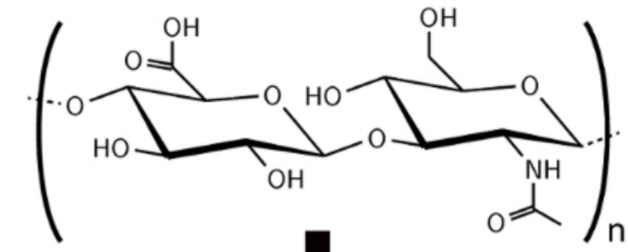
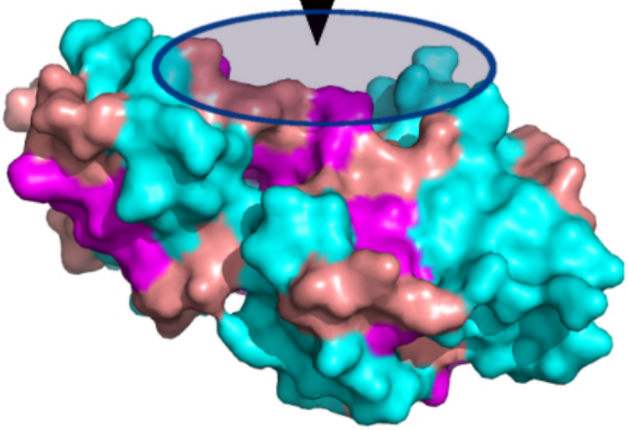
Docking software performance in protein-glycosaminoglycan systems

Urszula Uciechowska-Kaczmarzyk, Isaure Chauvot de Beauchêne, Sergey Samsonov



Docking



Dock

rDock

ClusPro

PLANTS

HADDOCK

Hex

SwissDock

ATTRACT

Docking software performance in protein-glycosaminoglycan systems

Urszula Uciechowska-Kaczmarzyk[1], Isaure Chauvot de Beauchene[2] and Sergey A. Samsonov[1*]

[1]Laboratory of Molecular Modeling, Department of Theoretical Chemistry, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdańsk, Poland.

[2] Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

*Corresponding author:
Sergey A. Samsonov
Faculty of Chemistry, University of Gdańsk,
ul. Wita Stwosza 63, 80-308 Gdańsk, Poland.
Tel.: +48 58 523 5166;
E-mail: sergey.samsonov@ug.edu.pl

Abstract

Glycosaminoglycans (GAGs) are a diverse group of linear anionic periodic polysaccharides that participate in many biological processes through the regulation of their protein partners activity. They are produced by many cell types and are found in the extracellular space as well as on cell surfaces, where they play an important role in mediation of cell-extracellular matrix interactions. Crystallization of protein-GAG complexes is difficult, therefore molecular docking can be a useful technique for predicting the binding conformation and understanding specific interactions in protein-GAG systems. At the same time, GAGs are also very challenging ligands for docking due to their high flexibility, periodicity and charged nature. Previously, we tested six different molecular docking software in terms of the performance on the protein-GAG complexes. In this study, we further performed docking simulations with other eight open access docking programs (Dock, rDock, ClusPro, PLANTS, HADDOCK, Hex, SwissDock and ATTRACT) for a dataset of 28 protein-GAG complexes with experimentally available structures, where a GAG ligand was longer than a trimer. Our results showed that Dock yielded the best prediction of a GAG binding pose, and its performance was independent of a GAG length. Overall, although the ligand binding poses could be correctly predicted in many cases by the tested docking programs, the ranks of the docking poses are often poorly assigned. Further comparison of the performance of fourteen docking programs, eight of which were analyzed in this study and six in the previous one, with the binding free energy components calculated for the corresponding experimental complexes allowed us to establish which binding free energy patterns define the success of each of these docking programs. Our work contributes to the evaluation of computational tools that could be used specifically to decipher protein-GAG interactions.

Keywords: molecular docking; protein-glycosaminoglycan interactions; modeling glycosaminoglycans; scoring function; binding pose.

Introduction

Glycosaminoglycans (GAGs), a class of long linear anionic periodic polysaccharides [1], play important roles in many crucial biological processes through their interactions with numerous protein partners [2, 3]. GAGs such as heparan sulfate (HS) are expressed by almost all types of cells and are anchored to the cell surface by covalent attachment to the membrane protein. They regulate cell proliferation, migration, differentiation, angiogenesis, axon guidance, response to CNS injury, viral invasion and immune response [3, 4]. GAGs can also exist as components of non-covalent macromolecular complexes in the extracellular matrix (ECM) [5]. Over the last two decades, a growing number of biological activities have been discovered to be regulated by the interaction of the heparin (HE) and HS with their proteins partners and so play important role in cancer, inflammatory process and infectious diseases [6]. GAGs have been shown to be also key players in a number of diseases associated with the pathologies of joints and bones [7]. Thus GAGs are very promising targets for the design of novel biomaterials to control and promote various biologically relevant processes for potential medical applications in field of bone and skin regeneration [8]. However, the experimental techniques alone are not sufficient to understand the complex biological processes mediated by GAGs and to gain insights into the interactions occurring at the molecular level. The crystallization of the protein-GAG complexes is challenging mostly due to the inhomogeneity of GAG fragments [9] and the nature of ionic interactions that may allow for multiple binding orientations [10]. Therefore, theoretical approaches beneficially contribute to the understanding of the role of GAGs interactions by bringing new and often experimentally inaccessible details [11,12]. Many computational methods have been frequently used for predicting the structures for specific GAG-complexes. Force fields and scoring functions were developed for carbohydrates and were proven to be useful for detailed characterization of the structure-function relationships and understanding the mechanism of protein-GAG interactions [13-16] . In particular, many protein-GAG systems were characterized by using computational approaches, which included docking GAGs to endostatin [17], IL-8 [18], BMPs [19, 20], SDF-1 [21, 22], cathepsins [23], lysil oxidase propeptide [24], VEGF [25]. Despite these successes, there are still many challenges for molecular docking studies of protein-GAG systems. The structural data of protein-GAG complexes in the Protein Data Bank (PDB) are limited – less than 100 complexes are available – which substantially restricts the dataset used in calibration and testing studies. The periodic nature of the GAGs, their linearity and functional groups disposition in relation to the reducing and non-reducing ends make it difficult to distinguish different binding poses when these molecules are bound at the protein

surface [26]. For this reason, multiple pose binding of GAGs was investigated in details for several systems [10, 27, 28]. Nevertheless, even in case the binding site of a GAG on a protein is known, it is still challenging to predict a binding pose properly by using docking approaches. This is due to the fact that: i) positively charged residues participating in GAG binding have long side-chains and, therefore, exploring their conformational space is crucial for predicting GAG binding [29], and the choice of the receptor structure used for docking could be decisive for the obtained binding poses [30], ii) protein-GAG complexes reveal poor geometric complementarity between the receptor and the ligand in their interfaces [31], iii) water molecules can play a crucial role for GAG binding and they should be therefore taken into account properly [32], iv) ring conformational space of GAG monomeric units could be crucial for docking results [33], v) there is still a lack of specific docking tools containing scoring schemes developed for GAGs. Recent evaluation of six widely-used docking programs in terms of their general applicability for protein-GAG systems for local docking performance showed that only a free docking software Autodock 3 and a commercial docking program Glide yielded relatively good results for those systems despite certain limitations [34]. Because of the named challenges in molecular docking for protein-GAG systems, implementation of experimental data-based restraints can be particularly beneficial for guiding docking approaches [35, 36]. Also, application of molecular dynamics-based protocols for docking GAGs was shown to be a promising approach since it allows to account for the full flexibility of these systems and to consider the effect of solvent explicitly [37].

In this work, several docking programs (Dock, rDock, ClusPro, PLANTS, HADDOCK, Hex, SwissDock and ATTRACT) have been evaluated by comparing their ability to generate and rank docking poses of a protein-GAG complex with the reference to the corresponding experimental structure. A dataset of 28 protein-GAG complexes were used where the GAG length was higher than dp3 (dp: degree of polymerization). Statistical analyses have been applied to the docking results in order to meaningfully differentiate the performance of these programs. Our study contributes to knowledge on general applicability of molecular docking tools that can be more effectively and specifically used for protein-GAG biologically relevant systems.

1. Materials and methods

Dataset

In this study, 28 protein-GAG complexes with a GAG longer than dp3 were retrieved from the PDB (Supplementary Table 1) to be used in the docking simulations with different software.

Water molecules were removed, protein receptor structure was optimized using ff99SB forcefield implemented in AMBER 16 [38]. Two energy-minimization steps in explicit TIP3P solvent octahedral box were carried out: 0.5×10^3 steepest descent cycles and 10^3 conjugate gradient cycles with harmonic force restraints on solute atoms, then 3×10^3 steepest descent cycles and 3×10^3 conjugate gradient cycles without any constraints.

Docking

Specific details on how each docking program was run are given in the following section. All docking programs were used in their default configuration with no turning on optional parameters, unless otherwise noted. For the ultimate evaluation of the obtained docking solutions, we used RMSatd (root mean squared atom type deviation) metric as the root-mean-square of pairwise atomic distances while pairing up the spatially closest atoms of the same type. Apart from being physically more sound than classical RMSD in terms of properly treating the equivalent atoms in chemical groups, this distance metric accounts for the periodicity in GAGs and considers two GAGs shifted by periodic units as structurally similar [37].

Dock

Dock 6 (version 6.1) uses an anchor-and-grow algorithm, and it is one of the first programs which involved shape complementarity through a set of spheres in the determination of the ligand-protein interactions [39]. The volume occupied by the ligand depends on the diameter of the spheres inside the binding pocket of the protein [40]. Receptors were protonated at pH 7.0 and assigned corresponding ff99SB parameters [37]. The types for the GAG atoms and hydrogens were added using Chimera [41]. Subsequently, AM1-BCC partial electrostatic charges were calculated for the GAGs using the Antechamber package distributed with AMBER [42, 43]. No additional optimization of the protein structure was carried out. The GRID program of Dock 6 was used to pre-calculate the potential grids [44]. All parameters were set to default parameters. For matching, the dms program was used to generate a molecular surface for each receptor. The SPHGEN package of Dock was used to create a negative image of the surface using spheres [45, 46]. We selected all spheres found within 10 Å of the corresponding ligand atoms from the crystal structure. The receptor box restricting the active site was calculated with the program SHOWBOX using the box length of 8 Å. All docking runs described in this section involved both rigid and flexible ligand docking

procedure. For each GAG-protein complex, 50 conformations were generated and considered for analysis.

rDock

rDock is a molecular docking program developed at Vernalis (http://www.vernalis.com) and initially designed for virtual screening of RNA targets. The rDock platform consists of command-line programs, scripts and uses flexible docking procedure. The main procedures are performed by the programs *rbcavity* (cavity generation) and *rbdock* (docking). The program contains intermolecular scoring functions (vdW, polar, desolvation terms) validated against protein and RNA targets [47]. rDock uses a combination of stochastic and deterministic search techniques to generate low energy ligand poses such as Genetic Algorithm search, low temperature Monte Carlo (MC) and Simplex minimization (MIN). rDock implements several pseudo-energy scoring functions that are added to the total scoring function under optimization and restricted search protocols (pharmacophoric restraints, tethered template). For docking studies, the receptor is provided in MOL2 format with standard amino acid atom types, the amino acids close to the binding pocket should be defined, as the rDock scoring function depends on formal charge assignments. The volume for placement of a ligand is defined by the *rbcavity* script that provides two mapping algorithms, the accessible volume within a specific distance of a reference ligand and a two probe spheres method [48]. For the protein-GAG complexes in this work, the reference GAG is used within a distance of 10 Å in order to define a binding pocket. The ligands were converted to SDF format (SDF) using *openbabel* software [49, 50]. For each GAG-protein complex, 100 poses were obtained.

ClusPro

ClusPro [51, 52] is a web-based server (https://cluspro.org) initially developed for protein-protein docking. The server performs three computational steps as follows: (1) rigid body docking by sampling billions of conformations, (2) root-mean-square deviation (RMSD) based clustering of the 1000 lowest energy structures generated to find the largest clusters, (3) refinement of selected structures using energy minimization. The rigid body-docking step uses PIPER [51] docking program based on the Fast Fourier Transform (FFT) correlation approach. This protocol was revealed to represent a good approximation for several classes of protein complexes [52]. In ClusPro, only parameters for HE are available, so this ligand was used for docking procedure for all complexes in the dataset. ClusPro docking server first

clustered 1,000 GAG positions with the lowest energy score within a 9 Å interfaces RMSD cut-off, and then ranked the best solutions based on the cluster centers scores. We applied the HE docking method that is available as an advance option of the ClusPro server. The numbers of poses obtained for further analysis was dependent on a particular protein-GAG complex.

PLANTS

The PLANTS program is based on an algorithm called Protein-Ligand ANT System, which is based on ant colony optimization (ACO) and uses flexible docking procedure [53]. An artificial colony is used to find a minimum energy conformation of the ligand in the binding site. The ACO algorithm is based on MAX-MIN Ant System (MMAS) [54]. The artificial ants construct solutions by choices based on the pheromone values and heuristic information, where one value corresponds to a single degree of freedom [53]. PLANTS treats the ligand as flexible, the flexibility of the protein is partially considered by the optimization of the positions of the hydrogen atoms involved in hydrogen bonding [53]. The search space with respect to the ligand's translational and rotational degrees of freedom is defined by the size and position of the binding site provided for each protein. [53]. The empirical scoring function used in PLANTS is a combination of intermolecular scores based on the piecewise linear potential scoring function [55] and hydrogen bonding interactions between both complex partners as published in GOLDS's CHEMSCORE implementation [56]. The intramolecular ligand scoring function consists of a simple clash term and torsional potential [57]. The clustering algorithm in PLANTS sorts all solutions according to increasing scoring function values. After that, the obtained structures form *rankingStructures* parameter, which is set to 10 by default. It is used to obtain the condition that minimal RMSD between extracted solutions is larger than 2 Å [58]. We used SPORES program (Structure PrOtonation and Recognition System) in order to obtain the correct format file for docking studies [59, 60]. For each GAG-complex the program generated 10 docking solutions.

HADDOCK

HADDOCK (High Ambiguity Driven Docking) [61, 62] web server was used to perform flexible docking simulations. The docking protocol of HADDOCK consists of three stages: (1) rigid-body docking by energy minimization from random orientations of the starting ligand conformations, (2) semi-flexible refinement of the interface by simulated annealing in torsion angle space, (3) short molecular dynamics refinement in explicit solvent. A

HADDOCK score is defined to rank the structures after each docking stage. It is a weighted sum of intermolecular electrostatic (Elec), van der Waals (vdW), desolvation (Dsolv) buried surface area (BSA), and optional user-defined restraints energies [61, 62]. The number of random starting structures for rigid docking and for semi-flexible docking with explicit solvent refinement was set to 10000 and 400, respectively. The RMSD was used as a clustering metric with a cutoff of 2.0 Å. The scoring parameters were set to 1.0 for Evdw1 and 0.1 for Eelec3 and the initial temperature for second and third cooling steps were 500 K and 300 K, respectively (default parameters for protein-ligand docking procedure). Different numbers of poses were obtained for different complexes.

Hex

HexServer (http://hexserver.loria.fr/) applies an SPF approach (Spherical Polar Fourier), utilizes rotational correlations [63], which reduces execution times to a matter of minute [64]. The Hex SPF algorithm has been validated for protein-protein docking in CAPRI (Critical Assessment of Predicted Interactions) blind docking experiment [65], and acceptable rigid-body predictions have been found within the top 100 orientations in CAPRI scoring sections [65]. The docking proceeds by rotating the receptor and ligand about their centroids at each of a range of intermolecular distances [63]. The correlation type was set to Shape Only and the FFT Mode (3D GAG). 10 poses were obtained for each complex.

SwissDock

SwissDock is a webserver-based software (http://www.swissdock.ch/docking) for predicting the molecular interactions between a target protein and a small molecule. It is based on the EADock DSS engine, combined with setup scripts used for protein and ligand preparation [66]. EADock dihedral space sampling (DSS) is built on the most efficient features of EADock2, namely its hybrid sampling engine and multiobjective scoring function [66]. The CHARMM force field based energies are determined on a grid, and the most favourable one are evaluated with FACTS and clustered, a rigid docking procedure is used.

ATTRACT

ATTRACT was primarily developed for coarse-grained protein-protein docking. It can also perform all-atom docking of proteins and nucleic acids [67] and more recently of any ligand by creating on-the-fly protein-ligand interaction parameters based on the ligand topology and non-bonded parameters (CNS or AMBER format) provided by the user. In this study, we

apply for the first time this procedure to protein-GAG systems, using charges, van der Waals parameters and atom types from the GLYCAM06 force field [68], and performing rigid docking by minimization with interface restraints. HADDOCK-like AIR restraints [69] were applied between all binding residues of the receptor defined as passive and all ligand residues defined as active. The receptor binding site was defined as all residues with at least 1 heavy atom at less than 5 Å from a ligand heavy atom in the crystal structure. The ligand was placed at 100,000 random starting positions and orientations around the receptor, the AIR restraints energy was first minimized alone in order to orientate the two molecules, then the interaction energy and AIR energies were minimized using precomputed energies stored on a receptor grid, followed by a scoring step without grids. Redundant poses (with RMSD < 0.1 Å to another pose) were removed, and the 100 best-scored poses were retained.

Data analysis and its graphical representation were carried out with the R package [70].

Results

In this study, 28 protein-GAG complexes available in the PDB with a GAG length longer than dp3 were used to evaluate the performance of 8 docking programs: Dock, rDock, ClusPro, PLANTS, HADDOCK, Hex, SwissDock and ATTRACT (**Table 1**). The results revealed that Dock and ATTRACT generated docking poses with the lowest $RMSatd_{top}$ and $RMSatd_{best}$ with the best values of 0.4 (for 1AXM) and 0.1 Å (for 2AXM and 3UAN), respectively (**S: Table 2, 9, 10)**. In case of the RMSatd for all obtained solutions, program Dock showed the best values with the range of 4.8-10 Å. The rDock and PLANTS programs predicted the GAG poses with the RMSatd values for all solutions within the ranges of 1.4-8.0 Å and 2.7-10.6 Å, respectively (**S: Table 3** and **S: Table 5)**. Significant differences between the docked poses and crystal structures were obtained by ClusPro where the range of the $RMSatd_{top}$ and $RMSatd_{best}$ is between 3.0 to 56.7 Å. In this case, HE was the docked ligand for all 28 complexes, because there no parameters for other GAGs than HE available within this program. As some of the analyzed complexes did not contain HE as a GAG ligand (2JCQ, 3C9E, 3H7D, 3UAN, 1E03), we removed them from the statistical analysis in ClusPro-HE subset in **Table 1.** However, the RMSatd values did not improve enough to state that the lack of the parameters for CS4 or HA could have influence on the overall performance of this program for our dataset.

HADDOCK and Hex yielded better results than ClusPro, $RMSatd_{top}$ and $RMSatd_{best}$ were in the ranges of 1.5 Å to 46 Å and 0.1 Å to 29 Å, respectively (**S: Table 4**, **S: Table 6** and **S:**

Table 7). This is especially surprising for Hex because this program, as it was used with the "Shape Only" parameter, searches for docking solutions based only on geometrical complementarity of the protein and ligand, which is in general expected to be a poor approach for electrostatically-driven protein-GAG complexes [63, 64]. SwissDock performed well for some complexes and worse for others, yielding the ranges of $RMSatd_{top}$ and $RMSatd_{best}$ of 3.9 Å to 46.1 Å and of 2.5 to 27.9 Å. ATTRACT was able to model some of the complexes very successfully, which is indicated by the value of $RMSatd_{best}$ lower than 1 Å for ten complexes, while it completely failed for others. Such a good performance of ATTRACT is at least partly due to the rigidity of the ligand kept throughout the docking process, while in other programs as HADDOCK, for example, flexible refinement was additionally performed. Therefore, ATTRACT results revealed an advantage of bound-bound docking, especially for long ligands. Hex also was used with rigid bound ligand, explaining its quite good results despite the above mentioned shape-complementarity. However, the RMSadt values obtained by Hex are not as low as the ones obtained with ATTRACT, because no additional minimization is performed. In general, Fast Fourier Transform programs are good at exhaustively sampling local minimum areas but not at optimizing the obtained solutions. ATTRACT produced docking solutions with $RMSatd_{top}$ and $RMSatd_{best}$ ranging from 0.1 Å to 38.6 Å and from 0.1 Å to 18.5 Å, respectively. All in all, numerous failures in GAG docking predictions were obtained when using ClusPro, HADDOCK, SwissDock (despite a rigid-docking procedure) and Hex programs.

In **S: Table 2-10**, we also summarize the Pearson and Spearman correlation coefficients ($r_{Pearson}$) and ($r_{Spearman}$) between the docking scores and RMSatd values for all obtained docking solutions. The higher correlation is the better is the performance of the respective scoring function. Based on this, the docking program with the best scoring function is SwissDock, which yielded $r_{Pearson}$ of 0.95 and $r_{Spearman}$ of 0.82 obtained for all solutions. The next two top-ranked scoring functions are the ones from Dock and rDock which yielded $r_{Pearson}$ and $r_{Spearman}$ of 0.76 and 0.84, 0.47 and 0.40, respectively. ATTRACT, HADDOCK and PLANTS yielded poor scoring performance despite producing binding poses of quite high quality. Also, we observed that there is an obvious difference between the top scored poses and the best poses for most docking programs except for Dock (**S: Table 2**). This is due to the fact that the good performance of a placement algorithm does not guarantee the good performance of the corresponding scoring function to rank properly the obtained binding poses. In particular, for ClusPro, rDock, HADDOCK, Hex and ATTRACT, the differences between the $RMSatd_{top}$ and $RMSatd_{best}$ values are quite significant for some GAG-complexes

(**Figure 1**). The RMSatd values for ClusPro program did not decrease significantly for the best poses in comparison to the top poses as it was the case for other programs suggesting its poor performance for both placement and scoring. The $RMSatd_{best}$ range for ClusPro is still up to 50 Å, whereas for other programs it significantly improved in comparison to $RMSatd_{top}$ (**Figure 1**).

The analyzed docking programs can be ranked by their performance in terms of $RMSatd_{top}$ and $RMSatd_{best}$ as following (from the best to the worst): Dock-r, Dock-f, rDock, PLANTS, ATTRACT, SwissDock, HADDOCK, ClusPro, Hex; Dock-r, Dock-f, rDock, ATTRACT, PLANTS, HADDOCK, SwissDock, ClusPro, Hex, respectively. According to the values summarized in **Table 1**, Dock, rDock, PLANTS can be useful for docking GAGs yielding acceptable results, while ATTRACT performed particularly well on some and failed on other complexes, and HADDOCK, SwissDock, ClusPro and Hex are rather not helpful when docking in these specific type of molecular systems. However, in practice, it is highly relevant to know how many of the obtained docking solutions are expected to be within a certain RMSatd range when compared to the experimental structure. We calculated the corresponding fractions of the obtained binding poses for RMSadt lower than 3 Å and 5 Å (**Table 2, S: Tables 11-18**). Based on this performance criteria, the docking programs were ranked in the following order (from the best to the worst): For $RMSatd_{top}$ < 3 Å: Dock-r, Dock-f ATTRACT, Hex, rDock, PLANTS=SwissDock=ClusPro= HADDOCK; For $RMSatd_{top}$ < 5 Å: Dock-r, Dock-f, ATTRACT, Hex, rDock PLANTS, SwissDock, ClusPro, HADDOCK; For $RMSatd_{best}$ < 3 Å: Dock-r, ATTRACT, Dock-f, rDock, Hex, HADDOCK, PLANTS, SwissDock, ClusPro, HADDOCK; For $RMSatd_{best}$ < 5 Å: rDock, Dock-f, ATTRACT, PLANTS, Dock-r, ClusPro, HADDOCK, Hex, SwissDock.

The success rate for Dock independently of the ligand flexibility for the top pose prediction is about 40% and 60% for RMSadt lower than 3 Å and 5 Å, respectively, which supports the previous conclusion that this program can be appropriate for the GAG docking. Other programs besides ATTRACT demonstrated relatively rather poor success rate in the prediction of the top pose. In terms of the placement ($RMSDadt_{best}$) quality, Dock, rDock, ATTRACT and Hex performed well.

For the longest GAG ligand in our dataset (HE dp10, 1E0O complex), the best predictions were obtained by ATTRACT with the $RMSatd_{top}$ and $RMSatd_{best}$ values of 0.3 Å, while Hex yielded the corresponding values of 3.0 Å and 0.5 Å. In comparison to this good performance, the Dock program produced solutions with $RMSatd_{best}$ and $RMSatd_{top}$ values of 7.2 Å and 3.6 Å, respectively (**S: Tables 2, 8 and 9)**. In case of the 3INA, complex containing HE dp8 and

HE dp6, the best results were obtained by Hex and Dock with both $RMSatd_{best}$ and $RMSatd_{top}$ of 0.3 Å and 0.4 Å, respectively. ATTRACT yielded $RMSatd_{best}$ and $RMSatd_{top}$ values of 18.5 Å and 24.7 Å, respectively. This is surprising when we take into account the otherwise good results obtained by this program for the longest GAG. This finding can be attributed to the sampling challenge for ATTRACT for this particular complex, which in comparison to all others complexes in the dataset has a cavity topology of the binding site: it makes more difficult to insert a ligand rigidly in it from a remote starting position by simple gradient minimization. The docking performance for 1AXM and 2AXM complexes, which differ only by the length of HE (dp5 and dp6, respectively), consistently increases for Dock, rDock, PLANTS, Hex and Swissdock for a longer HE molecule. At the same time, ATTRACT showed a slight improvement in both $RMSatd_{top}$ and $RMSatd_{best}$ from 0.3 Å to 0.1 Å upon the HE elongation in this system.

In order to further analyze how much the docking performance depends on the size of the GAG, Pearson and Spearman correlations were calculated between the $RMSatd_{top}$, $RMSatd_{best}$ and the corresponding length of the GAGs (**Table 1**). rDock, Plants, ClusPro and ATTRACT revealed the dependence of their performance on the length of the GAG ligand, whereas other programs performed independently of this factor.

To summarize, Dock, both for rigid and flexible ligand, performs best in terms of the pose placement, has the most suitable scoring function, and its performance did not depend on the ligand size. This suggests that the implementation of the electrostatic treatment as well as the shape complementarity procedure in Dock are the most suitable for docking the GAG ligands among the programs analyzed in this study. The challenges for the docking software observed could be related to GAGs high flexibility (all used GAG ligands in this study contain more than 10 rotatable bonds), which should be properly handled in the stage of placement, and their highly charged nature, which requires an appropriate treatment of electrostatics and, therefore, affects scoring.

Furthermore, we analyzed which free binding energy components in the complexes from the analyzed dataset, calculated previously by the application of the MD-based MM-GBSA free energy decomposition analysis for the corresponding experimental structures [34], are important for the performance of the 8 docking programs analyzed in this study and of the 6 docking from our previous study (Autodock 3 [71], Autodock Vina[72], MOE[73], FlexX[74], eHiTs[75], Glide[76]) [34]. For this, we calculated the Pearson correlations between the $RMSatd_{best}$ and $RMSatd_{top}$ values obtained for all these 14 programs and the

following free energy components: total free energy (ΔG_{total}), electrostatic component (ΔG_{ele}), van der Waals component (ΔG_{vdw}), full electrostatic energy (ΔG_{ele+gb}), total free energy (ΔG_{total}) (**Table 3**). Such analysis allows for a more detailed understanding of why some programs perform better or worse for complexes with different binding free energy patterns. Correlations obtained for $RMSatd_{best}$ and $RMSatd_{top}$ are not significantly different between each other for all programs suggesting that the energetic pattern of a complex has similar impact on both placement and scoring performance. Dock, which according to our analysis performed the best for this protein-GAG dataset in the actual study, also yielded the highest correlation coefficients: 0.47 and 0.59 for the van der Waals free energy component and 0.38 and 0.55 for the total energy for rigid and flexible docking, respectively. The correlation for the electrostatic component *in vacuo* is 1.5-2.0 times lower for both $RMSatd_{top}$ and $RMSatd_{best}$ than for the whole electrostatic component (0.30 and 0.48). This suggests that not only the electrostatic component but also van der Waals and solvent contributions to the complex binding energy define how well this program performs. For Glide, which also performed well in our previous study [34], the calculated correlations are still significant but lower, pointing in general to the potential ability of Glide to be more successful for low affinity complexes than Dock. The next top ranked docking tools after Dock were HADDOCK, Hex, SwissDock, which yielded significant correlations for all analyzed binding free energy components. This means that these three programs, which demonstrated medium and poor docking performance, perform better for complexes with stronger protein-ligand interactions. In particular, for Hex, which bases both its conformational sampling and scoring on geometry complementarity between a receptor and a ligand, such high dependence on the van der Waals component and the absence of the dependence on electrostatic component *in vacuo* are expected. All other programs (except for FlexX for $RMSadt_{best}$ for ΔG_{ele}) independently of their performance yielded insignificant correlations with free energy components suggesting that their performance is similar for the complexes with high or low affinities and is independent of the free energy pattern in a complex. Considering the fact that protein-GAG interactions are usually transient and relatively weak, this makes the usage of the programs with low correlations (Autodock 3, ATTRACT) more attractive for protein-GAG docking.

Conclusions

In this study, we analyzed the performance of 8 docking programs in terms of their docking pose sampling and its scoring on a dataset of 28 protein-GAG complexes. Our results suggest

that Dock program has advantages over other used programs, while ATTRACT performed very well for some and failed for other complexes. The performance of these two programs is also independent of the GAG length suggesting that their sampling and scoring functions are suitable for docking longer GAG molecules of known conformation. Furthermore, the analysis of the influence of protein-GAG binding free energy pattern on the performance of these and 6 other programs tested in our previous study showed that Dock performs better for the complexes with stronger binding, while Autodock 3 performs well for a wider variety of protein-GAG complexes. In practice, a combination of different molecular docking methods as well as with other modeling approaches such as molecular dynamic (MD) and binding free energy calculations may be a better strategy for predicting and evaluating GAG binding poses than using a single molecular docking approach alone. To sum up, we performed an updated docking benchmark comparison of the sampling and scoring power of different molecular docking approaches for protein-GAG complexes, and we believe that this work will serve as a good reference for the choice of a docking tool to be applied for this particular type of biologically relevant systems.

Acknowledgements

This work was supported by the National Science Center of Poland (Narodowy Centrum Nauki, grant UMO-2016/21/P/ST4/03995; This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665778) and BMN (Badania Młodych Naukowców) programme grant from the Department of Chemistry, University of Gdańsk. The computational resources were provided by cluster at the Faculty of Chemistry, University of Gdańsk.

References:

[1] J. Esko, D. Kimata, U. Lindahl, Essentials of Glycobiology, (2009), Cold Spring Harbor Laboratory Press.
[2] V. Veverka, A. Henry, P. Slocombe, A. Ventom, B. Mulloy, F. Muskett, M. Muzylak, K. Greenslade, A. Moore, L. Zhang, J. Gong, X. Qian, C. Paszty, R. Taylor, M. Robinson, Carr M., J. Biol. Chem. 284 (2009), 10890-10900.
[3]K. Sharma, M. Selzer, S. Li, Exp. Neurol. 237 (2012), 370-378.
[4] D. Xu, J. D. Esko, Annu. Rev. Biochem. 83, (2014), 129-157.
[5] F. Peysselon, S. Ricard-Blum, Matrix Biology, *35, (*2014), 73-81.
[6] S. Wang, X. Lu, P. Liu, Y. Fu, L. Jia, S. Zhan, Y. Lup, Mol. Cancer Ther., 14(2015), 1-9.
[7] C. M. Simonaro, M., D'Angelo, X. He, E. Eliyahu, N. Shtraizent, M. E. Haskins, E H. Schuchman, Am. J. Pathol. 172 (2008), 112-122.

[8] D. Scharnweber, L. Hubner, S. Rother, U. Hempel, U. Anderegg, S. A. Samsonov, M. T. Pisabarro, L. Hofbauer, M. Schnaberlauch, S. Franz, J. Simon, V. Hintze, JMSM, *26* (2015), 232-237.
[9] A. Imberty, H. Lortat-Jacob, S. Perez. Carb Res. 342 (2007), 430-439.
[10] P. R. B. Joseph, P. D. Mosier, U. R. Desai, K. Rajaratham, Biochem. J., 472 (2016), 131-133.
[12] A. Almond, Curr. Opin. Struct. Biol. 50 (2018), 58-64.
[11] N. Y. Sankaranarayanan, B. Nagarajan, U. R. Desai, Curr. Opin. Struct. Biol. 50 (2018), 91-100.
[13] A. Kerzmann, D. Neumann, O. Kohlbacher, J. Chem. Inf. Model. 46 (2006), 1635-1642.
[14] L. Foley, M. Tessier, R. Woods, R., Wiley Int. Rev. Comp. Model. Sci., 2(2012), 652-697.
[15] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. Daniels, R. R. Foley, R. J. Woods, J. Comp. Chem. 4 (2008), 622-655.
[16] A. Kerzmann, J. Fuhrmann, O. Kohlbacher, D. Neumann, J. Chem. Inf. Model. 48 (2008), 1616-1625.
[17] S. Ricard-Blum, O. Feraud, H. Lortat-Jacob, A. Rencurosi, N. Fukai, F. Dkhissi, A. Imberty, B. R. Olsen van der Rest M., J. Biol. Chem. 277 (2004), 33864-33869.
[18] N. Gandhi, R. Mancera, J. Chem. Inf. Model. 51 (2011), 335-358.
[19] N. Gandhi, R. Mancera, Acta BBA- Proteins Proteomics. 1824 (2012), 1374-1381.
[20] V. Hintze, S. A. Samsonov, M. Anselmi, S. Moeller, J. Becher, M. Schnaberlauch, D. Scharnweber, M. T. Pissabaro, Biomacromol. 15 (2014), 3083-3092.
[21] N. Sapay, E. Cabannes, M. Petitou, A. Imberty, Glycobiol. 21 (2011), 1181-1193.
[22] N. Panitz, S. Thesingen, S. A. Samsonov, J. P. Gehrcke, L. Bauman, K. Bellmann-Sickert, S. Kohling, M. T. Pisabarro, J. Rademann, D. Huster, A. G. Beck-Sickinger, Glycobiol. 26 (2016), 1209-1221.
[23] J. I. Sage, F. Mallevre, F. Barbarin-Costes, S. A. Samsonov, J. P. Gehrcke, M. T. Pisabarro, E. Perrier, S. Schnebert, A. Rogel, T. Livache, C. Nizard, G. Lalmanach, F. Lecaille , Biochem. 52 (2013), 6487-6498.
[24] Vallet
[25] U. Uciechowska-Kaczmarzyk, S. Babik, F. Zsila, K. K. Bojarski, T. Beke-Somfai, S. A. Samsonov, J. Mol. Graph. Mod. (2018). Epub.
[26]M. Forster, B. Mulloy, Biochem. Soc. Trans. 34 (2006), 431-434.
[27] Atkowska
[28] A. Singh, W. Kett, I. Severin, I. Agyekum, J. Duan, J. Amster, A. Proudfoot, D. Coombe, R. Woods, J. Biol. Chem. 290 (2015), 15421-15436.
[29] K. Mobious, K. Nordsieck, A. Pichert, S. A. Samsonov, L. Thomas, J. Schiller, S. Kalkhof, M. T. Pisabarro, A. Beck-Sickinger, D. Huster, Glycobiol. 23 (2013), 1260-1269.
[30] S. Babik, S. A. Samsonov, M. T. Pisabarro, J. Glycoconj. 34 (2017), 427-440.
[31] C. Taroni, S. Jones, J. Thornton, Protein Eng. 13 (2000), 89-98.
[32] S. A. Samsonov, J. Teyra, M. T. Pisabarro, J. Comp. Aid. Model. Des. 25 (2011), 477-489.
[33] S. A. Samsonov, M. T. Pisabarro, Carb. Res. 381 (2013), 133-137.
[34] S. A. Samsonov, M. T. Pisabarro, Glycobiol. 26 (2016), 850-857.
[35] N. Seyfried, J. Atwood, A. Yongye, A. Almond, A. Day, R. Orlando, R. Woods, Rap. Comp. Mass. Spec. 21 (2007), 121-131.
[36] T. Hofmann, S. A. Samsonov, A. Pichert, K. Lemmnitzer, J. Schillera, D. Huster, M. T. Pisabarro, von Bergen M. S. Kalkhof, Methods, 89 (2015), 45-53.
[37] S. A. Samsonov, J. P. Gehrcke, M. T. Pisabarro, J Chem Inf Model. 54 (2014), 582-92.
[38] D. A. Case, P. A. Kollman, et al. (2014), University of California, San Francisco.

[39] D. T. Moustakas, P. T., Lang, S. Pegg, E. Pettersen, I. D. Kuntz, N. Brooijmans, R. Rizzo, J. Comput. Aided Mol. 20 (2006), 601–619.
[40] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, J. Mol. Biol. 161 (1982), 269.
[41] E. F. Pettersen T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, J. Comp. Chem. 13 (2004), 1605-12.
[42] J. Wang, W. Wang, P. A. Kollman, D. A. Case, J. Mol. Graph. Mod. 25 (2006), 247-260.
[43] J. Wang, R. M. Wolf, J. Caldwell, P. A. Kollman, D. A. Case, J. Comp. Chem. 25 (2004), 1157-1174.
[44]R. L. DesJarlais, R. P. Sheridan, G. L. Seibel, J. S. Dixon, I. D. Kuntz, R. Venkataraghavan, J. Med. Chem. 31 (1988), 722-728.
[45] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin, J. Mol. Biol. 161 (1982) 269-275.
[46] M. Ester, H. Kriegel, J. Sander, X. Xu, AAAI Press. 226 (1996), 94-99.
[47] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A. Beatriz, S. J. Garmendia-Doval, D. Morley, PLOS Comp. Biol. 10 (2014) , 4-10.
[48] S. D. Morley, M. Afshar, J. Comput. Aided Mol. Des. 18 (2004), 189–208.
[49] M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. J. Chem. Inf. 3 (2011), 33-38.
[50] The Open Babel Package, version 2.3.1 http://openbabel.org (accessed Oct 2011).
[51] D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov, S. Vajda, Nat. Protoc. 12 (2017), 255-278.
[52] D. Kozakov, D. Beglov, T. Bohnuud, S. Mottarella, B. Xia, D. R. Hall, S. Vajda, Proteins. 13 (2013), 267-274.
[53] O. Korb, T. Stützle, T. Exner, Lecture Notes in Computer Science, 4150 (2006), 247-258.
[54] T. Stützle, T. E. Exner, Swarm Intell. 1 (2007), 115-134,
[55] D. K. Gehlhaar K .E. Moerder D. Zichi, S. J. Sherman, R. C. Ogden, S. T. Freer, J. Med. Chem. 38 (1995), 466-472.
[56] M. L. Verdonk, J. C. Cole, M. J. Harshorn, C. W. Murray, R. D. Taylor. Proteins. 53 (2003), 609-623.
[57] M. Clark, R.D. Cramer III, J. Comput. Chem. 10 (1989), 982-988.
[58] M. Dorigo, M. Birattari, T. Stutzle, IEEE Comp. Intell. Mag. 1 (2006), 4-6.
[59] T. Brink, T. E. Exner, J. Chem. Inf. Model. 49 (2009), 1535-1546.
[60] T. Brink, T. E. Exner, J. Comput. Aided Mol. Des. 24 (2010), 935-942.
[61] C. Dominguez, R. Boelens, A. M. J. J. Bonvin, J. Am. Chem. Sec. 125 (2003), 1731-1737.
[62] G. C. P. Zundert, J. P. G. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, S. J. de Vries and A. M. J. J. Bonvin. J. Mol. Biol. 428 (2016), 720-725.
[63] D. W. Ritchie, V. Venkatraman, Bioinformatics, 26 (2010), 2398-2405.
[64] D. W. Ritchie, D. Kozakov, S. Vajda. Bioinformatics, 24 (2008), 1865-1873.
[65] A. W. Ghoorah, M. Smail-Tabbone, M. D. Devignes, D. W. Ritchie. Proteins. (2013), 2150-2158.
[66] A. Grosdidier, V. Zoete, O. Michielin, J. Comput. Chem. 32 (2011), 2149-2159.
[67] S. J. de Vries C. E. Schindler, I. Chauvot de Beauchêne, M. Zacharias. Biophysical journal, 108 (2015), 462-465.
[68] K. Kirschner, A. Yongye, S. Tschampel, J. González-Outeiriño, C. Daniels, L. Foley, R. Woods. J. Comput. Chem. 29 (2008), 622-655.
[69] C. Dominguez, R. Boelens, A. M. Bonvin, J. Am. Chem. Soc. 19 (2003), 1731-1737.
[70] R Development Core Team. 2006. A language and environment for statistical computing.
[71] G. Morris, D. Goodsell, R. Halliday, R. Huey, W. Hart, R. Belew, A. Olson. J. Comput.

Chem. 19 (1999), 1639-1644.
[72] L. Thompson, M. Pantoliano, B. Springer, Biochemistry, 33 (1994), 3831-3840.
[73] Chemical Computing Group Inc M.C. MOE v2005.06.:2006–2006.
[74] R Development Core Team. 2006. R. A language and environment for statistical computing.
[75] X. Xiong, Z. Chen, B. Cossins, Z. Xu, Q. Shao, K. Ding, W. Zhu, W. Shi, J. Carb. Res. 401 (2015), 73-81.
[76] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin. J. Med. Chem. 47 (2004), 1739-1749.
[77] P. Wessa, 2017, Cronbach alpha (v1.0.5) in Free Statistics Software (v1.2.1), Office for Research Development and Education, URL https://www.wessa.net/rwasp_cronbach.wasp/

Program	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	$RMSatd_{all}$ (Å)	$r_{Pearson\ top}$	$r_{Spearman\ top}$	$r_{Pearson\ best}$	$r_{Spearman\ best}$
Dock-r	3.5±2.9	2.0±1.5	7.0±1.4	0.15	0.06	0.14	0.06
Dock-f	4.5±3.8	2.5±2.1	7.2±1.2	-0.19	-0.17	-0.03	-0.13
rDock	5.2±1.6	3.0±0.6	5.7±0.9	0.23	0.23	0.47*	0.49*
ClusPro	17.5±12.9	9.3±10.6	17.5±11.7	0.13	0.12	0.08	0.40*
ClusPro HE	17.8±13.6	8.9±11.1	17.7±12.2	0.08	0.37	0.14	0.15
PLANTS	6.4±2.2	4.4±1.7	5.8±1.7	0.55*	0.52*	0.60*	0.55*
HADDOCK	15.4±9.8	7.8±5.0	17.0±7.7	0.06	0.14	0.17	0.24
Hex	18.8±12.1	9.1±7.5	19.2±9.5	-0.01	-0.05	-0.21	-0.27
SwissDock	14.4±12.3	8.5±7.1	21.3±9.7	-0.15	-0.19	-0.12	-0.19
ATTRACT	13.1±11.0	3.4±4.4	20.3±7.4	-0.02	-0.02	0.14	-0.02

Table 1. Performance of docking programs for protein-GAG complexes. Dock-r and Dock-f represent rigid and flexible docking procedure, respectively. $RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure, $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure, $RMSatd_{All}$: the structural difference of the all obtained docking poses to the ligand structure, $r_{Pearson\ top}$, $r_{Spearman\ top}$, $r_{Pearson\ best}$, $r_{Spearman\ best}$: correlation coefficients for the RMSatd of Top and Best poses to the length of the GAG, respectively. *The statistical significance for the dataset of 28 cases corresponding to the values of Pearson and Spearman coefficients with p-values < 0.05.

Program	$RMSadt_{top}$<3 Å (%)	$RMSadt_{top}$<5 Å (%)	$RMSadt_{best}$<3 Å (%)	$RMSadt_{best}$<5 Å (%)
Dock-r	46	64	71	64
Dock-f	43	61	61	93
rDock	11	42	60	100
ClusPro	0	7	0	53
PLANTS	0	25	11	86
HADDOCK	0	7	14	42
Hex	14	27	28	35
SwissDock	0	25	11	28
ATTRACT	32	32	53	67

Table 2. The number of top and best poses obtained for the $RMSadt_{top}$ and $RMSadt_{best}$ lower than 3 Å and 5 Å, respectively.

Program	$r^{\Delta Gtotal}_{top}$	$r^{\Delta Gele}_{top}$	$r^{\Delta GvdW}_{top}$	$r^{\Delta Gele+gb}_{top}$	$r^{\Delta Gtotal}_{best}$	$r^{\Delta Gele}_{best}$	$r^{\Delta GvdW}_{best}$	$r^{\Delta Gele+gb}_{best}$
Dock-r	0.38*	0.19	0.47*	0.30*	0.38*	0.20	0.47*	0.30*
Dock-f	0.55*	0.39*	0.59*	0.48*	0.52*	0.37*	0.55*	0.48*
rDock	-0.19	-0.11	-0.22	-0.16	-0.17	-0.19	-0.08	-0.21
ClusPro	-0.11	-0.36	-0.03	-0.16	-0.46	-0.62	-0.37	-0.48
PLANTS	-0.41	-0.28	-0.42	-0.38	-0.38	-0.16	-0.46	-0.32
HADDOCK	0.37*	0.32*	0.32*	0.39*	0.29	0.23	0.26	0.28
Hex	0.26	0.07	0.36*	0.19	0.29	0.10	0.38*	0.22
SwissDock	0.30*	0.12	0.38*	0.24	0.12	-0.02	0.22	0.04
ATTRACT	0.03	-0.07	-0.05	0.08	-0.14	-0.09	-0.25	-0.06
AD3	-0.06	0.03	-0.20	0.02	-0.25	-0.10	-0.40	-0.14
Autodock Vina	0.11	0.008	0.25	0.01	0.16	0.02	0.25	0.09
MOE	0.04	-0.006	0.03	0.04	-0.004	-0.03	0.09	-0.07
eHITS	-0.40	-0.42	-0.26	-0.44	-0.04	-0.34	-0.07	-0.11
FlexX	-0.12	0.25	-0.40	0.05	-0.05	0.31*	-0.32	0.10
Glide	0.19	0.22	0.17	0.19	0.13	0.16	0.12	0.14

Table 3. Pearson correlation coefficients between the values of $RMSatd_{best}$, $RMSatd_{top}$ obtained by the 8 docking programs in this study and 6 docking programs in the previous study [34], and the binding free energy components of the corresponding experimental complexes analyzed by the MM-GBSA free energy decomposition approach: ΔG_{total}: total energy, ΔG_{ele},: electrostatic component *in vacuo*, ΔG_{vdw}: van der Waals component, ΔG_{ele+gb}: full electrostatic component in solvent that is a sum of electrostatic component *in vacuo* and Generalized Borne reaction field. Dock-r and Dock-f represent rigid and flexible docking procedure, respectively. *The statistical significance for the dataset of 28 cases corresponding to the value of Pearson coefficients with p-values < 0.05.

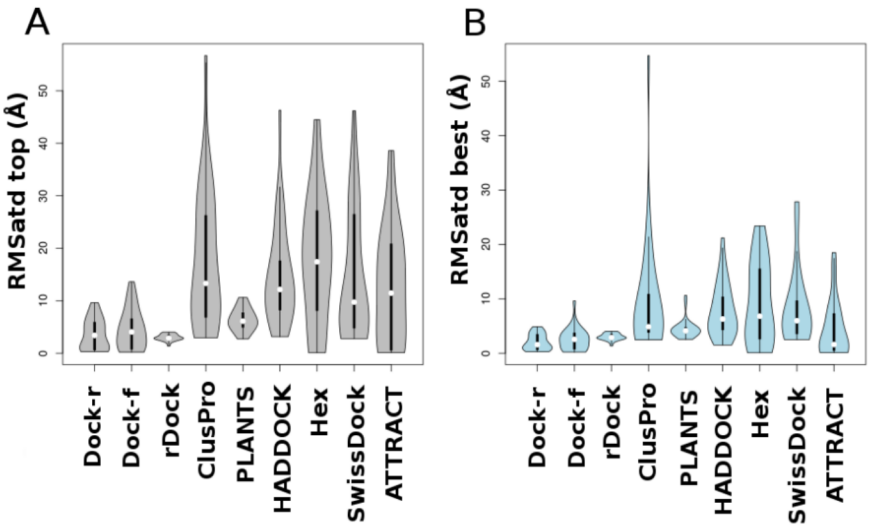


Figure 1. Violin plots of (A) $RMSatd_{top}$ and (B) $RMSatd_{best}$ values obtained with the analyzed docking programs on the whole protein-GAG complex dataset: Dock-r (rigid ligand in Dock), Dock-f (flexible ligand in Dock), rDock, ClusPro, PLANTS, HADDOCK, Hex, SwissDock and ATTRACT.

PDB ID	Resolution (Å)	GAG	Length (dp)	Protein
1AXM	3.00	HE	5	FGF-1
1BFB	1.90	HE	4	FGF-2
1BFC	2.20	HE	6	FGF-2
1E03	2.90	HE-like	5	Anthithrombin-III
1E0O	2.80	HE	10	FGF-1/FGFR-2
1FQ9_1	3.00	HE	8	FGF-2/FGFR-1
1FQ9_2	3.00	HE	6	FGF-2/FGFR-1
1G5N	1.90	HE	4	Annexin V
1GMN	2.30	HE	5	NK1 (HGF)
1QQP	1.90	HE	5	FMDV-receptor
1RID	2.10	HE	8	VCP
1T8U	1.95	HE	4	3-O-Sulfotransferase
1XMN	1.85	HE	6	Thrombin
1XT3	2.40	HE	6	Cobra Cardiotoxin A3
2AXM	3.00	HE	6	FGF-1
2HYU	1.42	HE	4	Annexin 2A
2HYV	1.42	HE	5	Annexin 2A
2JCQ	1.25	HA	7	CD44
2LVZ	NMR	HE	3	ECP
3C9E	1.80	CS4	6	Cathepsin K
3DY0	1.55	HE	5	Protein C inhibitor
3H7D	2.24	CS4	6	Cathepsin K mutant
3INA	1.90	HE	8	Heparinase I
3MPK	2.81	HE	6	VFT2
3QMK	2.21	HE	4	E2 domain of APLP1
3UAN	1.84	HS	6	3-O-sulfotransferase
4AK2	1.35	HE	6	BT4661, a Suse-like surface located polysaccharide binding protein
4C4N	2.36	HE	6	Hedgehog morphogen

S: Table 1. Protein-GAG complex dataset used in this docking study.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	0.4	0.4	50	4.8±1.5	0.38	0.30
1BFB	3.0	1.2	50	5.3±2.3	0.45	0.51
1BFC	5.9	3.5	50	7.4±1.9	0.30	0.36
1E03	0.6	0.6	50	6.2±2.0	0.13	0.33
1E0O	7.2	3.6	50	10.1±4.1	0.30	0.53
1FQ9_1	0.9	0.9	50	7.9±2.7	0.30	0.34
1FQ9_2	0.6	0.6	50	7.3±3.01	0.45	0.40
1G5N	6.0	4.0	50	10.2±3.0	0.25	0.79
1GMN	1.3	1.3	50	6.2±1.9	0.28	0.19
1QQP	6.1	4.2	50	6.7±1.6	0.02	0.00
1RID	9.6	4.6	50	7.8±1.5	-0.13	-0.17
1T8U	0.3	0.3	50	5.8±1.7	0.70	0.68
1XMN	0.5	0.5	50	6.3±2.2	0.37	0.59
1XT3	8.3	4.9	50	9.6±2.1	0.20	0.24
2AXM	0.6	0.6	50	6.6±2.9	0.76	0.84
2HYU	4.1	1.7	50	5.9±1.7	0.46	0.47
2HYV	6.6	2.9	50	6.3±1.7	0.37	0.40
2JCQ	1.4	1.4	50	7.3±2.5	-0.14	0.01
2LVZ	3.9	1.8	50	4.2±2.1	0.17	0.23
3C9E	1.1	1.1	50	7.2±2.9	0.12	0.30
3DY0	5.6	3.5	50	6.5±1.6	0.18	0.26
3H7D	5.5	2.2	50	8.0±2.9	-0.08	-0.03
3INA	0.4	0.4	50	7.4±2.7	-0.24	-0.32
3MPK	8.9	3.9	50	7.5±1.9	0.30	0.49
3QMK	1.4	1.4	50	6.0±3.1	0.16	0.50
3UAN	3.8	2.7	50	8.0±3.4	0.23	0.38
4AK2	0.6	0.6	50	6.6±2.2	0.16	0.45
4C4N	4.0	2.5	50	7.0±3.5	0.41	0.75

S: Table 2. Results obtained by the Dock software for protein-GAG complexes using rigid docking procedure. $RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	2.9	2.9	100	4.9±1.5	0.42	0.40
1BFB	4.2	2.6	100	4.5±1.3	0.47	0.40
1BFC	3.2	2.9	100	5.3±1.6	0.12	0.16
1E03	3.8	2.6	100	5.6±1.6	0.23	0.28
1E0O	6.2	4.0	100	6.4±1.2	0.26	0.38
1FQ9_1	6.4	2.9	100	5.4±1.7	0.04	-0.09
1FQ9_2	5.6	3.6	100	7.2±1.9	0.18	0.14
1G5N	5.7	4.0	100	6.5±1.8	-0.34	-0.33
1GMN	3.4	2.9	100	5.4±1.5	0.24	0.26
1QQP	8.0	2.6	100	5.3±1.4	0.03	-0.01
1RID	7.6	3.9	100	6.1±1.7	0.18	0.20
1T8U	6.1	2.6	100	5.2±1.5	0.19	0.20
1XMN	4.5	2.8	100	5.4±1.6	0.27	0.27
1XT3	7.6	3.6	100	5.5±1.3	0.07	0.09
2AXM	6.8	4.0	100	5.0±1.4	0.45	0.44
2HYU	3.5	2.8	100	5.0±1.4	0.33	0.31
2HYV	2.9	2.9	100	5.7±1.6	0.17	0.20
2JCQ	6.1	3.3	100	7.2±2.3	0.14	0.12
2LVZ	7.0	2.2	100	5.1±1.3	0.12	0.14
3C9E	4.0	3.5	100	5.5±1.4	0.12	0.14
3DY0	6.2	3.7	100	5.6±1.2	-0.14	-0.18
3H7D	2.8	2.8	100	5.2±1.4	0.07	0.12
3INA	7.0	2.8	100	8.9±2.4	0.09	0.09
3MPK	5.8	2.1	100	6.2±2.0	0.07	0.03
3QMK	6.0	1.4	100	6.2±2.0	0.42	0.42
3UAN	3.8	3.5	100	4.9±1.4	0.26	0.24
4AK2	6.4	2.9	100	5.2±1.7	0.22	0.26
4C4N	3.6	3.0	100	5.3±1.6	0.24	0.23

S: Table 3. Results obtained by the rDock software for protein-GAG complexes.
$RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	5.9	3.6	4	6.3±1.8	0.24	0.40
1BFB	5.9	3.4	8	4.7±0.9	0.38	-0.11
1BFC	8.0	3.9	10	6.5±1.7	0.35	0.11
1E03	7.2	6.3	7	7.2±0.7	-0.06	0.07
1E0O	21.2	6.5	9	14.3±6.7	0.16	0.08
1FQ9_1	14.4	14.4	4	16.6±2.5	0.18	0.20
1FQ9_2	56.7	54.7	4	45.8±10.0	-0.01	0.40
1G5N	32.5	4.5	30	26.2±14	-0.10	-0.08
1GMN	29.8	3.0	8	23.0±11.5	-0.74	-0.56
1QQP	34.0	10.8	21	33.5±14.0	0.44	0.25
1RID	29.7	4.9	30	47.0±31	-0.12	-0.11
1T8U	9.5	2.5	7	4.6±1.7	-0.11	-0.28
1XMN	10.9	10.9	4	11.5±1.0	0.21	0.00
1XT3	25.8	24.8	4	25.8±0.9	-0.35	-0.80
2AXM	4.0	4.0	4	4.4±0.3	0.41	0.50
2HYU	21.7	16.8	14	26.6±6.2	-0.20	-0.02
2HYV	34.9	7.5	15	27.3±8.5	0.30	0.26
2JCQ	12.4	6.4	15	10.4±2.5	-0.10	-0.06
2LVZ	3.0	3.0	20	15.1±9.4	0.34	0.28
3C9E	21.9	18.9	7	23.3±2.7	-0.15	-0.53
3DY0	11.2	8.9	14	18.1±9.7	-0.21	-0.13
3H7D	24.8	18.1	10	25.0±4.4	-0.24	-0.34
3INA	5.8	4.6	7	6.5±1.0	0.65	0.78
3MPK	27.5	4.3	12	18.9±10.2	0.13	0.27
3QMK	5.3	3.3	10	9.4±6.8	-0.35	-0.35
3UAN	4.8	4.8	6	6.1±1.7	-0.09	0.25
4AK2	8.5	3.4	16	9.8±6.8	0.68	0.50
4C4N	14.2	4.6	10	16.6±12.7	0.10	0.09

S: Table 4. Results obtained by the ClusPro software for protein-GAG complexes.
$RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	4.7	3.8	10	4.6±0.3	0.05	-0.02
1BFB	5.5	3.8	10	4.9±1.1	-0.40	-0.30
1BFC	7.0	3.0	10	4.9±1.4	-0.35	-0.33
1E03	7.6	3.7	10	6.4±1.5	-0.26	0.01
1E0O	9.7	9.4	10	10.6±1.7	0.61	0.38
1FQ9_1	8.9	4.4	10	6.5±1.8	-0.50	-0.34
1FQ9_2	5.7	4.0	10	4.9±0.6	-0.48	-0.58
1G5N	3.7	3.7	10	4.7±1.5	0.00	-0.15
1GMN	5.5	4.9	10	6.0±0.8	-0.02	-0.05
1QQP	6.5	4.9	10	5.8±1.2	0.02	-0.06
1RID	8.3	3.5	10	6.5±2.1	-0.38	-0.28
1T8U	7.4	4.7	10	6.0±0.8	-0.17	0.07
1XMN	8.9	2.9	10	6.7±1.7	0.22	0.20
1XT3	7.4	4.6	10	6.1±1.0	-0.51	-0.53
2AXM	5.7	4.4	10	5.4±0.5	-0.24	-0.17
2HYU	4.3	3.8	10	4.3±0.6	-0.02	-0.11
2HYV	3.1	2.6	10	4.2±0.8	0.01	-0.15
2JCQ	3.2	3.2	10	5.2±1.4	0.51	0.50
2LVZ	5.8	3.8	10	5.0±1.0	0.53	0.40
3C9E	6.7	4.3	10	5.8±0.9	0.01	0.10
3DY0	4.8	4.3	10	5.5±0.8	-0.17	-0.22
3H7D	8.1	3.8	10	5.9±1.3	-0.15	0.03
3INA	10.6	10.6	10	11.8±0.9	0.03	-0.01
3MPK	10.6	4.9	10	7.18±1.8	-0.60	-0.61
3QMK	2.7	2.7	10	3.0±0.4	0.47	-0.48
3UAN	5.0	4.6	10	5.3±0.4	0.32	0.22
4AK2	5.3	4.9	10	5.1±0.5	0.08	0.25
4C4N	6.4	5.3	10	6.2±0.7	0.32	0.36

S: Table 5. Results obtained by the Plants software for protein-GAG complexes.
$RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	12.7	5.7	16	9.9±2.5	-0.33	-0.17
1BFB	12.8	9.4	40	18.6±5.4	0.07	-0.06
1BFC	10.4	6.9	16	11.4±6.6	-0.14	-0.30
1E03	3.2	2.0	40	21.3±17.8	-0.03	-0.02
1E0O	6.2	3.6	16	14.3±4.8	0.02	-0.03
1FQ9_1	7.5	5.2	12	8.2±5.3	0.01	-0.02
1FQ9_2	6.9	4.9	12	6.2±1.2	0.01	-0.04
1G5N	46.3	21.2	12	29.2±8.8	-0.86	-0.80
1GMN	11.6	1.5	12	19.2±11.5	0.02	0.10
1QQP	27.1	2.1	40	21.2±10.2	-0.04	0.11
1RID	28.4	13.2	20	27.6±8.3	-0.07	-0.16
1T8U	5.5	4.3	12	5.43±1.2	0.48	0.46
1XMN	10.2	4.3	20	17.9±12.2	0.46	0.53
1XT3	21.5	12.7	40	18.33±3.3	-0.28	-0.15
2AXM	10.1	10.0	12	9.5±1.2	0.46	0.44
2HYU	4.3	4.3	8	10.63±2.3	0.36	1.16
2HYV	8.3	8.3	4	8.5±0.5	-0.25	-0.40
2JCQ	10.5	3.8	40	10.6±7.0	0.35	0.35
2LVZ	16.1	4.4	40	13.9±4.8	-0.07	-0.13
3C9E	16.8	16.8	8	21.6±3.9	-0.18	-0.11
3DY0	10.3	9.4	40	30.2±13.7	0.20	0.26
3H7D	17.0	15.6	8	18.8±2.3	0.27	0.17
3INA	15.4	9.2	20	17.9±4.8	-0.50	-0.80
3MPK	26.3	4.6	16	11.3±8.6	-0.70	-0.60
3QMK	19.5	7.1	20	22.1±11	-0.10	0.30
3UAN	28.3	12.7	10	24.3±6.7	-0.20	-0.60
4AK2	14.6	11.4	40	29.8±9.3	0.24	0.18
4C4N	8.2	4.6	12	6.2±1.3	-0.57	-0.40

S: Table 6. Results obtained by the HADDOCK software for protein-GAG complexes. $RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	1.32	0.43	10	1.1±0.3	0.45	0.39
1BFB	17.07	13.43	10	20.4±5.3	-0.06	0.04
1BFC	18.01	4.67	10	16.8±9.7	-0.19	-0.25
1E03	17.77	14.56	10	27.5±10.6	0.20	-0.32
1E0O	3.00	0.49	10	14.0±11.3	-0.21	-0.07
1FQ9_1	16.27	13.92	10	21.6±5.5	0.50	0.60
1FQ9_2	26.34	2.96	10	18.5±7.7	-0.02	0.23
1G5N	44.50	15.60	10	29.0±9.0	-0.10	-0.06
1GMN	27.70	23.40	10	27.0±2.0	-0.07	0.18
1QQP	41.40	22.00	10	32.0±8.4	-0.50	-0.60
1RID	17.80	17.80	10	38.9±13.9	0.15	0.26
1T8U	0.08	0.08	10	8.7±5.8	0.40	0.20
1XMN	16.60	3.04	10	20.0±8.7	0.30	0.70
1XT3	9.22	8.86	10	16.8±7.6	0.50	0.30
2AXM	4.78	0.59	10	2.9±1.2	-0.08	0.12
2HYU	33.60	10.20	10	20.7±8.8	-0.25	-0.2
2HYV	15.50	15.50	10	29.2±6.7	0.28	-0.07
2JCQ	14.30	7.30	10	16.3±7.4	0.08	0.15
2LVZ	15.50	15.50	10	18.4±2.1	0.60	0.50
3C9E	29.30	5.40	10	24.5±8.8	-0.43	-0.30
3DY0	18.70	18.70	10	23.3±3.0	-0.12	0.07
3H7D	28.00	5.70	10	24.0±9.0	-0.26	-0.44
3INA	0.28	0.28	10	5.0±7.8	-0.19	0.32
3MPK	26.80	21.80	10	28.4±2.8	0.12	0.34
3QMK	3.18	1.56	10	4.4±2.2	0.70	0.80
3UAN	0.24	0.24	10	4.9±1.7	0.42	-0.60
4AK2	27.10	5.80	10	24.5±10.2	-0.38	0.04
4C4N	27.30	6.30	10	18.5±9.0	-0.40	-0.34

S: Table 7. Results obtained by the Hex software for protein-GAG complexes.
$RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	5.5	5.3	256	17.3±5.5	0.52	0.36
1BFB	9.7	8.3	256	11.2±4.7	0.35	0.38
1BFC	10.2	7.2	256	15.4±4.6	0.28	0.30
1E03	4.7	3.5	256	21.5±14.0	0.44	0.48
1E0O	4.2	2.5	256	18.5±6.7	0.32	0.36
1FQ9_1	9.8	7.4	256	32.7±19.4	0.77	0.70
1FQ9_2	15.3	12.6	256	32.9±15.3	0.52	0.27
1G5N	46.2	14.7	256	38.4±10.4	0.10	0.09
1GMN	28.5	27.6	256	29.4±2.0	0.48	0.19
1QQP	35.5	22.7	256	43.3±7.4	0.19	0.13
1RID	35.0	9.3	256	36.3±28.0	0.35	0.10
1T8U	4.3	4.0	256	12.6±10.1	0.64	0.86
1XMN	4.8	3.4	256	9.96±8.5	0.85	0.78
1XT3	26.4	6.4	256	21.2±9.5	0.24	0.04
2AXM	5.2	2.9	256	4.8±1.1	0.28	0.27
2HYU	17.2	8.4	256	25.3±9.8	0.52	0.50
2HYV	17.6	10.6	256	22.6±8.4	0.54	0.58
2JCQ	9.9	8.9	256	14.7±7.8	0.60	0.70
2LVZ	6.7	2.9	256	12.3±8.3	0.83	0.72
3C9E	3.5	3.5	256	23.8±6.9	0.73	0.45
3DY0	26.7	14.5	256	24.1±3.6	-0.69	-0.59
3H7D	26.8	5.3	256	24.0±7.1	0.58	0.19
3INA	2.8	2.7	256	8.9±9.7	0.95	0.82
3MPK	28.9	27.9	256	28.9±0.6	-0.04	-0.08
3QMK	5.0	3.6	256	11.3±7.2	0.77	0.78
3UAN	3.7	3.7	256	20.1±10.9	0.80	0.67
4AK2	3.9	3.9	256	24.0±12.2	0.83	0.60
4C4N	6.1	5.6	256	11.0±5.1	0.13	0.32

S: Table 8. Results obtained by the SwissDock software for protein-GAG complexes. $RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	0.3	0.3	100	11.9±2.8	0.18	0.24
1BFB	6.4	0.5	100	11.6±5.5	0.18	0.14
1BFC	0.5	0.5	100	12.9±7.6	0.31	0.31
1E03	18.4	3.2	100	12.1±3.9	-0.06	0.05
1E0O	0.3	0.3	100	20.2±5.8	0.49	0.17
1FQ9_1	0.4	0.4	100	23.2±10.5	0.38	0.42
1FQ9_2	38.6	13.9	100	33.6±8.0	-0.08	0.39
1G5N	12.3	7.2	100	26.1±9.8	-0.14	-0.16
1GMN	0.6	0.6	100	17.5±8.9	0.16	0.24
1QQP	10.9	8.1	100	27.3±12.4	-0.09	-0.09
1RID	21.5	8.3	100	37.6±20.2	-0.01	0.05
1T8U	11.8	5.7	100	16.2±7.4	0.07	0.19
1XMN	15.5	3.9	100	14.3±3.8	0.08	0.09
1XT3	15.1	7.5	100	19.5±7.2	0.16	0.19
2AXM	0.1	0.1	100	12.7±3.3	0.13	0.04
2HYU	35.1	7.0	100	27.6±10.6	0.10	0.18
2HYV	21.5	0.3	100	21.2±10.2	-0.02	0.01
2JCQ	19.5	1.3	100	21.8±7.1	0.32	0.38
2LVZ	6.8	1.5	100	3.67±9.3	0.20	0.19
3C9E	21.8	1.7	100	24.7±8.5	-0.13	-0.12
3DY0	26.0	8.2	100	25.4±9.5	0.02	0.07
3H7D	20.6	16.3	100	28.2±6.2	-0.17	-0.11
3INA	24.7	18.5	100	24.3±4.5	0.53	0.14
3MKP	10.9	0.7	100	17.6±8.2	-0.05	0.06
3QMK	11.1	3.3	100	25.5±8.9	0.19	0.18
3UAN	0.1	0.1	100	17.9±9.4	0.18	0.13
4AK2	0.5	0.5	100	18.5±9.9	0.18	0.20
4C4N	1.5	1.0	100	15.8±7.9	0.25	0.2

S: Table 9. Results obtained by the ATTRACT software for protein-GAG complexes. $RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

PDB ID	$RMSatd_{top}$ (Å)	$RMSatd_{best}$ (Å)	N_{poses}	$RMSatd_{all}$ (Å)	$r_{Pearson}$	$r_{Spearman}$
1AXM	0.2	0.2	81	5.0±.1.5	0.46	0.34
1BFB	2.8	2.8	86	5.6±2.4	0.52	0.55
1BFC	5.9	2.6	97	7.3±2.0	0.11	0.25
1E03	2.6	2.0	97	6.3±2.3	0.19	0.46
1E0O	0.2	0.2	33	7.9±2.5	0.11	0.17
1FQ9_1	0.6	0.6	25	8.5±3.0	0.73	0.62
1FQ9_2	0.3	0.3	95	7.6±2.7	0.09	0.57
1G5N	11.3	3.9	97	8.2±1.8	-0.37	-0.34
1GMN	1.3	1.3	96	6.2±1.9	0.37	0.36
1QQP	7.0	3.7	96	7.0±1.8	0.30	0.24
1RID	9.6	9.6	1	9.6±0	-	-
1T8U	11.3	5.9	19	8.6±1.8	0.27	0.16
1XMN	0.5	0.5	78	6.3±2.2	0.66	0.61
1XT3	8.3	4.9	19	9.6±2.1	0.06	0.18
2AXM	0.8	0.8	37	6.9±2.7	0.19	0.42
2HYU	4.6	1.0	93	5.9±1.7	0.44	0.43
2HYV	5.8	4.0	44	7.2±2.1	0.37	0.40
2JCQ	5.3	2.4	73	7.0±3.0	0.10	0.15
2LVZ	3.6	1.6	64	4.4±1.2	0.18	0.25
3C9E	3.7	2.6	55	7.4±2.8	0.16	0.33
3DY0	6.4	3.0	66	6.2±1.4	0.26	0.20
3H7D	13.6	3.8	61	8.0±2.9	-0.11	-0.07
3INA	0.3	0.3	5	9.5±4.9	0.91	0.7
3MPK	8.9	3.7	84	7.4±1.9	0.20	0.45
3QMK	0.4	0.4	90	6.6±3.5	-0.14	0.02
3UAN	4.6	4.0	18	7.4±3.7	0.45	0.57
4AK2	2.5	2.5	99	6.5±2.4	0.03	0.6
4C4N	4.4	3.1	84	7.2±3.3	0.3	0.48

S: Table 10. Results obtained by the Dock software for protein-GAG complexes using flexible docking procedure. $RMSatd_{top}$: structural difference of the top-scored docking pose to the ligand in crystal structure; $RMSatd_{best}$: the lowest structural difference among all the analyzed docking poses to the ligand in experimental structure; N_{poses}: the number of poses obtained from the docking program; $RMSatd_{all}$: the structural difference of the all obtained docking poses to the ligand structure. $r_{Pearson}$ and $r_{Spearman}$: correlation coefficients between the scores of docking poses and their structural difference to the ligand in crystal structure.

		Dock-r			Dock-f	
PDB ID	N_{poses}	RMSadt<3 Å (%)	RMSadt<5 Å (%)	N_{poses}	RMSadt<3 Å(%)	RMSadt<5 Å (%)
1AXM	50	4.2	100.0	81	2.1	48.0
1BFB	50	6.7	95.0	86	11.6	49.0
1BFC	50	12.5	0.0	97	1.0	11.0
1E03	50	3.2	37.0	97	3.2	28.7
1E0O	50	6.0	6.0	33	2.0	4.0
1FQ9_1	50	9.0	18.0	25	7.4	14.8
1FQ9_2	50	1.0	14.7	95	2.0	8.0
1G5N	50	0.0	4.2	97	0.0	6.2
1GMN	50	2.1	27.0	96	3.2	33.6
1QQP	50	0.0	10.8	96	0.0	10.4
1RID	50	0.0	2.0	1	0.0	0.0
1T8U	50	5.3	21.0	19	3.6	37.8
1XMN	50	6.2	32.0	78	3.8	19.2
1XT3	50	0.0	2.0	19	0.0	0.0
2AXM	50	3.1	41.0	37	2.7	21.6
2HYU	50	2.1	32.6	93	2.2	22.0
2HYV	50	1.3	8.0	44	4.5	9.0
2JCQ	50	2.7	17.0	73	4.3	29.3
2LVZ	50	2.0	7.0	64	3.0	8.0
3C9E	50	1.0	25.0	55	1.8	20.0
3DY0	50	0.0	17.3	66	0.0	13.4
3H7D	50	1.0	8.3	61	0.0	23.0
3INA	50	20.0	20.0	5	6.5	17.1
3MPK	50	0.0	10.7	84	0.0	9.5
3QMK	50	8.5	45.0	90	4.5	47.8
3UAN	50	1.0	14.2	18	0.0	6.0
4AK2	50	3.1	25.7	99	2.0	27.3
4C4N	50	2.0	38.4	84	0.0	20.2

S: Table 11. The number of docking poses (%) with RMSadt lower than 3 Å and 5 Å for Dock program: Dock-r (rigid procedure) and Dock-f (flexible procedure).

	rDock		
PDB ID	RMSadt<3 Å (%)	RMSadt<5 Å (%)	N_{poses}
1AXM	4.0	68.0	100
1BFB	4.0	78.0	100
1BFC	2.0	45.0	100
1E03	1.0	42.0	100
1E0O	0.0	8.9	100
1FQ9_1	1.0	47.0	100
1FQ9_2	0.0	8.4	100
1G5N	0.0	20.0	100
1GMN	3.0	47.0	100
1QQP	2.0	44.0	100
1RID	0.0	32.0	100
1T8U	6.0	52.0	100
1XMN	3.0	44.0	100
1XT3	0.0	37.0	100
2AXM	0.0	58.0	100
2HYU	7.0	57.0	100
2HYV	3.0	59.0	100
2JCQ	0.0	17.0	100
2LVZ	4.0	46.0	100
3C9E	0.0	23.0	100
3DY0	0.0	36.0	100
3H7D	1.0	54.0	100
3INA	1.0	1.0	100
3MPK	2.0	38.0	100
3QMK	3.0	32.0	100
3UAN	0.0	62.0	100
4AK2	4.0	52.0	100
4C4N	3.0	50.0	100

S: Table 12. The number of docking poses (%) with RMSadt lower than 3 Å and 5 Å for rDock program.

	ClusPro		
PDB ID	RMSadt<3 Å (%)	RMSadt<5 Å (%)	N_{poses}
1AXM	0.0	25.0	4
1BFB	0.0	12.5	8
1BFC	0.0	20.0	10
1E03	0.0	0.0	7
1E0O	0.0	0.0	9
1FQ9_1	0.0	0.0	4
1FQ9_2	0.0	0.0	4
1G5N	0.0	3.4	30
1GMN	25.0	25.0	8
1QQP	0.0	0.0	21
1RID	0.0	3.4	30
1T8U	14.2	42.0	7
1XMN	0.0	0.0	4
1XT3	0.0	0.0	4
2AXM	25.0	75.0	4
2HYU	0.0	0.0	14
2HYV	0.0	0.0	15
2JCQ	0.0	0.0	15
2LVZ	5.0	35.0	20
3C9E	0.0	0.0	7
3DY0	0.0	0.0	14
3H7D	0.0	0.0	10
3INA	0.0	14.0	7
3MPK	0.0	16.7	12
3QMK	0.0	25.0	10
3UAN	0.0	60.0	6
4AK2	0.0	31.2	16
4C4N	0	7.6	10

S: Table 13. The number of docking poses (%) with RMSadt lower than 3 Å and 5 Å for ClusPro program.

	PLANTS		
PDB ID	RMSadt<3 Å (%)	RMSadt<5 Å (%)	N_{poses}
1AXM	0.0	90.0	10
1BFB	0.0	40.0	10
1BFC	10.0	40.0	10
1E03	0.0	2.0	10
1E0O	0.0	0.0	10
1FQ9_1	0.0	30.0	10
1FQ9_2	0.0	40.0	10
1G5N	0.0	70.0	10
1GMN	0.0	10.0	10
1QQP	0.0	40.0	10
1RID	0.0	30.0	10
1T8U	0.0	10.0	10
1XMN	10.0	10.0	10
1XT3	0.0	10.0	10
2AXM	0.0	20.0	10
2HYU	0.0	70.0	10
2HYV	10.0	90.0	10
2JCQ	0.0	60.0	10
2LVZ	0.0	30.0	10
3C9E	0.0	10.0	10
3DY0	0.0	40.0	10
3H7D	0.0	30.0	10
3INA	0.0	0.0	10
3MPK	0.0	20.0	10
3QMK	30.0	100.0	10
3UAN	0.0	20.0	10
4AK2	0.0	50.0	10
4C4N	0.0	0.0	10

S: Table 14. The number of docking poses (%) with RMSadt lower than 3 Å and 5 Å for PLANTS program.

	HADDOCK		
PDB ID	RMSadt<3 Å (%)	RMSadt<5 Å (%)	N_{poses}
1AXM	0.0	0.0	16
1BFB	0.0	0.0	40
1BFC	0.0	0.0	16
1E03	7.5	35.0	40
1E0O	0.0	12.5	16
1FQ9_1	0.0	0.0	12
1FQ9_2	0.0	25.0	12
1G5N	0.0	0.0	12
1GMN	16.7	16.7	12
1QQP	5.0	5.0	40
1RID	0.0	0.0	20
1T8U	0.0	41.0	12
1XMN	0.0	5.0	20
1XT3	0.0	0.0	40
2AXM	0.0	0.0	12
2HYU	0.0	37.5	8
2HYV	0.0	0.0	4
2JCQ	0.0	5.0	40
2LVZ	0.0	7.5	40
3C9E	0.0	0.0	8
3DY0	0.0	0.0	40
3H7D	0.0	0.0	8
3INA	0.0	0.0	20
3MPK	0.0	18.7	16
3QMK	0.0	0.0	20
3UAN	0.0	0.0	10
4AK2	0.0	0.0	40
4C4N	0.0	16.7	12

S: Table 15. The number of docking poses (%) with RMSadt lower than 3 Å and 5 Å for HADDOCK program.

	Hex		
PDB ID	RMSadt<3 Å (%)	RMSadt<5 Å (%)	N_{poses}
1AXM	100	100	10
1BFB	0	0	10
1BFC	0	10	10
1E03	0	0	10
1E0O	30	40	10
1FQ9_1	0	0	10
1FQ9_2	10	10	10
1G5N	0	0	10
1GMN	0	0	10
1QQP	0	0	10
1RID	0	0	10
1T8U	10	30	10
1XMN	0	10	10
1XT3	0	0	10
2AXM	60	100	10
2HYU	0	0	10
2HYV	0	0	10
2JCQ	0	0	10
2LVZ	0	0	10
3C9E	0	0	10
3DY0	0	0	10
3H7D	0	0	10
3INA	50	60	10
3MPK	0	0	10
3QMK	10	70	10
3UAN	10	60	10
4AK2	0	0	10
4C4N	0	0	10

S: Table 16. The number of docking poses (%) with RMSadt lower than 3 Å and 5 Å for Hex program.

	SwissDock		
PDB ID	RMSadt<3 Å (%)	RMSadt<5 Å (%)	N_{poses}
1AXM	0.0	0.0	256
1BFB	0.0	0.0	256
1BFC	0.0	0.0	256
1E03	0.0	6.3	256
1E0O	5.0	7.4	256
1FQ9_1	0.0	0.0	256
1FQ9_2	0.0	0.0	256
1G5N	0.0	0.0	256
1GMN	0.0	0.0	256
1QQP	0.0	0.0	256
1RID	0.0	0.0	256
1T8U	0.0	9.3	256
1XMN	0.0	15.6	256
1XT3	0.0	0.0	256
2AXM	9.3	50.0	256
2HYU	0.0	0.0	256
2HYV	0.0	0.0	256
2JCQ	0.0	0.0	256
2LVZ	0.8	37.5	256
3C9E	0.0	6.3	256
3DY0	0.0	0.0	256
3H7D	0.0	0.0	256
3INA	31.3	59.4	256
3MPK	0.0	0.0	256
3QMK	0.0	27.7	256
3UAN	0.0	28.1	256
4AK2	0.0	12.5	256
4C4N	0.0	0.0	256

S: Table 17. The number of docking poses (%) with RMSadt lower than 3 Å and 5 Å for SwissDock program.

	ATTRACT		
PDB ID	RMSadt<3 Å (%)	RMSadt<5 Å (%)	N_{poses}
1AXM	10.0	10.0	100
1BFB	30.0	70.0	100
1BFC	12.0	17.0	100
1E03	0.0	50.0	100
1E0O	30.0	40.0	100
1FQ9_1	10.0	20.0	100
1FQ9_2	0.0	0.0	100
1G5N	0.0	0	100
1GMN	20.0	30.0	100
1QQP	0.0	0	100
1RID	0.0	0.0	100
1T8U	0.0	0.0	100
1XMN	0.0	20.0	100
1XT3	0.0	0.0	100
2AXM	1.0	1.0	100
2HYU	0.0	0.0	100
2HYV	6.0	6.0	100
2JCQ	2.0	2.0	100
2LVZ	4.0	10.0	100
3C9E	3.0	3.0	100
3DY0	0.0	0.0	100
3H7D	0.0	0.0	100
3INA	0.0	0.0	100
3MPK	3.0	3.0	100
3QMK	0.0	1.0	100
3UAN	1.0	1.0	100
4AK2	3.0	7.0	100
4C4N	9.0	9.0	100

S: Table 18. The number of top and best poses obtained for the top and best ranked RMSadt all for lower than 3 and 5 Å for ATTRACT program.