Published in final edited form as:

J Comput Chem. 2014 March 15; 35(7): 526–539. doi:10.1002/jcc.23517.

The Importance of Ligand Conformational Energies in Carbohydrate Docking: Sorting the Wheat from the Chaff

Anita K. Nivedha¹, Spandana Makeneni¹, B. Lachele Foley¹, Matthew B. Tessier¹, and Robert J. Woods^{1,2,*}

¹Complex Carbohydrate Research Center, 315 Riverbend Road, University of Georgia, Athens, GA 30606 ²School of Chemistry, National University of Ireland, Galway, University Road, Galway, Ireland

Abstract

Docking algorithms that aim to be applicable to a broad range of ligands suffer reduced accuracy because they are unable to incorporate ligand-specific conformational energies. Here, we develop internal energy functions, **Carbohydrate Intrinsic (CHI)**, to account for the rotational preferences of the glycosidic torsion angles in carbohydrates. The relative energies predicted by the CHI energy functions mirror the conformational distributions of glycosidic linkages determined from a survey of oligosaccharide-protein complexes in the Protein Data Bank. Addition of CHI energies to the standard docking scores in Autodock 3, 4.2, and Vina consistently improves pose ranking of oligosaccharides docked to a set of anti-carbohydrate antibodies. The CHI energy functions are also independent of docking algorithm, and with minor modifications, may be incorporated into both theoretical modeling methods, and experimental NMR or X-ray structure refinement programs.

Introduction

Protein-carbohydrate interactions are crucial in numerous aspects of biology, including metabolism, gene expression, cell-cell communication, growth, development, and immune response¹. *In vivo*, complex carbohydrates (glycans) are found on cell surfaces as glyconjugates (glycoproteins/glycolipids) or polysaccharides, mediating biological function by their direct interaction with proteins, such as receptors (lectins), enzymes, and antibodies. Cancer is marked by aberrant glycosylation which can serve as a disease-related marker, or as a target for therapeutic intervention^{2–5}. Conversely, endogenous cell-surface glycans are

*To whom correspondence should be addressed: Phone: +1 (706) 542-4454, FAX: +1 (706) 542-4412, rwoods@ccrc.uga.edu.

Individual author contributions:

- Anita K. Nivedha: Authored portions of the paper and prepared figures for the paper; designed docking protocols and the antibody alignment algorithm; performed the dockings; developed the CHI energy functions and applied the functions to docking results; provided tools for analysis, analyzed and interpreted the data.
- Spandana Makeneni: Authored portions of the paper; helped in the design of docking protocols and the antibody alignment algorithm; performed binding site volume calculations and provided tools for the analysis of data.
- B. Lachele Foley: Contributed to the design of the antibody alignment algorithm and the development of the CHI energy functions.
- Matthew B. Tessier: Contributed to the design of preliminary docking protocols, provided PREP files for the non-standard sugar residues, and scripts for the collection of quantum mechanical data.
- Robert J. Woods: Authored the paper; conceived and designed the experiment, and contributed to the analysis and interpretation of data.

frequently exploited by infectious agents, as in the hemagglutinin-mediated adhesion of influenza A virus^{6–8}. A physical understanding of carbohydrate-protein interactions aids in the development of therapeutic agents designed to block such interactions,^{9–12} such as antibodies which target specific glycans^{13,14}. A better understanding of the immune system's response to carbohydrate-based vaccines^{15–18}, facilitates the prediction and rationalization¹³ of hazardous or misleading cross-reactivities between antibodies against disease-related carbohydrates, and endogenous glycans^{19,20}.

The challenges involved in obtaining co-complexed carbohydrate-protein structures using experimental methods such as X-ray crystallography and NMR spectroscopy include, production and purification of the protein, isolation or synthesis of the glycan, and co-crystallization of the complex²¹. Therefore, there is a long-standing interest in applying theoretical modeling methods (automated docking) to aid in the characterization of the 3D structure of carbohydrate-protein complexes^{13,22–27}. However, these methods also have limitations. Automated docking faces the triple challenge of accurately predicting 1) the ligand orientation in the binding site (pose); 2) the ligand conformation in the binding site (shape); and 3) the relative affinity of the optimal pose (interaction energy). Ligand internal energies are only approximately modeled within docking algorithms by mainly considering energies associated with internal steric repulsion²⁸. Such an approximation inherently degrades the accuracy of docking predictions as various ligand classes have specific conformational properties. The glycosidic torsion angles between individual monosaccharides forming glycans are crucial in defining their 3D structure and dynamics. The accurate prediction of oligosaccharide conformations requires the additional consideration of stereo-electronic properties responsible for the anomeric, *exo*-anomeric, and *gauche* effects²⁹. Their omission frequently leads to the incorrect prediction of docked oligosaccharide conformations^{30–32}.

Docking programs treat interaction energy terms as empirically-adjustable components, which may be tuned for a particular ligand class, such as carbohydrates³³. Inclusion of carbohydrate conformational energies in the docking energy function would likely require reoptimization of the empirical weighting resulting in a non-transferable carbohydrate-specific implementation of the algorithm. Alternatively, we wished to develop a carbohydrate-specific conformational energy function which predicts oligosaccharide energies independent of docking algorithm, and could potentially also be employed to evaluate the conformational energies of experimentally-determined oligosaccharide structures. We focused on modeling conformational properties *intrinsic* to glycosidic linkages between pyranoses, with the criterion that the method should also be generalizable to other carbohydrate ring forms, such as furanoses, as well as to other linkages, such as 1–6, 2–3, 2–6, etc. Tetrahydropyran, and related analogs, have long been employed as representative carbohydrates in quantum mechanical calculations for this purpose^{34–41}. The assumption being that any additional effects on the conformational properties, for example from hydrogen bonding, overlay the intrinsic properties of the linkages between pyran rings. Quantum mechanical calculations were employed on a set of glycosidically-linked tetrahydropyrans representing all two-bond linkages between pyranoses. The rotational energy profiles for these linkages were used to derive the desired carbohydrate *intrinsic* (CHI) energy functions. Given a 3D oligosaccharide structure, the CHI energy functions may be employed to estimate the energy arising from any distortion of the glycosidic linkages, relative to their lowest energy conformations.

Because of the important roles of anti-carbohydrate antibodies in therapeutic and diagnostic applications, and the challenges associated with experimentally defining their 3D structures, they have been the subject of numerous automated docking studies^{42–49}. We chose six crystallographically-determined antibody-carbohydrate complexes to evaluate the ability of

CHI energy functions to improve predicted rankings of the docked poses. These systems were selected based on the diversity of the antibody binding site topologies (canyon, valley, crater)⁵⁰, and size variations of the carbohydrate ligands (tri- to penta saccharides including linear and branched sequences).

Methods

System selection and docking protocol

Docking was performed using AutoDock 3.0.5 (AD3)⁵¹, 4.2 (AD4.2)⁵² and Vina 1.1.2 (ADV)⁵³. Details of the reference systems, including PDB IDs, ligand sequences and biological origin are presented in Table 1. In each case, the protein chain containing the ligand with the lowest average B-factor was selected for docking. The carbohydrate ligands in systems 1UZ8, 1S3K and 1M7I were built using the Carbohydrate Builder on GLYCAM-Web (www.glycam.org)⁵⁴. The remaining ligands contain the nonstandard sugar residues abequose and 2-deoxy-rhamnose. Oligosaccharides containing these deoxy residues were assembled using the tLEaP⁵⁵ module from the AMBER package employing GLYCAM06i force field parameters and PREP residue structure files, available for download at www.glycam.org (SM11). The antibody structures were obtained from the PDB (www.rcsb.org)⁵⁶. All protein and ligand files were prepared for docking using AutoDock Tools 1.5.4 (ADT)⁵². The choice of partial charge was based on the method used to calibrate the scoring functions of the individual docking programs; Kollman charges⁵⁷ were added to the protein for docking with AD3, while Gasteiger charges⁵⁸ were used to prepare proteins for docking with AD4.2 and ADV, and in each case Gasteiger charges were assigned to the ligands. AutoDock distributes any non-zero residual net charge across the macromolecule. Hydrogen atoms were added to the protein using ADT, whereas GLYCAM hydrogens were retained in the ligands. A standard grid box (dimensions: 26.25 × 26.25 × 37.50 Å) was employed for all runs, centered relative to the complementarity determining regions (CDRs) of the antibody (Figure 1a). Before docking, the ligand was translated to the center of mass (CoM) of the CDRs but maintained in the default GLYCAM orientation and conformation. VMD⁵⁹ was used for molecular visualization and image-rendering.

In all ligands, the hydroxyl groups and glycosidic torsion angles were defined as being flexible, while the C₅–C₆ bonds were restrained at the orientation present in the reference crystal structures. The protein was maintained rigid. In AD3 and AD4.2, 100 runs of the Lamarckian Genetic Algorithm were employed, with 800,000 energy evaluations per run, and a population size of 200. The translation step size was 2 Å, while the quaternion and dihedral step sizes were each 50°. The ADV source code was modified to increase the total number of output structures from 20 to 100 (Supplementary Material, SM1). The maximum energy difference between the best and worst binding modes was set at 10 kcal/mol while the exhaustiveness value was 8. The complete set of docking parameters used is given in SM2, SM3 and SM4.

Antibody and docking grid box alignment

Consistent grid box placement on the CDRs was achieved by positioning the box relative to three points defined by specific CoM's within the CDRs. The CDRs were identified using the AbM definition^{65,66}, based on both the Kabat⁶⁷ and Chothia⁶⁸ numbering schemes. To ensure consistent orientation of the antibody surface relative to the box grid points, the protein coordinates were transformed with respect to a set of internal coordinate axes, as shown in Figure 1. This protocol removes any issues arising from the fact that the grid is cubic and not spherical, which can otherwise result in varied regions of each antibody being included within the grid.

Quantum mechanical calculations

Quantum mechanical calculations were performed using Gaussian09⁶⁹. Structures were optimized at the HF/6-31G++(2d, 2p) level of theory, and single-point energies calculated at the B3LYP/6-31G++(2d, 2p) level, consistent with the approach used in the GLYCAM force field development³⁸. Rotational energy profiles were computed at 15° increments, allowing complete relaxation of other coordinates.

Shape, and pose, RMSD values

Pose RMSD (PRMSD) values were obtained by calculating the RMSD between the ring atoms of the crystal ligand maintained in its native co-crystallised position, and the corresponding ring atoms in the docked ligand maintained in its docked position (Figure 2a). A pose with a PRMSD $\leq 2\text{\AA}$ was considered to have been successfully docked. Shape RMSD (SRMSD) values were obtained by first superimposing the crystal and docked ligands followed by calculating the RMSD between their respective ring atoms (Figure 2b). The SRMSD is a quantification of the dissimilarity in the 3D conformations of the docked and crystal ligands, irrespective of their relative positions on the protein surface.

Results and Discussion

I. Assessment of current docking methodologies

The six ligands extracted from their co-crystal structures could successfully be docked back rigidly into the same structure of the protein (results not shown); this is an outcome observed previously in studies of carbohydrate-protein docking^{49,70}. Although necessary, this docking experiment is not a sufficient prerequisite for any docking method, since both molecules in a co-crystallized complex are already in the correct conformation for binding, and do not require induced fit to occur during docking.

Independently-generated oligosaccharide 3D structures were employed as ligands to test the performance of the docking methodologies in predicting bound conformations of unknown carbohydrate-protein complexes. These starting structures were generated using GLYCAM, known to produce low-energy conformations of carbohydrates; the structures generated were found to be essentially equivalent to the same ligands found in the co-crystal structures, as indicated by their SRMSDs (Table 1), and by a comparison of their glycosidic torsion angles (SM5). The average SRMSD between the crystallographic ligands and theoretical structures was 0.53Å. The preliminary SRMSD analysis also showed that the ligand in each antibody complex adopted a low energy conformation, similar to that expected for the free ligand.

A second requirement for a general docking protocol is to permit the ligands a reasonable level of freedom by allowing their glycosidic torsion angles and hydroxyl groups complete flexibility. This approach enables comparisons to be made between structures of the experimental and theoretical ligands, facilitating an assessment of the impact of induced fit in the ligand on the outcome from docking analysis.

After docking, the ϕ ($O_5'-C_1'-O_x-C_x$) and ψ ($C_1'-O_x-C_x-C_{x-1}$) glycosidic torsion angles of the docked poses (Figure 4a) were measured, and compared to the torsion angles of corresponding linkages in the experimental co-crystal structure, and in the initial GLYCAM theoretical structure. The analysis indicated that the distribution of the torsion angle values amongst the docked poses frequently deviated considerably from both the crystal and GLYCAM reference values (SM5). Five examples of this analysis are highlighted in Figure 3. Presented in Figure 3a is an instance in which all three docking programs identified the lowest energy pose correctly (that is, with the glycosidic angles falling within 30° of the

corresponding torsion angles in the crystal structure). Presented in Figures 3b, 3c, and 3d are cases in which only one of the docking programs identified the correct pose, and finally an example is shown in which all three programs failed to produce the correct torsion angles (Figure 3e). All of the methods were able to generate some number of conformations that were within 30° of the crystallographic ϕ and ψ values, however, these were often not the poses that had the best docking energy. Thus, in a routine application of docking, they would not be identified as the most likely (highest-ranked) pose. Overall, a very broad range of torsion angles (and therefore 3D shapes) were generated by each algorithm, indicating a potential opportunity to employ a conformational energy function as an additional filter to identify unlikely conformations in the docking data.

II. Development and validation of the CHI energy functions

Quantum mechanical conformational energies for a variety of model disaccharides were obtained by employing tetrahydropyran (THP) as the minimal model of a carbohydrate ring. Two THP molecules were used to model each glycosidic linkage (1–2, 1–3 and 1–4) between pyranoses in the 4C_1 and 1C_4 configurations. Given that there are two anomeric configurations (α and β), and two hydroxyl configurations (axial (ax) and equatorial (eq)), associated with each linkage, the development of each CHI energy function required the analysis of the glycosidic rotational energies of at least four structures per linkage. For example, the different models used in modeling the 1–3 linkage are presented in Figure 4.

Individual rotational energy profiles were determined for both the ϕ ($O_5'-C_1'-O_x-C_x$) and ψ ($C_1'-O_x-C_x-C_{x-1}$) glycosidic torsion angles of the various disaccharide models (Figure 5). A similar approach has been employed by A. D. French to examine the properties of various disaccharides and disaccharide analogs^{40,41,71,72}. Models with similar local symmetries gave rise to similar torsional energy profiles and were grouped together. Average energy curves were then obtained for each group. Based on similar energy profiles, two average energy curves for the Φ -linkage were computed: one, for all models with an α -linkage (Figure 5a), and the other for all models with a β -linkage (Figure 5b). Similarly, two average curves for the Ψ -linkage were computed, based on division of the linkages into the following two groups: 1-2ax, 1-4ax, 1-3eq (Figure 5c); and 1-2eq, 1-4eq, 1-3ax (Figure 5d).

The CHI energy functions (SM6) were generated by fitting Gaussian expansions (Eq. 1) to the average energy values for each of the curves in Figure 5 using the default fitting routine in Gnuplot ver. 4.0⁷³:

$$f(x) = \sum_{i=1}^N a_i e^{-\frac{(x-b_i)^2}{c_i}} \quad (\text{Eq. 1})$$

where, N is the number of individual Gaussian functions used for each CHI energy equation, x refers to the glycosidic torsion angle (ϕ or ψ), and a_i , b_i , and c_i refer to the magnitude, width, and mid-point of the distribution respectively. All curves (SM7) were adjusted to a minimum value of 0 kcal/mol, and may therefore be considered conformational energy penalty functions. In order to apply the energy curves shown in Figure 5 to linkages containing L-sugars, it is simply necessary to employ the mirror images of the relevant energy curve.

The experimental distribution of glycosidic angles in carbohydrate-protein crystal structures in the PDB provides an independent metric for comparison with the predicted CHI energies. Glycosidic torsion angle data for over 13,000 glycosidic linkages were extracted using the GlyTorsion web-tool⁷⁴ (SM8), binned, and plotted against the corresponding CHI energy curves (Figure 6). The comparison leads to the important conclusion that the majority of proteins that recognize oligosaccharides select low energy (solution-like) conformations of

the glycosidic linkage. This has considerable importance for carbohydrate docking, as it supports the view that biasing selection toward low energy linkage conformations should enhance the likelihood of correct pose prediction.

III. Refinement of the docking results using the CHI energy functions

An assessment of the performance of each of the docking algorithms can be made by plotting the difference between the conformations of the ligands, relative to that in the co-complex (SRMSDs), against the predicted interaction energies. Ideally, poses with correct ligand shapes should have lower interaction energies than seen for incorrect shapes. Plots of interaction energy *versus* SRMSD were generated for AD3, AD4.2 and ADV (Figure 7), and the coefficient of determination (R^2) computed by linear regression. In each case, only weak linear relationships between ligand shape and interaction energy were observed ($R^2 = 0.19$), and in the case of ADV a slight negative slope was observed. Following rescoring of the docked poses by addition of the CHI energy from each glycosidic angle to the docked energy of the structure, a clear enhancement of the R^2 values was observed, across all three programs ($0.60 < R^2 < 0.68$). It should be reiterated here that none of the three docking algorithms include internal rotational energies (torsion terms), and at best account for ligand internal energies in a general steric sense. In the case of glycosidic linkages, this internal energy was found to be less than approximately 0.2 kcal/mol. Thus, while some double counting of internal energy is introduced by adding the CHI energy directly to the total docking energy, it does not result in a significant error.

Prior to inclusion of the CHI energies, all poses from AD3 and ADV and a majority of those from AD4.2 were predicted to have favorable (negative) interaction energies; a result of the nearly horizontal slope of the SRMSD-versus-interaction-energy curves. Addition of the CHI energies led to positive slopes and frequently unfavorable interaction energies (positive) for high-energy ligand conformations. Therefore, an intuitive interaction energy cut-off of 0 kcal/mol could be defined as a convenient filter for eliminating the most unlikely structures.

For all six antibody complexes, the poses that are most similar to the co-crystal (lowest PRMSD poses) also have CHI-adjusted interaction energies > 0 kcal/mol, with the single exception being the AD4.2 results for 1M7I (Figure 7b). All 100 docked poses of that pentasaccharide received positive rescored interaction energies, reflecting the sub-optimal quality of the conformations produced by AD4.2 for this system. In this case, the pose closest to the co-complex displayed a PRMSD = 3.4 Å, and a CHI-corrected interaction energy of 14.7 kcal/mol; rescoring can't correct for the absence of a correct pose. Thus, the addition of the CHI energy to the docked energy scores provides a cutoff (0 kcal/mol), below which all poses may be considered possible binders.

Presented in Figure 8, are the ϕ and ψ torsion angles for the docked poses from all 6 antibody-carbohydrate systems, overlaid onto the corresponding CHI energy curves. They provide a clear indication that the docking algorithms sample a disproportionately large number of high-energy ligand conformations, particularly evident for AD4.2 and ADV. Several low energy regions, particularly for the ψ angles, are also not well-represented. In quantitative terms, for AD3 >45% of the poses contain ligands with at least one bond in a high energy conformation (CHI energies > 2 kcal/mol); the numbers for AD4.2 and ADV being 73 and 77 %, respectively.

Pose ranking after including the CHI energy—In 9 of the 18 cases (6 antibodies x 3 docking algorithms), the top-ranked pose remained the same before and after inclusion of the CHI energies (Figure 9), with an average SRMSD of 0.3 Å. That the ranking of these poses did not change is unsurprising, given that inclusion of the CHI energy function does

not greatly alter the interaction energy if the ligand is already in a low-energy conformation. However, in 7 of the 9 remaining cases, the SRMSD of the top-ranked pose improved by an average of 0.8 Å, after rescoring and reranking.

Prior to rescoring, from the 100 docking runs, poses with PRMSDs ≤ 1 Å were obtained in 17 out of the 18 cases, however, they were not necessarily lowest energy poses, highlighting the challenge in recognizing a correctly docked pose amongst all poses produced by a docking run. The impact of the CHI energy on the ability of docking to both produce a correctly docked pose and rank it as the lowest energy structure is indicated in terms of PRMSDs in Figure 9b. In several instances in which the lowest energy pose produced by the docking program was incorrect (PRMSD > 2 Å), reranking after including the CHI energy led to lowest energy structures having both PRMSD and SRMSD < 1 Å.

The impact of rescoring on the conformations (SRMSDs) and orientations (PRMSDs) of the top-ranked poses are presented for several examples in the following section. Docking of the tetrasaccharide ligand onto the 1S3K antibody, using AD3 (Figure 10), and docking of the trisaccharide ligand onto the 1M7D antibody, using AD4.2 yielded top-ranked poses with PRMSDs > 5 Å. Both these structures obtained high CHI energy scores of 7.0 kcal/mol and 11.6 kcal/mol, respectively. The lowest energy poses after reranking had PRMSDs of 0.6 Å (1S3K/AD3), and 0.5 Å (1M7D/AD4.2), with lower CHI energies of 1.0 kcal/mol and 0.9 kcal/mol respectively.

Prior to rescoring, lowest energy structures obtained for 1MFD from all three programs had PRMSDs > 5 Å, with CHI energies > 4 kcal/mol for the poses from AD4.2 and ADV, and 1.3 kcal/mol for the pose from AD3 (Figure 11a, SM9). After rescoring, the lowest energy pose from AD3 remained unchanged, whereas, the corresponding pose from AD4.2 was replaced by a pose with a lower CHI energy score, however, the newly top-ranked pose still had a high PRMSD.

Even though rescoring did not result in correctly docked lowest energy poses in either of these cases, it improved the overall ranking of the lowest PRMSD structures (PRMSDs < 1 Å) from 18 to 9 in AD3, and 13 to 2 in AD4.2 (SM10). It should also be noted that the second lowest energy pose in AD3 (PRMSD = 1 Å) remained unchanged in ranking after rescoring. In contrast, the relatively high CHI energy score of the lowest energy pose from ADV contributed to this pose being replaced by a correctly docked structure, with a lower CHI energy score, after rescoring (Figure 11b, SM9, SM10).

The ligand in 1MFD is a branched trisaccharide comprised of mannose (Man), galactose (Gal), and abequose (Abe). Abe is an analog of Gal (3,6-dideoxyGal), and the anchoring residue for the trisaccharide in the crystal structure⁷⁵ (Figure 11c). An examination of the docking results indicated that all three docking programs consistently generated better scores for poses in which the Gal residue replaces Abe in the binding site (Figure 11d), with little increase in the SRMSD for the incorrect pose. That is, the trisaccharide can fit equally well into the binding site in the two possible orientations effectively flipped by 180°. The theoretical preference for Gal in the binding site appears to be a consequence of its ability to make additional hydrogen bonds with the protein relative to the more hydrophobic Abe. This observation suggests that the balance between contributions from hydrogen bonding *versus* hydrophobic interactions is imperfect in these docking algorithms. In addition, the 1MFD crystal structure reveals the presence of a water molecule within the binding pocket, mediating hydrogen bond interactions between the antibody and the ligand's Abe residue. Given that explicit waters are not generally included in docking studies, the algorithms may be compensating for their absence by placing the more polar Gal inside the binding pocket. This conclusion is supported by the observation that one of the hydroxyl groups of the Gal

residue (O-4) occupies a position in close proximity to this water molecule (PDB residue name: WAT 601) originally found in the crystal complex (Figure 11d).

The flipping of the carbohydrate ligand that was observed in 1MFD, was not observed in the case of its scFv counterpart (1MFA); instead, all three lowest energy poses (AD3; AD4.2; ADV) for 1MFA had orientations similar to that of the crystal ligand (PRMSDs < 2 Å). Since the ligands being docked to both antibodies are identical, we can infer that the two binding sites are not identical (Table 1). To facilitate a better understanding of the difference between the two binding pockets, their volumes were calculated using Fpocket⁷⁶; the volume of the 1MFA binding pocket was calculated to be 423.01 Å³, while that of 1MFD was 582.51 Å³. The 1MFD binding pocket, being 150 Å³ larger, is able to accommodate the flipped orientation of the Gal residue, whereas, the smaller 1MFA binding pocket is not as accommodating of this ligand orientation, due to possible steric clashes. This potential steric clash was confirmed by superimposing the coordinates of the Gal residue onto those of Abe in 1MFA (Figure 11e, f).

The known challenge associated with docking large, flexible molecules using AD4.2^{53,77} was encountered with the linear pentasaccharide ligand in 1M7I. None of the 100 poses were correctly docked (all PRMSDs > 2 Å); the lowest energy pose had a PRMSD of 3.9 Å and a CHI energy of 18.5 kcal/mol (Figure 12a). After rescoring, the lowest energy pose had a considerably improved CHI energy score of 4.3 kcal/mol, however, it still had a high PRMSD (Figure 12b). It has been suggested that the maximum number of rotatable bonds be limited to 10 when employing AD4.2⁷⁷; the ligand in 1M7I has nearly double that number at 19, making this quite a challenging system to dock using AD4.2. In AD3, although only 4 of the 100 output poses were correctly docked, they occupied the top 4 ranks, before and after rescoring. In ADV, 7 of the 100 output poses were correctly docked, of which 5 were amongst the 8 top-ranked poses, before and after rescoring. Although both AD3 and ADV seem to have had difficulty in finding the correct pose for the pentasaccharide, whenever such a pose was found, both programs scored them favorably. As these poses also had low SRMSD values, they were identified as lowest energy poses after rescoring.

Conclusions

A solution to a major challenge encountered in flexible carbohydrate docking has been presented in this study by the development of intrinsic energy terms for carbohydrates, which quantify the relative energy of their glycosidic torsion angles. In 7 of the 18 cases (6 systems x AD3/AD4.2/ADV), the lowest energy poses generated by the docking programs had PRMSDs > 2 Å, however, after rescoring using the CHI energy functions, the PRMSDs in 4 of the 7 cases improved, with correctly docked poses (PRMSDs ≤ 2 Å) replacing incorrect poses, and increasing the total count of correctly docked lowest energy poses to 15 out of 18. Rescoring also led to lowest energy poses that had SRMSDs ≤ 1 Å in 16 out of 18 cases, and SRMSDs ≤ 1.5 Å in the two remaining cases. Among the three docking programs employed in this study, ADV was most successful in producing and appropriately ranking the correct ligand pose, with a success rate of 83% before rescoring, and 100% after rescoring. Inclusion of the CHI energy term in rescoring docked poses enabled the filtering of poses based on their conformations, increasing the chances of finding the correct pose amongst all output poses generated.

In most docking applications, locating the correctly docked pose amongst the numerous output poses largely depends on the ranking of these poses based on their energy scores. The CHI energy functions may in principle be used in the assessment of carbohydrate structures obtained from any theoretical or experimental method. By favoring energetically reasonable ligand conformations, the CHI energies significantly improve the pose ranking for structures

obtained from docking algorithms, making the rescored energy a better predictor of the quality of the docked pose. This improvement was observed across all three programs indicating that the CHI energy functions may be employed independently of the scoring functions. The CHI energy functions could also be incorporated directly within docking programs as a component of the scoring function, although that might require a reoptimization of the scoring functions. Application to crystallographic data leads to the conclusion that proteins primarily recognize low-energy conformations of carbohydrates. This final observation has considerable relevance to the design of carbohydrate-based inhibitors and vaccines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to the National Institutes of Health (GM094919 (EUREKA)) and the Science Foundation of Ireland (08/IN.1/B2070) for financial support.

References

1. Varki A. Glycobiology. 1993; 3:97–130. [PubMed: 8490246]
2. Ghazarian H, Idoni B, Oppenheimer SB. Acta Histochem. 2011; 113:236–247. [PubMed: 20199800]
3. Hakomori S. Annu Rev Immunol. 1984; 2:103–126. [PubMed: 6085749]
4. Fukuda M. Cancer Res. 1996; 56:2237–2244. [PubMed: 8625291]
5. Hakomori SI. Adv Cancer Res. 1989; 52:257–331. [PubMed: 2662714]
6. Eisen MB, Sabesan S, Skehel JJ, Wiley DC. Virology. 1997; 232:19–31. [PubMed: 9185585]
7. Suzuki Y, Nagao Y, Kato H, Matsumoto M, Nerome K, Nakajima K, Nobusawa E. J Biol Chem. 1986; 261:17057–17061. [PubMed: 3782153]
8. Wiley DC, Skehel JJ. Annu Rev Biochem. 1987; 56:365–394. [PubMed: 3304138]
9. Magnani JL, Ernst B. Nat Rev Drug Discov. 2009; 8:661–677. [PubMed: 19629075]
10. Dreitlein WB, Maratos J, Brocavich. J Clin Ther. 2001; 23:327–355.
11. Moscona A. N Engl J Med. 2005; 353:1363–1373. [PubMed: 16192481]
12. Kevin, HM. In Galectins and Disease Implications for Targeted Therapeutics. American Chemical Society; 2012. p. 61-77.
13. Tessier MB, Grant OC, Heimbürg-Molinari J, Smith D, Jadey S, Gulick AM, Glushka J, Deutscher SL, Rittenhouse-Olson K, Woods RJ. PloS One. 2013; 8:e54874. [PubMed: 23365681]
14. Woods, R.; Yongye, A. In Anticarbhydrate Antibodies. Kosma, P.; Müller-Loennies, S., editors. Springer; Vienna: 2012. p. 361-383.
15. Kadirvelraj R, Gonzalez-Outeriño J, Foley BL, Beckham ML, Jennings HJ, Foote S, Ford MG, Woods RJ. Proc Natl Acad Sci. 2006; 103:8149–8154. [PubMed: 16705032]
16. Yongye AB, Gonzales Outeriño J, Glushka J, Schultheis V, Woods RJ. Biochemistry. 2008; 47:12493–12514. [PubMed: 18954144]
17. Calarese DA, Scanlan CN, Zwick MB, Deechongkit S, Mimura Y, Kunert R, Zhu P, Wormald MR, Stanfield RL, Roux KH, Kelly JW, Rudd PM, Dwek RA, Kattinger H, Burton DR, Wilson IA. Science. 2003; 300:2065–2071. [PubMed: 12829775]
18. Dyekjær, JD.; Woods, RJ. In NMR Spectroscopy and Computer Modeling of Carbohydrates Recent Advances. Vliegthart, JFG.; Woods, RJ., editors. American Chemical Society; Washington: 2006. p. 203-219.
19. Gildersleeve J, Roach TA, Li Z, Gildersleeve JC. Glycobiol. 2008; 18:746–756.
20. Pincus SH, Moran E, Maresh G, Jennings HJ, Pritchard DG, Egan ML, Blixt O. Vaccine. 2012; 30:4849–4858. [PubMed: 22634296]

21. Woods RJ, Tessier MB. *Curr Opin Struct Biol*. 2010; 20:575–583. [PubMed: 20708922]
22. Cooke RM, Hale RS, Lister SG, Shah G, Weir MP. *Biochemistry*. 1994; 33:10591–10596. [PubMed: 7521209]
23. Mahmoudian M. *J Mol Graphics Modell*. 1997; 15:149–153.
24. Laederach A, Dowd MK, Coutinho PM, Reilly PJ. *Proteins: Struct Funct Genet*. 1999; 37:166–175. [PubMed: 10584063]
25. Goodsell DS, Morris GM, Olson AJ. *J Mol Recognit*. 1996; 9:1–5. [PubMed: 8723313]
26. Sotriffer CA, Flader W, Winger RH, Rode BM, Liedl KR, Varga JM. *Methods*. 2000; 20:280–291. [PubMed: 10694451]
27. Jorgensen WL. *Science*. 2004; 303:1813–1818. [PubMed: 15031495]
28. Goodsell DS, Olson AJ. *Proteins*. 1990; 8:195–202. [PubMed: 2281083]
29. Foley BL, Tessier MB, Woods RJ. *Wiley Interdiscip Rev Comput Mol Sci*. 2011:1–69.
30. Laederach A, Reilly PJ. *Proteins: Struct Funct Genet*. 2005; 60:591–597. [PubMed: 16001416]
31. Sapay N, Nurisso A, Imberty A. *Methods Mol Biol (Clifton, NJ)*. 2013; 924:469–483.
32. Bras NF, Fernandes PA, Ramos M. *J Theor Chem Acc*. 2009; 122:283–296.
33. Laederach A, Reilly PJ. *J Comput Chem*. 2003; 24:1748–1757. [PubMed: 12964193]
34. Hwang MJ, Ni X, Waldman M, Ewig CS, Hagler AT. *Biopolymers*. 1998; 45:435–468. [PubMed: 9538697]
35. Woods, RJ.; Edge, CJ.; Wormald, MR.; Dwek, RA. *Complex Carbohydrates in Drug Research*. Bock, K.; Clausen, H.; Krogsgaard-Larsen, P.; Kofod, H., editors. Munksgaard; Copenhagen, Denmark: 1993. p. 15-36.
36. Weldon AJ, Vickrey TL, Tschumper GS. *J Phys Chem A*. 2005; 109:11073–11079. [PubMed: 16331953]
37. Woodcock HL, Moran D, Pastor RW, MacKerell AD, Brooks BR. *Biophys J*. 2007; 93:1–10. [PubMed: 17554075]
38. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. *J Comput Chem*. 2008; 29:622–655. [PubMed: 17849372]
39. Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, Jamison FW, MacKerell AD. *J Chem Theory Comput*. 2011; 7:3162–3180. [PubMed: 22125473]
40. French AD, Kelterer AM, Johnson GP, Dowd MK, Cramer CJ. *J Comput Chem*. 2001; 22:65.
41. French AD, Dowd MK. *J Mol Struct (Theochem)*. 1993; 286:183–201.
42. Talavera A, Eriksson A, Okvist M, Lopez-Requena A, Fernandez-Marrero Y, Perez R, Moreno E, Krengel U. *Mol Immunol*. 2009; 46:3466–3475. [PubMed: 19748674]
43. Paula S, Monson N, Ball WJ Jr. *Proteins*. 2005; 60:382–391. [PubMed: 15971203]
44. Blaszczyk-Thurin M, Murali R, Westerink MAJ, Steplewski Z, Sung Co M, Kieber-Emmons T. *Protein Eng*. 1996; 9:447–459. [PubMed: 8795045]
45. Vyas NK, Vyas MN, Chervenak MC, Bundle DR, Pinto BM, Quijcho FA. *Proc Natl Acad Sci USA*. 2003; 100:15023–15028. [PubMed: 14645714]
46. Agostino M, Sandrin MS, Thompson PE, Ramsland PA, Yuriev E. *Pep Sci*. 2011; 96:193–206.
47. Agostino M, Sandrin MS, Thompson PE, Yuriev E, Ramsland PA. *Glycobiol*. 2010; 20:724–735.
48. Agostino M, Sandrin MS, Thompson PE, Yuriev E, Ramsland PA. *Glycobiol*. 2009; 47:105–115.
49. Agostino M, Jene C, Boyle T, Ramsland PA, Yuriev E. *J Chem Inf Model*. 2009; 49:2749–2760. [PubMed: 19994843]
50. Lee M, Lloyd P, Zhang X, Schallhorn JM, Sugimoto K, Leach AG, Sapiro G, Houk KN. *J Org Chem*. 2006; 71:5082–5092. [PubMed: 16808494]
51. Huey R, Morris GM, Olson AJ, Goodsell DS. *J Comput Chem*. 2007; 28:1145–1152. [PubMed: 17274016]
52. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. *J Comput Chem*. 2009; 30:2785–2791. [PubMed: 19399780]
53. Trott O, Olson AJ. *J Comput Chem*. 2010; 31:455–461. [PubMed: 19499576]

54. Woods, R.J., editor. WoodsGroup. Complex Carbohydrate Research Center, The University of Georgia; Athens, GA: 2005–2013.
55. AmberTools Version 1.0. 2008. p. 29
56. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. *J Mol Biol.* 1977; 112:535–542. [PubMed: 875032]
57. Singh UC, Kollman PA. *J Comput Chem.* 1984; 5:129–145.
58. Gasteiger J, Marsili M. *Tetrahedron Letters.* 1978; 19:3181–3184.
59. Humphrey W, Dalke A, Schulten K. *J Mol Graphics.* 1996; 14:33–38.
60. Zdanov A, Li Y, Bundle DR, Deng SJ, MacKenzie CR, Narang SA, Young NM, Cygler M. *Proc Natl Acad Sci USA.* 1994; 91:6423–6427. [PubMed: 7517550]
61. Bundle DR, Baumann H, Brisson JR, Gagne SM, Zdanov A, Cygler M. *Biochemistry.* 1994; 33:5183–5192. [PubMed: 8172893]
62. van Roon AMM, Pannu NS, de Vrind JPM, van der Marel GA, van Boom JH, Hokke CH, Deelder AM, Abrahams JP. *Structure.* 2004; 12:1227–1236. [PubMed: 15242599]
63. Vyas NK, Vyas MN, Chervenak MC, Johnson MA, Pinto BM, Bundle DR, Quirocho FA. *Biochemistry.* 2002; 41:13575–13586. [PubMed: 12427018]
64. Ramsland PA, Farrugia W, Bradford TM, Mark Hogarth P, Scott AM. *J Mol Biol.* 2004; 340:809–818. [PubMed: 15223322]
65. Martin AC, Cheetham JC, Rees AR. *Proc Natl Acad Sci.* 1989; 86:9268–9272. [PubMed: 2594766]
66. Martin ACR, Cheetham JC, Rees AR. *Methods Enzymol.* 1991; 203:121–153. [PubMed: 1762554]
67. Wu TT, Kabat EA. *J Exp Med.* 1970; 132:211–250. [PubMed: 5508247]
68. Chothia C, Lesk AM. *J Mol Biol.* 1987; 196:901–917. [PubMed: 3681981]
69. Frisch, MJ.; GWT; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, GA.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, HP.; Izmaylov, AF.; Bloino, J.; Zheng, G.; Sonnenberg, JL.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, JA., Jr; Peralta, JE.; Ogliaro, F.; Bearpark, M.; Heyd, JJ.; Brothers, E.; Kudin, KN.; Staroverov, VN.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, JC.; Iyengar, SS.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, JM.; Klene, M.; Knox, JE.; Cross, JB.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Martin, RL.; Morokuma, K.; Zakrzewski, VG.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Dapprich, S.; Daniels, AD.; Farkas, Ö.; Foresman, JB.; Ortiz, JV.; Cioslowski, J.; Fox, DJ. Gaussian Inc; Wallingford CT: 2009.
70. Hill A, Reilly PJ. *J Comput Chem.* 2008; 29:1131–1141. [PubMed: 18074341]
71. French, AD.; Tran, VH.; Pérez, S. In *Computer Modeling of Carbohydrate Molecules*. French, AD.; Brady, JW., editors. American Chemical Society; Washington, DC: 1990. p. 191-212.
72. French AD, Johnson GP, Cramer CJ, Csonka GI. *Carbohydr Res.* 2012; 350:68–76. [PubMed: 22265378]
73. Thomas Williams, CKea. <http://www.gnuplot.info>
74. Lütke T, Frank M, von der Lieth CW. *Nucl Acids Res.* 2005; 33(suppl 1):D242–D246. [PubMed: 15608187]
75. Cygler M, Rose DR, Bundle DR. *Science.* 1991; 253:442–445. [PubMed: 1713710]
76. Le Guilloux V, Schmidtke P, Tuffery P. *BMC Bioinform.* 2009; 10:168.
77. Chang MW, Ayeni C, Breuer S, Torbett BE. *PloS One.* 2010; 5:9.

Figure 1.

(a) Illustration of an antibody with its variable fragment (Fv) aligned to the grid box. The yellow dot represents the CoM of the CDRs (0,0,0), and the green dot represents the center of the grid box (0,0,11). (b) Aligned orientation of an antibody antigen-binding fragment (Fab) in complex with the antigen (yellow), with respect to the internal reference axes. The region in red + pink represents the V_H domain (CDRs (red) and framework regions (pink) of the heavy chain) of the antibody, while the region in blue represents the V_L domain (CDRs (dark blue) and framework regions (cyan) of the light chain). The X-axis for the alignment was defined by a vector passing through the CoM of the variable light chain (V_L domain, which contains the light chain CDRs and framework sequences), and the CoM of the variable heavy chain (V_H domain). The Z-axis was defined as a vector normal to the X-axis, and passing through the CoM of the entire variable region, or variable fragment (Fv). The antibody was then translated so that the CoM of the CDRs was placed at the origin. The Y-axis was defined as a vector perpendicular to the XZ-plane, and passing through the origin. The docking grid box was aligned to the internal co-ordinate axes with its center offset from the origin by 11 Å along the Z-axis, so as to optimally encompass the CDR loops, while also permitting adequate volume for the movement of the ligand during docking. Such a definition enabled the docking grid box to be consistently aligned with respect to the CDRs.

Figure 2. PRMSD and SRMSD calculation. Shown in (a) and (b) are the PRMSD and SRMSD, respectively, of a representative docked pose with respect to its crystal ligand. (a) The PRMSD is the RMSD between the ring atoms of a representative docked structure (pink) and the corresponding crystal structure (green). (b) The SRMSD is the RMSD value obtained after the docked structure (pink) is superimposed on the crystal structure (green).

Figure 3.

The ϕ and ψ angle distributions from 100 docked structures, for selected linkages, as indicated by the dashed rectangle. Data are presented, in order, for AD3 (black bars), AD4.2 (white bars) and ADV (grey bars). The bin containing the experimentally-determined values is highlighted with a light blue outline. The bin containing the structure with the lowest docked energy is indicated as follows: AD3, yellow; AD4.2, orange; ADV, green.

Figure 4.
Representation of the 8 model disaccharides pertinent to the development of CHI energy functions.

Figure 5. Individual (solid colored lines) and average (dotted black line) rotational energy curves for models (see Figure 4) whose linkages have similar local geometries.

Figure 6. Comparison of the CHI energy functions (solid line) to the glycosidic torsion angle distributions of carbohydrates from experimental co-crystal structures (histograms).

Figure 7.

Scatter plots demonstrating improvement in the linear correlation between SRMSD and docked energies after rescoring, for each of the three docking programs. Points before rescoring are shown in dark grey and points after rescoring are shown in light grey. Shown in the insets are SRMSD vs. docked energy plots of only the overall lowest PRMSD structure for each of the six antibody systems before (dark grey) and after (light grey) rescoring. The black rectangles in all insets enclose plot areas with SRMSD ≤ 1 Å and energies ≥ 0 kcal/mol.

Figure 8.

Graphs showing the distribution of conformations produced by AD3 (●), AD4.2 (▲) and ADV (■) plotted onto the corresponding CHI energy curves for each of the representative linkage combinations; the curves are offset from each other by 6 kcal/mol.

Figure 9.

(a) SRMSDs of the lowest energy poses for all six systems from AD3, AD4.2 and ADV, before (dark grey) and after (light grey) rescoring. (b) PRMSDs of the lowest energy poses for all six systems from all three docking programs, before (dark grey) and after (light grey) rescoring.

Figure 10.

(a) AD3 lowest energy pose for 1S3K before rescoring (orange) compared to the crystal ligand (green); PRMSD = 5.7 Å. (b) Lowest energy pose after inclusion of the CHI energy (pink) compared to the crystal ligand (green); PRMSD = 0.6 Å.

Figure 11.

Docking the trisaccharide to the Salmonella antibody (in 1MFD and 1MFA). (a) Lowest energy pose from ADV for 1MFD before rescoring (orange) compared to the crystal ligand (green); PRMSD = 5.5 Å. (b) Lowest energy pose from ADV for 1MFD after rescoring (pink) compared to the crystal ligand (green); PRMSD = 1.0 Å. (c) and (d) show the 1MFD antibody in transparent surface representation along with the oxygen atom belonging to the water molecule from the crystallographic co-complex, WAT 601; in (c) the crystal ligand from 1MFD is shown in CPK representation, and in (d) the lowest energy pose from ADV for 1MFD before rescoring (in CPK representation) showing the Gal residue replacing Abe within the binding pocket is shown. (e) The Gal residue from the ligand in 1MFD (in van der Waals representation) after being superimposed onto the Abe residue from the ligand in 1MFA is shown within the 1MFA binding site. A cross-section of the 1MFA antibody is represented as a transparent surface with potential steric clashes visible between the Gal residue and the antibody. (f) Same as (e) but with the 1MFA antibody represented as an opaque surface thus more clearly depicting potential steric clashes between the O-3 and O-6 groups of the Gal residue and the interior of the binding pocket.

Figure 12.

Docking to the antibody in 1M7I using AD4.2. (a) Lowest energy pose before rescoring (orange) compared to the crystal ligand (green); PRMSD = 3.9 Å. (b) Lowest energy pose after rescoring (pink) compared to the crystal ligand (green); PRMSD = 10.7 Å.

Overall, after including the CHI energies, ADV delivered the best results amongst the three docking programs tested in this study as established by the SRMSD and PRMSD values of top-ranked structures (Figure 9). The SRMSDs of all lowest energy poses from ADV after rescoring were 1.1 Å, with all corresponding PRMSDs being 2 Å.

Table 1

PDB IDs and ligand sequences employed in the study, including the shape RMSD (SRMSD) values for the ligands generated by GLYCAM, relative to the crystallographic ligands.

PDB ID: Chain ID (Resolution) ^a	Ligand (average B-factor) ^b	Graphic representation of the ligand	SRMSD ^{a,b}	Biological Origin
1MFA ^{69,d} : L/H (1.7)	DAbe α 1-3[DGal α 1- 2]DMan α -OMe (25.1)		0.6	<i>Mus musculus</i>
1MFD ^{70,d} : L/H (2.1)	DAbe α 1-3[DGal α 1- 2]DMan α -OMe (30.1)		0.5	<i>Mus musculus</i>
1UZ8 ⁷¹ : A/B (1.8)	DGal β 1-4[LFuc α 1- 3]DGlc β NAc β -OMe (41.8)		0.3	<i>Mus musculus</i>
1M7D ⁷² : A/B (2.3)	LRhap α 1-3(2- deoxy)LRhap α 1- 3DGlc β NAc β -OMe (39.8)		0.3	<i>Mus musculus</i>
1S3K ⁷³ : L/H (1.9)	LFuc α 1-2DGal β 1- 4[LFuc α 1- 3]DGlc β NAc α -OH (26.6)		0.4	<i>Homo sapiens</i> , <i>Mus musculus</i>
1M7I ⁷² : A/B (2.5)	LRhap α 1-2LRhap α 1- 3LRhap α 1- 3DGlc β NAc β 1- 2LRhap α -OMe (35.4)		1.1	<i>Mus musculus</i>

● = Mannose (Man)

● = Galactose (Gal)

▲ = Fucose (Fuc)

▴ = 2-Deoxy Rhamnose

● = Abequose (Abe)

■ = N-Acetyl Glucosamine(GlcNAc)

▴ = Rhamnose (Rha)

┐ = Aglycon (OME/OH)

^a In Å.

^b In Å².

^c SRMSD defined in Section *Shape, and pose, RMSD values*.

^d 1MFA and 1MFD, consisted of the trisaccharide antigen from *Salmonella* serotype B. In 1MFD, the trisaccharide is bound to a Fab antibody fragment, while in 1MFA the trisaccharide is bound to a single-chain Fv fragment of the antibody. Although the antigen-binding site in both the Fab and scFv fragments are essentially the same, and bound to the same trisaccharide antigen, in the Fv-complex a water molecule has become inserted into an internal hydrogen bond within the trisaccharide, leading to a perturbation of the trisaccharide conformation⁶⁹.