.; Stach, M.; Soares, T. A.; Darbre, T.; Reymond, J.-L. L.; Cascella, M. Electrostatics and Flexibility Drive Membrane Recognition and Early Penetration by the Antimicrobial Peptide Dendrimer Bh1. *Chem. Commun. (Cambridge, U. K.)* **2013**, *49*, 8821−8823.

(20) Novoa, A.; Eierhoff, T.; Topin, J.; Varrot, A.; Barluenga, S.; Imberty, A.; Römer, W.; Winssinger, N. A LecA Ligand Identified from a Galactoside-Conjugate Array Inhibits Host Cell Invasion by Pseudomonas Aeruginosa. *Angew. Chem., Int. Ed.* **2014**, *53*, 8885−8889.

(21) Kadam, R. U.; Bergmann, M.; Hurley, M.; Garg, D.; Cacciarini, M.; Swiderska, M. A.; Nativi, C.; Sattler, M.; Smyth, A. R.; Williams, P.; Cámara, M.; Stocker, A.; Darbre, T.; Reymond, J.-L. L. A Glycopeptide Dendrimer Inhibitor of the Galactose-Specific Lectin LecA and of Pseudomonas Aeruginosa Biofilms. *Angew. Chem., Int. Ed.* **2011**, *50*, 10631−10635.

(22) Kilpatrick, D. C. Animal Lectins: a Historical Introduction and Overview. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, *1572*, 187−197.

(23) Mitchell, E.; Houles, C.; Sudakevitz, D.; Wimmerova, M.; Gautier, C.; Pérez, S.; Wu, A. M.; Gilboa-Garber, N.; Imberty, A. Structural Basis for Oligosaccharide-Mediated Adhesion of Pseudomonas Aeruginosa in the Lungs of Cystic Fibrosis Patients. *Nat. Struct. Biol.* **2002**, *9*, 918−921.

(24) Filipe, L. C. S.; Machuqueiro, M.; Baptista, A. M. Unfolding the Conformational Behavior of Peptide Dendrimers: Insights from Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 5042−5052.

(25) Filipe, L. C.; Machuqueiro, M.; Darbre, T.; Baptista, A. n. M. Unraveling the Conformational Determinants of Peptide Dendrimers Using Molecular Dynamics Simulations. *Macromolecules* **2013**, *46*, 9427−9436.

(26) Johansson, E. M.; Kolomiets, E.; Rosenau, F.; Jaeger, K.-E. E.; Darbre, T.; Reymond, J.-L. L. Combinatorial Variation of Branching Length and Multivalency in a Large (390625 Member) Glycopeptide Dendrimer Library: Ligands for Fucose-Specific Lectins. *New J. Chem.* **2007**, *31*, 1291−1299.
(27) Fritz, D.; Koschke, K.; Harmandaris, V. A.; van der Vegt, N. F. A.; Kremer, K. Multiscale Modeling of Soft Matter: Scaling of Dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10412−10420.
(28) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(29) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(30) Zhao, J.; Zhao, C.; Liang, G.; Zhang, M.; Zheng, J. Engineering Antimicrobial Peptides with Improved Antimicrobial and Hemolytic Activities. *J. Chem. Inf. Model.* **2013**, *53*, 3280−3296.
(31) Zhao, X.; Yu, H.; Yang, L.; Li, Q.; Huang, X. Simulating the Antimicrobial Mechanism of Human β-Defensin-3 with Coarse-Grained Molecular Dynamics. *J. Biomol. Struct. Dyn.* **2015**, *33*, 2522−2529.
(32) Horn, J. N.; Sengillo, J. D.; Lin, D.; Romo, T. D.; Grossfield, A. Characterization of a Potent Antimicrobial Lipopeptide via Coarse-Grained Molecular Dynamics. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 212−218.
(33) Cohen, F. E.; Sternberg, M. J. On the Prediction of Protein Structure: the Significance of the Root-Mean-Square Deviation. *J. Mol. Biol.* **1980**, *138*, 321−333.
(34) Yazici, B.; Yolacan, S. A Comparison of Various Tests of Normality. *J. Statist. Comput. Simulation* **2007**, *77*, 175−183.
(35) Frey, B. J.; Dueck, D. Clustering by passing Messages Between Data Points. *Science* **2007**, *315*, 972−976.
(36) Schneider, T. D.; Stephens, R. M. Sequence Logos: a New Way to Display Consensus Sequences. *Nucleic Acids Res.* **1990**, *18*, 6097−6100.
(37) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(38) Pall, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. *Solving Software Challenges for Exascale*; Springer: Cham, Switzerland, 2014; pp 3−27.
(39) López, C. A.; Rzepiela, A. J.; De Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.
(40) Fuhrmans, M.; Sanders, B. P.; Marrink, S.-J.; de Vries, A. H. Effects of Bundling on the Properties of the SPC Water Model. *Theor. Chem. Acc.* **2010**, *125*, 335−344.
(41) Cioci, G.; Mitchell, E. P.; Gautier, C.; Wimmerová, M.; Sudakevitz, D.; Pérez, S.; Gilboa-Garber, N.; Imberty, A. Structural Basis of Calcium and Galactose Recognition by the Lectin PA-IL of Pseudomonas Aeruginosa. *FEBS Lett.* **2003**, *555*, 297−301.
(42) Van Der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: a Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22−30.
(43) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(44) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(45) Crooks, G. E.; Hon, G.; Chandonia, J.-M. M.; Brenner, S. E. WebLogo: a Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188−1190.
(46) Huerta-Cepas, J.; Dopazo, J.; Gabaldón, T. ETE: a Python Environment for Tree Exploration. *BMC Bioinf.* **2010**, *11*, 24−30.
(47) Hunter, J. D. Matplotlib: a 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(48) Kolomiets, E.; Swiderska, M. A.; Kadam, R. U.; Johansson, E. M. V.; Jaeger, K.-E. E.; Darbre, T.; Reymond, J.-L. L. Glycopeptide Dendrimers with High Affinity for the Fucose-Binding Lectin LecB from Pseudomonas Aeruginosa. *ChemMedChem* **2009**, *4*, 562−569.
(49) Zamyatnin, A. A. Protein Volume in Solution. *Prog. Biophys. Mol. Biol.* **1972**, *24*, 107−123.
(50) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105−132.
(51) Visini, R.; et al. Structural Insight into Multivalent Galactoside Binding to Pseudomonas Aeruginosa Lectin LecA. *ACS Chem. Biol.* **2015**, *10*, 2455−2462.
(52) Sicard, D.; Chevolot, Y.; Souteyrand, E.; Imberty, A.; Vidal, S.; Phaner-Goutorbe, M. Molecular Arrangement Between Multivalent Glycocluster and Pseudomonas Aeruginosa LecA (PA-IL) by Atomic Force Microscopy: Influence of the Glycocluster Concentration. *J. Mol. Recognit.* **2013**, *26*, 694−699.
(53) Stark, A. C.; Andrews, C. T.; Elcock, A. H. Toward Optimized Potential Functions for Protein-Protein Interactions in Aqueous Solutions: Osmotic Second Virial Coefficient Calculations Using the MARTINI Coarse-Grained Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 4176−4185.
(54) Bohacek, R. S.; McMartin, C.; Guida, W. C. The Art and Practice of Structure-Based Drug Design: a Molecular Modeling Perspective. *Med. Res. Rev.* **1996**, *16*, 3−50.
(55) Bergmann, M.; Michaud, G.; Visini, R.; Jin, X.; Gillon, E.; Stocker, A.; Imberty, A.; Darbre, T.; Reymond, J.-L. L. Multivalency Effects on Pseudomonas Aeruginosa Biofilm Inhibition and Dispersal by Glycopeptide Dendrimers Targeting Lectin LecA. *Org. Biomol. Chem.* **2016**, *14*, 138−148.
(56) Gestwicki, J. E.; Strong, L. E.; Kiessling, L. L. Tuning Chemotactic Responses with Synthetic Multivalent Ligands. *Chem. Biol.* **2000**, *7*, 583−591.
(57) Mullen, D. G.; Fang, M.; Desai, A.; Baker, J. R.; Orr, B. G.; Banaszak Holl, M. M. A Quantitative Assessment of Nanoparticle-Ligand Distributions: Implications for Targeted Drug and Imaging Delivery in Dendrimer Conjugates. *ACS Nano* **2010**, *4*, 657−670.
(58) Schreiber, G.; Haran, G.; Zhou, H.-X. X. Fundamental Aspects of Protein-Protein Association Kinetics. *Chem. Rev.* **2009**, *109*, 839−860.

DOI: 10.1021/acs.jcim.6b00146
J. Chem. Inf. Model. 2016, 56, 1193−1204