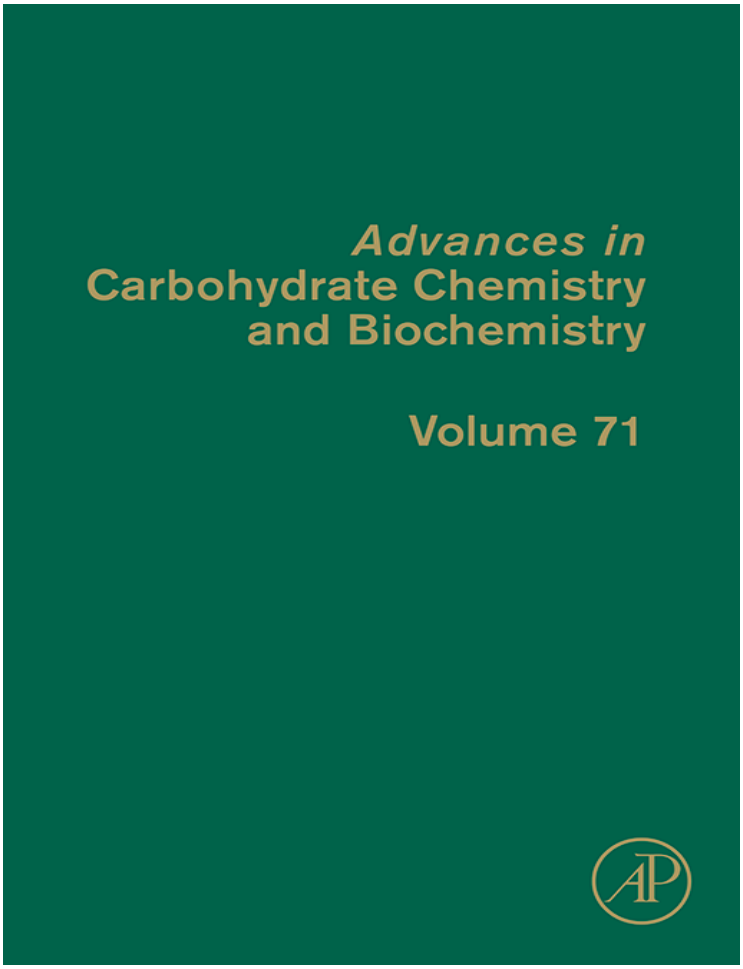

From Serge Pérez and Igor Tvaroška, Carbohydrate–Protein Interactions: Molecular Modeling Insights. In: Derek Horton, editor, *Advances in Carbohydrate Chemistry and Biochemistry, Vol. 71*, Burlington: Academic Press, 2014, pp. 9-136.
ISBN: 978-0-12-800128-8

ADVANCES IN CARBOHYDRATE CHEMISTRY AND BIOCHEMISTRY, VOL. 71

CARBOHYDRATE–PROTEIN INTERACTIONS: MOLECULAR MODELING INSIGHTS

SERGE PÉREZ[a,1] and IGOR TVAROŠKA[b,c,1]

[a]Department of Molecular Pharmacochemistry, CNRS, University Grenoble-Alpes, Grenoble, France
[b]Department of Chemistry, Slovak Academy of Sciences, Bratislava, Slovak Republic
[c]Department of Chemistry, Faculty of Natural Sciences, Constantine The Philosopher University, Nitra, Slovak Republic

I. Introduction 12
II. Specific Features of Carbohydrate Modeling 15
1. Nomenclature and Structural Representation 15
2. Stereoelectronic Effects 17
3. Conformational Flexibility 21
III. Experimental Data on Protein–Carbohydrate Interactions 22
1. Crystallography 22
2. Nuclear Magnetic Resonance 26
3. Thermodynamics and Other Biophysical Methods 26
IV. Features of Protein–Carbohydrate Interactions 29
1. van der Waals and Electrostatic Interactions 29
2. CH/π Interactions 29
3. Solvation–Desolvation 30
V. Computational Tools for Docking Carbohydrates on Proteins 30
1. Force Fields Designed for Carbohydrates 30
2. QM Methods 35
3. QM/MM Methods 38
4. MD Methods 41
5. Free-Energy Calculations 43
6. Metadynamics Calculations 45
7. Molecular Robotics 47
8. Molecular Docking 49
VI. Carbohydrate Biosynthesis and GTs 51
1. Inverting GTs 52
2. Retaining GTs 63
3. Transition-State Structures 68

[1]Corresponding authors: E-mail: serge.perez@cermav.cnrs.fr; igor.tvaroska@savba.sk

ISBN: 978-0-12-800128-8
http://dx.doi.org/10.1016/B978-0-12-800128-8.00001-7.

VII. Recognition 71
1. Lectins 71
2. Antibodies 78
3. GAG-Binding Proteins 83
4. Transport 86
VIII. Glycoside Hydrolases 89
1. GHs on a Single Carbohydrate Chain 89
2. Retaining GHs 95
3. Inverting GHs 99
4. Beyond the Catalytic Mechanism 100
5. GHs on a Solid Substrate: The Cellulase Case 104
IX. Conclusion and Perspectives 107
Acknowledgments 110
References 110

ABBREVIATIONS

AMBER, assisted model building with energy refinement, package of molecular simulation programs; AutoDock, automated docking, a molecular modeling simulation program; B3LYP, Becke, 3-parameter, Lee–Yang–Parr, hybrid exchange-correlation functional; B3LYP/6-31+G*, Becke, three-parameter, Lee–Yang–Parr, hybrid exchange-correlation functional using 6-31+G* Pople split-valence double-zeta basis set; BiOligo, a database of three-dimensional structures of bioactive oligosaccharides; CBM, cellulose-binding module; CC, coupled clusters; CCSD(T), coupled clusters calculations, using double substitutions both single and double substitution, in which the triples contribution is calculated approximately; CHARMM, Chemistry at Harvard Molecular Mechanics program; COMFA, Comparative Molecular Field Analysis; CPMD, Car–Parrinello molecular dynamics; CV, collective variable; DC-SIGN, dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin; DFT, density-functional theory; DFT-D, density functionals with Grimme's corrections for van der Waals interactions; DOCK, docking program; EHiTs, a software for molecular docking; ELLA, enzyme-linked lectin assay; ESP, electrostatic potential; FGFs, fibroblast growth factors; FimH, fimbriae protein; FlexX, software package to predict protein–ligand interactions; GAG, glycosaminoglycan; GBSA, generalized Born and surface area solvation; GGA, generalized gradient approximation; GH, glycoside hydrolase; Glide, docking program from Schrödinger; GLYCAM06, a generalizable biomolecular force field; GOLD, software package to predict protein–ligand interactions; GPI, glycosylphosphatidylinositol;

GRID, force field and associated software with main application in medicinal chemistry; GROMOS-45A4, Groningen Molecular Simulation computer program package; GT, glycosyltransferase; HBGAs, human histo-blood group antigens; HF, Hartree–Fock; HIA, inhibition of hemagglutinin assay; ITC, isothermal titration calorimetry; IUPAC–IUBMB, International Union of Pure and Applied Chemistry–International Union of Biochemistry and Molecular Biology; KIE, kinetic isotope effect; LecB, *Pseudomonas aeruginosa* lectin II; LgtC, lipopolysaccharide β-1,4-galactosyltransferase C; LSDA, local spin-density approximation; M05-2X, empirical exchange-correlation functionals developed by Zhao and Truhlar; M08-SO, empirical exchange-correlation functionals developed by Zhao and Truhlar; MFS, major facilitator superfamily; MM3, molecular modeling force field; MOLCAD, a software package; MP, Moller–Plesset; MPW1K, exchange-correlation functionals developed by Zhao and Truhlar; MPWB1K, exchange-correlation functionals developed by Zhao and Truhlar; NA, viral neuraminidase; NMR, nuclear magnetic resonance; NOESY, standard nuclear Overhauser effect spectroscopy; NVE, the microcanonical ensemble, in which the system is isolated from changes in moles (N), volume (V), and energy (E); NVT, the canonical ensemble, in which the system is isolated from changes in moles (N), volume (V), and temperature (T); OGT, O-linked β-*N*-acetylaminyltransferase; OPLS-AA, optimized potentials for liquid simulations force field; OPLS-AASEI, an improved version of the original OPLS-AA force field for carbohydrates; OtsA, trehalose-6-phosphate synthase; PA-IL, *Pseudomonas aeruginosa* lectin I; PARM94, AMBER force-field parameters; PBSA, molecular mechanics/Poisson–Boltzmann surface area method; PES, potential energy surface; PM3CARB, parameterized model number 3 is a semiempirical method parameterized for carbohydrates; POPE, phosphatidylethanolamine (PE) and phosphatidylglycerol (PG) bilayer; QCI, quadratic configuration interaction; QM, quantum mechanics; QM(DFT)/MM, quantum mechanics (DFT)–molecular mechanics; QM/MM, quantum mechanics–molecular mechanics; QPLD, quantum method-polarized ligand docking; QSAR, quantitative structure–activity relationship; RESP, restrained electrostatic potential method; RMSD, root mean-square deviation; RSL, *Ralstonia solanacearum* lectin; SCF, self-consistent field; SPR, surface plasmon resonance; STDR-NMR, saturation transfer difference NMR; SWEET-II, SWEET is a program for constructing 3D models of saccharides from their sequences using standard nomenclature; SYBYL, software package of TRIPOS; TI, thermodynamic integration; TIP3P, three-site water model for molecular dynamics simulations; TrNOE, transferred nuclear Overhauser effect; UDP, uridine 5′-diphosphate; UDP-GlcNAc, uridine 5′-(2-acetamido-2-deoxy-α-D-glucopyranosyl diphosphate)

I. Introduction

In Nature, carbohydrates form an important family of biomolecules. Carbohydrates, in the form of polysaccharides, glycopeptides, glycolipids, glycosaminoglycans (GAGs), proteoglycans, or other glycoconjugates, have long been known to participate in many biological processes. Most of the earlier studies on carbohydrates focused on such plant polysaccharides as cellulose, starch, pectins, and the like, largely because of their wide range of functionalities and subsequent applications. Living cells, from bacteria to mammals, are covered with a carbohydrate coating (glycocalyx) that presents the first information about the cell to the outside environment. Carbohydrate–carbohydrate, carbohydrate–protein, and carbohydrate–nucleic acid interactions constitute the molecular basis of viral entry, signal transduction, inflammation, cell–cell interactions, bacteria–host interactions, fertility, and development, among other processes. The discovery that a large part of the biological information is encoded in carbohydrate structures (or glycocode) rapidly became a central concept in glycobiology, promoting the beginning of the modern biological discipline termed glycomics, which explores the role of carbohydrates in biological processes. Deconvoluting the importance of sugars in these biological events is a major challenge, owing to various factors. These include the structural complexity of the glycans, the complex biosynthesis of the sugar component of glycoproteins (a pathway not under direct gene control), the multivalent nature of biological recognition of the glycan, and the subtle phenotypes of glycan manipulation that often require multicellular environments to manifest. Of particular interest are the carbohydrate-mediated recognition events that are important in biological phenomena, and this accords a central role to the study of protein–carbohydrate interactions. The binding partners of carbohydrates encompass a wide range of macromolecules, involved in such functions as recognition, biosynthesis, modification, hydrolysis, and others (Fig. 1). The interactions between proteins and carbohydrates play a role in numerous biological processes, such as protein specificity in antibody–antigen recognition, cell–cell adhesion, enzyme–substrate specificity, molecular transport, and others. They play an essential role in the onset, detection, and, potentially, also the prevention of such human diseases as cancer, inflammation, diabetes, neurodegenerative diseases, and bacterial and viral infections. The interactions between proteins and carbohydrates are also involved in the biosynthesis and biodegradation of the principal polysaccharide raw materials on Earth.

Determination of the three-dimensional (3D) structural and dynamical features of complex carbohydrates, carbohydrate polymers, and glycoconjugates, along with an understanding of the molecular basis of their associations and interactions, constitute the main challenges in structural glycoscience.[1]

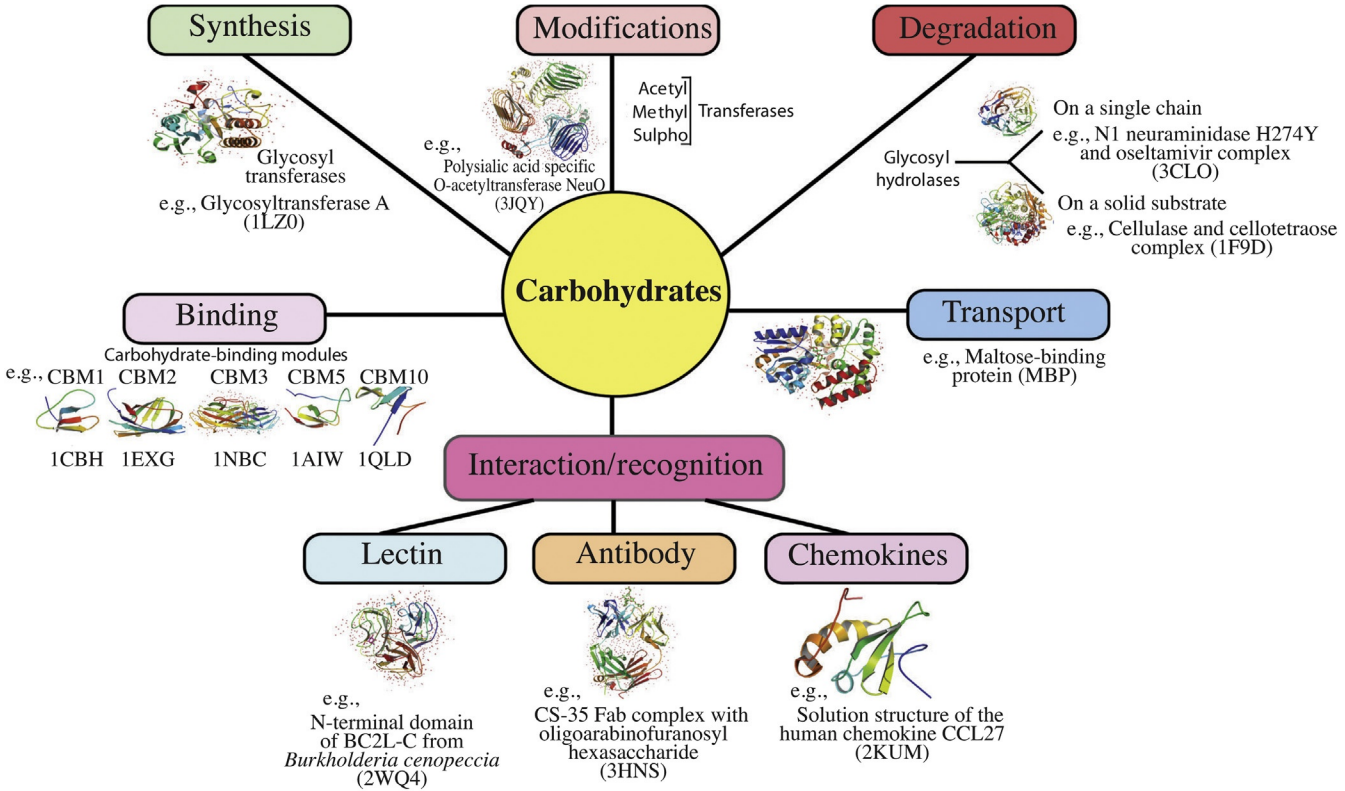

FIG. 1. Synopsis of the families of proteins interacting with carbohydrates, along with three-dimensional depictions of some representative crystal structures of proteins taken from the Protein Data Bank (PDB). The functions of the proteins are for synthesis,[478] modification (acetyltransferase[479]), and hydrolysis (a) of a single oligo-/polysaccharide chain[480] and (b) of a (semi)crystalline polysaccharide substrate.[481] Binding throughout carbohydrate modules (CBMs),[482–486] transport,[487] and interaction/recognition: (a) lectin,[488] (b) antibody,[489] and (c) chemokines.[490] Reprinted with permission from Ref. 491. Copyright 2013 CRC Press, Taylor & Francis.

Elucidation of the 3D structures and the dynamic properties of oligosaccharides is a prerequisite for a better understanding of the relationships between structure and function, involving the biochemistry of recognition processes and the subsequent rational design of carbohydrate-derived drugs. It is evident that elucidation and understanding of the different structural levels of polysaccharides are required to relate structure to properties. Furthermore, some polysaccharides are also carriers of biological information that can be deciphered only if their interactions with other biological macromolecules are understood. Unfortunately, oligosaccharides, either in their free form or as components of glycoconjugates, are inherently difficult to crystallize, and therefore structural data from X-ray studies are limited.[2,3] In solution, the flexibility of individual glycosidic linkages leads to multiple conformations that coexist in equilibrium. Application of various spectroscopic methods, with proper time resolution, is required for analysis of the conformational behavior of such molecules.[4,5] As regards polysaccharides, they differ from other biological macromolecules because the diffraction data that can be obtained from them are not sufficiently detailed to permit crystal structure elucidation based on the data alone. Hence, procedures for molecular modeling of carbohydrates and carbohydrate polymers have been devised as important tools for structural studies of these compounds. The 3D structures of polysaccharides established from various experimental methods (X-ray, neutron, and electron diffraction, molecular modeling, and high-resolution NMR spectroscopy) have been incorporated into an annotated database, Polysac3DB.[6] Various methods for molecular modeling have been developed[7] and have been widely used for the determination of oligosaccharide and polysaccharide conformations.[3,8] The progress made in algorithms and computational power allows for the simulation of carbohydrates in their natural environment, namely, solvated in water or in an organic solvent, in concentrated solution. These developments along with their applications have been thoroughly reviewed in a previously published article.[7]

Experimental assessment of carbohydrate–protein interactions by X-ray crystallography is impeded by difficulties of cocrystallizing proteins and carbohydrates. Nevertheless, highly resolved protein–carbohydrate complexes gathered by using X-ray synchrotron techniques have accumulated to the point where it has been possible to compare the experimentally derived structures with those predicted from computational methods. Some general features governing the protein–carbohydrate interactions have been delineated, and computational tools have evolved and been improved accordingly. These tools provide efficient ways to increase our understanding of the various contributions to the binding energy. These developments allow efficient searching of the conformational space and yield reliable estimates of the binding

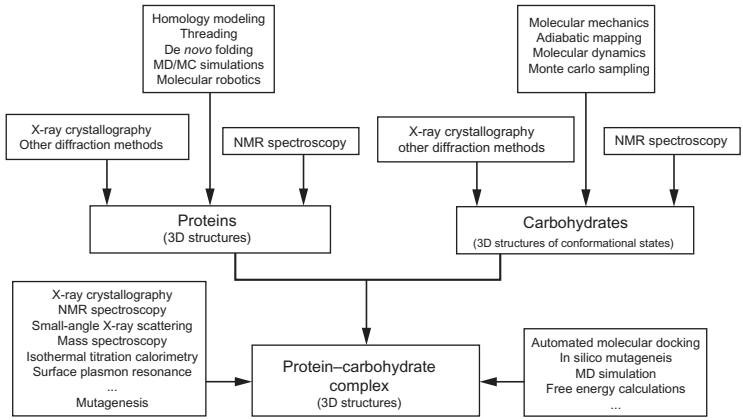

FIG. 2. The interplay between experimental and theoretical methods required to decipher the structural basis of protein–carbohydrate interactions.

free energy. They allow exploring *in silico* examples where the experimental data are lacking, and provide sound structural information for the rational design of bioactive carbohydrates or carbohydrate mimetics.

In this article, we survey the significant contributions to, and the current status of, the application of computational methods to the characterization and prediction of protein–carbohydrate interactions. It takes into account the interplay between several experimental and theoretical methods, as is required for unraveling the structural basis underlying the diverse biological roles of carbohydrates (Fig. 2).

II. Specific Features of Carbohydrate Modeling

1. Nomenclature and Structural Representation

Carbohydrates may be divided into several categories: monosaccharides, oligosaccharides, polysaccharides, and compounds derived from monosaccharides. The monosaccharides are the monomeric constituents of glycan chains. Oligosaccharides were once defined as carbohydrates that comprise 2–10 monosaccharide units, linked together in a linear or branched way, while the term “polysaccharide” was used for glycan chains built up from more than 10 monosaccharides. Nowadays the term oligosaccharide generally denotes a single defined structure, while a polysaccharide

may be a polymer-homologous series of structures based on a particular repeating unit, but no discrete molecular weight. The collective term glycoconjugate covers a large family of substances the main types being glycoproteins, glycopeptides, peptidoglycans, proteoglycans, GAGs, glycolipids, and lipopolysaccharides. Glycoproteins are proteins bearing covalently attached glycan chains that result from a cotranslational or posttranslational modification. In the N-glycoproteins, the glycan chain is attached through the amide nitrogen atom of the side chain of asparagine, whereas O-glycoproteins have the glycan chain attached to oxygen of the hydroxyl group on the side chain of hydroxylysine, hydroxyproline, serine, or threonine. Glycolipids are glycosyl derivatives of such lipids as acylglycerols, ceramides, and prenols. The term glycome may be defined as the entire set of glycans in an organism, which may vary from mono- to poly-saccharides, either free or linked to such aglycone moieties as proteins or lipids.

Carbohydrates have a potential information content that is several orders of magnitude higher than any other biological macromolecule (Table I). This variety of carbohydrate structures results from the wide range of different monosaccharides (>100 if all variations are considered) from which they are composed, and the different ways in which these monomers can be glycosidically joined. Thus, even a small number of monosaccharide units can provide a large number of different oligosaccharides (also termed glycans), including branched structures (a unique feature among biomolecules). For example, the number of all possible linear and branched isomers of a hexasaccharide exceeds[9] 10^{12}.

The mechanism of carbohydrate recognition depends on (i) the nature of the individual monosaccharides in the glycan (thus glucose versus mannose), (ii) the configuration at the anomeric centers (thus α or β), and (iii) the inter-residue linkage positions (thus $1\rightarrow3$ versus $1\rightarrow4$), and (iv) chemical modifications of the

TABLE I
Comparison of the Possible Structural Isomers for Nucleic Acids, Polypeptides, and Carbohydrates Found in Mammals[a]

Size	Nucleotides	Peptides	Carbohydrates
1	4	20	20
2	16	400	1360
3	64	8000	126,080
4	256	160,000	13,495,040
5	1024	3,200,000	1,569,745,920

[a] The numbers are calculated considering both the α and β configurations for the 10 most common mammalian monosaccharides, in their pyranose form.

monosaccharide components (amination, oxidation, sulfation, phosphorylation, methylation, acetylation, and others).

The strength of this interaction is also determined by the conformation of the carbohydrate and its orientation with respect to the binding site.

Various methods have been formulated to depict carbohydrate sequences in a linear code format.[10–13] For a text-based representation of such sequences, an earlier tentative 1971 IUPAC document[14] has evolved into the currently accepted 1996 IUPAC–IUBMB recommendations.[15] The latter document provides a "standard" or "extended" version that employs three-letter abbreviations for the monosaccharide components, along with ring-size designators, anomeric and configurational symbols, and the inter-residue linkage positions. An alternative "short" form compresses the sequence by omitting configurational symbols and ring-size designators; the latter have to be understood within the context.

These forms are widely adopted within the carbohydrate community and are satisfactory for describing complex sugar sequences. The recommendations are also applicable for the depiction of polysaccharides, glycoproteins, and other glycoconjugates, but become unwieldy for the depiction of large structures. For the latter, various "cartoon"-type representations for the individual monosaccharide components have been proposed.[16,17] These are well accepted in the glycoprotein field, where the number of individual monosaccharide types is small, but extension to such areas as bacterial polysaccharides, where a large range of monosaccharide constituents is encountered, presents significant difficulties. An IUPAC–IUBMB panel working on an update to the 1996 recommendations is expected to address both the use of iconic symbols in depicting saccharide sequences as well as encoding procedures for such sequences applicable in the field of glycoinformatics.

Various ways of depicting saccharide sequences in complex carbohydrates are shown in Fig. 3.

2. Stereoelectronic Effects

Carbohydrate molecules are considered to be especially difficult to model because of their highly polar functionalities and the differences in electronic arrangements, such as the anomeric and *exo*-anomeric effect and *gauche* effects that occur during conformational changes and configurational variation. The tendency of electronegative substituents at the anomeric carbon atom, C-1, of an aldopyranose ring to populate the axial position more highly than could be expected from analogy with cyclohexane derivatives is termed the anomeric effect.[18] The name had been extended

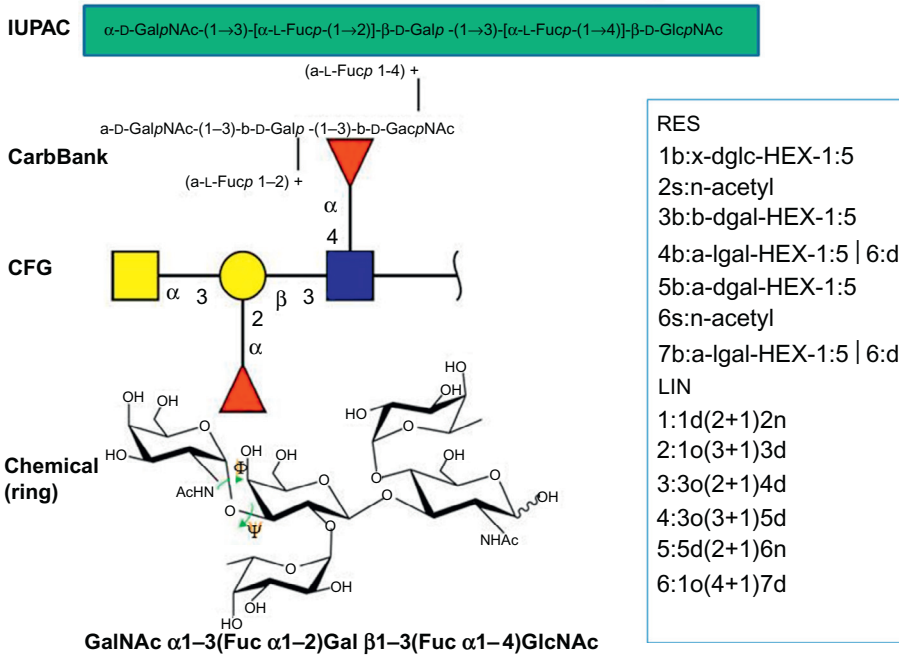

FIG. 3. Nomenclature and popular schematic representations of oligosaccharides.

to describe the preference of the C—X dihedral angle of R—X—C—Y fragments for synclinal (*gauche*) over antiperiplanar (*trans*) orientation, when X = N, O, or S and Y = Br, Cl, F, N, O, or S, and it has been termed the generalized anomeric effect.[19] The orientational preference of the aglycon around the glycosidic C-1—O-1 glycosidic bond was called the *exo*-anomeric effect[20] and the reverse anomeric effect[21]. Similarly, the *gauche* effect was defined[22] as the tendency for a molecule to adopt the structure with a number of synclinal interactions between adjacent electron pairs or polar bonds in a molecular segment X—C—C—Y, where X and Y are electronegative atoms (Fig. 4). The primary hydroxymethyl group in hexopyranoses is a typical example of such a segment (O-5—C-5—O-6—C-6). The *exo*-anomeric and *gauche* effects are of prime importance because of their impact on conformations around the glycosidic linkage of oligo- and polysaccharide structures. General aspects of these stereoelectronic effects have been discussed in several reviews,[23–28] and here we briefly discuss results from the molecular modeling of carbohydrates.

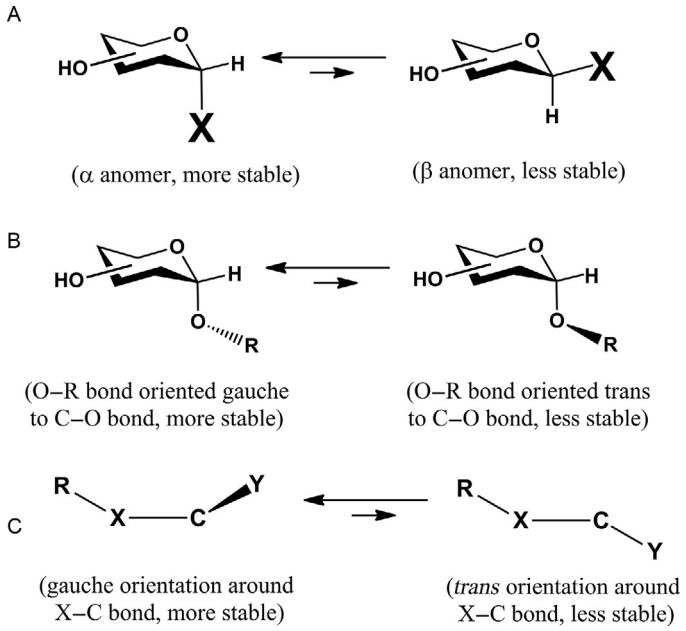

FIG. 4. Illustration of stereoelectronic effects. (A) The anomeric effect, (B) the *exo*-anomeric effect, and (C) the *gauche* effect.

a. Origin of the Anomeric Effect.—There have been different rationalizations of the origin of the anomeric effect. It is clear that any explanation should clarify both the conformational preference and variations in geometrical parameters. Two of the most accepted rationalizations are electrostatic[18] and delocalization (hyperconjugation).[29] In the first, the anomeric effect is explained in terms of electrostatic repulsions between a carbon–heteroatom dipole and the resulting dipole of electron lone pairs. These electrostatic concepts predict the stability of conformers in qualitative agreement with experimental data. For example, in 2-methoxytetrahydropyran having the methoxy group equatorial, the dipoles are oriented nearly parallel. In contrast, in the axial conformer, the dipoles are oriented quasi antiparallel. As a consequence, electrostatic repulsion is larger in the equatorial conformer. This interpretation is supported by experimental observations that polar media decreased the magnitude of the anomeric effect.[30] However, electrostatic rationalization fails to explain the equilibrium of conformers quantitatively and, more importantly, the characteristic geometrical changes between the axial and equatorial conformers.

In the second interpretation, a stabilization of the *gauche* orientation is attributed to delocalization (or hyperconjugation) of the lone-pair orbital on oxygen to antibonding orbital σ^*_{C1X} of the carbon–heteroatom (C-1—X) bond.[29,31] This delocalization interaction is maximized when the lone pair at oxygen is aligned in an antiparallel geometry with the antibonding orbital. In such an orientation, the C-1—X bond is elongated, and the O—C-1 bond shortened. It has been shown that delocalization interactions in the C—O—C—O—C moiety are conformationally dependent.[32] There is clear structural and spectroscopic evidence for delocalization interactions. For example, the C-1—H-1 equatorial bond in saccharides is shorter than the axial one, because of the better orientation of an oxygen lone pair with antibonding orbital σ^*_{C1H}. Therefore, the C-1—H-1 equatorial bond has a larger $^1J_{CH}$ coupling constant by 10 Hz as compared to C-1—H-1 in the axial position.[33,34] The structural aspects are manifested in variations between the α- and β-glycosides and have been attributed to the lone-pair interactions associated with the anomeric effect.[24,35] An important role of hyperconjugation in the anomeric effect has been nicely documented by laser spectroscopy and DFT calculations.[36] Since neither the electrostatic or delocalization explanations account for all aspects of the anomeric effect alone, it was suggested that both contribute to the anomeric effect,[37] and the balance of both contributions depends on the character of the atoms in the C—O—C—X moiety. Recently, other rationalizations have been proposed.[38–41]

b. Calculation of the Anomeric Effect.—The first QM calculations on the anomeric and *exo*-anomeric effects consisted of conformational studies on small acyclic models for the glycosidic linkage, using *ab initio* methods,[42–44] and acyclic and cyclic models using semiempirical methods (see, for instance, Refs. 24,25). Later, a variety of *ab initio* and DFT calculations were performed on the axial and equatorial anomers of the cyclic carbohydrate models, to assess the energy of the anomeric, *exo*-anomeric, reverse anomeric, and *gauche* effects.[38,45–56] In the majority of studies, 2-substituted tetrahydropyran (2-oxane) derivatives were used to avoid "contamination" of the stereoelectronic effects by interactions of hydroxyl groups present in saccharides.

The calculated energies showed that magnitude of the anomeric effect[45–47,49] decreases in the order: $Cl > F > OCH_3 > SCH_3 > OH$; and the reverse anomeric effect increases in the order $NHCH_3 > CH_2CH_3 > NH_2{}^+CH_3$. Similarly, calculations predicted that the *exo*-anomeric effect decreases in the order: $OCH_3 > NHCH_3 > OH > SCH_3$, while the $NH_2{}^+CH_3$ group did not show an *exo*-anomeric effect. It is necessary to emphasize that QM-calculated conformational energies refer to isolated molecules in vacuum. The solvent effect on conformational equilibrium can be substantial; thus, more than 77% of 2-methoxytetrahydropyran exists in the axial form in nonpolar solvents, but only 52% in water.[30] The calculations

predicted,[46,56,57] in agreement with experimental data, that a population of the axial conformer decreased with increasing solvent polarity as determined by the dielectric constant. Similarly, a population of the *trans* conformation around the glycosidic linkage C-1—O-1 was increased with increasing solvent polarity. Since an overall dipole moment of the C—O—C—O—C moiety decreases with a number of C—O bonds in the *gauche* conformation, this effect is predictable.

The calculations predicted distinct geometrical changes with rotation around the glycosidic linkage for both anomers. The C—O—C and O—C—O bond angles varied by up to 11° and the C-1—O-1 and C-1—O-5 bond lengths varied by up to 0.04 Å.[46] These variations can be rationalized by a perturbation analysis of hyperconjugation interactions of lone-pair orbitals.[32] The energy and geometrical parameters of conformers calculated for 2-substituted tetrahydropyran derivatives provided appropriate values for parameterization of force-field methods for carbohydrates (see, for example, Refs. 58–60).

3. Conformational Flexibility

Carbohydrates and their derivatives possess many hydroxyl groups and thus a large number of rotatable bonds. Due to the many hydroxyl groups, these compounds are usually water soluble and their log P values are often negative. The surface of carbohydrates and their derivatives is composed of hydrophobic and hydrophilic patches formed by nonpolar aliphatic protons and polar hydroxyl groups. This leads to anisotropic solvent densities around carbohydrate molecules. In aqueous environments, favorable interactions of water molecules with the hydrophilic patches result from electrostatic interactions and hydrogen bonding. Conversely, the interaction of water with hydrophobic patches on the surface is unfavorable. An equilibrium between hydrophobic and hydrophilic patches forms the basis for such properties as carbohydrate solubility in water or such functions as molecular recognition.

Another essential feature of carbohydrates is their conformational flexibility.[61] As compared to conventional drug-like molecules, carbohydrates are typically much more flexible. The relative orientation of two contiguous monosaccharides, linked by a glycosidic bond in a disaccharide, is characterized by the Φ and Ψ torsion angles. In the so-called Heavy Atom Definition commonly used in crystallography, Φ is the torsion angle Φ = O-5—C-1—O—C-x and Ψ is the torsion angle Ψ = C-1—O—C-x—C-x+1, where x is the number of the carbon atom of the second monosaccharide with which the $1 \rightarrow x$ glycosidic bond is formed. An alternative definition, of use in

NMR spectroscopy, refers to the hydrogen atoms about the glycosidic bond in a way such as: Φ_H = H-1—C-1—O—C-*x* and Ψ_H = C-1—O—C-*x*—H-*x*. For two monosaccharides linked by a 1 → 6 linkage, another parameter (ω) is required to describe the orientation about the exocyclic bond C-5—C-6. Its orientation is usually described by the torsion angles O-5—C-5—C-6—O-6 and C-4—C-5—C-6—O-6, which combination defines the so-called *gauche–trans* (gt), *gauche–gauche* (gg), and *trans–gauche* (tg) conformations.[62] The energetically favorable conformations of a carbohydrate dimer may be easily shown on energy plots termed (Φ, Ψ) maps (Fig. 5), which are quite similar to the Ramachandran plots used to visualize the backbone dihedral angles of the essential amino acids in proteins. These plots feature multiple minima, with the separating energy barriers being over 10–15 kcal mol^{-1}.

However, carbohydrates in complexes have been found to adopt conformations belonging to different minima. Figure 6 depicts the distributions of glycosidic torsion angles within two disaccharide segments: αMan*p*-(1 → 3)-Man*p* and βGlcNAc*p*-(1 → 2)-Man*p*, as experimentally observed in crystal structures of protein–carbohydrate complexes. These observations emphasize the necessity for careful sampling of the conformational space of carbohydrate oligomers during docking. While this may be possible for glycosidic bonds, the number of degrees of freedom increases rapidly when, in addition to this, the orientation of the hydroxyl groups is taken into account.

III. Experimental Data on Protein–Carbohydrate Interactions

1. Crystallography

It is estimated that over 50% of all eukaryotic proteins are glycosylated,[63] but only 5% of the 3D structures deposited in such structural databases as Protein Data Bank[64] include proteins having N- or O-linked carbohydrates. Only around 7% of all PDB entries contain information on protein–carbohydrate systems (with the carbohydrate being covalently or noncovalently bound to the protein). There are even fewer examples of high-resolution structures where the associated carbohydrate components have been fully resolved.[65] Despite there being such a limited amount of experimental information, some general features of protein–carbohydrate complexes can be established. In general, carbohydrate-binding sites are at the surface of proteins and form cavities or grooves. Most of the amino acids participate in binding to carbohydrates, although there is a frequent overrepresentation of amino acids having polar, charged, and aromatic side chains linking the carbohydrate (Fig. 7). In particular,

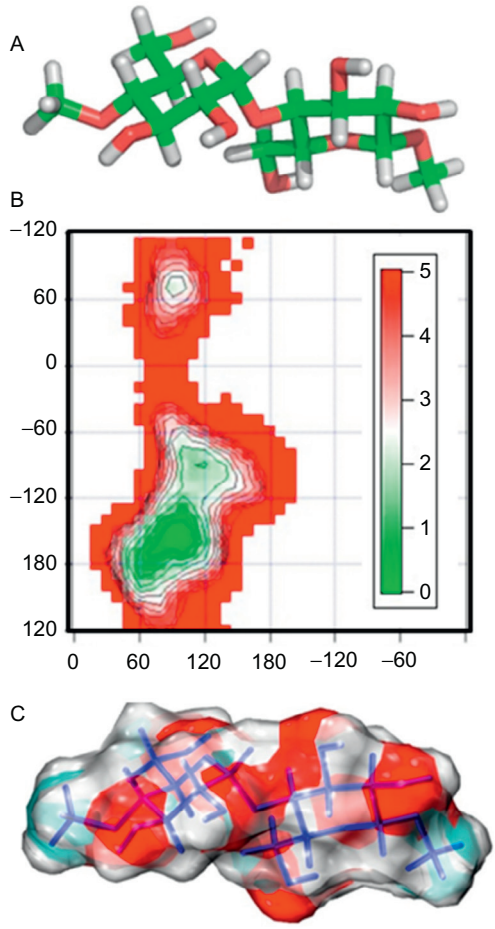

FIG. 5. Molecular representation of the disaccharide αGlc*p*-(1 → 4)-Glc*p* along with the torsion angles Φ and Ψ at the glycosidic linkage. The potential energy surface shows conformational energy with respect to the Φ and Ψ torsion angles. The favored low-energy Φ/Ψ combinations are shown in light color, whereas the high-energy regions are shown in red and the inaccessible regions are shown in white. The surface of the disaccharide is composed of hydrophobic (green) and hydrophilic (red) patches, formed by nonpolar aliphatic protons and polar hydroxyl groups.

hydrophobic interactions, typically between the aromatic residues of Tyr and Trp side chains and the hydrophobic faces of carbohydrate rings, contribute to the affinity of carbohydrate–protein interactions through the CH–π effects (see Section IV.2).

Two common topologies of binding sites have been repeatedly observed in the crystal structures of protein–carbohydrate complexes: an end-on insertion and a

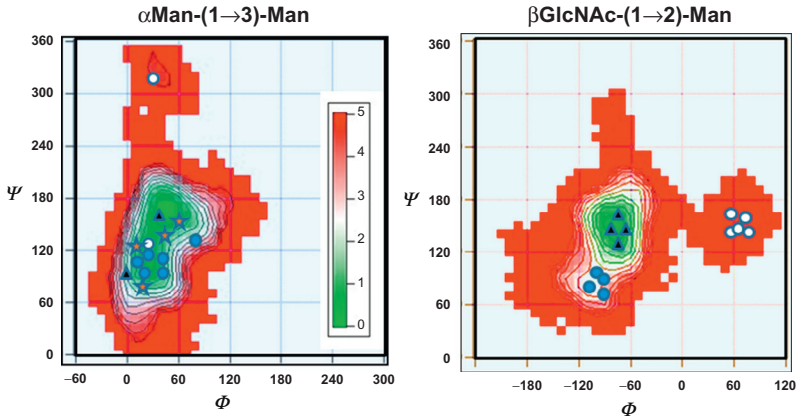

FIG. 6. Isoenergy contours of three disaccharide segments as calculated with the MM3 program, along with the glycosidic conformations observed in crystalline complexes with lectins. (A) αMan*p*-(1→3)-Man*p*; (B) βGlcNAc*p*-(1→2)-Man*p*; and (c) αNeuAc*p*-(2→3)-Gal*p*. In the case of the αMan*p*-(1→3)-Man*p* segment, the observed conformations are essentially located around a Φ value of 80°, with an excursion of Ψ in the vicinity of 330°. More interesting is the observation that a remote low-energy area (located at $\Phi = 90°$ and $\Psi = 310°$) can be occupied, as observed in the crystalline complex between *Lathyrus ochrus* and a biantennary glycan.[492] The study of the dispersion of conformations observed for the disaccharide segment βGlcNAc*p*-(1→2)-Man*p* provides another illustration of the occurrence of conformations in remote energy wells of the potential energy surfaces. The location of this well is 120° away from what would correspond to the stable conformation driven by the *exo*-anomeric effect in the case of an equatorially oriented linkage. Such examples are observed in the crystalline complexes involving the isolectin II of *L. ochrus*, complexed with high-molecular-weight oligosaccharides, such as a biantennary octasaccharide,[492] a glycopeptide, or an N2 fragment of lactotransferin.[492] The αNeuAc*p*-(2→3)-Gal*p* offers an extreme case of conformational flexibility, as it can be 10-fold more flexible than the other disaccharides. Here again, the conformation corresponding to the establishment of the *exo*-anomeric effect $\Phi = 60°$ is adopted in several cases, but such a stabilizing influence can be easily overridden, as exemplified by the occurrence of several low-energy conformations having Φ in the vicinity of −60°, for the GM1 pentasaccharide interacting in the combining site of cholera toxin.[493]

groove-like binding. In the groove-type binding, proteins can interact with internal carbohydrate moieties, a situation most commonly found for glycosyl hydrolases. The terminal groups of the carbohydrate ligands (frequently the nonreducing residues) enter first and most deeply into the combining site. Two major types of anchored binding are those that are metal ion mediated (typically by calcium) and those by charge neutralization or compensation of terminal sialic acid residues. This charge neutralization occurs through the creation of ion pairs between the sialic acid carboxylate group and basic residues of the carbohydrate-binding proteins. The role of water in protein–carbohydrate interactions is critical in terms of carbohydrate binding contributing to the specificity of the interaction within the carbohydrate-binding pocket. There are well-documented examples wherein key water molecules are

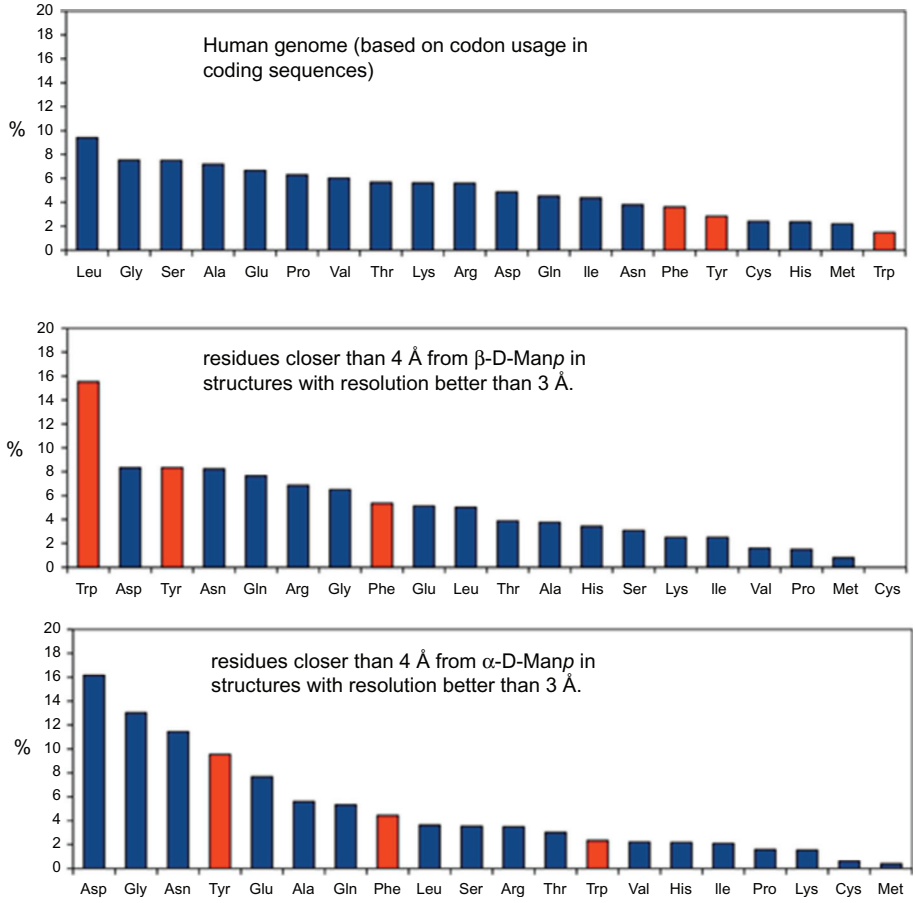

FIG. 7. Occurrence of spatial interactions between aromatic residues (Phe and Trp) and carbohydrates. (A) Occurrence of aromatic residues in the human genome, as based on codon usage coding sequences; (B) occurrence of amino acid residues at a distance closer than 4.0 Å from β-D-mannose residues in crystal structures of proteins having resolution better than 3.0 Å; (C) same as in (B) for α-D-mannose residues (http://www.glycopedia.eu/Maltoporin).[494]

maintained in what appears to be a preconfigured binding site ready to engage carbohydrate ligands. Mimicking the carbohydrate ligands, water molecules clearly occupy the same location as oxygen atoms in the bound carbohydrate ligands and are displaced upon entry of the carbohydrate into the binding site.

In the case of GAG-binding proteins, the interaction is established via salt bridges between basic groups of amino acid side chains and sulfated or acidic groups of the

ligand. The involvement of bridging calcium ions that is frequently observed in several lectins occurs very rarely. Only the annexins, a family of calcium-dependent, membrane-binding eukaryotic proteins, require calcium for binding to heparan sulfate. In other cases, examples are observed where the calcium does not interact directly with the ligand, but induces the conformation of protein loops necessary for binding.[66]

For antibodies and lectins, there appears to be a modular type of assembly and a capacity for protein subunits to associate as in a multimeric fashion (ranging from 2 to 12). The multivalency of carbohydrate–protein interaction (Fig. 8) results in high "avidity," and it is a way to overcome the low to moderate affinities of most carbohydrate interactions with individual binding sites. The resulting interaction is thus a combination of several sequential, or almost simultaneous, binding effects. The overall binding is significantly greater than the simple combination of the individual binding events, that is, the "whole" of the interaction is greater than the sum of its parts. With reference to carbohydrate interactions, this effect is referred to as multivalence or the cluster glycoside effect. It is well documented for lectin–carbohydrate interactions.[67–69] Such a propensity for carbohydrate-binding proteins to form oligomers is thought to correspond to some requirement for biological functions.

2. Nuclear Magnetic Resonance

NMR spectroscopy is a powerful technique for investigating protein–carbohydrate interactions. It possesses the unique ability to address systems with weak and transient interactions as well as tight complexes. Four main methods include (i) chemical shift perturbation studies, (ii) saturation transfer difference NMR (STDR), (iii) transferred nuclear Overhauser effect (TrNOE), and (iv) standard nuclear Overhauser effect spectroscopy (NOESY) experiments. Several reviews have been written in this field[17,70–77] that cover the four approaches that are the most frequently used. A brief summary of the NMR methods used for the characterization is provided in Table II.[78]

3. Thermodynamics and Other Biophysical Methods

A wide range of assays has been used to characterize protein–carbohydrate interactions, allowing access to various types of quantitative information, including such thermodynamic data as stoichiometry of binding, binding constant, enthalpy of

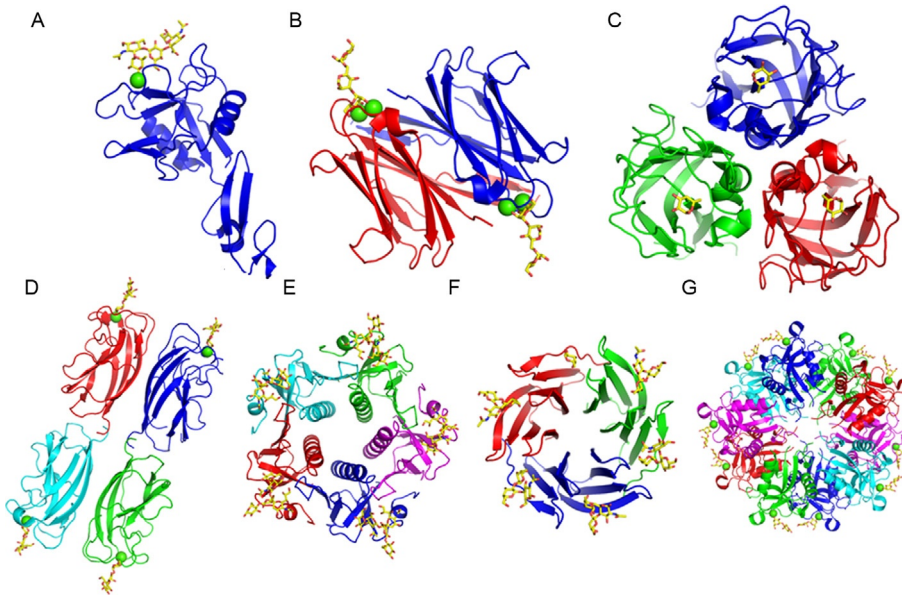

FIG. 8. The multivalent binding of carbohydrate to lectins from X-ray crystal structures. (A) Monovalent E-selectin with bound sialyl Lewis X: αNeuAc*p*-(2→3)-βGal*p*-(1→4)-[αFuc*p*-(1→3)]-GlcNAc*p* (PDB 1G1T).[495] (B) Linear dimeric *Burkholderia cenocepacia* lectin (Bc2la) bound to αMan*p*-(1→6)-[αMan*p*-(1→3)]-Man*p* (PDB 2WRA).[496] (C) Trimeric lectin from *Erythrina corallodendron*, complexed with *N*-acetyllactosamine [βGal*p*-(1→4)-GlcNAc*p*] (PDB 1K12).[497] (D) Tetrahedral tetrameric *Pseudomonas aeruginosa* PA-IIL lectin bound to isogloboside [αGal*p*-(1→3)-βGal*p*-(1→4)-Glc*p*] (PDB 2VXJ).[360] (E) Pentameric cholera toxin B subunit (CTB5) bound to βGal*p*-(1→3)-βGalNAc*p*-(1→4)-[αNeuAc*p*-(2→3)]-βGal*p*-(1→4)-Glc*p* (PDB 3CHB).[498] (F) Hexameric *B. cenocepacia* lectin complexed with H-type 2 trisaccharide αFuc*p*-(1→2)-βGal*p*-(1→4)-GlcNAc*p* (PDB 3ZZV).[499] (G) Decameric rattlesnake venom lectin complexed with lactose [βGal*p*-(1→4)-Glc*p*] (PDB 1JZN).[500] Courtesy: Dr. Anne Imberty, http://lectin3d.cermav.cnrs.fr.

binding, and also kinetics and mechanistic information. Careful control of experimental conditions is necessary in order to avoid misinterpretation, because of the complexity of carbohydrate–protein interactions as an inherent consequence of multivalent interactions. Historically, the strength of these interactions, as for instance with sugar–lectin interactions, has been quantified by inhibition of hemagglutination assay (HIA),[79] and enzyme-linked lectin assays (ELLA),[80] although various spectroscopic techniques (NMR,[81,82] UV spectroscopy,[82] and fluorescence polarization[83,84]) have also been utilized. Finally, among the advanced methods, isothermal titration calorimetry (ITC) and surface-plasmon resonance (SPR) constitute two

TABLE II
NMR Methods Used for the Characterization of Protein–Carbohydrate Interactions[78]

Scheme	Detection Site	Affinity Range	Exchange Regime	Carbohydrate to Protein Ratio	Protein Concentration (μM)	Protein Size (kDa)	Isotope Labeling
CSP	Protein	nM–mM	Fast and slow	>1	>50	<40	Yes
TrNOE	Carbohydrate	μM–mM	Fast	10–50	<10	The larger the better	No
STD-NMR	Carbohydrate	nM–mM	Fast	>50	>1	–	No
NOESY	Both	nM–mM	Slow	1–2	>500	<40	Yes
Water LOGSY	Carbohydrate	μM–mM	Fast	10–50	1–10	–	No
Diffusion	Carbohydrate	μM–mM	Fast	>5	50–100	<40	No

complementary and powerful techniques. Elucidation of the mechanism whereby sugar binding takes place is being slowly unraveled by the design of well-characterized synthetic glycoconjugates[85–92] which, apart from giving insight into the binding parameters that govern these complex processes, offer considerable interest for diagnostic and therapeutic applications.

IV. Features of Protein–Carbohydrate Interactions

As with other types of macromolecular interactions, the formation of a complex is driven by favorable changes in enthalpy (ΔH) and entropy (ΔS). Thermodynamic measurements have indicated that the free energy of binding of monosaccharides to proteins is relatively small. The ΔG value increases significantly whenever disaccharides or higher oligosaccharides interact with proteins. Whenever such proteins interact with carbohydrates, a high "avidity" is observed as a consequence of a multivalent effect. The binding free energy (ΔG) between a carbohydrate molecule and a protein partner is indeed the variable of interest to be assessed. It is assumed to be composed of independent contributions from van der Waals forces, electrostatic interactions (with or without encompassing hydrogen bonding), the hydrophobic effect, and other factors.

1. van der Waals and Electrostatic Interactions

The large number of hydrogen-bond donors and acceptors present in carbohydrates leads to complex and dense networks of hydrogen bonds with proteins. The complexity of such networks is augmented by competition for hydrogen bonds from the water molecules present. The overall enthalpic gain from hydrogen bonding may be counterbalanced by some entropic cost.

2. CH/π Interactions

These interactions characterize the enthalpy of binding of carbohydrates to proteins. They involve a type of hydrogen bond between a hydrogen atom attached to a carbon atom and the π systems of arenes. Typically, this effect is weaker than other interactions. Despite the full recognition of this effect, its computational treatment requires a high level of theory and is not fully taken into account in conventional computational procedures.[93,94]

In many crystal structures of protein–carbohydrate complexes, aromatic residues of the proteins are often observed stacked against certain faces of the carbohydrate molecules. Such an arrangement results from dispersion/hydrophobic interactions wherein small hydrophobic moieties of the solute induce an ordering of the water molecules at the solvent interface. The resulting decrease of the hydrophobic surface area induces a decrease in solvent ordering and a consequent favorable change in entropy. Alternatively, a nonclassical hydrophobic effect has also been documented to occur in lectin–carbohydrate complexes. Here the complex formation is driven by enthalpy because of favorable interactions between the solute forming the complex as well as favorable interactions between the solvent molecules.[95,96]

3. Solvation–Desolvation

The docking of carbohydrates onto proteins is considered to maximize the number of atomic contacts between ligand and protein, and the subsequent structure shows that the carbohydrate lies more or less flat on the protein surface.[97] However, X-ray crystal structures frequently show contradictory features, with carbohydrate residues extending into the surrounding solvent. These structures might be accurately computed if the impact of solvation and desolvation on the binding free energy were properly taken into account.

V. Computational Tools for Docking Carbohydrates on Proteins

1. Force Fields Designed for Carbohydrates

The study of carbohydrate structures and properties by molecular modeling techniques requires molecular mechanics potential energy functions and parameters specific for this class of molecules. Appropriate force fields for carbohydrate systems have been developed, for the purpose of reproducing those specific effects that affect their overall structural properties in solution.[98]

To simulate the behavior of carbohydrates *in vacuo* or in solution (for instance, to study ring-puckering[99] or rotational barriers in oligosaccharides), either established force fields or special parameterizations (Fig. 9) may be used.[100–104] Such force fields permit investigation of the deformation of carbohydrate rings as well as predicting their behavior.

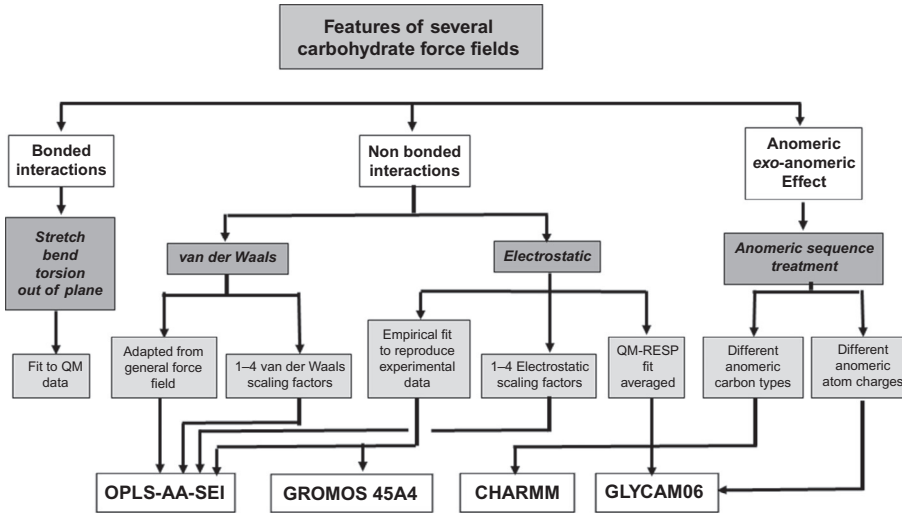

FIG. 9. Parameterization protocol comparison between the carbohydrate force fields: GLYCAM06, GROMOS 45A4, CHARMM, and OPLS-AA-SEI.

These special force fields (as well as already established ones) have been employed regularly for molecular dynamics simulations of protein–carbohydrate complexes.[105,106] In some instances, the simulations were successfully used for estimating binding free energies.[107–110]

Despite the many potential advantages of established force fields, they were not designed to predict binding free energies or enthalpies in protein–ligand docking. Since solvent molecules are usually modeled explicitly, force fields do not need to include additional terms for hydrophobic effects. The special CH/π interactions are not taken into account.[111,112]

Some force fields do model explicitly for hydrogen bonds, while others consider it as part of the electrostatic interactions. Irrespective of the approach, displacement of water molecules competing for hydrogen bonds is not taking into account.

Some force fields correlate well with *ab initio* calculations for *ab initio*-optimized geometries.[113] A recent comparison of the results of *ab initio* and force-field calculations underlines the difficulties in predicting binding enthalpies in protein–carbohydrate complexes using existing force fields. For example, the stabilizing interaction energy for the interaction between fucose and tryptophan is seriously overestimated by the AMBER force field.[114]

GLYCAM06 is a force field widely used for modeling carbohydrates, glycoproteins, glycolipids, and also protein–carbohydrate complexes.[100,115] It can be used to describe the physicochemical properties of complex carbohydrate derivatives, and it is fully compatible with the AMBER force field. Parameters have been developed that take into account a test set of 100 molecules from the chemical families of hydrocarbons, alcohols, ethers, amides, esters, carboxylates, molecules of diverse functional groups, as well as simple ring systems related to cyclic carbohydrates, and fit these to quantum-mechanical data. GLYCAM06 may be used in simulation packages other than AMBER through the utilization of appropriate file-conversion tools.

To facilitate transferability of the parameters, all atomic sequences have an explicitly defined set of torsion terms, with no generic terms, and PARM94 parameters, the same ones used in AMBER, are used for modeling the carbohydrate van der Waals terms.[116] No scaling factors for treating 1,4 interactions are introduced for reproducing the *gauche* effect on rotamers about the ω angle.[117]

In GLYCAM06, the stereoelectronic effects that influence bond and angle variations at the anomeric carbon atom are included in a unique anomeric atom type. This feature permits it to mimic the ring inversion (flipping) observed in glycosidic monomers that occur, for example, during catalytic events.[118] Comparison with experimental data confirmed that the force field is able to reproduce correctly the rotational energies and carbohydrate features if combined with an appropriate charge set. This is true, except for highly polar molecules, for which empirical terms have been introduced to correct energetic torsion errors.[100] The atomic partial charges are calculated residue by residue. For each residue, 50–100 ns MD simulation is performed, 100–200 snapshots are extracted, and charges are calculated by fitting to the averaging quantum mechanics molecular electrostatic potential (ESP). This strategy is adopted for incorporating the dependence of molecular conformations on partial charges. Restraints are employed in the ESP fitting procedure (RESP) to ensure that the charges on all aliphatic hydrogen atoms are zero, since C—H aliphatic hydrogen atoms are not significant for reproducing dipole moments.[119,120] An optimal RESP charge-restraint weight of 0.01 is applied, based on simulations of carbohydrate crystal lattices.[121]

The orientation of the exocyclic hydroxymethyl group is defined by the ω-angle (O-5—C-5—C-6—O-6) and its preference for *gauche* states can be reproduced by introducing scaling factors that slightly modify the 1,4 nonbonded interactions.[117] The 1,4 nonbonded interactions define the influence, in terms of electrostatic and van der Waals potentials. These 1,4 nonbonded interactions are not treated in the same manner in all force fields, and this could be a problem when simulating complex

systems in which two different force fields have to be used. In these cases, the separate treatment of 1,4 nonbonded interactions can assure full compatibility among the force fields. The potential consequences of choosing the 1,4 scaling factors often become irrelevant when glycans bind to proteins, because their freedom in the binding site is usually diminished. The literature has several reviews that describe and compare the performance of carbohydrate force fields used in glyco-modeling.[58,122]

The alternative carbohydrate force fields GROMOS-45A4, CHARMM, and OPLS-AA are used in computational chemistry, together with GLYCAM06, to describe conformational properties of carbohydrates. The GROMOS force field was earlier developed for MD simulations of proteins, nucleotides, or sugars in aqueous or apolar solutions or in crystalline form, but later it was modified to include the anomeric effects for mono- and oligoglucopyranoses.[123] As in GLYCAM06, quantum mechanics (QM) methods are used for calculating bond-angle force constants, whereas derivation of dihedral parameters and van der Waals terms are taken directly from previous GROMOS versions.[124,125] An ESP fitting procedure, with restraints on aliphatic hydrogen atoms and averaging over atom types, is chosen for reproducing the ESP, using a trisaccharide as a model for charge development.[123] No distinction is made between α and β monomers in terms of charges and type of anomeric atom, and electrostatic–van der Waals 1,4 scaling factors are not introduced, so that the *gauche* effects on ω angles are correctly reproduced. A 20-ns-long MD simulation in explicit water[126] was used for validating the force field, demonstrating the ability to predict the stereoelectronic effects and the most stable ring conformations correctly, but sometimes failing to reproduce their correct energies. Recently, GROMOS was proposed as the most adapted force field to mimic the transition from $^4C_1$ chair to skew conformations of the iduronic acid residues in MD simulations of heparin.[127]

The CHARMM force field has been extended to glucopyranose and its aldohexopyranose diastereomers.[128] Several revisions for carbohydrates have been proposed in order to extend this force field to five-membered sugar rings and oligosaccharides.[129,130] The same hierarchical parameterization procedure and treatment of 1,4 nonbonded interactions are used to ensure full compatibility with other CHARMM biomolecular force fields.[131,132] Preliminary parameter sets are created using small-molecule models corresponding to fragments of pyranose rings, and these are then successively applied to whole pyranose monosaccharide structures. Missing dihedral parameters are developed by fitting over 1800 quantum-mechanical conformational energies for hexopyranoses. Both partial atomic charges and Lennard–Jones parameter values, taken from previous CHARMM versions, are adjusted to reproduce scaled

quantum-mechanical carbohydrate–water interaction energies and distances. These are further refined to reproduce experimental heats of vaporization and molecular volumes for liquids. The force field, with different atom types for α and β anomers, was shown to be valid, as it reproduces calculated quantum-mechanical and experimental properties using MD simulations with TIP3P water molecules.

The OPLS force field has been expanded to include carbohydrates.[133] In the OPLS-AASEI (Scaling Electrostatic Interactions) force field, 1,4, 1,5, and 1,6 scaling factors are introduced to enhance the prediction of Φ/Ψ conformational properties, as well as anomeric effects and relative energies.[133] Unique charge sets and atom types for α and β anomers are used. All nonbonded parameters are imported directly from the parent force-field OPLSAA.[134] Charges are derived, as was done for previous force-field versions,[134,135] from standard alcohols and acetals to simply reproduce consistent energetic properties, and these are then transferred to carbohydrates.

Other force fields are employed to understand carbohydrate properties *in silico*. In particular, MM3, a force field originally meant for hydrocarbons, is now considered appropriate for a wide range of compounds. The MM3 force field for amides, polypeptides, and proteins[136,137] is widely used for the construction of adiabatic maps of disaccharides. The TRIPOS molecular mechanics force field is designed to simulate both peptides and small organic molecules,[138] but parameter development for oligosaccharides has included sulfated GAG fragments and glycopeptide–carbohydrate interactions.[139,140] The TRIPOS force field is implemented in the molecular package Sybyl[141] and commonly used for geometry optimizations. These computational and graphical tools are used to calculate and display different properties on molecular surfaces. The most common are (i) the lipophilicity potential [evaluated as the partition coefficient ($\log P$) in polar/apolar heterogeneous systems]; (ii) the ESP; (iii) the ESP as evaluated from the Poisson–Boltzmann equations; and (iv) the hydrogen-bonding sites. The exploration of the (macro)molecular surface can be also performed using a GRID-type procedure,[142] which can be used to locate the positions near a molecular system (the target) at which a small chemical entity or molecule (the probe) would interact favorably.

Complex carbohydrates is generally difficult to crystallize in the form of single crystals, and the number of X-ray crystallographic structures for protein–carbohydrate complexes that have been resolved is still limited. As a consequence, molecular modeling methods have been extremely helpful and widely used to describe the conformation of complex carbohydrates. Developments of sustainable methods and tools, accompanied by significant advances in computing performance, are opening the way to high-throughput molecular modeling, where hundreds of complex glycan structures can be investigated during the course of time-limited investigations.

Web-based tools have been developed to build preliminary 3D structures, starting from a sequence as implemented in SWEET-II[143] and the GLYCAM carbohydrate builder. The outcome of using such tools has shown significant success. About 250 glycan determinants (bioactive oligosaccharides) have been subject to systematic conformational sampling and their conformational preferences determined.[144] They originate from a wide variety of sources, including blood group antigens, core structures, fucosylated oligosaccharides, Lewis antigens, sialylated oligosaccharides, GPI anchors, N-linked oligosaccharides, globosides, and others, and have been systematically organized into an open-source database (http://glyco3d@cermav.cnrs.fr). The constituent disaccharides and monosaccharide entries (currently ~120 and 80 entries, respectively) of these bio-oligosaccharides have also been characterized and made available through a subset within the database.

2. QM Methods

A vast number of QM methods[145–148] are available, and their detailed description is beyond the scope of this article. Many of them have been successfully applied to carbohydrates. A brief overview is presented here, but without a claim to completeness. The selection of citations is also biased by the authors' experiences.

a. Molecular Orbital Methods.—In contrast to force fields, the *ab initio* or nonempirical QM methods do not use empirical parameters. Until recently, the Hartree–Fock (HF) method [also known as the self-consistent field (SCF) approximation] was assumed as a suitable *ab initio* approximation and was used in QM computations. A variational character of the HF method suggests that an improvement of the basis set of atomic orbitals provides, within the HF approximation, significantly better results. In the end, the procedure converged to the basis set limit of the method. However, this method does take into account mutual correlation of electron motions. In contrast, empirical force field and semiempirical QM methods may behave in unpredictable ways when they are applied for the systems different from those considered for their parameterization.

In general, two factors affect the quality of *ab initio* QM calculations: (i) the quality of the atomic orbitals used to build the molecular orbitals and (ii) the inclusion of electron-correlation effects. To achieve high-quality results, a basis set of high quality together with inclusion of electron correlation are required. Various post-HF methods were developed to include electron correlation, such as the perturbation Moller–Plesset (MP, usually second-order MP2), coupled clusters (CC), and quadratic configuration interaction (QCI) methods.[148] Unfortunately, although they lead to

improved results for the structure and behavior of molecules, their use for large systems is restricted by the requirement for extremely large computer resources.

The application of QM methods to carbohydrates lagged behind that for other biomolecules because of their structural complexity. In fact, the first QM calculations on saccharides dealt with conformational studies of small model compounds representing the glycosidic linkage, using *ab initio*[42–44] and semiempirical methods.[149] These and other early QM calculations were focused on understanding the stereoelectronic effects,[45–50] and the results obtained provided appropriate values for parameterization force-field methods for carbohydrates.[58–60]

Semiempirical QM methods are founded on the same HF theory as *ab initio* methods, but some interactions are approximated or entirely omitted. To adjust for the inaccuracy from neglecting some interactions, parametric forms are introduced for some aspects of the calculations. The parameters involved are chosen by fitting the experimental or *ab initio* results. Semiempirical methods provide a significant advantage as they are significantly faster than *ab initio* QM methods, but their reliability in the carbohydrate field is severely limited.[150] In spite of their limitation, semiempirical QM methods clearly showed that mono- and oligosaccharides are flexible molecules and their conformational behavior is affected by the solvent used.[56,61,151] Recently, the PM3CARB semiempirical method was developed for carbohydrate modeling.[152,153]

Although the *ab initio*-calculated structures[154] for disaccharide analogues lacking hydroxyl groups were in close agreement with the conformations observed in crystals, it soon became obvious that orientations of hydroxyl groups play a key role in determining the conformations of mono- and oligosaccharides. This was documented by extensive QM calculations on β-D-glucopyranose[155] and also on a number of various mono- and oligosaccharides.[156–159] During the past decade, density-functional theory (DFT) methods have replaced standard HF methods for calculations of carbohydrates.

The role of ring distortion of glycosides in the mechanism of their hydrolysis has been investigated by using MP2 *ab initio* calculations.[160,161] The calculations showed that protonation of the $^1C_4$ chair conformation leads to a stable oxocarbenium ion, whereas all other conformations studied formed oxocarbenium ion–water complexes. It was suggested that ring distortion decreases the energy for stretching the glycosidic bond, thereby lowering the transition-state energy for bond cleavage.

b. DFT Methods.—During the past two decades, the DFT method has become one of the most widely used techniques in investigations of biomolecular systems. The reasonable performance and low cost as compared to post-HF methods are behind its widespread popularity. The DFT method describes many-electron systems by means of the

electron density as the basic variable, instead of the wave function used in molecular orbital methods.[162] The accuracy of a DFT calculation depends upon the quality of the exchange-correlation functional, and part of the success of DFT lies in the remarkable progress made in development of new, more reliable functionals.[163–166] The simplest functionals, which used the local spin-density approximation (LSDA), were not useful for chemical problems. The generalized gradient approximation (GGA) uses more complex functionals that depend on density and their gradients. The GGA functionals have provided surprisingly reasonable predictions, but have shortcomings in modeling a chemical reaction; in particular they underestimated energy barriers and the transition-state geometries. Development of hybrid functionals that mix certain portions of the local GGA functionals with nonlocal HF exchange permitted considerable improvement; these functionals are usually termed hybrid functionals.

In particular, the B3LYP functional[167,168] became the most popular functional for calculation of conformational properties of mono- and oligosaccharides, and many articles have been published in recent years on this subject (see, for example, Refs. 156,157,159,169–182). It is beyond the scope of this article to consider all of the results in detail, and we just mention here some methodological aspects. The conformational sampling and optimization of geometry is usually performed at the B3LYP/6-31+G* level, followed by single-point calculations of the final set of conformers at the B3LYP/6-311++G** level. Solvation corrections are then usually added to the energy, by various continuum models.[145] The results obtained at this level of theory describe ground-state properties of carbohydrates, quite satisfactorily, but it has been clearly documented that diffuse functions must be used to provide reliable conformations.[183,184] The DFT calculations were used to obtain a mechanistic insight into the glycosylation reaction.[185–193] These calculations found the $^{1}S_{3}$ transition state for a reaction pathway connecting the most stable D-glucopyranosyl oxocarbenium ions, namely conformers $^{4}H_{3}$ and $^{5}S_{1}$. The calculations also showed that glycosylations are dissociative, with the scission glycosidic linkage being greater than 2 Å before any nucleophilic attack, and that nucleophilic attack required some preassociation of the nucleophile.

In spite of the great success of the B3LYP functional, there are reports questioning its consistency. It was noted that B3LYP underestimates reaction barriers[194,195] and is inadequate for describing noncovalent complexes.[196,197] A variety of different specialized empirical functionals have been developed[164,194–201] that offer more reliable results. The ability to predict both the reaction barrier and transition-state structure is of greatest importance for modeling reaction mechanisms.

Standard functionals fail to describe dispersion interactions accurately.[202] Therefore, various empirical correction schemes were developed to account for noncovalent interactions.[198,199] Such interactions contribute significantly to carbohydrate–protein

interactions and were investigated by using the DFT-D method.[198] Comparison of DFT-D and CCSD(T) calculations showed that the DFT-D method described interaction energies reasonably well, and it can be used for calculations of carbohydrate–aromatic residue interactions.[203,204]

A variety of databases were used to determine the performance of functionals. Several hybrid functional, such as M08-SO,[200] M06-2X,[165] M05-2X,[164] MPW1K,[205] and PWB6K,[206] were recommended for calculating barrier heights and transition-state geometry.[166,195] Four different structural models of the α-glycosidic linkage in sugar nucleotides were applied to predict the structure and dissociation of the C—O glycosidic linkage.[207] A comparison with the MP2 results showed that M05-2X, MPW1K, and MPWB1K predicted the thermochemical kinetic parameters reasonably well. Based on calculations of the geometries of 19 transition states and using 12 different functionals,[208] a strategy for large systems was recommended. In this procedure, hybrid GGA functionals, such as B3LYP, are used in geometry optimization and then single-point calculations are performed using more expensive hybrid meta-GGA functionals. The DFT calculations are less basis set sensitive than the electron-correlated *ab initio* methods, but have a similar basis set requirement as HF calculations. On the other hand, the use of polarization functions is a must, and diffuse functions are recommended.

3. QM/MM Methods

While MM methods have been widely used in the description of complex carbohydrates and protein–carbohydrate interactions,[60] they are unable to describe bond-forming or bond-breaking processes. In contrast, QM methods consider the electronic structure of a system and, therefore, provide a description of a chemical reaction. Unfortunately, good quality QM methods are computationally extremely demanding, and their use is still limited to relatively small systems. Therefore, for modeling the active sites of enzymes using QM methods, a relatively small but carefully selected region of the enzyme must be chosen and treated by the QM method. This so-called cluster approach has provided valuable insight into enzymatic reactions.[209] However, these models obviously ignore the effect of the enzyme's surroundings. Also, a slow convergence was observed for the cluster approach with respect to the size of the cluster.[210,211] The calculated energies can still vary more than 10 kcal mol^{-1} for models of size up to 230 atoms.[212] Clearly, a more realistic description of a reaction mechanism requires the inclusion of a full enzyme environment into the computations. The combined quantum mechanics–molecular mechanics (QM/MM) approach, first outlined by Warshel and Levitt,[213] provides a solution to the limitations of MM and QM. These QM/MM methods combine reliable

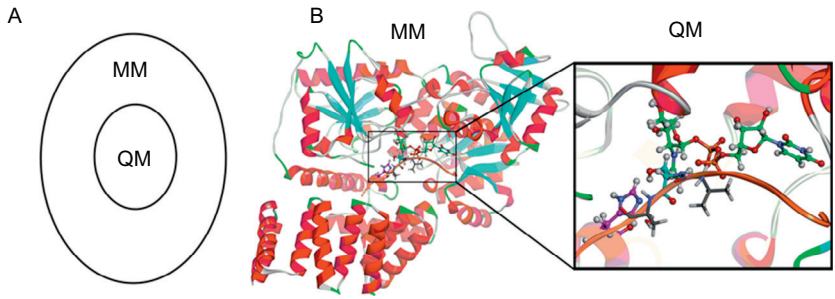

FIG. 10. Illustration for the QM/MM method: (A) the system is divided into QM and MM regions and (B) the MM region of the overall structure of human OGT–UDP-GlcNAc–CKII peptide complex is shown in ribbon representation and residues included in the QM region are shown in stick representation. For clarity, water molecules are not shown. Reprinted with permission from Ref. 329. Copyright 2012 American Chemical Society.

calculations of electronic structure with a realistic description of the enzyme environment at the atomic level. Thus, large system can be described by methods that combine the precision of a QM description and the low cost of MM. In QM/MM methods, the relevant part of the system, such as the active site of an enzyme, is described at the electronic level with the QM method, whereas the remaining part of the system is described at the atomic level using the MM method. Recent studies have demonstrated that the QM/MM treatment converges faster to the "full QM limit" than the pure QM treatment.[210,211,214] The development and application of QM/MM methods have been presented and discussed in a number of recent reviews,[215–221] to which readers should refer for further details.

The general idea of the QM/MM method is illustrated in Fig. 10. The QM/MM methods divide the whole system into two (or more) localized regions that are treated at different levels of theory. In studies of enzymatic reactions, the QM region usually describes the active site, where a chemical reaction proceeds, and relevant amino acid residues nearby. This region is treated with QM methods. The surrounding MM region is calculated at the MM level and influences the QM region by electrostatic and steric constraints. The total internal energy of the QM/MM system can be formally written as:

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM-MM}} \tag{1}$$

In this Eq. (1), the E_{QM} term represents the QM energy of the QM region and the E_{MM} term is the MM energy of the MM region. The $E_{\mathrm{QM-MM}}$ term describes the

interaction energy between the QM and MM regions. Two different schemes, known as the additive and subtractive approaches, were developed to calculate the total energy ($E_{QM/MM}$) of the QM/MM system.[217] Subtractive and additive schemes are equivalent in principle, differing only in the technical details of their implementation. Various combinations of QM and MM methods can be used in the QM/MM method. In the past, semiempirical methods were generally used for QM calculations, but DFT methods are currently employed more frequently. Force fields used for the calculation of the MM region ignore such nonadditive effects as polarization of atoms in the electric field from the surroundings. Some attempts have been made to develop polarization force fields for carbohydrates[222] but their employment in QM/MM calculations remains to be seen.

The critical points in the QM/MM approach are the size of the QM region and treatment of the boundaries between the QM and MM. The accuracy of the calculation is considerably affected by both of them, and the size of the QM region and its boundaries must be selected carefully.[214] In studies of large biomolecules, the boundary between the QM and MM region frequently passes through covalent bonds. This is the well-known and difficult problem of QM/MM methods, and it has been extensively discussed.[216,217,220]

Various QM/MM schemes have been developed,[218,219] and the boundary region may contain additional atoms (link atoms) that cap the QM system and are not part of the entire system, or it may consist of atoms with special features that are calculated by use of both the QM and MM methods.[223,224] Another approach[225–227] uses localized orbitals (frozen orbitals) between the QM and MM regions that are determined by calculations made on small molecules. Nevertheless, both link atoms and frozen orbitals significantly perturb the system under study. It is, recommended,[210] therefore, to move any QM/MM boundary at least one residue away from all active-site residues.

Another key point of QM/MM calculations is the description of interactions between the QM and MM regions, often termed QM/MM coupling. Various approaches have been developed to handle the interactions between the QM and MM regions. In general, both bonded (bond stretching, bond bending, and torsion terms) and nonbonded (van der Waals and electrostatic) interactions contribute to the coupling energy term. Bonded interactions are relevant in cases when a covalent bond crosses through the boundary region. The electrostatic interactions between the QM charge density and the atomic charges in the MM region constitute the crucial part of coupling. Based on the treatment of these interactions, there are three basic approaches for QM/MM methods: mechanical embedding, electrostatic embedding, and polarized embedding. These differ, fundamentally, by the amount of mutual

polarization of two regions. In the mechanical embedding approach, electrostatic interactions between the QM and MM regions are treated at the MM level, with the same charge model being applied to the QM region. Mechanical embedding constitutes the simplest, but also the most approximate QM/MM coupling. The QM calculation of the QM region is performed in the entire absence of its MM surrounding, and thus the polarization of the QM wave function is totally neglected. This neglect might cause a problem, especially when charge distribution of the region varies considerably, as during a reaction. In the electrostatic or electric embedding approach, the MM charges are included into QM Hamiltonian as one-electron terms. Thus, the QM calculations are performed in the presence of MM charges that polarize the QM wave function. The charge distribution of the MM region is usually represented by a set of the atom-centered point charges. However, since the electrostatic interactions between the QM and MM regions in this approach are treated at the QM level, the computational demands are substantially increased. This polarization might be significantly overestimated if the simplified atom-centered MM point charges are positioned close to the QM wave function. To avoid significant inaccuracies, an enlargement of the QM region should be utilized in such a case. Moreover, current investigations have demonstrated[210] that the best results were obtained with mechanical embedding.

The QM/MM methods provide the energy for a given structure and can be combined with any technique that can scan the configuration space available to large biomolecules. This can be any optimization, molecular dynamics (MD), metadynamic, or Monte Carlo method. When properly applied, the QM/MM calculations have a high potential for elucidating fundamental atomistic details of the mechanisms involved in the catalysis. However, QM/MM results must be considered with caution because of the limitations of sampling and accuracy. Cooperation of experimentalists and theorists is, therefore, highly desirable.

4. MD Methods

In MD simulations, an ensemble of configurations is generated by applying the laws of motion to the atoms of the molecule. The concept behind MD simulation involves calculating the displacement coordinates in time (trajectory) of a molecular system at a given temperature. Integration of Newton's equation of motion in time yields the location of the positions of a set of particles, along with their velocities, as a function of time. Molecular simulations can be performed either as a microcanonical (constant NVE) or as a canonical (constant NVT) ensemble. Consequently, all other

thermodynamic quantities must be determined by ensemble averaging. In a classical system, Newton's equations of motion conserve energy and thereby provide a suitable scheme for calculating a microcanonical ensemble. A canonical ensemble can readily be performed by coupling the molecular system to a constant temperature bath, which rescales the atomic velocities according to the desired temperature. In a similar manner, constant pressure simulations can be performed by scaling through coupling to a constant temperature position, as the pressure can be calculated from the virial theorem.

Several algorithms have been developed for MD simulations. Such simulations monitor a system for a limited time. Physically observed properties are computed as the appropriate time averages through the collective behavior of individual molecules. For the results to be meaningful, the simulations must be sufficiently long so that the critical motions are statistically well sampled. Experimentally accessible spectroscopic and thermodynamic quantities can be computed, compared, and related to microscopic interactions. It should be noted that MD is severely limited by the available computer power. With currently available computers, it is possible to perform a simulation with several thousand explicit atoms for a total time of up to a few hundred nanoseconds. To explore the conformational space adequately, it is necessary to perform many such simulations. In addition, it may be possible that carbohydrate molecules undergo dynamical events on longer timescales. These motions cannot be investigated with standard MD techniques. Another way is to use high-temperature dynamics to allow the molecule to assume high-energy conformations. Caution has to be taken while using this approach, as the molecule may acquire "nonphysiological" conformations. A microsecond timescale is becoming accessible through the development of novel simulation algorithms and specialized hardware. It must be noted that the longer the simulation, the higher the possibility of force-field imperfections to appear.

Recent examples of the use of MD to study carbohydrates and carbohydrate–protein interactions have been discussed and properly illustrated (Fig. 11).[60] Among these applications are the validations of binding modes obtained from automated molecular docking. In conjunction with NMR experiments (primarily NOE- and residual dipolar coupling-based experiments), MD can also be used to generate probable carbohydrate structures that undergo dynamical events as a result of their high conformational flexibility. In such cases, ways may be found to create penalty functions with respect to violations of measured NOEs in the generation of structures consistent with NMR data. The so-called time-averaged restrained molecular dynamics simulations (tar-MD) add to the energy of the system (E_{system}) an artificial term (E_{penalty}) in the force field in such a way that the total energy ($E_{\text{system}}+E_{\text{penalty}}$)

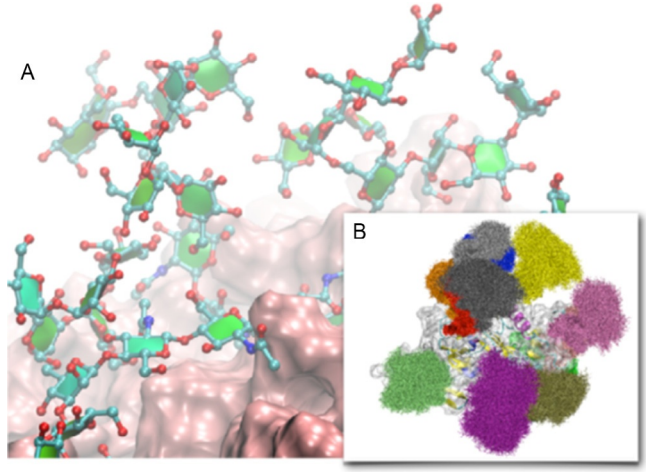

FIG. 11. MD simulation of HIV gp120 glycoprotein.[501] Complete N-glycans were modeled at 13 glycosylation sites, based on the X-ray crystal structure (PDB 2BF1).[502] The macromolecular system has more than 100,000 atoms: 4832 protein atoms, 3432 carbohydrate atoms, 30,665 water molecules, and 4 chloride ions. (A) Solvation shell of a selected N-glycan on the protein surface. (B) A significant surface area of the protein is shielded by the N-glycans. Reprinted with permission from Ref. 501. Copyright 2010 Springer.

increases as do the violations. This methodology[228–233] has been extensively and successfully applied for studying different molecular systems,[234] in particular for the investigations of iduronate-containing carbohydrates[235]

5. Free-Energy Calculations

The absolute energies of ligand–receptor interactions can be obtained by performing average MM-PBSA calculations on an ensemble of uncorrelated snapshots in an implicit water environment, collected from an equilibrated MD simulation. MM-PBSA is a method (Fig. 12) that approximates the average free-energy of binding (ΔG) between the ligand L and the receptor R in an implicit aqueous environment as:

$$\Delta G = \Delta G_{RL} - \Delta G_{R} - \Delta G_{L} \tag{2}$$

Each term of Eq. (2) is further broken down as follows:

$$\Delta G_{RL} = \Delta E_{MM} + \Delta G_{PBSA} - T\Delta S_{MM} \tag{3}$$

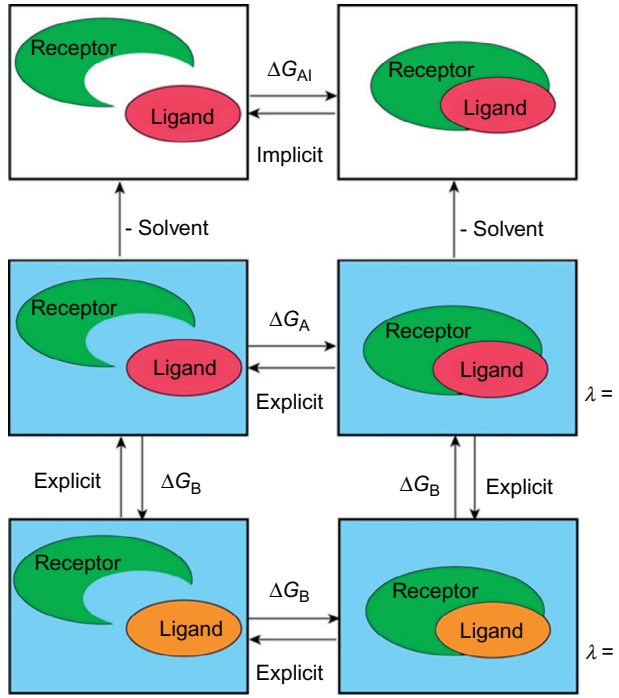

FIG. 12. MM-PBSA calculations determine the absolute free energy of binding of a ligand to a receptor (ΔG_{AI}) in an implicit solvent environment, whereas thermodynamics integration methods calculate the difference in free energy of binding between receptor–ligand complexes ($\Delta\Delta G = \Delta G_C - \Delta G_D = \Delta G_A - \Delta G_B$), where only the ligand is changed.

$$\Delta G_R = \Delta E_{MM} + \Delta G_{PBSA} - T\Delta S_{MM} \tag{4}$$

$$\Delta G_L = \Delta E_{MM} + \Delta G_{PBSA} - T\Delta S_{MM} \tag{5}$$

where ΔE_{MM} is the average molecular mechanical energy incorporating the bond angle, torsion angle, van der Waals, and electrostatic energy terms described in the force field.

The solvation free-energy term ΔG_{PBSA} incorporated the electrostatic and nonpolar solvent contributions:

$$\Delta G_{PBSA} = \Delta G_{PB}^{el} + \Delta G_{SA}^{np}. \tag{6}$$

The Poisson–Boltzmann equation is solved for determining the solvent polar effects[236] ΔG_{PB}^{el}, whereas the solvent-accessible surface area is used to determine

the nonpolar energetic term[237] ΔG_{SA}^{np}. Finally, $T\Delta S_{MM}$ defines the entropic term due to the loss of degrees of freedom upon association. Evaluation of this term raises an issue in computational chemistry; it is commonly performed by using a quasi-harmonic method or by normal-mode analysis.[238–240] The high computational cost, combined with a very slow convergence and the approximations, introduces significant uncertainty in the result. Thus, the entropy contribution can be neglected when a comparison of states of similar entropy is desired, such as in a series of similar ligands binding to the same protein receptor.[241]

Thermodynamic integration (TI) calculations determine the free-energy difference between two closely related systems A and B by gradually transforming the initial state A to the final state B. The two states are coupled via a parameter λ that serves as an additional, nonspatial coordinate. This parameter describes the transformation from the reference system A to the target system B and allows the free-energy difference between the states to be computed as:

$$\Delta G_{TI} = \int_0^1 d\lambda \langle \delta V(\lambda)/\delta(\lambda) \rangle_\lambda \tag{7}$$

In this equation, λ represents the coupling parameter that corresponds to the potential energy $V(A)$ for $\lambda = 0$ and $V(B)$ for $\lambda = 1$. The integration is carried out over the average of the λ derivative of the coupled potential function at given λ values. Thus, MD simulations in explicit water at different discrete λ points are performed, and the value of the integral is calculated numerically. For TI calculations, the system should not undergo significant conformational changes during the transformation; otherwise, MD simulations will most probably not sample sufficient phase space for obtaining converged results.[242]

6. Metadynamics Calculations

The major disadvantage of MD simulation is that the timescales accessible by standard MD simulations are usually much shorter than those of interest. In principle, it is possible to determine free-energy differences by MD simulations. However, transitions between structural families of carbohydrates are often too slow to be observed in a reasonable timescale, and their populations usually cannot be measured within an acceptable degree of accuracy. Metadynamics is a recently developed technique for free-energy modeling that enhances sampling of configuration space within a molecular dynamic simulation. Metadynamics belong to a class of

computational methods commonly referred as enhanced sampling techniques.[243,244] Since its introduction,[245] metadynamics has evolved significantly, and an overview of these developments can be found elsewhere.[246–249]

In metadynamics, the system is treated using a standard MD simulation to which an external history-dependent bias potential is added to the Hamiltonian of the system. The bias potential is a function of the collective variables (CVs) and is typically written as the sum of Gaussian hills along the trajectory:

$$V_{\text{bias}}(q) = \sum_{t_i<t} \prod_j w_{t_i} \exp\left[\frac{-\left(q_j(X) - q_j^{t_i}\right)}{2\delta q_j^2}\right] \tag{8}$$

where q are CVs (parameters that describe the development of the process studied), X are the Cartesian coordinates of the system, and w and δq are the height and width of a Gaussian hill, respectively. This bias potential accumulates during the simulation; it drives the system out from configurations already explored and moves the system to visit unexplored regions by the crossing of free-energy barriers. Moreover, an estimate of the free-energy surface of the system can be obtained from the sum of bias potentials accumulated over the entire simulation. The main advantage of the metadynamics algorithm is that it automatically searches for the biasing potential without any *a priori* knowledge about the free-energy profile.[249–252]

The choice of the appropriate CVs is the critical factor for obtaining meaningful result from metadynamics simulation. For example, the CVs should be able to distinguish between reactants, transition states, intermediates, and products when studying reaction mechanisms. Metadynamics requires a set of CVs (typically one to three). The bias potential then allows efficient exploration of the space of these variables. The resulting free-energy surface is also the function of these variables. The selection of CVs tested in metadynamics included distances,[245,250,251] dihedral angles,[245,250] coordination numbers,[250,252] and ring-puckering coordinates.[253,254]

The great enhancement in conformational sampling by metadynamics permits the study of processes where changes in the electronic structure play fundamental roles, such as chemical reactions.[248] The processes are described by means of the first-principle dynamics, in *ab initio* molecular or Car–Parrinello molecular dynamics (AIMD or CPMD).[255] Development of the hybrid Car–Parrinello QM/MM metadynamics[245] which combines CPMD with classical force-field methods opened an exciting area of application on biomolecular systems, particularly in the field of enzyme catalysis.

Although metadynamics is a relatively new technique, it has found application in carbohydrates. Metadynamics using the force-field method has been applied in

modeling the solvent effect on conformational equilibrium of the primary hydroxymethyl group in hexopyranoses,[256] in the modeling of free-energy surface of sialic acid[257] and iduronic acid,[258] and for an evaluation of the carbohydrate force field to the conformations of the glucopyranose ring.[254] Car–Parrinello metadynamics has been applied in the modeling of ring puckering of pyranose sugars in a vacuum[253,259–262] and the ring-puckering coordinates.[253,254] It contributed significantly in the understanding of the catalytic mechanism of glycosidases[263] and glycosyltransferases (GTs)[264] (see Sections VI and VIII for details).

7. Molecular Robotics

Because macromolecular structures are dynamic rather than static, information regarding their dynamics is required to establish the link between structure and function. A number of simulations have been performed to date using MD calculations. Current computational power permits MD simulations of only a few microseconds which, as already noted, is insufficient, since many processes occur over the range of seconds. Because these simulations are usually performed on short timescales, they allow modeling of the dynamic properties of equilibrium states, but do not allow the capture of an entire conformational event, especially with explicit simulation of solvent. Alternative methods are being developed with the aim of simulating molecular motions that can occur on larger spatial and temporal scales. Among them, normal-mode analysis has reemerged and has been successfully applied in structural biology, following the observation that the predicted collective motions for folded structures are highly robust and bear fundamental significance. Normal-mode analysis has been successfully applied to enhance conformational sampling selectively along specific directions of motion.[265] The analysis can identify large collective motions that may occur in the protein upon binding to a ligand. Application to the field of GTs involved in the biosynthesis of blood group antigens has pioneered this important field of research.[266]

Among the alternative methods that are being developed, some are inspired by the field of robot motion planning, with a proper extension to compute molecular motions. Robotics-based algorithms have been applied to the study of such different problems as ligand docking and accessible pathways in flexible receptors, or conformational changes of proteins due to loop motions and domain motions.

The so-called molecular robotics separates the search for conformational pathways into two stages. In the first stage, the robotics-based approach is performed to explore geometrically realistic motions. The second stage uses molecular mechanics for

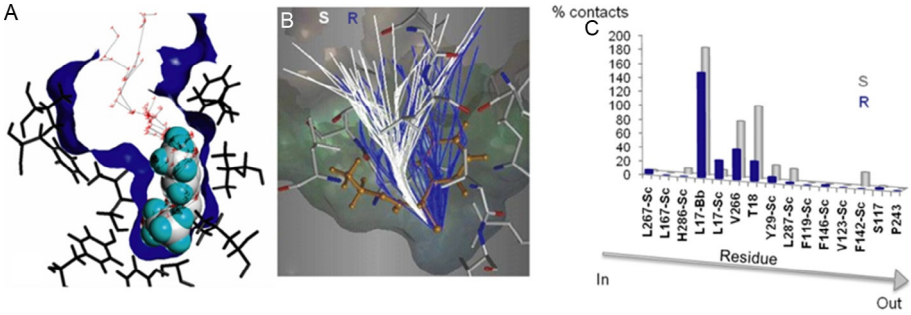

FIG. 13. Illustration of the molecular robotics approach for investigating the role of substrate accessibility to the active site on *Burkholderia cepacia* lipase and its enantioselectivity. (A) Conformational exploration of the active-site pocket, using Path-Planning algorithms in order to search exit paths of the ligand from its catalytic position. (B) Exit paths are computed for the *R*- and *S*-enantiomers (50 paths for each enantiomer). The distribution obtained for the *R*-enantiomer (blue) clearly appears larger and less constrained than for the *S*-enantiomer (white). (C) Histogram representing for each enantiomer the relative frequency of interatomic contacts (averaged among the 50 paths) with amino acid residues This automated analysis of ligand–protein contacts permits highlighting the amino acid hindering the displacement of enantiomers, and thus providing target residues for engineering enantioselectivity.[269] Reprinted with permission from Ref. 157. Copyright 2011 Wiley-VCH Verlag GmbH & co, KGaA.

evaluation of solutions found in the previous stage, while taking into account the explicit simulation of solvents. Large molecular motions can be handled in a continuous way and within very short computing times.

The key advantage of the robotics-based approach (Fig. 13) is that it enables fast exploration of high-dimensional conformational space, thanks to the combination of a geometrical treatment of the key molecular constraints together with the performance of path-planning algorithms. Combined with methods in computational physics, such as normal-mode analysis, or by using appropriate multiscale molecular models,[267] robot path-planning algorithms relying on a mechanistic modeling of (macro)molecules are able to compute large-amplitude conformational transitions in proteins with several orders of magnitude faster than such standard simulation methods as MD.[268,269] These robotics-inspired methods have also been developed with very low computational cost to simulate ligand displacement inside an active-site pocket of a protein, considering both partners as flexible molecular models.[268,270,271] Another novel feature lies in the possibility for estimating the "escape path" and "escape time" for the ligand to escape from the "funnel of attraction" at the binding site.

Such methods, which can predict binding sites and compute access/exit ligand pathways, have already been successfully applied for rational enzyme engineering.[272,273] They show the effectiveness and the potential of molecular robotics methods to guide the engineering of enzyme mutants having enhanced activity, selectivity, and specificity. New developments are being made wherein these robotics-inspired methods are combined with coarse-grained normal-mode analysis.[274]

8. Molecular Docking

Molecular docking is a computational procedure that aims to predict the favored orientation of a ligand to its macromolecular target (receptor), when these are bound to each other to form a stable complex.[275] Although each docking program operates slightly differently, they share common features that involve ligand and receptor, sampling, and scoring. Sampling entails conformational and orientational location of the ligand within the constraints of the receptor-site binding. A scoring function selects the best ligand conformation, orientation, and translation (referred to as poses), and classifies ligands in rank order. A successful docking exercise must accurately predict either or both ligand structure (pose prediction) and its binding propensity (affinity prediction). Available docking programs differ essentially in ligand placement in the "combining" site, exploration of conformational space, and scoring or binding estimate. The interaction with the ligand relies both on the protein backbone fold in the region of the binding site and on the orientation of the side chains in that binding site. One of the most significant limitations in docking is that it is typically performed while keeping the protein surface rigid, which prevents consideration of the effects of induced-fit within the binding site.

Difficulties in molecular docking are largely due to the high number of degrees of freedom characterizing a protein–ligand system, and this increases the computational cost of the calculations. Thus, several approximations concerning the flexibility states may be introduced in molecular docking experiments. The simplest approximation (rigid docking) considers only the three translational and three rotational degrees of freedom of protein and the ligand, treating them as two separate rigid bodies. The most widely used algorithms at present allow the ligand to fully explore its conformational degree of freedom in a rigid-body receptor.[276,277]

The docking algorithms can be grouped into deterministic and stochastic approaches. Deterministic algorithms are reproducible, whereas stochastic algorithms include random factors that do not allow full reproducibility. Incremental construction algorithms involve the division of a ligand into rigid fragments. One of the

fragments is selected and situated in the protein binding site. The reconstruction of the ligand is then performed *in situ*, adding the remaining ligand fragments. The program DOCK,[278] for example, is based on this algorithm. A genetic algorithm is a stochastic searching approach that uses techniques inspired by evolutionary biology for obtaining reliable results. It mimics the process of evolution by manipulating a set of data structures called chromosomes. AutoDock[279] uses this algorithm for obtaining reliable docking results. A variety of other sampling methods have been implemented in docking programs. Some of them include simulated annealing protocols and Monte Carlo simulations. The algorithm used in Glide,[280] for example, can be defined as a hierarchical algorithm. It uses an exhaustive systematic search for discovering the most favored ligand conformations in the combining site of the protein, with a screening based on progressively decreased energetic cut-offs. Energy-scoring functions are required to evaluate the free energy of binding between proteins and ligands. Some sophisticated techniques for predicting free energies of binding are currently too slow to be used in molecular docking of large sets of compounds.

Consequently, fast scoring functions have been developed. Empirical scoring functions use a set of parameterized terms describing properties known to be decisive in protein–ligand binding to formulate an equation for predicting affinities. Multilinear regression is used to optimize these terms, using a set of known protein–ligand complexes. These terms generally describe polar–apolar interactions, loss of ligand flexibility (entropy), and desolvation effects. The Glide Score[280] is a regression-based scoring function. Force field-based scoring functions (AutoDock, DOCK) are based on the nonbonded terms of the force fields of classical molecular mechanics. A Lennard–Jones potential describes van der Waals interactions, whereas the Coulomb energies describe the electrostatic interactions. Several developed docking approaches use knowledge-based scoring functions based on statistical observations of intermolecular close contacts in protein–ligand X-ray databases, which are used to derive potentials of mean force. This method assumes that the amount of close intermolecular interactions between individual ligand and protein atoms contributes favorably to the binding affinity. In this approach, no fitting to experimental affinities is required and solvation and entropic terms are treated implicitly.[281]

It has been recognized that various docking programs and scoring functions behave differently for different targets[282,283] as well as for different ligand types.[284] In this situation, docking studies, for such ligands as carbohydrates, provide novel ways for examining mechanisms of ligand interactions with proteins and furnish some impetus for development of scoring functions to be established for these systems. There are some peculiarities to protein–carbohydrate interactions, as compared to typical ligand–protein interactions, which make carbohydrate docking particularly

challenging. The extreme flexibility, large number of hydroxyl groups, extensive hydrogen-bonding networks, and the formation of CH–π contacts[112] may be inappropriately accounted for by widely used docking programs and docking functions. Carbohydrate-specific scoring functions[285,286] and docking algorithms aim to address these specific issues associated with carbohydrate–protein docking.[287,288] Existing software and functions need to be evaluated for these systems before they can be used reliably and routinely,[289] and with the understanding that a perfect docking program does not yet exist. Results of molecular docking are often considered in conjunction with MD and simulation or NMR-based approaches.

VI. Carbohydrate Biosynthesis and GTs

A myriad of glycan structures found in nature is derived from the enzymatic formation and the breakdown of glycosidic linkages achieved by carbohydrate-processing enzymes, such as glycoside hydrolases (GHs) (glycosidases) and GTs. Glycosylation proceeds in a stepwise manner, and therefore, the expression and specificity of the enzyme constitute key regulatory factors in defining the repertoire of biosynthesized glycans. The covalent addition of glycan to proteins and lipids comprises not only the most abundant posttranslational modification but also by far the most structurally diverse. Structural changes in cell-surface glycans accompany many physiological and pathological cell processes. The functional significance of these changes is still not fully understood. The understanding of the mechanisms utilized by these enzymes is, therefore, of a great interest.

GTs (EC 2.4.x.y in the Enzyme Classification system) are carbohydrate-processing enzymes that transfer glycosyl residues from a donor to other molecules.[290–293] The glycosyl donors are mostly sugar nucleotides, such as UDP-GlcNAc, UDP-Gal, and GDP-Man. However, lipid-linked sugars, such as glycosyl dolichol phosphates and unsubstituted glycosyl phosphates, are also utilized. Acceptor substrates are carbohydrates, proteins, lipids, and DNA, as well as antibiotics, or other small molecules. The GTs dependent on sugar nucleotides are often referred to as Leloir enzymes. GTs display low sequence homology,[294] and they have recently been classified into 90 families, GT-1 to GT-90, based on amino acid sequence similarities of over 87,000 GTs.[295,296] Based on the large number of GT families and their functions, it might be expected that there are, as in the case of glycosidases, many different GT folds. Surprisingly, the 3D architectures of Leloir-type GTs are remarkably conserved, and their X-ray structures exhibit only two general types of folds, termed GT-A and GT-B. The number of folds was recently extended by the prediction

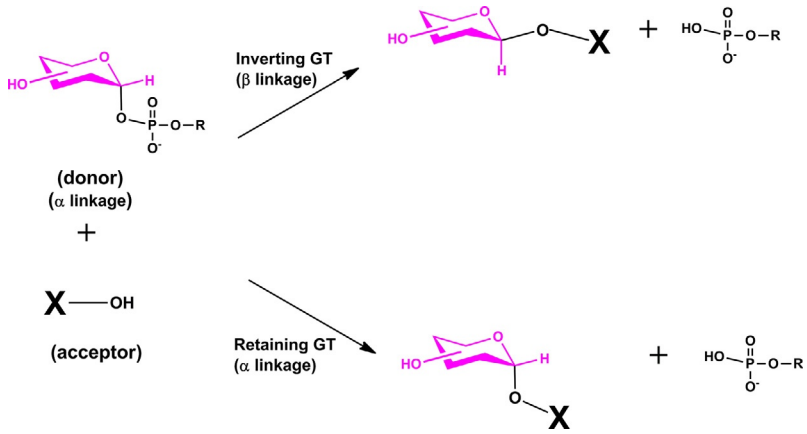

FIG. 14. Schematic diagram of overall reaction catalyzed by glycosyltransferases.

of a GT-C fold,[297] and the first structure with this fold[298] has been solved. The chemistry of the catalytic reaction can be regarded as a nucleophilic displacement of the substituted phosphate leaving group (such as the UDP group) at the anomeric carbon atom (C-1) of the transferred saccharide residue of a donor by a hydroxyl group of a specific acceptor. The formation of a new glycoside linkage during this reaction can proceed mechanistically with either inversion or retention of stereochemistry at the anomeric carbon (C-1) of the donor sugar (Fig. 14). Thus, GTs can be classified into either retaining or inverting enzymes, depending on the stereochemical outcome.

Several detailed reviews of mechanistic and structural studies of GTs have been published.[299–311] Experimental data and theoretical calculations support an S_N2-like direct displacement reaction for inverting GTs (Fig. 15), with the amino acid side chain at the active site functioning as a catalytic base that deprotonates the attacking nucleophile of the acceptor. For retaining GTs, both a double-displacement mechanism involving a covalent glycosyl–enzyme intermediate, and an alternative S_Ni-like mechanism for GTs lacking an appropriately positioned nucleophile in the active site were suggested (Fig. 16).

1. Inverting GTs

a. Cluster Models.—The first applications of modeling methods to explore the catalytic mechanism of GTs were performed by using a cluster model of active sites.[312–315] The basic idea of the cluster approach for modeling the catalytic

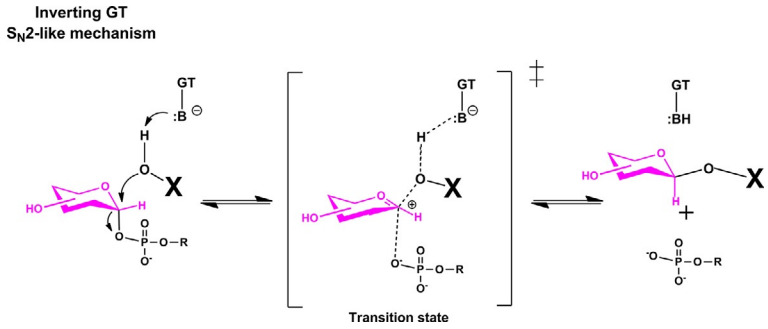

FIG. 15. Schematic representation of the reaction mechanism proposed for inverting glycosyltransferases. The reactions proceed in a single-displacement S_N2-like mechanism, with formation of an oxocarbenium ion transition state. A catalytic amino acid (B) serves as a general base that deprotonates the nucleophile hydroxyl group of the acceptor (HOX).

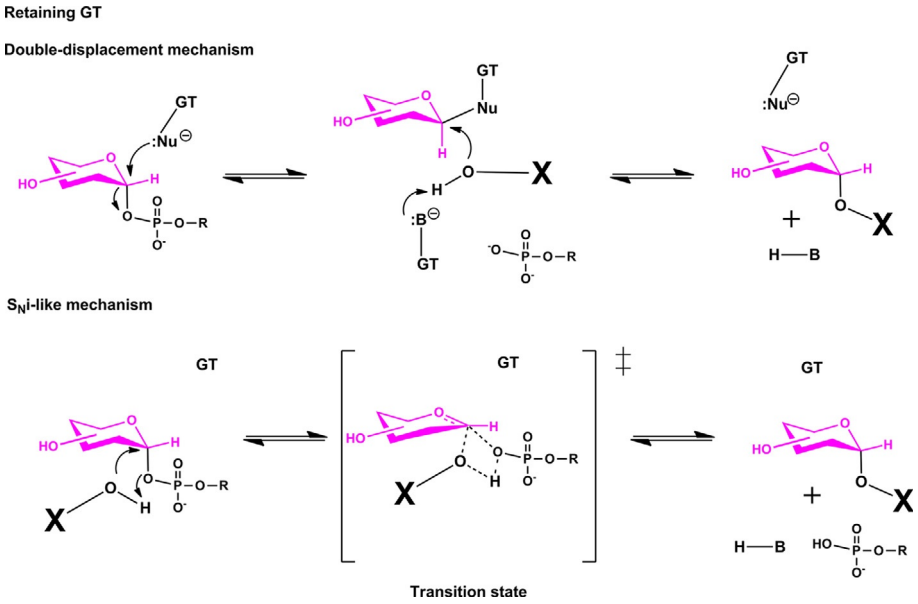

FIG. 16. Schematic representation of reaction mechanisms proposed for retaining glycosyltransferases. Double-displacement mechanisms involve two successive S_N2-like steps with nucleophilic attack of an amino acid of a glycosyltransferase on the anomeric center of the donor substrate, leading to the formation of a covalent glycosyl–enzyme intermediate. In the second step, the glycosyl–enzyme intermediate is attacked by a hydroxyl group of the acceptor (HOX), facilitated by its deprotonation by a catalytic base, resulting in net overall retention of configuration. For the S_Ni-like mechanism, front-side nucleophilic attack proceeds in a single step with the formation of an enzyme-stabilized oxocarbenium ion. The deprotonation of the acceptor nucleophile is facilitated by interaction with the departing phosphate and the incoming acceptor nucleophile.

mechanism of enzymes is to select a relatively small part of the enzyme and treat it with quantum chemical methods as accurately as possible.[209] Because of the lack of X-ray structures of GTs, the early investigations[312,313] used the arrangement based on the crystal structures of GHs. The active-site models were quite large for that time (86 atoms) and consisted of all the essential molecules or their fragments implicated in the catalytic reaction. Various possible mechanisms for the reaction were investigated by calculating energy as a function of predefined reaction coordinates that represent the formation of a new glycosidic linkage, cleavage of the glycosidic linkage of the donor, and transfer of a proton from the acceptor hydroxyl group to a catalytic base. Potential energy surfaces (PESs) calculated using high-accuracy electronic structure methods (from HF/6-31G** to B3LYP/6-311++G**) revealed 19 transition states and 9 intermediates related to various reaction pathways. Surprisingly, it was found that even these crude models reproduce rather well many of the experimental features of the catalytic mechanism. In particular, the prediction that the enrollment of only a catalytic base (in contrast with GHs) is required for the catalytic reaction of GTs was unexpected and is supported by all experimental data up to now. Furthermore, the calculations provided the first structures of the transition-state models for these reactions.

These results were quite optimistic, however, and a subsequent investigation[314] using a more complex model based on the X-ray structure of UDP-*N*-acetylglucosamine:α-3-D-mannoside β-1,2-*N*-acetylglucosaminyltransferase I (*N*-acetylglucosaminyltransferase I, GnT-I, EC 2.4.1.101) showed some limitations of this approach. These studies have been already reviewed,[304,305,308] but a few comments as background may be useful. There are several factors that influence the calculated relative energies and, as a consequence, the predicted mechanism. The first one is associated with the quality of the quantum chemical method used for calculations, as discussed in a previous section. The second is associated with the cluster approach. The environment of the enzyme that is not explicitly included can affect the model by imposing steric constraints on various regions of the model. The protein environment can also provide long-range polarization, which can affect the computed energies. In the following example, we illustrate how the number of atoms and their arrangement in a model can influence the predicted mechanism.[1]

[1]The names assigned in this article to individual monosaccharides, oligosaccharides, and sequences in polysaccharides are in conformity with the 1996 IUPAC–IUBMB Nomenclature of Carbohydrates recommendations.[15] The names used for individual enzymes are, for the most part, those coined by authors and are not necessarily officially recommended names.

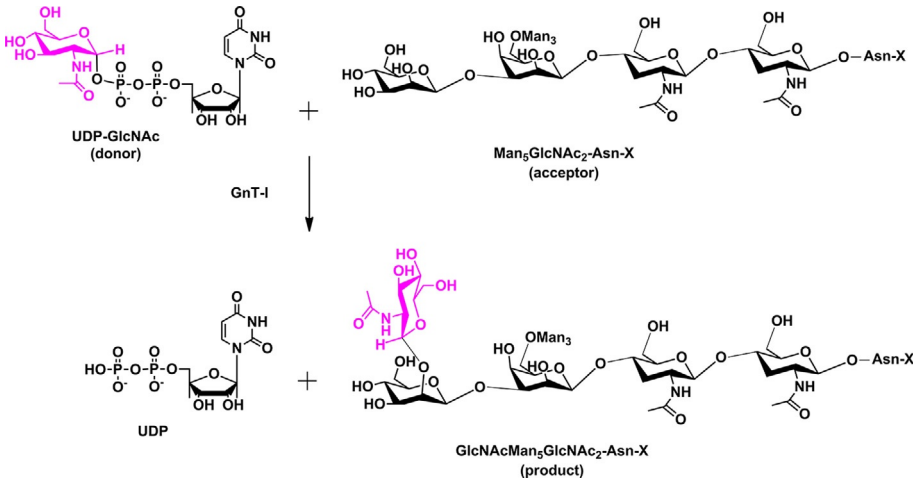

FIG. 17. Schematic diagram of the enzymatic reaction catalyzed by GnT-I.

GnT-I catalyzes the transfer of GlcNAc from the donor UDP-GlcNAc [uridine 5′-(2-acetamido-2-deoxy-α-D-glucopyranosyl diphosphate)] to the acceptor, which is the 2-hydroxyl groups of a mannose residue of the trimannosidic core of the $Man_5GlcNAc_2$ oligosaccharide, and is the first step in the biosynthesis of hybrid and complex N-glycans. This reaction occurs in the *cis*-Golgi apparatus and can be regarded as a nucleophilic displacement of the UDP group at the anomeric carbon atom (C-1) of the UDP-GlcNAc donor by a 2-hydroxyl group of the oligosaccharide acceptor (Fig. 17). The enzymatic reaction of GnT-I leads to inversion of the anomeric configuration.

The first model was based on analogy with X-ray structures of GHs[312] and consisted of a complete sugar-donor molecule, a hydroxyl group of the oligosaccharide acceptor modeled by methanol, a divalent metal cofactor represented by Mg^{2+}, as well as the essential parts of the catalytic acid and catalytic base as modeled by acetic acid and acetate molecules, altogether 86 atoms. The second, more-elaborated model of 127 atoms[314] was based on the crystal structure of GnT-I complexed with UDP-GlcNAc[316] and contained (i) the complete sugar-donor molecule, UDP-GlcNAc, (ii) the oligosaccharide acceptor modeled by monosaccharide derivatives, (iii) a divalent metal cofactor modeled by Mg^{2+} chelated by three water molecules and glutamate, (iv) the essential fragment of another glutamate, which was assumed to be the catalytic base (B), and (v) the essential fragments of glutamate and aspartate that interact with the uridine part of the donor. All of these were also included in the

model. In this model, the relative positions of the α-carbon atoms representing the GnT-I were constrained to their X-ray positions. However, side-chain fragments were allowed to relax. Figure 18 shows the PES calculated in terms of the distance between C-1 and the oxygen atom (O_N) of the attacking hydroxyl-group nucleophile (formation of the glycosidic bond) and the distance between the H_N proton of the nucleophile and the oxygen atom (O_B) of the catalytic base (proton transfer). Clearly, the calculated PES for two reaction-site models differs and implies a different reaction mechanism. The PES given in Fig. 18A shows the presence of two distinct pathways that accord with the stepwise mechanism. Both pathways lead to the product complex (P), but differ in the order of the individual steps. The pathway (yellow dashed line) with the attack (along the horizontal axis) of the nucleophilic oxygen O_N on the anomeric carbon C-1 of UDP-GlcNAc as the first step and the proton (H_N) transfer (along the vertical axis) to the catalytic base (B) as the second one was found to be energetically favorable. The overall energy barrier predicted for this mechanism was approximately 15 kcal mol^{-1}, which was in reasonable agreement with experimental data. In contrast, the PES for the second model given in Fig. 18B shows the presence of one transition state (TS) and two maxima at corners of the map and implies a S_N2-like concerted mechanism. The energy barrier (42 kcal mol^{-1}) calculated for this mechanism is considerably higher than the range (15–25 kcal mol^{-1}) of experimentally determined barriers for GTs. The authors explained this difference by the loop structuring and long-range electrostatic interactions, which were not considered.[314] An estimate of these effects decreased the calculated energy barrier to a more

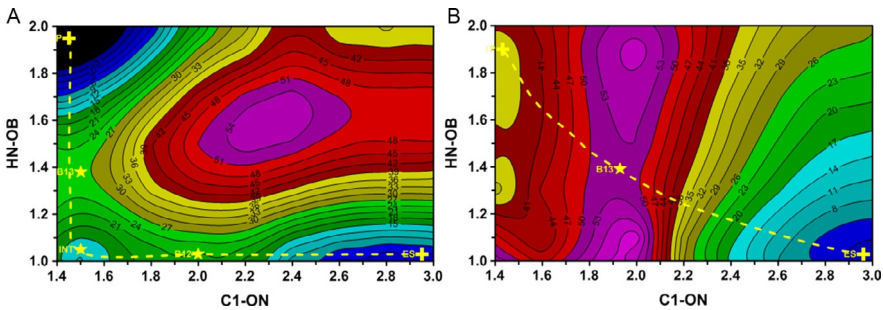

FIG. 18. PES-calculated level, using distances $r_{C1\text{-}ON}$ and $r_{HN\text{-}OB}$ as reaction coordinates for GnT-I using two different cluster models. (A) Structure of the 87-atom model was based on X-ray structures of glycoside hydrolases.[311] (B) The structure of the 127-atom model was based on the X-ray structure of GnT-I.[314] The yellow dashed line from the bottom left corner to the upper right corner indicates (a) the stepwise reaction pathway from the ES to the P via intermediate INT; (b) the S_N2-like reaction pathway from the ES to the P via the B13.

reasonable 20–25 kcal mol^{-1}. However, it is worth mentioning that the C-1—O-1 bond progressively dissociates with the creation of the new glycosidic bond between the anomeric carbon and nucleophilic oxygen.

Several relevant conclusions can be drawn from these results. First, it is evident that these cluster models clearly neglect the effects of the protein environment, and the calculated energies must be treated with caution. Another intriguing observation came from the simultaneous occurrence of the nucleophilic attack by O_B, and the dissociation of the leaving group. Interestingly, the geometry of the transition state was quite similar in both models. Here, it is particularly important to emphasize that arrangement of molecules in the model, and geometry optimization, is a key factor of the cluster approach. In spite of these limitations, the cluster approach has been quite a useful tool in the study of the GT reaction. However, to understand all mechanistic aspects of the catalytic mechanism of GTs, it is essential to take into account effects imposed by a whole protein environment, as by using QM/MM methods. These investigations will be surveyed next.

b. *N*-Acetylglucosaminyltransferase I.—This enzyme was the first GT studied using a QM(DFT)/MM method.[317] In QM(DFT)/MM calculations[317] of the catalytic mechanism of GnT-I, the trimannosyl core αMan-(1 → 3)-[3,6-OMe-αMan-(1 → 6)]-β Man of the $Man_5GlcNAc_2$-Asn-X oligosaccharide was used as the acceptor. This oligosaccharide comprises the minimal acceptor-binding determinant[318] and was docked into the X-ray structure of GnT-I in complex with UDP-GlcNAc.[316] In the QM/MM approach, the entire ternary complex of GnT-I, with donor, acceptor, and metal cofactor (Mn^{2+}), consisted of 5721 atoms and was divided into QM and MM parts (Fig. 19). The QM reaction-site region consisted of 88 atoms and included (i) the DP-GlcNAc portion of the sugar-donor molecule, (ii) the αMan-(1 → 3)-mannose residue of the trisaccharide acceptor, (iii) the catalytic base (aspartate D291), and (iv) the divalent metal cofactor Mn^{2+} fully coordinated by three water molecules and aspartate D213. The remaining part of the substrates and the enzyme, altogether 5633 atoms, was included in the MM region and treated by the AMBER95 all-atom force field. The reaction path calculated supported a concerted S_N2-type mechanism with the activation barrier of 19 kcal mol^{-1} and correlated well with available experimental data. The QM/MM model provides a reasonable activation barrier without any correction that was necessary in the truncated cluster model of the GnT-I complex. This result further supports the importance of including the protein environment in calculations of the enzymatic reactions. It was observed that the energy contribution from the surrounding enzyme decreased the reaction barrier by 9 kcal mol^{-1}. A low-barrier hydrogen bond between the O_a hydroxyl group and the carboxylate oxygen atom of the catalytic base D291 was found to be crucial for the mechanism by facilitating nucleophilic attack.

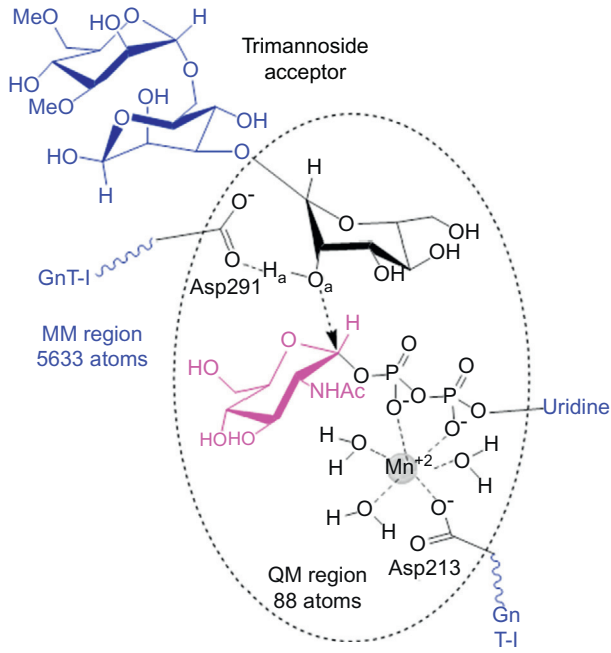

FIG. 19. Schematic diagram of the partitioning of the entire GnT-I complex into QM and MM regions.

A negative charge developing on the β-pyrophosphate was stabilized by the HO-6—β-phosphate oxygen interaction. The importance of these interactions supported the experimental data. They show that deletion of the C-6 hydroxyl group diminished the catalytic activity, and that an acceptor methylated at O-6 was not active, although it was a substrate.[318]

c. β-1,4-Galactosyltransferase-1.—The inverting enzyme β-1,4-galactosyltransferase-1 (β4Gal-T1; EC 2.4.1.38), in the absence of α-lactalbumin, catalyzes the transfer of the Gal residue from UDP-Gal [uridine 5′-(2-acetamido-2-deoxy-α-D-galactopyranosyl diphosphate)] to the O-4 oxygen atom of GlcNAc in the presence of a Mn^{2+} metal ion (Fig. 20). Experimental data support an S_N2-type direct displacement mechanism, with Asp318 serving as the catalytic base.[319] The structure of β4Gal-T1 was extensively investigated,[319–322] and the crystal structures of the catalytic domain of β4Gal-T1[323,324] were used recently to produce the structural model for a QM(DFT)/MM investigation.[325] The structural model of the active-site domain was built by superposing two crystal structures of β4GalT-1. The resulting model was

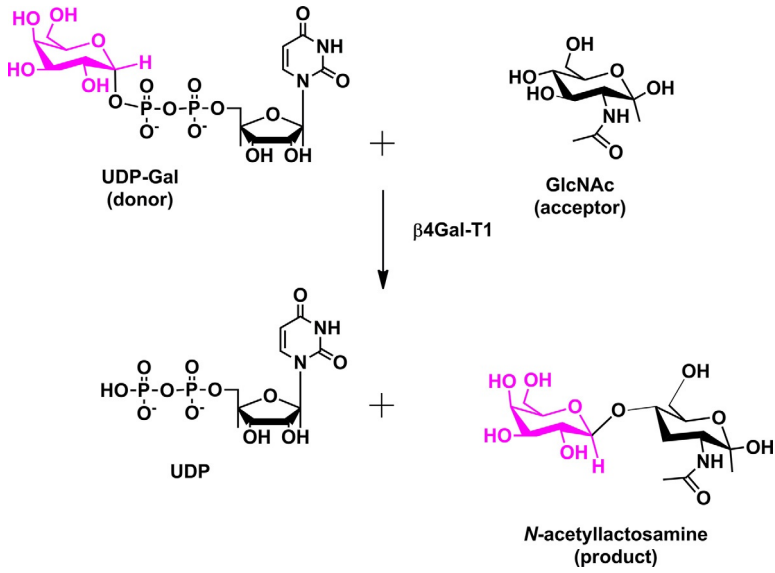

FIG. 20. Schematic diagram of the enzymatic reaction catalyzed by β4Gal-T1. Reprinted with permission from Ref. 325. Copyright 2009 American Chemical Society.

divided into two parts, the QM region with 253 atoms and the MM region having 4274 atoms. The QM region consisted of the acceptor and donor substrates, the metal cofactor, and the side chains of 11 amino acids implicated in binding the substrates.

The catalytic reaction was investigated by using the reaction coordinate describing the nucleophilic attack of the acceptor oxygen on the anomeric carbon of the donor. The QM/MM calculations suggested a concerted S_N2-type mechanism. The C-1—O-1 distance of 2.092 Å and the C-1—O-4 distance of 2.703 Å, together with a distorted ring shape of the donor, are representative of the TS structure. It is noteworthy that a rotation of the diphosphate group was found, similar to that observed in GnT-I. Whether this is characteristic for the concerted S_N2 mechanism of inverting GTs remains to be seen. The calculated barrier of 15 kcal mol^{-1} is in agreement with the experimental barriers. The QM(DFT)/MM calculations revealed a low-barrier hydrogen bond between the nucleophilic hydroxyl group of the acceptor (O-4) and the catalytic-base oxygen atom O(D318), supporting nucleophilic attack. The developing charge on diphosphate is stabilized, as with GnT-I, by the primary hydroxyl group (HO-6) of the acceptor. However, in the case of β4GalT-1, this interaction is mediated by water molecules.

d. O-GlcNAc Transferase.—GT β-*N*-acetylaminyltransferase (uridine diphospho-*N*-acetylglucosamine:polypeptide, OGT) transfers *N*-acetylglucosamine from UDP-GlcNAc onto the hydroxyl group of a serine or a threonine residue on protein substrates (Fig. 21). OGT is an inverting and metal-independent GT. It is noteworthy that the O-linked GlcNAc residue is not further modified into complex oligosaccharide structures.[326] The posttranslational modification of serines or threonine residues on proteins by O-linked GlcNAc is emerging as one of the most important cellular glycosylation processes. There is considerable evidence[327,328] to suggest that O-GlcNAcylation regulates a wide range of cellular processes and is implicated in several diseases, including diabetes, cancer, and neurodegeneration.

The catalytic mechanism of OGT has been studied by using QM/MM calculations.[329] The crystal structure of a truncated human OGT containing 4.5 tetratricopeptide repeats and in the complex with UDP[330] was used to build the structural model of the active site. Histidine 498 positioned between the acceptor hydroxyl and donor groups was suggested as the catalytic base.[330] The entire enzyme–substrate system (OGT-UDP-GlcNAc-CKII peptide ternary complex), containing 11,524 atoms, was divided into the QM and the MM subsystems. The QM region, consisting of 198 atoms, contained the donor substrate (UDP-GlcNAc), three residues of the acceptor (CKII peptide), the side chains of the catalytic base His498, and six amino acids relevant for catalytic activity. Also, three water molecules in the neighborhood

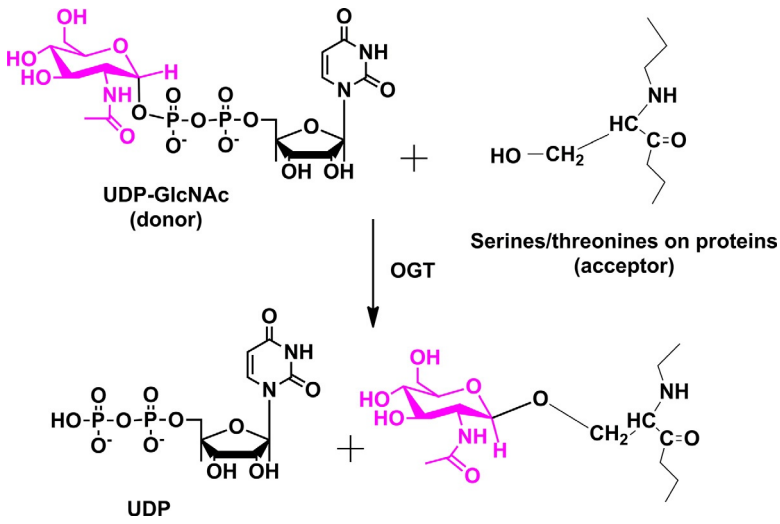

FIG. 21. Schematic diagram of the enzymatic reaction catalyzed by OGT.

of the donor were included into the QM region. The MM region consisted of 11,326 atoms from the remaining atoms of the complex. The enzymatic reaction was monitored by means of three reaction coordinates. They described the formation of a new glycosidic linkage between the serine acceptor and the anomeric carbon atom (C-1) of the donor GlcNAc, cleavage of the donor glycosidic linkage between GlcNAc and UDP, and a transfer of the proton from the acceptor hydroxyl group to a catalytic base. The reaction coordinates are given in Fig. 22, which shows only the QM region. The examination of PES required rather extensive calculations, and more than 600 structures were optimized to generate the three PES.

The PES, calculated as a function of the distance between C-1 and the O_{Ser} oxygen atom (r_1) and the distance between the H_{Ser} proton and N_{His} nitrogen atom (r_3), is given in Fig. 23. The horizontal axis represents the formation of the C—O glycosidic bond while the vertical axis represents the proton transfer to histidine. The calculated PES indicated the existence of only one reaction pathway, with a single transition barrier (B13) between the Michaelis complex (ES) and the products (P). This information pointed to a concerted, S_N2-like mechanism, in which the nucleophilic attack by the serine hydroxyl group and the dissociation of the leaving group all occur almost simultaneously. The nucleophilic attack is facilitated by the proton transfer to the histidine residue. The reaction barriers of 15.6, 19.6, and 15.5 kcal mol^{-1} calculated

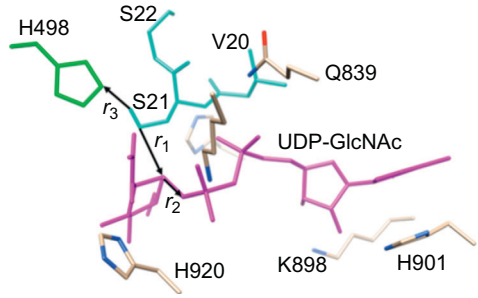

FIG. 22. View of the QM region of the OGT–UDP-GlcNAc–CKII peptide complex in stick representation. The QM region contains complete UDP-GlcNAc (in magenta), three whole residues, namely Val20, Ser21 (acceptor), and Ser22 of the CKII peptide (in cyan), and side chains of amino acids crucial for the catalytic activity: H498 (suggested catalytic base, in green), His558, Gln839, Lys842, Lys898, His901, and His920 (in tan). There are also three water molecules in the vicinity of UDP-GlcNAc (not shown for clarity). The QM region consists of 198 atoms. The reaction coordinates r_1, r_2, and r_3 are depicted as arrows. Reprinted with permission from Ref. 329. Copyright 2001 American Chemical Society.

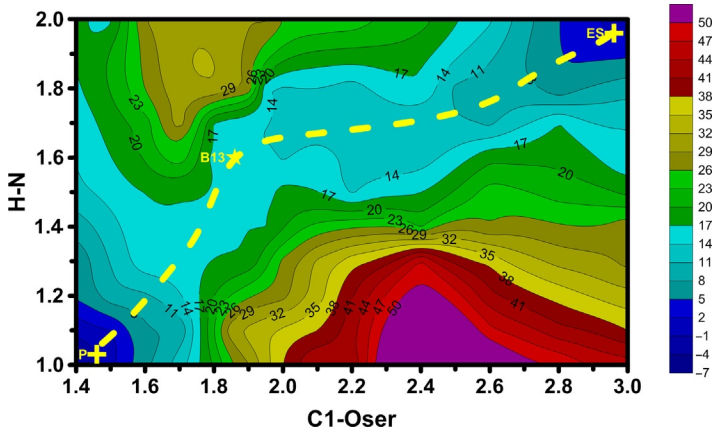

FIG. 23. PES calculated at the B3LYP/6-31G* level using distances r_1 and r_3 as reaction coordinates. The yellow dashed line from the upper right corner to the bottom left corner indicates the S_N2-like reaction pathway from the ES to the P via the B13. Reprinted with permission from Ref. 329. Copyright 2012 American Chemical Society.

at the B3LYP, MPW1K, and M06-2X level, respectively, are in reasonable agreement with the experimentally estimated activation barrier[331] of 21 kcal mol^{-1}.

The calculations also revealed that OGT employs the donor substrate acetamide to assist in cleavage of the glycosidic linkage. The stabilization of developing negative charge on the phosphate group (Fig. 24A) by the donor acetamido group in the catalytic process was observed for the first time with GTs. It is striking that a GH (OGA) involved in O-GlcNAcylation cycling uses an acetamido group as a nucleophile to cleave GlcNAc from serine/threonine.[332] Two independent groups have recently shed some light on the mechanism of OGT[331,333] by solving the structures of the complexes of OGT with the peptide acceptor and the slowly reacting UDP-GlcNAc analogue UDP-5-thioGlcNAc. Both articles support the role of the substrate acetamide in the enzymatic catalysis, as predicted by QM/MM calculations.[329] Interestingly, the two articles, regardless of the similarity of the structures solved, proposed quite different mechanisms. One group[333] proposed a mechanism wherein the α-phosphate group serves as the base catalyst (Fig. 24B). In contrast, the second group[331] explicitly rejects α-phosphate as a candidate base. Instead, a mechanism is favored in which a proton is transferred, probably to Asp554, via a chain of water molecules in a Grotthuus type of mechanism (Fig. 24C). In contrast, the role of Asp554 was rejected by the first group. There are some indications that 5-thio donors may influence the catalytic mechanism of inverting GTs.[334] The question is whether 5-thio donors may also influence the mechanism with OGT. Currently, thus, there are

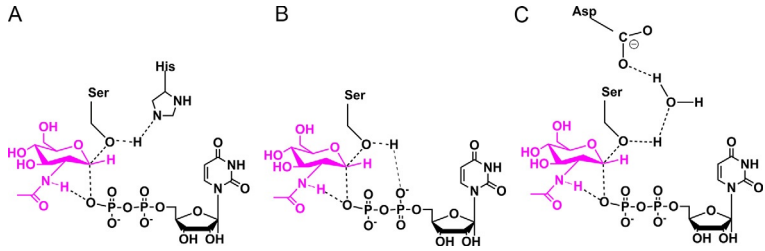

FIG. 24. Schematic diagram of the three proposed mechanisms for the enzymatic reaction catalyzed by OGT. (A) Histidine 291 as the catalytic base.[329] (B) The α-phosphate as the catalytic base.[333] (C) The Grotthuus water "wire" shunting mechanism.[331]

three different mechanisms that have been proposed, and further experimental and theoretical studies are certainly needed to solve this puzzle.

e. Core 2 Transferase.—The Golgi enzyme UDP-GlcNAc:Galβ1-3GalNAc-(GlcNAc to GalNAc) β1,6-GlcNAc-transferase (C2GnT) transfers GlcNAc to the core 1 structure on βGal-(1→3)-βGalNAc-O-Ser/Thr glycoproteins and forms the core 2 structure, βGlcNAc-(1→6)-[βGal-(1→3)]-βGalNAc-O-Ser/Thr. The QM/MM calculations performed on this inverting, metal-ion-independent, enzyme with GT-A fold support the substrate-assisted S_N2-like mechanism.[335] In the calculations, the entire enzyme–substrate system (C2GnT–UDP-GlcNAc–βGal-(1→3)-GalNAc ternary complex) was partitioned into the QM region containing 206 atoms and the MM subsystem consisting of 5914 atoms. The PESs calculated clearly revealed three simultaneous processes, namely the nucleophilic attack, the dissociation of the C-1—O-1 glycosidic linkage, and the transfer of a proton from the nucleophile oxygen to the catalytic base. As with the inverting GT OGT,[329] a metal-ion-independent enzyme with GT-B fold, it was found that the hydrogen-bond interaction between the H_{NAc} proton of GlcNAc and the glycosidic oxygen O-1 of the donor facilitates breaking the glycosidic linkage and the withdrawal of the leaving group (UDP). The transition state for the proposed reaction mechanism at the M06-2X/6-31G** level was located at C-1—O-6 = 1.74 Å and C-1—O-1 = 2.86 Å. The reaction barrier of between 20 and 29 kcal mol^{-1} was calculated for the C2GnT catalysis, depending on the functional used.

2. Retaining GTs

In contrast to inverting GTs, our understanding of the catalytic mechanism of retaining GTs is still a matter of debate.[305,307] Some retaining GTs have been

proposed to proceed via a double-displacement mechanism.[336,337] For retaining GTs that lack a suitably positioned nucleophile to form a glycosyl–enzyme intermediate, such as lipopolysaccharyl α-1,4-galactosyltransferase C (LgtC)[338] and trehalose-6-phosphate (OtsA),[339,340] an S_Ni-like mechanism was proposed. Theoretical analyses[305] of truncated cluster models have shown that both pathways are accessible, and that the S_Ni-like mechanism is favorable as the consequence of active-site geometrical constraints. Up to now, there have been few theoretical studies[264,313,315,341–345] on the catalytic mechanisms of retaining GTs, and they are discussed next.

a. α-1,3-Galactosyltransferase.—The enzyme UDP-galactose:*N*-acetyllactosaminide 3-α-D-galactosyltransferase (α-1,3-galactosyltransferase, α3GalT, EC 2.4.1.151) transfers galactose from UDP-Gal to the terminal *N*-acetyl-lactosamine component of glycans, with retention of the anomeric configuration, producing the αGal-(1 → 3)-βGal-(1 → 4)-GlcNAc oligosaccharide structure present in most mammalian glycoproteins (Fig. 25).

The crystal structures of the catalytic domain of bovine α3GalT (substrate-free and substrate-bound complex) have been solved.[346,347] These structural data show that α3GalT has a GT-A fold. The conserved residue E317 was proposed to function as the catalytic nucleophile involved in a double-displacement mechanism.[336,337,346] The role played by E317 in both possible mechanisms was investigated by constrained QM/MM MD simulation along predefined reaction coordinates, describing formation of a new glycosidic linkage and dissociation process, respectively.[342] The system studied contained 12,694 atoms and was divided into QM and MM regions. The QM

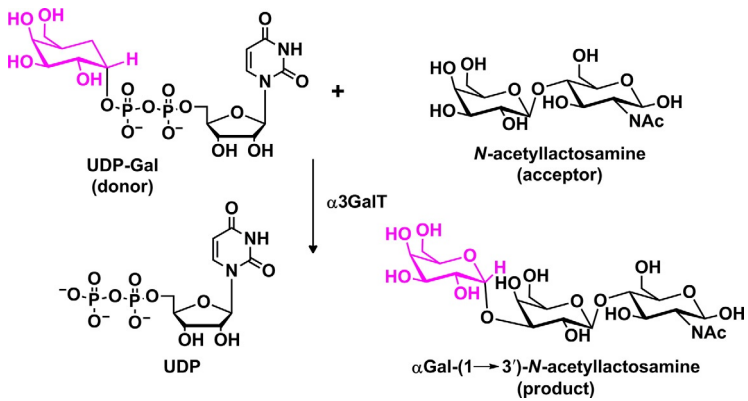

FIG. 25. Schematic diagram of the enzymatic reaction catalyzed by α3GalT.

region consisted of 84 atoms and included two galactose residues, the diphosphate group, side chains of the catalytic base, metal, and side chains of amino acids coordinating the metal cation. In the calculations, the active-site region (2089 atoms) was allowed to relax while the rest of complex was fixed. Random snapshots from 20 ps QM/MM MD simulation were selected for subsequent optimization of stationary points along the reaction paths. For the double-displacement mechanism, the results are in qualitative agreement with previous calculations using the cluster model[313] and support the role of UDP in deprotonating the nucleophile. For the S_Ni-like mechanism, the calculations supported a TSi character of the transition state, and UDP was found as the catalyst in the front-side attack. Interestingly, the authors, encountering similar barriers, did not reach a conclusion as to which mechanism is the dominant one. To deduce this would require more extensive calculations and ensemble-averaged energies. Another relevant conclusion to be mentioned is that E317 plays a crucial role in both mechanisms and, therefore, the interpretation of E317 mutation experiments is not simple.

b. α-1,4-Galactosyltransferase C (LgtC).—Lipopolysaccharide α-1,4-galactosyltransferase C (LgtC) from *Neisseria meningitides* is a retaining galactosyltransferase possessing a GT-A fold. LgtC is responsible for the transfer of α-galactose from a UDP-Gal donor to a galactose residue of the terminal lactose moiety in the bacterial lipooligosaccharides (Fig. 26). The X-ray structure of LgtC[338] with Mn^{2+} and a nonreactive donor was solved in the absence and presence of the acceptor analogue. Both of these X-ray structures revealed a catalytic site in which the only functional group appropriately placed to serve as the catalytic base is Gln189. However, the

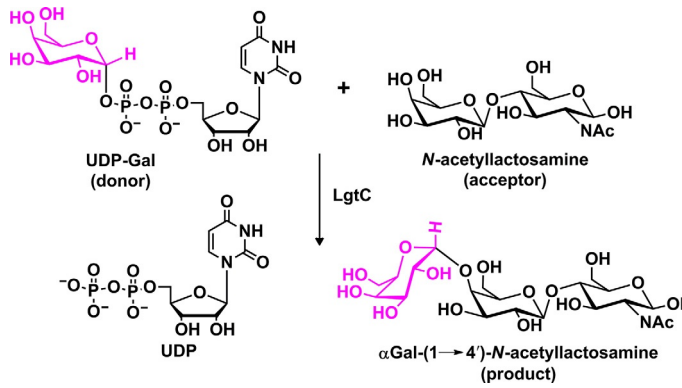

FIG. 26. Schematic diagram of the enzymatic reaction catalyzed by LgtC.

mutation experiments did not support its role. Furthermore, other electronegative groups located in the active site in the vicinity of the UDP-Gal donor were considered[348] as possible catalytic nucleophiles, but analysis of the results was inconclusive. The difficult identification of a catalytic base, as well as the absence of convincing evidence for the existence of a glycosyl–enzyme intermediate, led to the suggestion[338] that this enzyme might instead proceed through a single front-side displacement reaction, thus an S_Ni process. Extensive theoretical calculations[315,341,349] were used to shed some light on mechanistic details of the catalytic reaction of LgtC. The DFT calculations of the truncated cluster model (136 atoms) predicted[315] that the favored mechanism is a one-step process in which the acceptor oxygen attacks the anomeric carbon atom of the donor UDP-Gal from the side of the leaving group (UDP), with synchronized proton transfer to a phosphate oxygen atom.

The calculations also showed that Gln189 is involved in stabilization of the transition state by hydrogen bonding to the donor. The predicted structure of the transitions state was unique and characteristic for the S_Ni mechanism. It is noteworthy that recent experimental data showed that retaining GTs of this kind use the leaving-group phosphate oxygen as the catalytic base to deprotonate the acceptor nucleophilic oxygen.[339,340] The QM/MM calculations[341] of the full LgtC supported the foregoing findings. In the calculations, the authors used the ternary complex of LgtC with UDP-Gal and lactose based on X-ray structures,[338] solvated with a 24 Å radius sphere of water molecules. The manganese ion was modeled by Mg^{2+}, and the whole system consisted of 6728 atoms, with 101 atoms in a QM region. Reaction paths were scanned by performing constrained optimization along the reaction coordinate, which was defined as $r = d(\text{O-1—C-1}) - d(\text{O}_\text{N}\text{—C-1})$, where d(O-1—C-1) represented dissociation of the donor C-1—O-1 bond and $d(\text{O}_\text{N}\text{—C-1})$ represented the bond being formed. The calculations supported an S_Ni-like mechanism with highly dissociative character of the TS. Umbrella-sampling molecular dynamic simulations were used to estimate the free-energy barrier. The predicted barriers were in a reasonable agreement with the experimental value of 16 kcal mol^{-1}.[338,348]

c. Trehalose-6-Phosphate Synthase (OtsA).—The GT OtsA is a retaining GT of the GT-B fold. This enzyme catalyzes the transfer of a glucosyl residue from a donor substrate, UDP-Glc, to the 1-hydroxyl group of the acceptor substrate (glucose 6-phosphate), yielding the product α,α-trehalose-6-phosphate (Fig. 27). The catalytic mechanism of OtsA has been investigated by using QM/MM metadynamics simulations.[264] The Michaelis complex and the products of this reaction were generated by using the X-ray structure of the ternary complex of OtsA with UDP and validoxylamine 6-phosphate.[339] QM/MM calculations were performed by a method[245] that combines Car–Parrinello DFT molecular dynamics[255] with classical force-field molecular dynamic. The QM region contained 72 atoms and included the donor

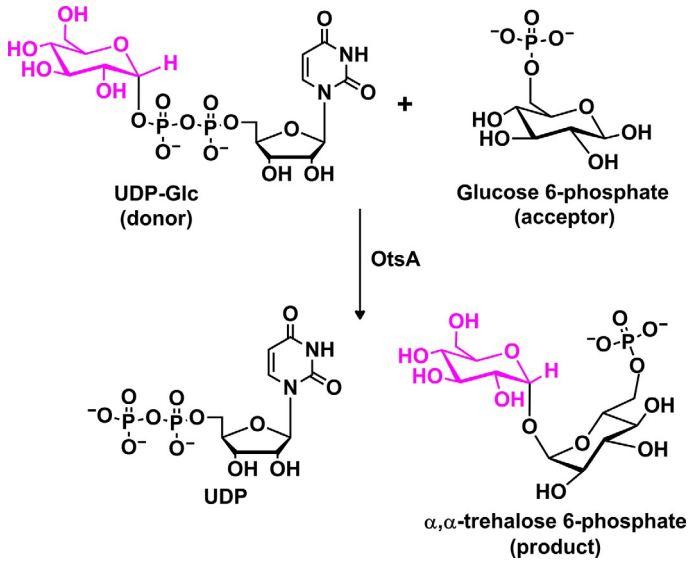

FIG. 27. Schematic diagram of the enzymatic reaction catalyzed by OtsA.

UDP-Glc, the acceptor Glc 6-phosphate, and the side chain of Arg262. Two CVs were employed to investigate glucose transfer by a metadynamics simulation. The first variable described the cleavage of the C-1—O-1 glycosidic linkage of the donor and the formation of the new glycoside linkage, while the second described the proton transfer from the acceptor to the phosphate group. These simulations supported the S_Ni or S_N1-like reaction, in agreement with recent experimental studies on OtsA.[339,340] The calculations also showed an important role of the donor phosphate group that serves as the catalytic base for the acceptor proton, which assists in cleavage of the glycosidic bond of the donor.

d. Others.—The retaining enzyme polypeptide UDP-GalNAc transferase (ppGalNAcT2) has been studied by two different research groups.[343,344] This enzyme catalyzes the transfer of a GalNAc residue from the donor UDP-GalNAc to threonine/serine as a first step in mucin biosynthesis. The ppGalNAcT2 transferase is the metal dependent of the GT-A fold. In the QM/MM calculations,[343] the QM region consisted of 80 atoms, and the natural metal cofactor (Mn^{2+}) was modeled by Mg^{2+}. The calculation revealed that the PES in the region corresponding to the energy maximum was very flat. The calculations using QM/MM method at the M05-2X/TZVP//BP86/SVP level supported a front-size attack mechanism with the estimated reaction barrier of 20 kcal mol^{-1}. However, the transition state for this mechanism was not found.

The mechanism of this enzyme was also studied by a combination of two different QM/MM-based approaches, namely a PES scan in two distance difference dimensions and a minimum-energy reaction path optimization using the nudged elastic band method.[344] The QM region was defined to include the essential parts of the substrates and those residues experimentally known to be crucial for reactivity, and it contained 252 atoms. It was found that ppGalNAcT2 catalyzes a same-face nucleophilic substitution with internal return (S_Ni). The optimized transition state for the reaction was 13.8 kcal mol^{-1} higher in energy than the reactant, whereas the energy of the product complex was 6.7 kcal mol^{-1} lower. The presence of a short-lived metastable oxocarbenium intermediate was likely, as indicated by the reaction energy profiles obtained using high-level density functionals. The transition states for the proposed reaction mechanism were located at C-1—O-1 = 2.35 Å and C-1—O_A = 2.97 Å for TS1 and at C-1—O-1 = 3.60 Å and C-1—O_A = 2.33 Å for TS2, respectively. It is noteworthy that the C-1—O_A distance in TS2 is almost the same as the distance of the cleaving C-1—O-1 bond in the TS2. This observation supports the previously proposed concept[340,350] of the two transition states involving each glycosidic bond being very similar, being almost "mirror images" of each other.

A catalytic mechanism of α-1,2-mannosyltransferase Kre2p/Mnt1p in the presence of Mn^{2+} and other ions (Mg^{2+}, Zn^{2+}, and Ca^{2+}) was modeled at the two hybrid DFT-QM/MM (M06-2X/OPLS2005 and B3LYP/OPLS2005) levels.[345] Kinetic and structural parameters of transition states and intermediates, as well as kinetic isotope effects, were predicted and compared with available experimental and theoretical data. The catalysis in the presence of the metal ions is predicted as a stepwise S_Ni-like nucleophilic substitution reaction ($D^*_{Nint}A^{\ddagger}_{N}D_hA_{xh}$) via oxocarbenium ion intermediates.

3. Transition-State Structures

Transition states of catalytic reactions are high-energy intermediates on reaction paths between reactants and products. They have lifetimes of $\sim 10^{-13}$ s, and their structure and location in catalytic sites can be only estimated. Although X-ray crystal structures of GTs are now available,[311] transition-state analogues that could provide information about the nature of the transition states are rare.[307,339,351,352] Thus, the modeled structures of transition states constitute important outcomes of the calculated PESs. The structures of transition states assist in understanding fundamental information about the interactions crucial for enzymatic reactions. Furthermore, the resulting structures of the transition states offer a blueprint for the design of transition-state inhibitor analogues.[353,354] In principle, the "entire" transition state

for a GT should contain a deformed nucleotide sugar and acceptor oligosaccharide that are linked in a specific orientation, along with amino acids involved in the catalytic reaction or in binding substrates, and for metal-dependent enzymes, also a metal cofactor. The following discussion focuses on a narrower concept of a transition state, dealing with only the structures and arrangement of donor and acceptor parts. Investigations of the catalytic mechanism for GTs revealed[264,312–315,317,325,329,341–345] the following general features of the transition-state models: (a) the transferred monosaccharide ring is flattened and its conformation resembles a deformed chair/envelope conformation having oxocarbenium character at the sp^2-hybridized anomeric carbon; (b) the C-1—O-1 bond is elongated as compared to the standard C—O bond length; (c) the length of a new glycosidic linkage is also longer than the standard bond length; and (d) both the forming and breaking bonds are oriented almost perpendicularly with respect to the plane defined by the C-2—C-1—O-5—C-5 atoms. It is also clear that the role of the acceptor to the character of transition state is significant. Therefore, the acceptor part of the transition-state structure should be incorporated in the design of stable transition-state analogues as potent inhibitors of GTs.

For GTs using the concerted S_N2 mechanism,[311,314,317,325,329,335] calculations led to more than 20 different transition-state models. The analysis of their structures provided information on possible structural variations of transition states. It was found that variations in the C-1—O and C-1—O-1 bond lengths can be as large as 1.3 and 1.7 Å, respectively. The sugar ring in the transition states is distorted from the standard $^4C_1$ conformation, and it adopts a half-chair form. The optimized TS structures were clustered into three groups using the values of the C-1—O and C-1—O-1 distances as structural geometrical criteria.[304] The clusters obtained characterize canonical models of transition-state structures for inverting GTs and are illustrated in Fig. 28. Long C-1—O bonds in the range of 2.4–2.7 Å and short C-1—O-1 distances between 1.5 and 2.1 Å characterize the first group (Fig. 28A). The geometrical parameters of this transition-state model resemble the structure of reactants, and this canonical form has been termed as the "early transition state." The "intermediate transition-state" structures represent the second canonical form (Fig. 28B). In this group, both C-1—O (2.1–2.4 Å) and C-1—O-1 (2.5–2.7 Å) distances are elongated as compared to their standard values, but the structures did not yet reach values representing the products. This type of TS corresponds to the so-called dissociative transition state. The geometrical characteristics of the third canonical form are similar to that of final products and therefore were named as the "late transition state." For this group, short C-1—O bonds within the range of 1.4–1.6 Å and long C-1—O-1 distances between 2.8 and 3.2 Å are typical. The

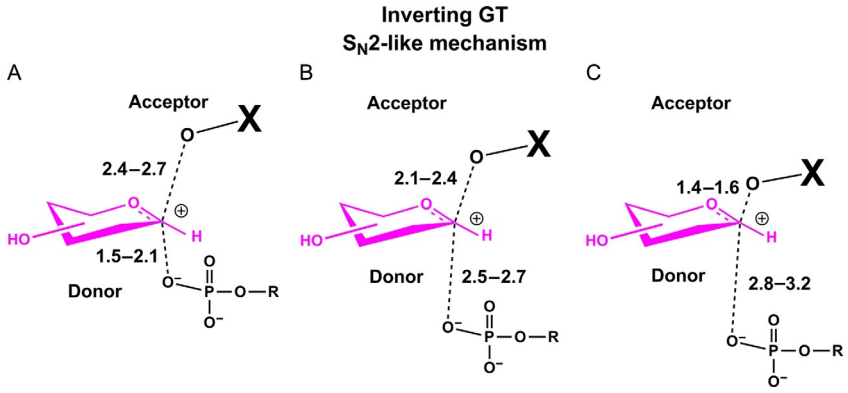

FIG. 28. Schematic representation of canonical forms of the transition states for glycosyltransferases. Canonical forms are defined by similarities in their C-1—O-1 and C-1—O distances. (A) Early transition state, (B) intermediate transition state (dissociative), and (C) late transition state.

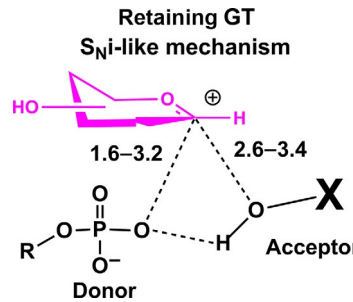

FIG. 29. Schematic representation of the transition state for retaining glycosyltransferases.

three canonical structures might represent a range of variations occurring in transition-state structures of GT reactions. Of course, the transition-state structure varies according to the structural circumstances of the catalytic site of a particular enzyme that imposes constrains on substrates.

The transition-state structures of GTs proceeding by the S_Ni mechanism clearly differ from those employing the S_N2 mechanism, and their geometry is unique (Fig. 29).[264,313,315,341–345] This structure is characterized by C-1—O-1 and C-1—O distances of about 3.0–2.4 and 2.3–3.0 Å, respectively. They indicate a weak bond

order to both the nucleophile and the leaving group. The main difference with the S_N2 transition-state structure is that both the nucleophilic attack by the acceptor oxygen O and departure of the leaving group occur on the same face of the transferred monosaccharide. The ring conformation of the transferred monosaccharide is a distorted ^4E envelope. Another interesting and unique feature of this transition state is the location of the transferred proton. The nucleophile proton is sandwiched between the oxygen atom from the phosphate group and the nucleophile. The H proton is located at the O—H and O-1—H distances of about 1.09 and 1.38 Å, respectively. The distance between the O and O-11 oxygen atom in the TS is about 2.4 Å.

VII. Recognition

1. Lectins

Lectins are proteins of nonimmune origin that bind to specific carbohydrates without modifying them. According to current knowledge, they act like molecular readers to decipher sugar-encoded information. They play biologically important roles in recognition processes involved in fertilization, embryogenesis, inflammation, metastasis, and parasite–symbiote recognition, in microbes and invertebrates to plants and vertebrates. In the plant kingdom, lectins have been demonstrated to play a role in defense against pathogens or predators and are hypothesized to be involved in establishing symbiosis with mushrooms and bacteria of the *Rhizobia* species. Among the proteins that interact noncovalently with carbohydrates, lectins bind mono- and oligosaccharides reversibly and specifically.

More than 1200 crystal structures of lectins have been solved, among which about 60% occurs in interactions with carbohydrate ligands.[355] These are fully documented in a 3D lectin database (www.glyco3D.cermav.cnrs.fr). Inspection of the database content indicates that most crystal structures have, up to now, been obtained for plant and animal lectins. Nevertheless, the number of structural investigations dealing with viral and bacterial materials has been increasing rapidly. Legume lectins and those intracellular animal lectins that are involved in quality control of glycoprotein synthesis share the same protein fold. The wealth of experimental data obtained from the crystallographic studies of oligosaccharide–lectin conjugates has provided characterization of the binding sites; these are usually relatively shallow, located near the surface, and therefore accessible to solvent. In several lectin families of different origins, one or two calcium ions are involved in the carbohydrate-binding site with

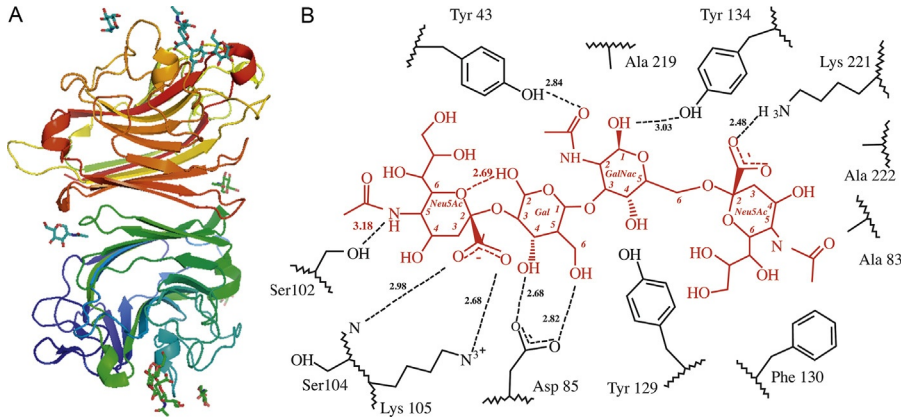

FIG. 30. Application of docking techniques to the prediction of lectin–oligosaccharide interactions as exemplified by the *Maackia* lectin. The *Maackia* lectin was built using the COMPOSER program[503] within the SYBYL software package.[141] A library of 3D crystallographic structures was created, containing 14 different legume lectins solved at high resolution. The several steps of the homology modeling procedure included 3D alignment of the 14 structures, sequence alignment of *Maackia amurensis* hemagglutinin (MAH) with these lectins, building of the structurally conserved regions, and building of the loops. Most of the MAH structure was built from *Maackia amurensis* leukoagglutinin (MAL). Five water molecules that are known to be conserved in all legume lectin structures[504] were incorporated in the model, together with Ca^{2+} and Mn^{2+} cations. Finally, hydrogen atoms were added and charges were calculated. A validation of the model stereochemistry was performed using the PROCHECK program.[505]

direct coordination to the hydroxyl groups of the sugar. These data were essential driving forces in the development of molecular modeling methods of complex oligosaccharides in their interactions with proteins.[356] They confirmed the flexible conformational behavior of oligosaccharides that had been anticipated from earlier calculations.

A wide variety of plant, mammalian, bacterial, and fungal lectins have been studied using molecular docking (for a review of earlier studies, see the study in Ref. 357). Accurate determination of carbohydrate–lectin complexes remains a nontrivial problem because of the shallow and multichambered binding sites of many lectins (Fig. 30). Selected examples dealing with carbohydrate–lectin recognition are discussed here.

One of the most comprehensive studies on the validation of carbohydrate–lectin interaction has been reported[358] in the course of an investigation comparing the abilities of AutoDock, DOCK, and Glide to reproduce the carbohydrate-binding modes of seven calcium-dependent lectins. On the one hand, Glide was found to perform best, in terms of pose (candidate binding mode), as assessed by values of

RMSD in comparison with the crystal structure. However, AutoDock was found to reproduce calcium-binding geometry most accurately. AutoDock and DOCK were also compared with respect to their performance in docking monosaccharide ligands to the calcium-dependent bacterial lectin PA-IL (from *Pseudomonas aeruginosa*) and its *in silico* mutants.[359]

The comparison was conducted to examine whether the programs could predict changes in lectin affinity and specificity resulting from minor mutations of the residues at the binding site that are essential to host recognition by the bacterial lectin. AutoDock and DOCK were evaluated in terms of both pose prediction and accuracy of binding energy prediction. The performance of each program was assessed by comparing the docked structures, and scoring function-based estimated binding energies, with the experimental data. AutoDock outperformed DOCK on both counts. Subsequently,[360] a protocol was set up such that AutoDock was followed by MD simulation. The selected example dealt with the recognition of αGal-(1→4)-Gal-terminating glycosphingolipids by lectin I of *P. aeruginosa* (PA-IL). The docking method was validated by docking αGal-(1→3)-Gal into the PA-IL binding site and demonstrating successful comparison with the results obtained.[361] The binding mode of several α-galactosyl disaccharides was investigated further by following the same protocol.[283]

The reliability of five docking programs, AutoDock3, AutoDock4, AutoDock Vina, DOCK (standard score and MM-GB/SA secondary score), and ICM Dock, in calculating binding affinity in mono and trisaccharide complexes with the lectin of *Ralstonia solanacearum*, has been evaluated and compared with experimental values.[362,363] Only AutoDock3 gave values within the range of experimental binding energies. None of the docking programs tested were able to separate clearly the nonbinders from weak- and high-affinity binders. The results of this study indicate that a contrived conversion of binding energies to apparent equilibrium constants, and the subsequent calculation of binding potencies can make the distinction of binders more visible. In a follow-up study,[94] *R. solanacearum* lectin–carbohydrate complexes were chosen as model systems to quantify the CH/π interactions, using a combined experimental approach, creating with a high-level computational method, single- and double-point mutants. The results suggest that the interaction between the lectin and monosaccharide (methyl α-L-fucopyranoside) is strongly driven by the dispersion interaction, whereas polar interactions of sugar hydroxyl groups in the combining site of the protein assist in counterbalancing the effect of carbohydrate desolvation. The prediction of binding free energies to lectins for monosaccharides and higher oligomers remains a subject of investigation. Models based on linear interaction energy show good performance for instances where no metal is part in the combining

site. However, systems where ligand binding involves two calcium ions still present considerable challenges.[362]

The recognition of the prominent role played by sialic acid-terminating glycans in their interactions with lectins provided the basis for investigating the ability of AutoDock, FlexX, and Glide to determine the binding mode. The example chosen concerned the lectin SHL-2.[364] The best poses obtained using Glide were in excellent agreement with the NMR data. A set of 15 high-resolution ($\leq$2.0 Å) human-derived carbohydrate–lectin complexes was selected to evaluate the pose-prediction performance of Glide, GOLD, AutoDock, and DOCK.[365] In each test case, the four evaluated programs were generally able to identify reasonable carbohydrate poses. However, the top ranked pose was rarely found to be the best one obtained by any of the programs. The best results were generated by the DOCK program, which correctly ranked the best pose in 6 out of the 15 cases. Glide was the second-best program in that comparison. The present imperfection of the scoring functions may explain this inability to rank the poses correctly. The best poses were generally found within the best 10 ranked poses as predicted from Glide and GOLD. The program AutoDock generally failed to identify the correct pose accurately. In terms of pose-prediction accuracy, regardless of scoring/ranking, GOLD was clearly the best program, showing an average RMSD for the best poses of 1.4 Å, as compared to 2.2–2.7 Å for the other three programs. This comparison of the application of docking programs to lectin–carbohydrate interactions has highlighted the fact that consideration of only the top-ranked docking pose is not suitable. Alternative and/or complementary approaches are required and need to be developed.

Adhesion of bacteria to glycosylated cells and their surfaces is typically facilitated through adhesive organelles protruding from bacterial surface. These are called fimbriae. The most important fimbriae protein, FimH, mediates adhesion in the α-mannoside mode. Mimics of α-mannosides for binding to FimH have been investigated and docked into the carbohydrate-recognition domain, using FlexX.[366] The structures obtained were used to optimize ligand interactions with FimH made by the pendant groups attached to mannose. It has recently been suggested that FimH contains multiple carbohydrate-binding sites. In a subsequent study,[367] molecular modeling was used to optimize the spacer length of bivalent glycopeptides, which targeted two putative carbohydrate-binding sites on FimH. However, it was found that the developed glycopeptide did not provide any significant enhancements in activity. Most recently, mannosides linked via squaric acid were evaluated for their ability to act as covalent inhibitors of FimH.[368] The assay results, in combination with docking, indicated that these compounds did not act as covalent inhibitors.

The DC-SIGN is a calcium-dependent (C-type) lectin that binds to Lewis X and high-mannose glycan. Pathogens that exhibit these carbohydrates on their surface utilize DC-SIGN to develop infection, making it a potentially valuable target for anti-infective agents. A mannobioside mimic of Man-α1-2-Man was generated and its complex with DC-SIGN was determined using QPLD.[369] The procedure gave rise to several poses having glycosidic torsion angles Φ in a conformation different from *exo*-anomeric conformations. In a subsequent study,[370] a series of mannosyl trisaccharides was docked to DC-SIGN using FlexiDock. The study showed the importance of multiple binding modes in carbohydrate complexes with DC-SIGN. A collection of mannobioside mimics with varying amide substituents on the second residue was prepared[371] and docked to DC-SIGN using FlexX. The docking model highlighted the importance of hydrophobic substituents in generating mannoside mimics.

Blood-group antigens on human erythrocytes are either carbohydrate-dependent or protein-dependent. The ABH(O) and Lewis blood group carbohydrate antigens, along with their interactions with monoclonal antibodies and lectins, have been the most widely studied. 3D-QSAR studies by comparative molecular field analysis (ComFA) have provided some insights about the features involved in the recognition.[372] Nevertheless, the number of docking studies between these antigens and lectins has been relatively sparse. As in the area of protein structural biology, incorporation of the heavy atom selenium in place of oxygen in selenoglycosides helps solve the phase problem in X-ray crystallography. In addition, the potential of selenoglycosides as active ligands for lectins is being explored. The effects of such a replacement in derivatives of the histo-blood group ABH antigens on their binding to anti-ABH lectins has been investigated,[373] based on the availability of experimental data for unmodified glycosides. Glide XP, with the OPLS2005 force-field parametrization for selenium, was used to dock eight test examples of selenoglycosides as lectin ligands. This docking protocol was evaluated by comparing the docked selenoglycosides to their unmodified counterparts, and docking scores to experimental binding energies. Examples were observed where docking failed to predict the bound poses. In such instances, a semi-manual approach was implemented in which the crystallized ligands were used as starting points for modification. The selenoglycosides were accommodated at the same site as their natural counterpart and, in most examples, the orientation of the key residues was conserved. Some instances were found where the conformations and/or orientations of key and remaining residues differed from the natural ligands. In such instances, the differences in the mode of binding were found to be compensated by more favorable interactions with proteins. Such a type of accommodation is found with PA-IL, where the terminal α-galactose residue of

most ligands docked in a manner similar to that in the crystal structure, whereas other carbohydrate residues assumed different orientations. This illustrates the ability of PA-IL to bind a variety of carbohydrates having a terminal α-galactose residue. This mechanism of specificity for a terminal anchor residue, combined with a diverse tolerance for the rest of the ligand, is similar to the mechanism proposed for the recognition of carbohydrate xenoantigens by anti-Gal antibodies.[374] This study[373] indicated that selenoglycosides are biologically active as lectin ligands, with a potential for enhanced affinity. Therefore, they can serve as nonhydrolyzable mimics of the histo-blood group ABH determinants and prove therapeutically useful.

The recognition that such viruses as noroviruses interact with the Secretor, Lewis, and ABO families of human histo-blood group antigens (HBGAs) in a strain-specific manner has established the relevance for understanding the evolution of virus-binding specificities and for *in silico* design of future antiviral therapeutics. Computational studies on the interaction of ABO-active carbohydrates with the norovirus VA387 capsid protein have been performed using MD simulations, with both explicit and implicit solvent models.[375] The modeling of the complexes with the histo-blood group A-active structures indicated some contacts, which provide insight into mutational data and opening some perspectives in structure-based design of adhesion inhibitors of noroviruses. Subsequent computer simulations suggested the possibility of two receptor binding modes, which could explain for the evolutionary changes that norovirus may have undergone to recognize Lewis antigens and non-Secretor saliva.[376]

The ability to distinguish between self and nonself carbohydrates confers to lectins a prominent role in the innate immunity systems of many invertebrates. Deciphering the molecular basis underlying such a crucial role has been attempted via molecular docking and site mapping.[365] The applicability of four popular molecular docking programs (Glide, GOLD, AutoDock, and DOCK) was checked on high-resolution 3D crystal structures of lectin–glycan complexes. Despite the fact that GOLD generated the most accurate binding modes, these could not be correctly ranked by the scoring functions. Alternatively, the site-mapping method, which takes into account multiple alternative modes, could identify the key amino acids of the lectins that were involved in the glycan recognition. The prediction was found to be of improved accuracy as compared to the top poses obtained from molecular docking.

High-resolution structures of human lectins cocrystallized with carbohydrates were used to verify the techniques. Four popular molecular docking programs (Glide, GOLD, AutoDock, and DOCK) were evaluated for their ability to reproduce the crystal-bound conformation of the carbohydrate in each instance. It was found that GOLD generated the most accurate binding modes; however, these generally could

not be ranked accurately by the scoring function. This observation highlighted the need for alternative scoring functions when considering carbohydrate–lectin interactions. The site-mapping technique (discussed in more detail later), which considers multiple alternative binding modes, was able to identify key lectin residues involved in carbohydrate recognition with improved accuracy over the top pose obtained from molecular docking.

Recent advances in glycobiology are revealing the essential role of lectins for deciphering the glycocode by specific recognition of carbohydrates. Integrated multiscale approaches are needed for characterizing lectin specificity: combining on the one hand high-throughput analysis by glycan array experiments and systematic molecular docking of oligosaccharide libraries, and on the other hand detailed analysis of the lectin–oligosaccharide interaction by X-ray crystallography, microcalorimetry, and free-energy calculations. The lectins LecB from *P. aeruginosa* and BambL from *Burkholderia ambifaria* are part of the virulence factors used by the pathogenic bacteria to invade the targeted host. These two lectins are not related, but both recognize such fucosylated oligosaccharides as the histo-blood group oligosaccharides of the ABH(O) and Lewis epitopes. A docking protocol of bioactive oligosaccharides that would be common to fucose-binding lectins was investigated so as to structurally rationalize large amounts of experimental affinity data, such as those arising from glycan array experiments.[377] Several computational protocols were evaluated for molecular docking. One of them used Glide (in standard precision) to generate the ligand poses and the Emodel scoring function for ranking them. Starting from several conformers derived from the BIOLIGO database (www.glyco3D.cermav.cnrs.fr) greatly enhanced the conformational sampling, as up to four conformers for each molecule were introduced in the docking procedure. This approach predicted the bound conformation of the oligosaccharide to the lectin, as well as the network of nonbonded contacts involved in stabilization of the complex. This procedure helped in rationalizing the preference of LecB for Lea oligosaccharides. The approach was also very efficient in predicting twisted or strained conformations of Lewis oligosaccharides in their interaction with BambL. Because of their branched structures on position 3 and 4 of GlcNAc, Lea and Lex are unusually rigid oligosaccharides, a feature that has been proposed to be important in understanding the recognition by antibodies. No other conformation has been observed in available crystal structures of lectin–Lewis oligosaccharide complexes (see www.glyco3D.cermav.cnrs.fr). BambL is therefore the first observed protein that binds the Lewis oligosaccharide in conformations that are very different than their shape in solution. The particular localization of one tryptophan residue excludes the low-energy conformation of Lewis oligosaccharide from the binding site. The strong stacking

interaction that fucose establishes with this tryptophan residue in the binding site of RSL, a lectin closely related to BambL, was recently quantified.[94] Solving the crystal structure of BambL–Lewis X complex demonstrated that indeed the fucose is bound with stacking to Trp, but that either the adjacent glycosidic linkages or the GlcNAc ring is distorted in order to adjust. The MM-GBSA rescoring calculations strongly increase the correlation between theoretical data and quantitative thermodynamic data. Even though lectin–oligosaccharide interactions are generally enthalpy-driven, incorporation of a solvation energy term, including polar component and surface area accessibility, is necessary for obtaining a reliable order of binding energies of oligosaccharides, on account of some entropy contribution in binding. The best correlation was achieved for LecB, for which the parameters used in the calculations were developed. The procedure selected for LecB was rather successful when applied to BambL, demonstrating that the procedure is valid for binding sites having different characteristics (local charges, solvent exposure). Deviations between glycan array data and MM-GBSA calculations are expected, and among the numerous sources of discrepancies may be cited the fact that the glycans are methylated and conformationally restricted because they are immobilized on a surface, usually through covalent interactions. In contrast, the oligosaccharides used in the microcalorimetry experiments have a free hydroxyl group at C-1 and can exchange dynamically between an open-chain form and a closed form, with either an axial or equatorial hydroxyl group at the anomeric position, while only the equatorial structures were considered in the calculations. Also, it should keep in mind that trisaccharides were used for docking studies, whereas experimental affinities were mostly performed on their tetrasaccharide analogues. Even though MD simulations in explicit water environment are better suited for estimating the binding free energies, the combined use of a database of oligosaccharide conformations and fast docking procedure appears as a medium-throughput screening approach for the analysis of glycan array data. This is the first study combining theoretical calculations and glycan array data in the expectation of dealing with a huge flow of information, such as that anticipated in the coming “Omics” era.

2. Antibodies

Many pathogens and aberrant malignant cells express unique carbohydrates on their surface, thereby presenting attractive targets for vaccine design. Novel carbohydrate-based vaccines have been identified, and some have reached clinical-phase studies. The success of several licensed carbohydrate-based vaccines against such bacterial pathogens as *Haemophilus influenza* type b, *N. meningitides*, or

Streptococcus pneumoniae demonstrates the great potential. As such, they offer promising or already successful vaccine components against various pathologies, or new opportunities toward cancer immunotherapy. Carbohydrate antigens recognized by preformed and elicited antibodies are also essential in blood transfusion and organ transplants. These carbohydrate determinants recognized by antibodies are expressed on the cell surface as glycolipids and glycoproteins. In many instances, the minimum carbohydrate epitopes are located at the terminal end of more complex carbohydrate chains, presenting a wide range of contexts, surface densities, and surroundings. Therefore, antibodies having similar specificities for individual carbohydrate epitopes can exhibit different selective cell profiling, depending upon the unique presentation of the carbohydrate on the target cells. This is typically the case of many tumor-associated carbohydrate antigens, which are expressed at very high densities on the cell surface of primary and metastatic tumors, but can be found at much lower levels in a few cells in healthy tissues. Progress is being made in addressing challenges posed by targeting the surface carbohydrates of bacteria, protozoa, helminthes, viruses, fungi, and cancer cells for vaccine purposes, as the identification and evaluation of unique carbohydrate epitopes present on a plethora of pathogens and malignant cells becomes available.[378]

Characterization of the structures of oligosaccharide antigen–antibody complexes is performed by such experimental techniques as NMR spectroscopy and X-ray diffraction. They can provide detailed insight at the atomic level; however, these studies are typically limited to systems involving antibody fragments, such as the antigen-binding fragment (Fab) or variable fragment (Fv), and to small oligosaccharides. These experimental techniques have yet to permit characterization of the structure of complexes involving large oligo- or polysaccharides, such as the capsular polysaccharides from bacterial surfaces. The conformational flexibility of carbohydrates in their unbound state is a common feature; it cannot be ruled out that part of an oligosaccharide may sustain significant flexibility in antibody–carbohydrate complexes. Elucidation of the molecular basis of the formation of the complexes also requires consideration of the balance between the enthalpic and entropic contribution involved in the binding. At present, only an appropriate combination of computational and experimental methods will help in establishing these features, which are needed to develop vaccines having broad serotype coverage. In comparison with the large number of docking studies carried out on carbohydrate–lectin and carbohydrate–enzyme recognition, there are relatively few computationally aided carbohydrate–antibody recognition studies published.

The major role of carbohydrates in blood-group transfusion and organ transplants dramatically highlights the importance of carbohydrate–protein interactions in key

biological processes. The two main histo-blood group carbohydrate determinants[379] are the antigen families, the so-called ABH(O) groups and the Lewis determinants. The majority of the ABO antigens are expressed on human erythrocytes, at the ends of long polylactosamine chains, while a minority of the epitope is expressed on neutral glycosphingolipids. Despite the key role played by these determinants, the description at the molecular level of the interactions occurring between the antigens and the antibodies is only beginning to be resolved and characterized, for instance, with the crystal structures of Fab against Lewis determinants.[380–382]

Exhaustive investigation of the cross-reaction patterns on 9 antibodies against 12 carbohydrate antigens has been conducted throughout by computationally based methods.[372,383] 3D descriptors of the molecular properties of the carbohydrate antigens were used in comparative molecular field analysis (COMFA). Processing of the QSAR data gave indications as to the carbohydrate epitopes necessary for antibody recognition, while yielding insights into the nature of the molecular recognition.

Transplantation of pig organs into humans (xenotransplantation) is not successful because of the occurrence of carbohydrate antigens on the surface of pig organs that are recognized by xenoreactive antibodies in the human bloodstream. The major carbohydrate xenoantigens terminate in αGal-(1 → 3)-Gal epitopes. An *in silico* protocol aimed at analyzing the interactions between these xenoantigens and the antibodies has been developed[384] and applied to the determination of the structures of these terminating carbohydrate antigens in their complexes with a panel of xenoreactive antibodies. An initial docking investigation was performed on an α-Gal disaccharide and α-Gal-terminating trisaccharides in the binding site of the anti-α-Gal monoclonal antibodies mAb 8.17 and mAb 15.01 using DOCK.[385,386] The large size of the antibody binding site was found able to accommodate αGal-(1 → 3)-βGal–GlcNAc/Glc trisaccharides by end-on insertion with the complementary-determining region residues participating in interactions with all three monosaccharide units. Subsequent investigations, using site-mapping techniques (combining both receptor-based and ligand-based *in silico* mapping techniques), were developed to further explore the antibody–carbohydrate interaction.[374] The favored conformation at the αGal-(1 → 3)-Gal linkage varies, depending on topography of the antibody binding site. The possibility that some of the antibodies recognize more than one αGal-(1 → 3)-Gal conformation may not be excluded. The binding modes determined indicate that each antibody may use different mechanisms in recognizing the target antigen. These results provide a confirmation and an extension of the conclusions provided by the molecular modeling investigation of the recognition of αGal epitope by a human xenoreactive antibody (IGHC3-11) induced in human patients mounting

an immune response to a bioartificial liver.[387] As an α-Gal epitope is recognized similarly by both mouse and human antibodies, the mouse antibodies constitute a viable model for the human immune-response xenoantigen.

Brucella is a genus of Gram-negative bacteria which cause brucellosis through ingestion of infected food or direct contact with an infected animal. These bacteria carry a smooth polysaccharide, a surface macromolecule that is a serious virulence factor and the most important serodiagnostic antigen. The O-polysaccharide (or O-antigen) section is a homopolymer of *N*-formyl-perosamine (D-Rha4NFo) in α-(1→2) and α-(1→3) linkages. The specificity of the monoclonal antibody (YsT9.1) against this repeating internal motif was the first attempt to model carbohydrate–antibody recognition[388] by performing a manual docking to a homology model of YsT9.1.

A majority of life-threatening cases of septicemia, meningitis, and pneumonia occur from the deleterious action of surface capsular polysaccharides on bacteria. Although these polysaccharides may have similar carbohydrate sequences, they may differ markedly in immunogenicity, antigenicity, virulence, and geographical dispersion. This is, for example, the situation with Group B *Streptococcus agalactiae* and *S. pneumoniae*. Generation of the antibody–carbohydrate antigens was performed through a combination of comparative antibody modeling and automated ligand docking. Subsequently, several 10-ns molecular dynamic simulations, within explicit hydration, were performed using the molecular mechanics–generalized Born surface area method, augmented by conformational entropy estimates. While providing detailed insight into the molecular details and the energy components involved in the formation of the complexes, the analysis offered a comprehensive interpretation of a large body of biochemical and immunological data related to antibody recognition of bacterial polysaccharides.[389]

Shigella flexneri is the primary causal agent of the endemic form of bacillary dysentery. The O-antigen is the polysaccharide moiety of the lipopolysaccharide; it is the major target of the serotype-specific protective humoral response elicited upon host infection by *S. flexneri*. The repeating unit of the O-antigen is a pentasaccharide. The availability of the X-ray structure of the Fab/$[AB(E)CD]_2$ complex, resolution 1.80 Å,[390] along with a sufficient amount of well-characterized pentasaccharides, and IgG monoclonal antibody, allowed a thorough analysis of the complexes by STD-NMR experiments and extensive MD simulations. The study brought to light information on the dynamics of the corresponding antibody:carbohydrate complexes that is not available from the X-ray structure nor from the NMR analysis (Fig. 31).[391] The proposed protocol uses MD simulations and STD-NMR. Such a combined approach facilitates the design of either ligands or carbohydrate-recognition domains, with the

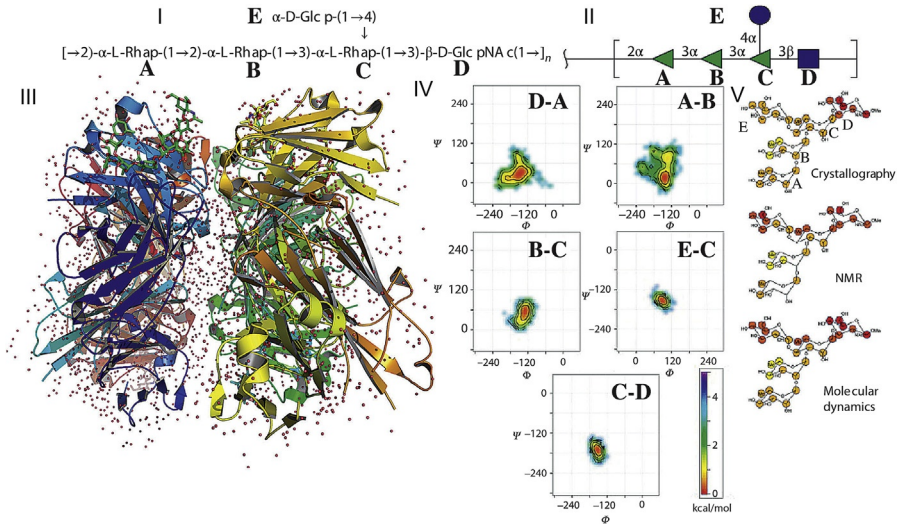

FIG. 31. Features of the *Shigella flexneri* O-antigen interacting with monoclonal antibody. (I) Primary structure of the *Shigella flexneri* SF2a O-Ag,[506] common AB(E)CD linear backbone repeat unit. (II) CFG representation of *Shigella flexneri*, common AB(E)CD linear backbone repeat unit, where the green triangles represent rhamnose, the circle denotes glucose, and the blue square denotes *N*-acetylgalatosamine. (III) Crystal structure of synthetic O-antigen decasaccharide from serotype 2a *Shigella flexneri* (PDB 3BZ4) in complex with a protective monoclonal antibody Fab F22-4. (IV) Φ, Ψ maps of MD simulations for the glycosidic linkages of two repeat units of the bound conformation of the *Shigella flexneri* O-antigen D_0 AB(E)CD pentasaccharide. (V) Comparison between the predicted saturation transfer difference (STD) values of the two repeat units of the truncated crystal structure of F22-4 and the measured STD-NMR intensities and the predicted values of the 50 MD simulation snapshots of AB(E)CD. Reprinted with permission from Ref. 491. Copyright 2013 CRC Press, Taylor & Francis.

aim of enhancing the binding properties displayed by the original carbohydrate–receptor interaction.

The limited number of antibody–carbohydrate docking studies is a reflection of the paucity of crystal structure complexes at high resolution. Analysis of the available structural data indicates some general trends about how such antibodies recognize different types of carbohydrates. Those antibodies that recognize a terminal carbohydrate motif display a cavity-like binding site, reminiscent of the combining sites found in lectins, where the insertion of one or more monosaccharide residues occurs in an "end-on" manner. Antibodies that recognize an internal carbohydrate motif (a single or several repeat units) of a polysaccharide generally display groove-like binding sites, or unusually large cavities which are "open" at both ends of the sites, allowing

for "side-on" entry of the antigen. While displaying some structural similarities with the catalytic sites of GHs, these features allow the carbohydrate to be recognized as a "conformational epitope" as it has been shown to occur in the case of α-(2 → 8)-linked sialic acid residues.[392]

3. GAG-Binding Proteins

The GAGs comprise a class of complex anionic polysaccharides which through their linkage to a core protein are components of macromolecules (proteoglycans) that are more complex. The GAG families include (1) glucosaminoglycans (heparin and heparan sulfate), (2) galactosaminoglycans (chondroitin sulfate and dermatan sulfate), and (3) hyaluronic acid and keratan sulfate. Assembled from disaccharide repeating units, GAGs manifest diverse patterns of sulfation. Among this group, hyaluronic acid is unique in not being attached covalently to a core protein, and it lacks sulfation. In addition to their participation in the physicochemical properties of the extracellular matrix, GAG fragments are specifically recognized by protein receptors, and they play a role in the regulation of many processes, such as hemostasis, growth factor control, anticoagulation, and cell adhesion.[393] Given the importance of protein–GAG interactions, oligosaccharide fragments are prime targets for drug design. GAG mimetics are soon likely to have clinical applications as modulators of cytokines, growth factors, or enzymes functioning in various diseases and pathologies.

Docking of GAG oligosaccharides or polysaccharides in protein–receptor binding sites presents several serious difficulties: (1) both the ligand and the protein display a high flexibility of side chains; (2) because of their negatively charged nature, GAGs require accurate consideration of electrostatic and water-mediated interactions; and (3) the binding site does not generally adopt a pocket or crevasse shape that would allow for easy identification. Furthermore, the development of proper conformational tools suffers from the paucity of structural data on GAG–protein complexes.

Analysis of the projection of the ESP on the Connolly surface of a protein, for example, with the MOLCAD (1985) program, which allows prediction of the most energetically favorable region (included in SYBYL, Tripos Associates), has proven to be useful. The GRID program[142] for binding of small probes on the protein surface is very successful in identifying sulfate-binding regions. For predicting the orientation of the oligosaccharide on the protein surface, the AutoDock program,[279] which considers flexibility at glycosidic linkages and pendant groups (hydroxyl groups,

hydroxymethyl, and so on), can be used for charged oligosaccharide fragments. Such an approach generally yields several families of conformations.

The conformational behavior of a pentasaccharide sequence in heparin responsible for high affinity to antithrombin III has been the subject of several investigations. This study is complicated by the fact that a conformational change occurs in the protein upon binding.[394,395] The first model obtained using homology modeling for the protein and hand-docking of the pentasaccharide allowed the determination of those basic amino acids involved in recognition of the sulfate and carboxylate groups.[396] A study making use of several newly developed docking programs arrived at the same prediction for the binding site.[397] In the crystal structures of the complex between antithrombin III and the pentasaccharide,[398] a cluster of basic amino acids has been demonstrated to interact with the oligosaccharide's sulfate and carboxylate groups. The conformation of the bound pentasaccharide is also subjected to induced-fit upon binding. At the present time, both X-ray crystallographic studies and NMR data coupled with molecular modeling[399] agree that the binding is accompanied by variations in the dihedral angle of two glycosidic linkages and conformational shift of the 2-*O*-sulfated iduronic residue.

Fibroblast growth factors (FGFs) are among the numerous proteins that bind heparin. They have received special attention because of their role in the control of cell proliferation, migration, and differentiation. Two FGFs and their coreceptors—FGFRs—have been cocrystallized with heparin fragments, and the minimal binding sequences could be determined.[400] Analysis of the crystal structures, together with molecular modeling, demonstrated that upon binding, the regular helical shape of heparin is distorted. A kink is formed at one point by both modifications of the conformation of one glycosidic linkage and the ring shape of one iduronate residue.[401] Such "induced-fit" of the ligand as a result of its interaction with a protein is very probable, as it is classically observed in lectin–glycan interactions. The vascular endothelial growth factor (VEGF) that plays a role in angiogenesis binds to heparan sulfate. Molecular modeling studies demonstrated that a highly sulfated heptasaccharide is the optimal binding sequence required for high-binding affinity.

Chemokines, derived from chemoattractant cytokines, constitute a large family of small proteins (8–12 kDa in their monomeric form). Based on their physiological features, they have been classified as "inflammatory" (or inducible) or "homeostatic" (or constitutive).[402] Their roles include events as diverse as development, angiogenesis, neuronal patterning, hematopoiesis, viral infection, wound healing, and metastasis. Given the importance of protein–GAG interactions (Fig. 32), oligosaccharide fragments are important targets for drug design. Chemokines interact with GAGs in general and with heparan sulfate in particular. This binding is thought to create a local

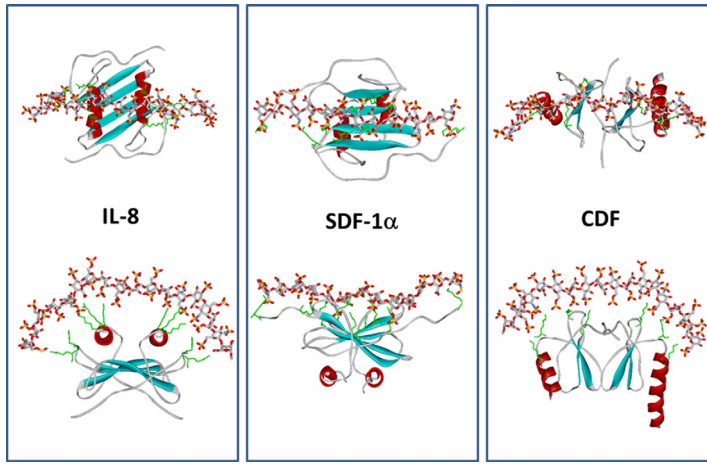

FIG. 32. General GAG–chemokines interactions. Orthographic representation of the lowest energy-binding mode of heparin with the chemokines IL-8, SDF-1α, and CDF.[507]

concentration, or a gradient, of chemokines on tissues where some GAGs are specifically expressed. Despite the large number of binding studies, crystallization of chemokine–oligosaccharide complexes has proved difficult, and the number of available crystal structures is limited.

Modeling studies have therefore been used for describing the interaction between chemokines and heparan sulfate. One intriguing structural feature is that chemokines may exist in solution as monomers or dimers (sometimes tetramers), but they bind GAGs in the dimeric or tetrameric state. Depending on the dimerization mode and the positions of basic amino acids in the peptide sequences, chemokines will exhibit positively charged clusters on their accessible surfaces that define several possibilities for binding heparan sulfate.[403,404]

As seen from the previous examples, the docking of GAGs has long been a very impractical task, and modeling of their complexes with proteins was usually performed in conjunction with other methods.[405–407] A method for the *de novo* placement of explicit water molecules has been developed with the purpose of improving the outcomes of docking studies.[408] A dataset of 11 protein complexes has been selected, based on high-resolution structures (no greater than 2.2 Å) containing ligands no higher than tetrasaccharides. The performance of these programs was found to be in the order of FlexX < MOE < eHiTs < AutoDock3. All of them produced better poses when coupled with explicit water molecules. This is a significant finding, as most docking programs, including the four tested ones, already include solvation effects

implicitly in their scoring functions. The program AutoDock 3 performed significantly better than the other three programs in the validation experiments. Specifically, it produced top poses with average 1.94 and 1.60 Å RMSDs for the binding sites without and with explicit water molecules, respectively. However, it failed to reproduce correctly those binding-site residues experimentally known (by site-directed mutagenesis) to be important for binding IL-8 to heparin/heparan sulfate disaccharides. The program FlexX performed significantly worse than the other three programs when tested with FlexX type 3 water molecules, but its performance was enhanced almost to the level of that of eHiTs when combined with crystallographic or GRID-generated water molecules.

Further simulations, including MD in the presence of explicit water and counter ions, have to be envisaged for a thorough investigation. To this end, a novel targeted MD-based protocol has been developed to address ligand and receptor flexibility, as well as the inclusion of explicit water molecules in local molecular docking. When applied to the protein–GAG system, this method exhibits high predictive significance for systems dominated by electrostatic interactions, and demonstrates its capacity to identify the receptor residues contributing most to the binding.[409]

4. Transport

Such carbohydrates as glucose, sucrose, lactose, malto-oligosaccharides, raffinose, fructo-oligosaccharides, L-fucose, trehalose, oligoalginate, oligogalacturonate, and others constitute a source of carbon for many organisms. These molecules have to be transported across the membrane channel and pores. Their motion is critically important for understanding the mechanism of many cellular processes. At the protein level, this is achieved by a family of proteins, collectively referred to as transporters. They constitute a wide group of transmembrane proteins that allow permeation of sugars and facilitate their transport. Their structures, along with the mechanistic transport model, are the subject of intense current research. The high-resolution structural elucidation of as yet only a limited number of transporters is enabling investigation into the MD of fundamental transport processes.

Transport across the membrane is mediated by channel-forming proteins, of which maltoporin has been the most extensively studied. Elucidation of the first high-resolution structure of maltoporin[410] revealed the general model of specific channel-forming membrane proteins: a β-barrel with 18 antiparallel strands. Like the general diffusion porins, the functional unit of maltoporin is a trimer with long loops exposed to the cell exterior and short turns exposed toward the periplasm.

A striking feature is a continuous stretch of aromatic residues in the channel arranged along a left-handed helical path, which has been described as the "greasy slide."

The translocation mechanism of malto-oligosaccharides across the maltoporin membrane channel has been investigated by MD calculations[411] (see the movie at http://www.iwr.uni-heildelberg.de/groups/biocomp/fischer/research/maltoporin.html). The first event is the binding of a sugar (such as a malto-oligosaccharide) to the first residue of the "greasy slide," and this occurs via van der Waals interactions to the hydrophobic face of the glucosyl ring. Deeper penetration into the channel occurs through guided diffusion of the oligosaccharide along the "greasy slide." A gradual dehydration of the malto-oligosaccharide favors the establishment of transitory hydrogen bonds between the hydroxyl groups of the sugars and the surrounding amino acids.

This is made possible by the conformational flexibility available at the glycosidic linkages and the primary hydroxyl groups. The presence of the charged side chains (referred to as "polar tracks") mimics the lost hydration shell to the sugar by providing hydrogen bonds to the hydroxyl groups of the carbohydrate. The polar tracks are divided into donor and acceptor lanes all along the greasy slide. The movement of the carbohydrate residues to the next binding site of the greasy slide, in combination with a rearrangement of hydrogen bonds, is referred to as the "register shift." The continuous making and breaking of hydrogen bonds induces motion of the oligosaccharide through the porin in a capillary-like fashion (Fig. 33).

An illustration of the architecture that governs the sliding of glucose throughout the GLUT1 transporter has been obtained throughout the application of AutoDock3 onto a 3D template model (Protein Data Bank code 1SUK). The study revealed nine hexose-binding clusters spanning the entire "hydrophilic" channel. Five of these cluster sites are within 3–5 Å in the vicinity of the location of 10 missense mutations of the glucose transporter-1 deficiency syndrome. Another three sites are within 8 Å of two other missense mutations. D-Glucose binds to five sites in the external channel opening, with increasing affinity toward the pore center, and then passes via a narrow channel into an internal vestibule containing four lower-affinity sites. An external site, not adjacent to any mutation, also binding phloretin but recognizing neither D-fructose nor L-glucose, may be the main threading site for uptake of glucose.[412]

Within the superfamily of carbohydrate transporters exists the major facilitator superfamily (MFS) that facilitates movement of small solutes across cell membranes in response to chemiosmotic ion gradients. These transporters are thought to use an alternating access mechanism to upload and download substrates. The elucidation of the 3D structure of a fucose transporter[413] opened the way to MD simulation of an L-fucose residue complexed to the transmembrane protein inserted into a POPE bilayer to mimic the bacterial membrane. Structural, biochemical, and computational

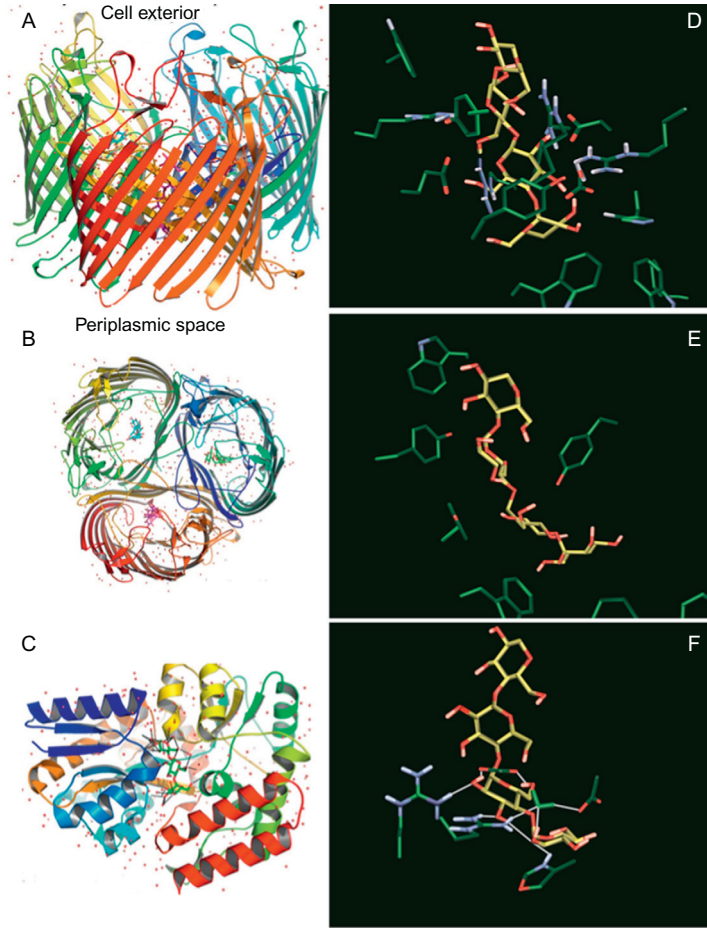

FIG. 33. Three-dimensional features of maltoporin structures (A–C), along with snapshots of the interaction of maltotetraose within the channel (D–F), showing the conformational changes undergone by maltotetraose during its journey through the "greasy slide".[508]

analysis provided insights into the function of the transporter, with the identification of key amino acids that play a vital role in the active-transport path.

As the structures of other unique transport systems are revealed, the power of computational methods in analysis and prediction of the transporter process is expected to grow.

VIII. Glycoside Hydrolases

The hydrolysis of glycosidic bonds in carbohydrates, polysaccharides, glycoproteins, glycolipids, and other glycoconjugates is effected by GHs. These enzymes are classified into *endo* and *exo* types. *Exo*-type glycosidases attack and hydrolyze simple glycosides into the free sugar and the aglycon. When acting on oligo- or polysaccharides, they liberate a monosaccharide unit from the nonreducing (upstream) end of the chain. *Endo*-type glycosidases act on oligo- and polysaccharides and catalyze the hydrolysis of an internal glycosidic linkage in the chain, thereby liberating two carbohydrate moieties or releasing an oligosaccharide (or polysaccharide) and a simple glycoside released from the reducing (downstream) end. Some glycosidases are capable of acting as both *exo* and *endo* types. The reactions resulting from the catalytic action of glycosidases can also be characterized by the anomeric configuration of the glycosidic bond of the substrate that the enzyme attacks, namely with retention or inversion of the anomeric configuration.

1. GHs on a Single Carbohydrate Chain

Glycosidases are enzymes that transfer glycosyl residues from their donor to water (Fig. 34). The glycosyl transfer can occur with either retention or inversion of configuration. Based on the similarities in their primary sequences, glycosidases have been grouped into over 110 families,[296] and for more than 75 families at least one structure has been solved. It appears that the members from the same family share the same general fold and function by a similar catalytic mechanism. Moreover, 3D structures are more conserved than primary sequences and, therefore, different families having similar structures have been clustered into superfamilies known as GH clans. Currently, there are 14 clans; each contains at least two families.

Two canonical mechanisms were proposed[414] for the catalytic cleavage of glycosidic linkages by GH (Fig. 34). Both mechanisms involve oxocarbenium ion-like transition states, and a pair of carboxylic acids that promote the departure of the aglycone as a leaving group and attack of water at the anomeric center.

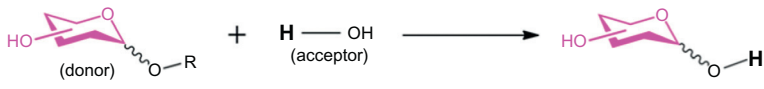

FIG. 34. Schematic diagram of a general reaction catalyzed by glycosidases.

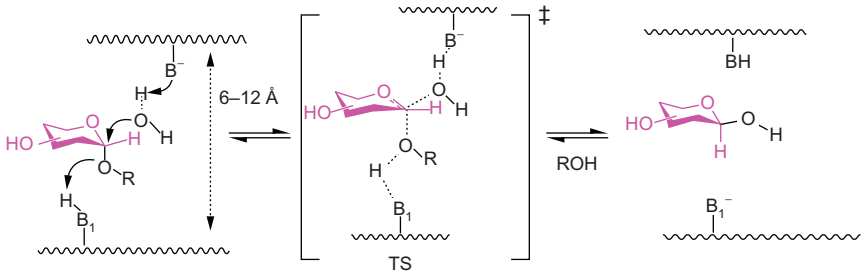

FIG. 35. Scheme of single-displacement mechanism of inverting glycosidases shown for attack on an α-D-glycoside. The hydrolysis reaction proceeds in a single step via direct displacement of the aglycone, and the transition state has oxocarbenium ion-like characteristics. One carboxylic acid (B) acts as the general base and activates a water molecule for nucleophilic attack at the anomeric center of the substrate. At the same time, the second carboxylic acid (B_1) facilitates the departure of the leaving group via general acid catalysis. Two key catalytic residues are typically separated by a distance of 6–12 Å.

In inverting glycosidases (Fig. 35), the active site contains a pair of carboxylic acids typically separated by a distance of 6–12 Å.[311] One of the carboxylic acid groups functions as a general acid catalyst and facilitates departure of the aglycone. The second functions as a general base catalyst, deprotonating the nucleophilic water molecule. The reaction is the nucleophilic attack of a water molecule to the saccharide substrate, activated by the catalytic base, with simultaneous protonation of the saccharide by the catalytic acid. The reaction is completed by release of the reducing subunit and by the inversion of configuration at the anomeric center of the glycosidic linkage that is hydrolyzed. The reaction proceeds in a one-step concerted mechanism.

For retaining glycosidases, a two-step, double-displacement mechanism has been proposed (Fig. 36). In retaining glycosidases, two carboxylic acid groups are separated by approximately 5 Å.[311] The shorter distance between the carboxyl groups than is seen in inverting glycosidases reflects their different modes of action. One carboxylic acid residue functions as the nucleophile and attacks the anomeric carbon atom directly. In this glycosylation step, a covalent glycosyl–enzyme complex is formed by a displacement reaction. The additional carboxylic acid functions as a general acid catalyst which protonates the departing glycosidic oxygen. The first step (glycosylation) corresponds to the attack of nucleophilic oxygen of the nucleophilic carboxylate to the substrate, activated by simultaneous protonation of the substrate by the catalytic acid, thus creating an enzyme–substrate complex and releasing the nonreducing subunit of the glycosidic linkage R. In the second, deglycosylation step, the reaction is completed by the general base-catalyzed attack of water at the anomeric

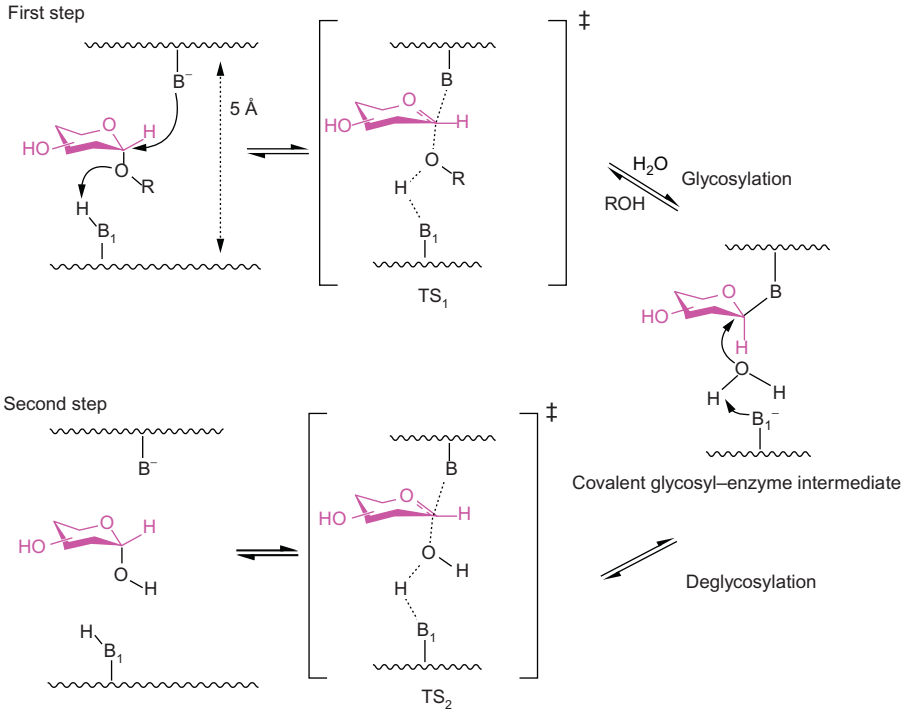

FIG. 36. Scheme of a double-displacement mechanism of retaining glycosidases shown for attack on an α-D-glycoside. Two nucleophilic displacement steps are required for hydrolysis. In the first glycosylation step, a covalent glycosyl–enzyme intermediate is formed via an oxocarbenium ion-like transition state. In the second deglycosylation step, the general acid catalyst is deprotonated and acts as a general base, by activating a water molecule for nucleophilic attack at the anomeric center of the glycosyl–enzyme intermediate. This step also proceeds via an oxocarbenium ion-like transition state. The two carboxylic acid residues (B, B_1) are usually separated by a distance of approximately 5 Å.

carbon with replacement of the enzyme's carboxylate group. The reaction is completed by release of the reducing saccharide with its anomeric configuration retained.

A majority of glycosidases act by these two mechanisms. Major exceptions are those that act on 2-acetamido sugars, where the oxygen atom of the acetamido carbonyl group functions as an intramolecular nucleophile in the so-called substrate-assisted mechanism.[332,415] In this mechanism, the acetamido group of the substrate plays the role of the enzyme nucleophile, attacking the anomeric carbon and forming an oxazolinium ion intermediate (Fig. 37).

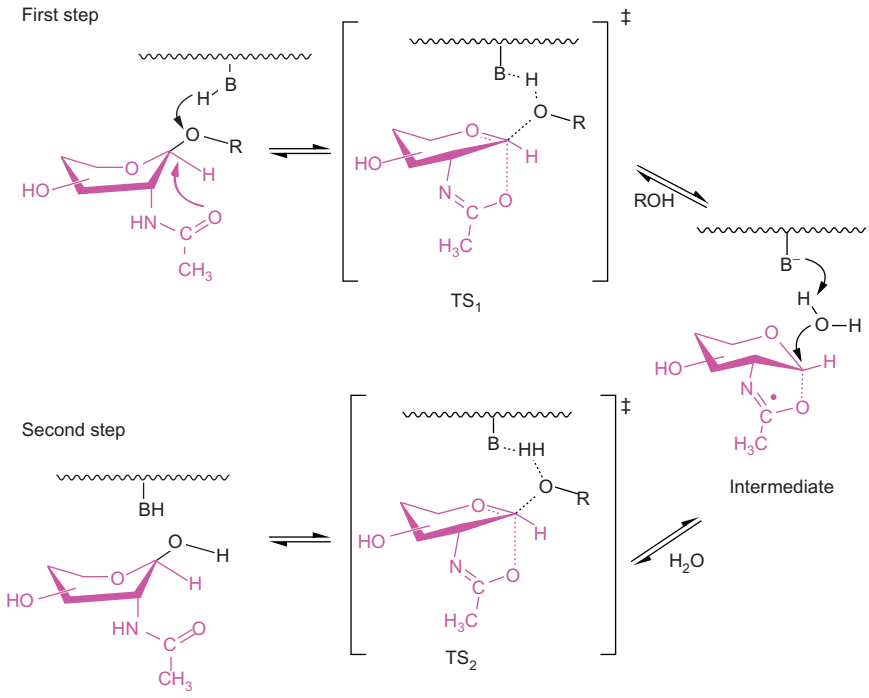

FIG. 37. Scheme of a substrate-assisted mechanism of retaining glycosidases operating via an oxazoline ion intermediate.

Advances in research on glycosidases and their mechanisms have been repeatedly reviewed.[299–302,311,416–420] This section offers an overview of modeling contributions to our understanding of the catalytic reaction of glycosidases at the atomic-detail level.

a. Reaction Coordinates of GHs.—One of the distinctive features in the catalytic mechanism of glycosidases is that the pyranose ring at subsite-1 is in a distorted conformation instead of the most stable $^4C_1$ conformation. This distortion was first proposed for lysozyme of hen egg white[421–423] and later observed in several complexes of retaining and inverting glycosidases.[261,417,424–429] Computational methods significantly contributed to the understanding of this substrate preactivation. It was suggested, based on a QM study of the model compound 2-oxanol,[160] that saccharide distortion determines the pathway of the glycosidase reaction. This proposal is supported by recently reported calculations of conformational free-energy surfaces,

which allowed a description of the catalytic itineraries used by glycosidases.[253,259,261,262]

The hybrid QM/MM CPMD[245,255] study of the enzyme–substrate complex of *Bacillus* 1,3-1,4-β-glucanase[259,430] showed that upon binding to the enzyme the substrate glucopyranose ring at the −1 subsite favors a distorted $^{1,4}B/^{1}S_3$ skew-boat conformation over the $^{4}C_1$ chair conformation. As compared to the $^{4}C_1$ ring form, the $^{1}S_3$ conformation possesses a longer C-1—O-1 linkage, a shorter C-1—O-5 linkage, and a higher positive charge on the anomeric carbon atom. All of these changes occur during the enzymatic reaction on going from the Michaelis complex to the transition state.[311] Therefore, it was suggested that the enzyme accelerates the catalytic reaction by distorting the substrate to a structure closer to that in the transition state.

The free-energy surfaces for β-D-glucopyranose,[253] α-L-fucopyranose,[260] and β-D-mannopyranose,[262] were calculated by using metadynamics simulation. The metadynamics simulations were performed using CPMD formalism and the Cremer–Pople puckering coordinates[431] as two CVs.[254] The calculated free-energy maps are shown in Fig. 38, and qualitative differences among them are obvious. The low-energy regions (the most stable minima) are located on one side of the diagram, but are shifted from southwest (β-D-glucopyranose) to northeast (α-L-fucopyranose) and northwest (β-D-mannopyranose). Comparison with the X-ray structures of complexes showed that experimental structures were located in low-energy regions (for example, see Fig. 38A) and suggested that corresponding glycosidases evolved to select preferentially those conformations that required less energy for distortion of the ring.

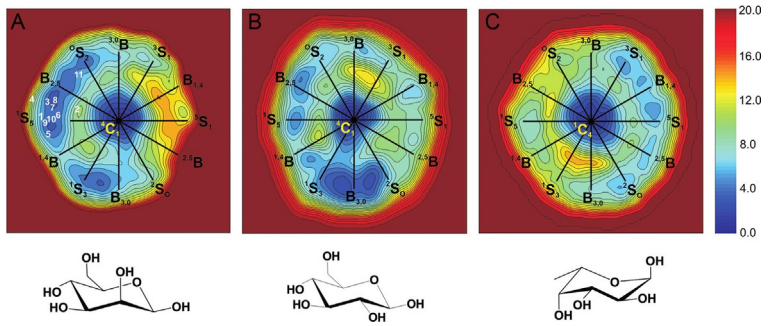

FIG. 38. Distribution of the canonical conformations on the computed free-energy surface of (A) β-D-mannopyranose,[262] (B) β-D-glucopyranose,[253] and (C) α-L-fucopyranose[260] (southern hemisphere). Reprinted with permission from Ref. 262. Copyright 2010 American Chemical Society.

The authors also defined[262] the preactivation index ξ, and the highest value of ξ represents ring conformations that are the most likely candidates to be in Michaelis complexes. Based on the analysis, they concluded that, for β-D-glucosidases, α-L-fucosidases, and β-D-mannosidases, the most suitable ring conformers are ${}^{2}S_{O}$ and ${}^{1}S_{3}$, ${}^{1}C_{4}$, and ${}^{1}S_{5}$, respectively (Fig. 39). In these ring conformations, the arrangement at the anomeric carbon allows an in-line nucleophilic attack into the antibonding orbital of the breaking glycosidic linkage. This could also explain why β-D-glucosidases favor a ${}^{1}S_{3} \rightarrow {}^{4}H_{3} \rightarrow {}^{4}C_{1}$ or ${}^{2}S_{O} \rightarrow {}^{2,5}B \rightarrow {}^{5}S_{1}$ itinerary. Similarly, these calculations justified the preference of α-L-fucosidases and β-D-mannosidases for a ${}^{1}C_{4} \rightarrow {}^{3}H_{4} \rightarrow {}^{1}C_{4}$ and ${}^{1}S_{5} \rightarrow B_{2,5} \rightarrow {}^{O}S_{2}$ itinerary, respectively.[424]

It is surprising that molecular modeling of the conformational preferences of isolated pyranosides successfully described potential conformational itineraries used by glycosidases. However, this correlation might suggest that the intrinsic conformational properties of the free hexopyranose residue are the main factors responsible for the distortion of the saccharide residue at the −1 subsite in glycosidases, and that active-site structures of the enzyme may have evolved to meet criteria for an efficient catalysis.

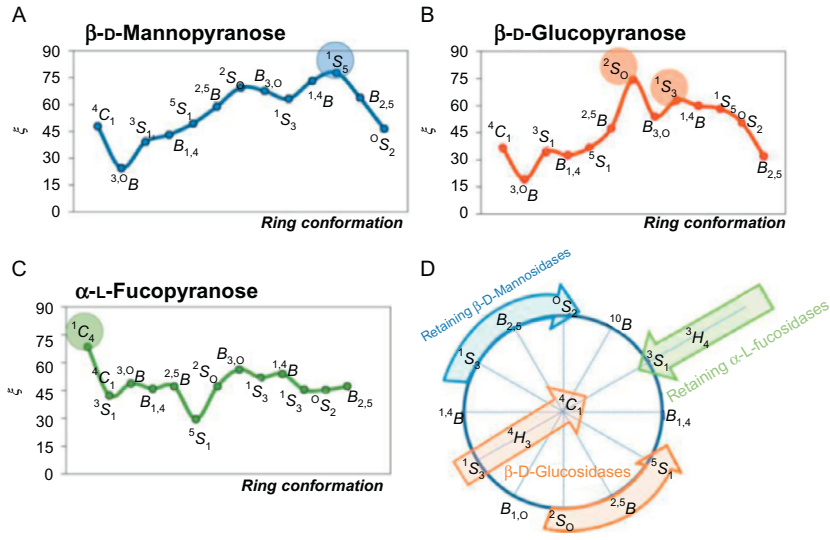

FIG. 39. (A–C) Variation of the values of the preactivation index ξ[262] as a function of ring conformation obtained for β-D-mannopyranose, β-D-glucopyranose, and α-L-fucopyranose. (D) Experimentally predicted[417] catalytic conformational itineraries for retaining β-D-mannosidases (blue), β-D-glucosidases (orange), and α-L-fucosidases (green). Reprinted with permission from Ref. 262. Copyright 2010 American Chemical Society.

2. Retaining GHs

The majority of the retaining glycosidases employ the canonical mechanism given in Fig. 36. This double-displacement mechanism proceeds through two transition states and a covalent intermediate. The retaining enzymes possess two essential amino acids, the catalytic nucleophile that forms a covalent glycosyl–enzyme intermediate and a general acid/base catalyst that in the first step assists departure of the leaving group and in the next step assists nucleophilic attack of water. Both catalytic residues are usually carboxylate groups.

a. Glucosidases.—The rate-limiting step of the hydrolysis of a tetrasaccharide catalyzed by bacterial 1,3-1,4-β-glucanases, namely formation of the glycosyl–enzyme intermediate, has been studied by means of QM/MM metadynamics.[430] The free-energy surface reconstructed from the metadynamics simulation is given in Fig. 40 as a function of two CVs. Stationary points on the free-energy surface showed that the reaction pathway itinerary can be described as $^{1,4}B/^{1}S_3$ (Michaelis complex) $\rightarrow E_5/^{4}H_3$(TS) $\rightarrow {}^{4}C_1$(P). The predicted reaction barrier of 35 kcal mol^{-1} is similar to the values obtained for other retaining glycosidases.[432,433] The transition state had oxocarbenium ion features, with sp^2 hybridization of C-1, the ring C-1—O-5 linkage shortened to 1.28 Å, and the C-2, C-1, O-5, and H-1 atoms lie in one plane. It is intriguing that, while the glycosidic linkage C-1—O-1 was increased to 3.39 Å, the nucleophile oxygen atom was still 3.61 Å away. This indicated a dissociative type of transition state.

b. Galactosidases.—The catalytic mechanism of the *E. coli* β-galactosidase was studied recently by various approaches.[434–436] This retaining enzyme catalyzes the hydrolysis of β-D-galactosides, the natural substrate being lactose. The 3D structure of β-galactosidase was determined, and it was found that the enzyme requires divalent Mg^{2+} cation for full catalytic efficiency.[437] The cluster approach was used to examine both steps of the double-displacement mechanism.[434,435] The small model systems contained the methyl β-D-galactopyranoside molecule, and the two catalytic residues were modeled by propanoic acid. The initial structure of the model was the crystal structure of *E. coli* β-galactosidase complexed with 2-deoxy-α-D-*lyxo*-hexopyranoside.[437] The structures calculated by DFT/B3LYP of the transition states showed a dissociative character. In the first step, the TS_1 was characterized by a distorted ring conformation; the length of the C-1—O-1 glycosidic linkage was 1.98 Å, the proton was transferred to the glycosidic oxygen, and the C-1—O-5 linkage was shortened to 1.30 Å. The next transition state had a more dissociative feature, with the glycosidic linkage being extended to 2.56 Å and the distance between the anomeric carbon and the attacking water oxygen being 1.94 Å. DFT calculations lead, depending on the

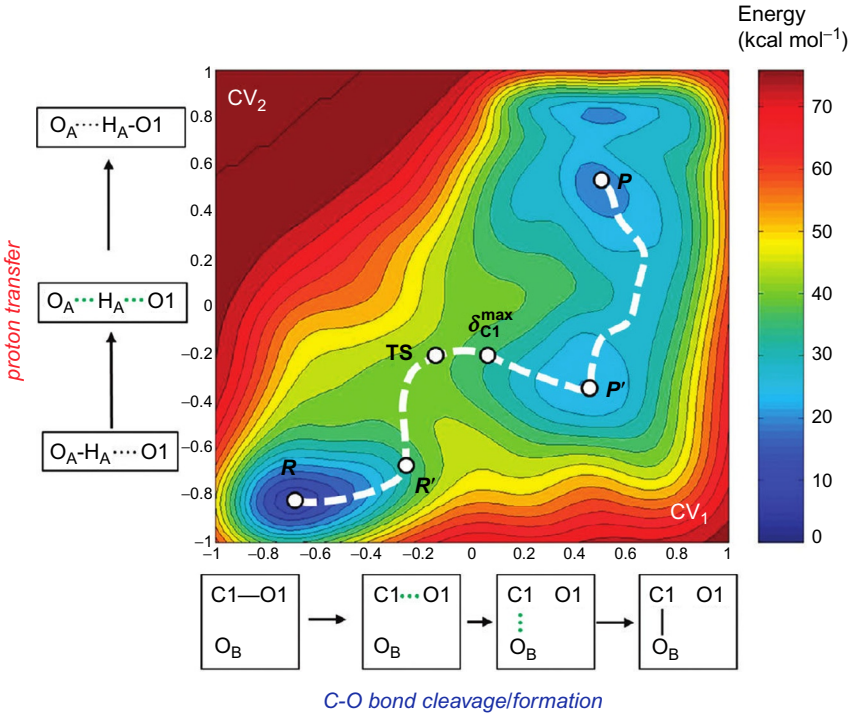

FIG. 40. Free-energy surface for the formation of the covalent glycosyl–enzyme intermediate in 1,3-1,4-β-glucanase. Reprinted with permission from Ref. 430. Copyright 2011 American Chemical Society.

functional used, to the activation barrier of 21–29 kcal mol^{-1} for the first step and 22–33 kcal mol^{-1} for the second step, respectively. The calculations support a role for the hydrogen bond between the nucleophile and the C-2 hydroxyl group as facilitating the first step. The results suggest that the deglycosylation step is rate-limiting in the model studied.

Later, the catalytic mechanism of the *E. coli* β-galactosidase was studied using QM/MM calculations and the extended model. The model consisted of 2707 atoms, contained docked lactose, and has been divided into two layers. The QM/MM calculations were performed using the ONIOM method in Gaussian software.[438] The higher-level layer contained 49 atoms and was treated at the B3LYP/6-31G(d) level, the rest of the system being treated at the molecular level using the Glycam04 force field.[439] The results of the QM/MM calculations differed from the cluster model calculations and predicted hydrolysis as an exothermic reaction with an activation

barrier around 15 kcal mol^{-1}. The glycosidic bond lengths are 2.45 and 2.25 Å in the TS_1 and TS_2, respectively. The calculations also supported stabilization of the TS by the C-2 hydroxyl group. Moreover, they clearly showed the importance of the Mg^{2+} ion in the catalytic reaction and estimated that the Mg^{2+} ion lowers the activation barrier by 15 kcal mol^{-1}.

c. Hexosaminidases: Substrate-Assisted Mechanism.—In general, hexosaminidases describe enzymes that cleave the glycosidic linkage of 2-acetamido-2-deoxy-β-D-glycosides. This group of enzymes includes, for example, lysozymes, chitinases, chitobioses, hyaluronidases, O-GlcNAcase, and others. The majority of these retaining enzymes operate by the substrate-assisted mechanism (Fig. 37), but some use the classic mechanism. The catalytic mechanism of some enzymes has been investigated by theoretical methods.[440–446]

The substrate-assisted catalytic mechanism of the enzyme β-hexosaminidase, which cleaves GlcNAc from a GM2 ganglioside, has been studied using the cluster model (48 atoms) of a substrate–enzyme complex and employing DFT and MP2/MP3 methods.[441] The calculation clearly showed that the rate-limiting step is the first step, with the energy barrier of 35.4 and 33.1 (36.8) kcal mol^{-1}, as calculated using the M06-2X functional and MP2 (MP3) methods, respectively. The barrier for the second step of 26.8 and 29.8 (34.0) kcal mol^{-1} was calculated using the M06-2X functional and MP2 (MP3) method, respectively. The B3LYP functional gave lower energy barriers up to 11 kcal mol^{-1}, namely 24.4 and 19.0 kcal mol^{-1}. The catalytic reaction proceeded from the Michaelis complex in the $^1S_3$ conformation through the transition state in the $^4H_3$ form to the intermediate in the $^4C_1$ conformation. As the water approached the intermediate, the ring adopted the $^4H_3$ form in the TS_2 and then back to the $^1S_3$ conformation in the product. A similar structure of both transition states is also documented by the lengths of glycosidic linkages. The distance between the anomeric carbon C-1 and nucleophilic oxygen was 2.28 and 2.26 Å in the TS_1 and TS_2, respectively. The distance between the anomeric carbon C-1 and the leaving group was 2.21 and 2.38 Å in the TS_1 and TS_2, respectively. Comparison with the classical double-displacement mechanism clearly shows that transition states in the substrate-assisted mechanism are less dissociative. This difference is understandable, considering the constraining influence of an intramolecular character of the nucleophilic attack.

The posttranslational modification of serine or threonine residues on the protein by O-linked GlcNAc regulates a wide range of cellular functions[327,328] and is balanced by two enzymes O-GlcNAcase and OGT. O-GlcNAcase trims O-linked GlcNAc from modified proteins and, together with OGT, controls the dynamic balance of cellular O-linked GlcNAc levels.[447] The alternation of O-linked GlcNAc levels has been associated with various diseases. Therefore, there has been considerable effort made

to develop potent and specific inhibitors.[417] Potential biological applications have generated interest in their mechanism. DFT calculations of a significantly larger cluster model (168 atoms) have been used[446] to support the substrate- assisted mechanism proposed based on experimental data for O-GlcNAcase.[332,448] The calculation provided detail about the structure of transition states and of the oxazoline intermediate. Formation of this intermediate was found to be the limiting step of the reaction, with the activation energy of 16.5 kcal mol^{-1}. Interestingly, based on their results, the authors pointed out that GlcNAc–thiazoline and PUGlcNAc inhibitors[449] could be considered as the mimetics of the oxazoline intermediate instead of the transition-state analogue.

d. Chitinases.—Chitinases of the GH18 family hydrolyze chitin, a β-(1 → 4)-linked GlcNAc polysaccharide that is a major structural component of fungal cell walls and exoskeletons of invertebrates, including insects and crustaceans. A substrate-assisted mechanism has been proposed for these hydrolases,[415] and various aspects of their enzymatic reaction have been studied by various QM methods.[440,444,445,450] All calculations support the substrate-assisted mechanism. A comparison of the transition state of the first step and the chitinase inhibitor allosamidin suggested that allosamidin functions as a transition-state analogue of an oxazoline intermediate.[440] The first step of the enzymatic reaction has also been studied by the QM/MM method.[444] The semiempirical method optimized for carbohydrates, PM3CARB,[152] was used for the QM region, which consisted of 73 atoms. The calculated energy barrier (15.8–19.8 kcal mol^{-1}) was in agreement with the experimental barrier of 16.1 kcal mol^{-1}. The results revealed that proton transfer from the catalytic acid to the glycosidic oxygen atom occurred concurrently with the nucleophilic attack of oxygen from the N-acetyl group on the anomeric carbon. The transition state had an oxocarbenium character. Interestingly, two stable intermediates were found on the reaction pathway $^{1,4}B$ (Michaelis complex) → $^{4}H_5/^{4}E$ (TS) → $^{4}C_1 \leftrightarrow B_{3,O}$ (intermediate).

The lysozyme of hen egg white was the first enzyme structure solved by X-ray crystallography.[422] The enzyme cleaves the β-(1 → 4) glycosidic linkage between *N*-acetylmuramic acid and *N*-acetylglucosamine. Different mechanisms have been proposed for the action of lysozyme [213,421,451] and have been investigated at different levels of theory.[213,442,443] The QM/MM calculations[443] do not support a mechanism involving the creation of an oxocarbenium ion and suggest that the reaction with the natural substrate proceeds via a covalent intermediate.

e. Others.—Golgi mannosidase II (GMII) is a carbohydrate-processing enzyme that cleaves the (1 → 3)- and (1 → 6)-linked mannose residue from a complex oligosaccharide. GMII is the retaining glycosidase and plays a vital role in the biosynthesis

of N-glycans.[452] An observation that swainsonine, an inhibitor of GMII, led to reduction of tumor growth and decreased the number of metastases[452] increased interest in the catalytic mechanism of GMII. It is noteworthy that this enzyme contains the Zn^{2+} cation in the active site. The QM/MM metadynamics simulations were used to examine the first step of the double-displacement mechanism of GMII.[433] The predicted free-energy barrier for the hydrolysis of a (1 → 6)-linkage is 23 kcal mol^{-1}. The calculations showed that GMII utilized an ${}^{O}S_2/B_{2,5} \rightarrow B_{2,5}(\mathrm{TS}) \rightarrow {}^{1}S_5$ itinerary for the glycosylation step. The $B_{2,5}$ TS had oxocarbenium ion character, with the glycosidic linkage being hydrolyzed lengthening to 2.0 Å. The nucleophilic oxygen of the catalytic base was 2.93 Å from the anomeric carbon atom, and the proton was transferred from the catalytic base to the leaving O-6 group.

Another group of retaining glycosidases comprises the sialidases. They catalyze cleavage of a sialic acid residue linked to glycoproteins, glycolipids, and polysaccharides, and there are numerous biological functions linked to these enzymes.[453] Despite their importance, particularly in relation to the influenza virus, the details of their catalytic mechanism have not been described. It was suggested[454] that the hydrolysis of sialosides follows an S_N1 type of mechanism of inverting glycosidases that is facilitated by a water molecule, and the final step is the mutarotation of the α anomer to the more stable β anomer in solution. Comparison of the calculated kinetic isotope effects (KIEs), using the semiempirical QM/MM methods, with experimental values led to the suggestion that the ring conformation of sialic acid in the TS was ${}^{4}H_5$.[455] Another semiempirical QM/MM study indicated that hydrolysis did not occur via a covalent intermediate and supported an S_N1-type mechanism.

The action mechanism of *endo*-β1-4-xylanase that catalyzes the hydrolysis of xylan and β-xylobiosides with retention of the anomeric configuration has been studied using semiempirical QM/MM MD methods.[456,457] The system consisted of 16,476 atoms, with 86 atoms in the QM region. Calculations indicated a stepwise mechanism for the glycosylation step. The rate-limiting step is the formation of the covalent xylosyl–enzyme intermediate formed by nucleophilic attack of the catalytic base without protonation of the leaving group. In the transition state TS_1, the bond lengths for breaking and forming glycosidic linkage are 2.13 and 2.11 Å, respectively.

3. Inverting GHs

As compared with retaining glycosidases, much less attention has been focused on molecular modeling of the catalytic mechanism of inverting enzymes, and only two of these enzymes have been subjected to detailed mechanistic analysis.[263,432] This situation is probably because of the relative simplicity of the catalytic

mechanism, since all of the inverting glycosidases so far studied proceed via a single-displacement mechanism (Fig. 35).

The reaction mechanism of the endoglucanase from family GH8 has been investigated with DFT QM/MM methods. This inverting enzyme catalyzes the cleavage of β-(1 → 4)-glycosidic linkages present in cellulose, chitosan, and xylan. The enzymatic reaction was monitored by means of two predefined reaction coordinates, using the metadynamics approach. The calculated free-energy surface for hydrolysis showed four stationary points: Michaelis complex (reactants), the protonated glycosidic oxygen atom, the transition state, and the product. The Michaelis complex showed that, in agreement with experimental data,[458] the glucose ring at subsite −1 favors a distorted $^{2}S_{O}/^{2,5}B$ conformation. The minimal energy pathway, β-$^{2}S_{O} \rightarrow {}^{2,5}B$ (TS) → $^{5}S_{1}$-α, proceeds through a transition state characterized by protonation of the glycosidic oxygen atom and a lengthening of the hydrolyzed glycosidic linkage. The calculations with the PBE functional[459] predicted hydrolysis of cellulose as an exothermic reaction with the reaction barrier around 36 kcal mol^{-1}.

A similar computational strategy was used for analysis of the hydrolysis of the β-(1 → 4) glycosidic linkage in xylobiose catalyzed by the inverting β-xylosidase[263] (XynB3) that belongs to family GH43. In the Michaelis complex, the pyranosyl ring is in the distorted $^{2}S_{O}$ conformation, and in the transition state the ring is close to the $^{2,5}B$ conformation. Thus, the conformational itinerary is similar to that found in an endoglucanase of the GH8 family. The predicted reaction barrier was almost 24 kcal mol^{-1}. The transition-state structure is characterized by lengthening of the hydrolyzed glycosidic linkage from 1.49 Å in the Michaelis complex to 2.62 Å, and shortening of the ring C-1—O-5 linkage from 1.40 Å in the Michaelis complex to 1.28 Å.

4. Beyond the Catalytic Mechanism

The identification of the active site of a protein is usually assessed through systematic mutations of the amino acids suspected to play a role in the recognition and catalytic events. Computational methods are then used to dock the respective oligosaccharides or polysaccharides into the active site. The majority of investigations focus on unraveling the binding modes and catalytic mechanisms, with the expectation of a providing a rational design for potential inhibitors. The proteins and their carbohydrate ligands are considered in their energetically most stable conformations. Usually, several models of interactions result from the docking procedure, and they are ranked according to the calculated energies of interactions. These

models provide a way to localize the substrate in the protein and start explaining conformer selectivity based on the dimensions of the active site. In many instances, the docking of inhibitors and substrates helps to delineate the dimensions of the binding site, which are usually large. They may extend over several monosaccharide residues beyond and toward the site of cleavage. The key amino acids that may be involved in the catalytic mechanism can be identified from these docking studies. Such computational protocols have been applied to the study of various classes of glycosidases.

Enzyme inhibitors can be classified into substrate analogues and transition-state analogues. Both types of analogues inhibit the enzyme by effectively competing with the substrate for binding to the active site of the enzyme, but are not affected by the enzyme. Substrate analogues mimic the structural features of the substrates, whereas transition-state analogues have some structural characteristics that are unique to the transition state.

Some recent examples incorporating state-of-the-art modeling tools have been used to investigate the features of heparanase interacting with heparin.[407,460–462] From these docking studies, the substrate specificity of heparanase toward heparin could be explained and could provide a way to the rational design of saccharide entities that might inhibit the enzymatic action of heparanase. However, the most striking examples can be found in studies on influenza A, for which recent cases of avian influenza H5N1 and those of swine origin (H1N1) have caused a great deal of concern. Viral transmission involves two critical interactions. The first step is the binding of hemagglutinin (HA), a glycoprotein on the viral surface, to the sialylated glycan receptors that are terminated by *N*-acetylneuraminic acid (Neu5Ac) groups expressed on the surface of the host cell. Glycans terminating in Neu5Ac that is α-(2→6) and α-(2→3) linked to the penultimate galactose serve as receptors for the human- and avian-adapted influenza A virus, respectively. The final stage of infection occurs when viral neuraminidase (NA) cleaves neuraminic acid from the cell surface and progeny virions facilitate release of virus from infected cells (Fig. 41). Structural analysis of the neuraminidase of influenza virus,[463] and neuraminidase in a complex with sialic acid,[464] led to the design of a potent inhibitor of neuraminidase activity, termed zanamivir.[465] Based on the efficacy of zanamivir (Relenza™), another neuraminidase inhibitor has been developed, namely oseltamivir phosphate (Tamiflu™).[466] They are both potent and clinically effective anti-influenza drugs.[467] It should be noted that Relenza is a carbohydrate-based drug, whereas Tamiflu is a carbocyclic mimetic. Nevertheless, major concerns remain regarding the development of resistance to these drugs, and this is already occurring. Point mutations in the neuraminidase of the influenza virus have been reported that lead to a

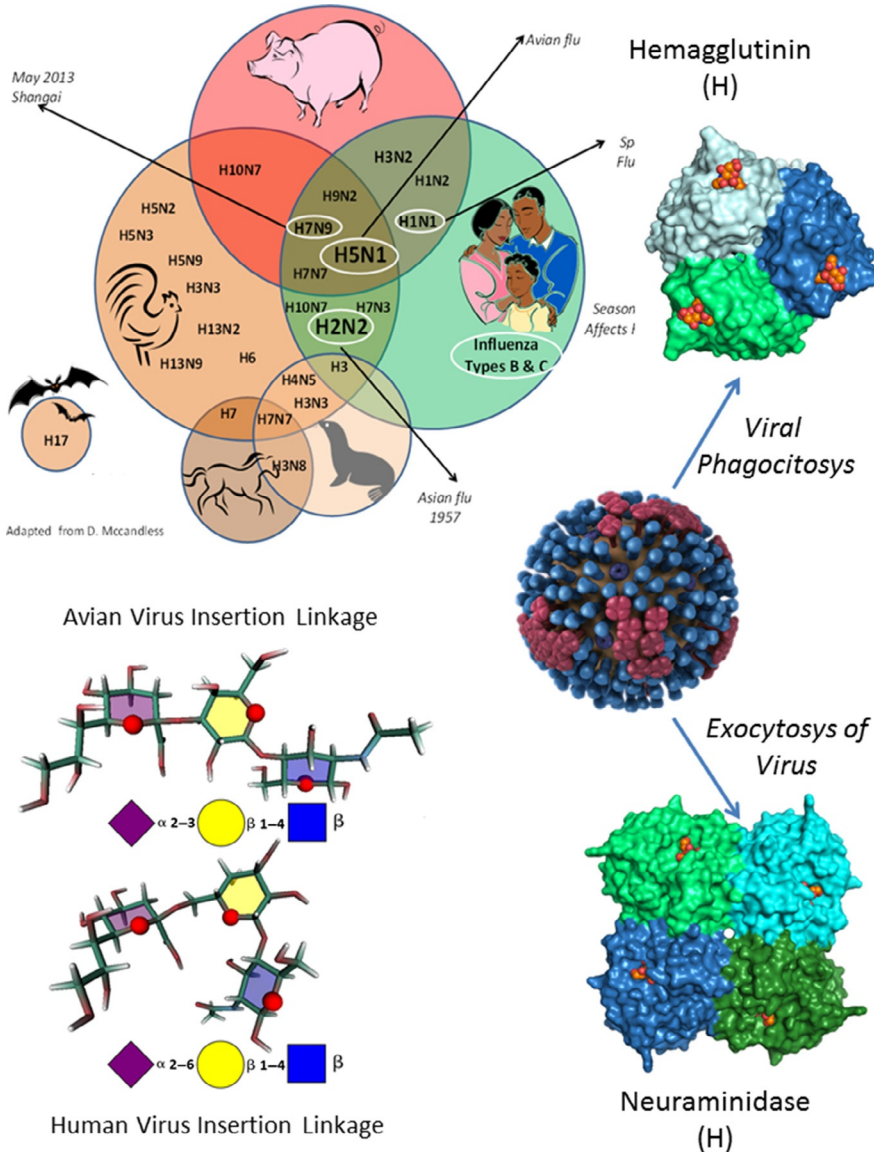

FIG. 41. Pictorial representation of the size and the interconnection of the reservoirs of influenza A viruses, along with their subtypes, (left) based on two proteins of the surface of the virus, (right) the hemagglutinin (H) and the neuraminidase (N), along with the structure of the oligosaccharides recognized by human viruses [αNeu5Ac*p*-(2→6)-βGal*p*-(1→4)-GlcNAc*p*] and in birds [αNeu5Ac*p*-(2→3)-βGal*p*-(1→4)-GlcNAc*p*] as recognized by avian viruses.

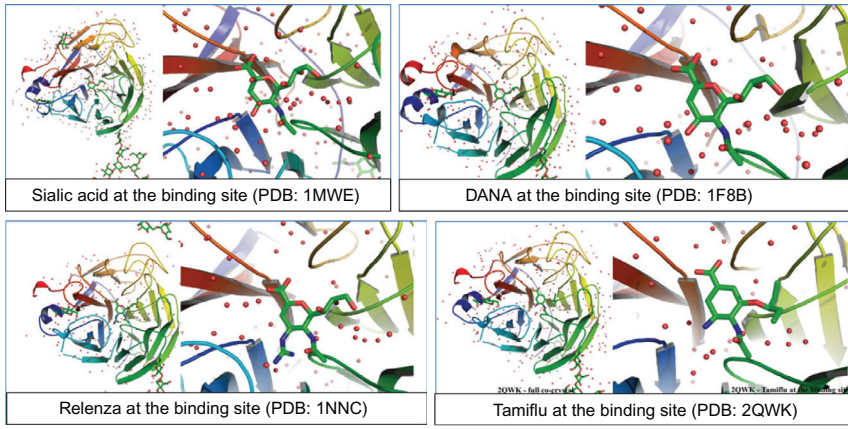

FIG. 42. Computational studies of the resistance of H5N1 influenza to neuraminidase inhibitors: Attachment of the various inhibitors of neuraminidase at the binding site of the cocrystal (i) sialic acid (PDB 1MWE),[509] (ii) 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid (DANA) (PDB 1F8B),[161] (iii) Relenza® (4-guanidino-Neu5Ac2en, or zanamivir) (PDB 1NNC),[510] and (iv) Tamiflu© [5-acetamido-3-(1-propoxy)-1-cyclohexene-1-carboxylic acid, or oseltamivir] (PDB 2QWK).[511]

dramatic loss of activity for the known neuraminidase inhibitors cited here (Fig. 42). A deeper elucidation of the molecular basis involved in such resistance is required for developing improved next-generation drugs. Modeling the binding of ligands with neuraminidase has been performed using explicit solvent all-atom MD simulations, free-energy calculations, and residue-based decomposition. The simulations predicted the effects of a known mutation at one amino acid (R292K) and provided clues as to the origins of resistance to the mutant. The results provide further rationalization of the experimental observations previously recorded.[468]

The possibility of future influenza pandemics (including the advent of highly pathogenic H5N1 strains) has highlighted the need for additional computational methods. The binding properties of the neuraminidase of H5N1 influenza virus have been inferred from molecular modeling.[469] The investigation, using molecular docking and MD simulations, aimed at characterizing the difference of the binding properties between sialic acid, methyl 3′-sialyl-lactoside, methyl 6′-sialyl-lactoside, and the neuraminidase of the H5N1 influenza virus. The results obtained indicate that, in the complex, sialic acid undergoes a conformational transition of the ring. Meanwhile, methyl 3′-sialyl-lactoside establishes only weak interactions with a key loop of the neuraminidase, in contrast to what is observed for the complex with methyl

6′-sialyl-lactoside. The differences could be attributed to the existence of different conformations about the glycosidic linkages. As these molecular modeling results are consistent with available experimental data on the specificity of neuraminidase, they provide reasonable structural information for the rational design of novel and specific inhibitors of H5N1 neuraminidase as potential therapeutics for the treatment of avian influenza.

5. GHs on a Solid Substrate: The Cellulase Case

The enzymatic breakdown and degradation of cellulose require a complex of enzymes working together, as the synergetic action of at least three types of enzymes is required for efficient digestion of cellulose into glucose (Fig. 43). These enzyme types are (i) endoglucanases (EG, EC 3.2.1.4), which cleave the cellulose chains randomly; (ii) cellobiohydrolases (CBH, EC 3.2.1.91), which recurrently cleave cellobiose from the chain end of cellulose; and (iii) β-glucosidases (EC 3.2.1.21), which hydrolyze cellobiose. Other complementarities are found among these enzymes, as exemplified by the synergistic action of two cellobiohydrolases namely, Cel6A and Cel7A, from such organisms as *Trichoderma reesei* and *Humicola insolens*. Cellulases also consist of a cellulose-binding modules (CBMs) which play a fundamental role in the hydrolysis of the solid substrate. Even if the exact function of this CMB is not totally elucidated, it is known to potentiate the action of cellulolytic enzymes on insoluble substrates and plays a significant role in the degradation of crystalline cellulose.[470] The CBM has a specific planar binding surface that allows the enzyme to adhere to the hydrophobic face of the sugar molecules, thus enhancing the activity of the catalytic module toward crystalline cellulose. The core domain contains an active center that hydrolyzes cellulose in a catalytic manner, together with subsites that interact with the cellulose chain close to the active site. The aromatic amino acids (tryptophan and tyrosine) provide strong binding to the cellulose chains via van der Waals interactions. The occurrence of the CMB, which plays a dual function of recognizing and adhering to the surface of the crystalline cellulose, is a key factor in the ability of the enzyme to degrade the insoluble substrate. The tight association of substrate with the enzymatic machinery (namely the catalytic module linked to the CBM) provides an efficient way to keep all of the components in proximity. Numerous studies have established that three aromatic residues are needed for binding onto cellulose crystals, and that tryptophan residues contribute to higher binding affinity than tyrosines. However, evidence has accumulated showing that different binding sites for the same cellulose-binding domains could occur. After a

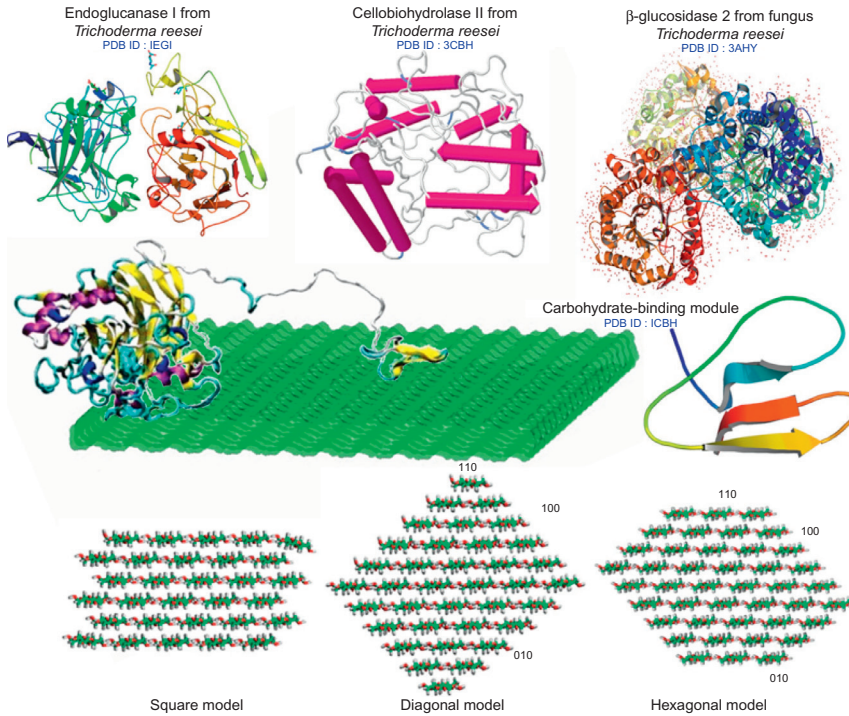

FIG. 43. The enzymatic digestion of cellulose. The top panel depicts the three-dimensional structures of the three main categories of enzymes (from *Trichoderma reesei*) that digest crystalline cellulose. The central panel provides a visual depiction as to how the cellulose-digesting enzyme (shown here as cellobiohydrolase I or Cel7A[477,512]) interacts with the crystalline arrangement of cellulose and breaks down cellulose into glucose, and it also illustrates a carbohydrate-binding module (CBM; PDB 1CBH). The bottom panel shows the three faces of the cellulose Iα crystal models, in projection with the Miller indices of their constituent crystal planes.[513] Reprinted with permission from Ref. 10. Copyright 2013 CRC Press, Taylor & Francis.

first step of adsorption onto the microfibrils, the cellulolytic enzymes are then able to leave cellulose chains, leading to morphological changes in the substrate. The reaction rate decreases dramatically with the extent of conversion; a slowdown could be explained by such enzyme-related factors as enzyme inactivation, product inhibition, or changes in substrate reactivity.

The morphology of the native cellulose crystals is an essential feature in the enzymatic degradation, and this is of profound concern, from both fundamental and applied aspects.[471] The binding of cellulose by GH enzymes (from Cazy family 6)

from *Hypocrea jecorina* and *Thermobifioda fusca* has been investigated using AutoDock 3.06. The results indicate that a β-glucopyranose residue binds most tightly to the enzyme subsite approximately two carbohydrate residues away from the glycosidic cleavage site (namely the so-called −2 subsite).[472] In a subsequent docking investigation,[285] the specificity of various GHs for β-glycosides was evaluated using AutoDock. The outcome of the simulation pointed toward the existence of eight active-site residues as being mainly responsible for carbohydrate specificity. The way in which cellulose is recognized by the CBM of the Cel9 enzyme from *T. fusca* has been explored through a docking study, using DOCK5 in conjunction with MD simulation. The results indicated the involvement of the CBM in the binding event, a finding that explains why removal of the CBM results in a decreased turnover of the hydrolysis of cellulose.[473]

A systematic study of the CBM protein of Cel7A of *T. reesei*, with the cellulose Iα crystal model has been performed using a combined Grid docking search and MD calculations (Fig. 42).[474] Three types of crystal models of cellulose Iα with infinite dimensions were constructed, each consisting of different crystallographic faces, namely (1 1 0), (1 0 0), and (0 1 0). The (1 1 0) complex models exhibited larger affinities at the interface than the other ones. It was found that the CBM was bound more stably to the (1 1 0) surface when it was placed in an antiparallel orientation with respect to the fiber axis of the cellulose. The predicted directional specificity of the CBM at the optimal positions was consistent with the observed processing direction of the Cel7A.[475] In the solvated dynamic state, the curved (1 1 0) surface resulting from the fiber twist assisted somewhat in a complementary fit with the CBM at the interface.

Much can be learned about the processivity by conducting carefully designed MD simulations of the binding of the catalytic domains of cellulases having different substrate configurations, solvation models, and thermodynamics protocols.[476] Processivity describes the action of an enzyme that binds a polymeric substrate and catalyzes a series of similar chemical reactions along that polymer before releasing the fully modified polymer to solvent.)

The computational model of cellobiohydrolase I (Cel7A) from *T. reesei* on a cellulose (1 0 0) surface displays a large catalytic domain (left), linker (middle single strand), and CBM (right small domain). A cellodextrin strand is shown peeled out of the surface of the cellulose and threaded into the catalytic tunnel of Cel7A. The solvating water and the lower portion of the cellulose fiber are not shown in this visualization.[477] The investigation requires the consideration of approximately 800,000 atoms. In order to encounter such a computational challenge, most of the numerical simulations will require significant modifications of existing code and algorithms.

IX. Conclusion and Perspectives

In the past few years, there has been an increase in the development and application of computational methods aimed at establishing the molecular features characterizing protein–carbohydrate interactions. Quite naturally, these computational methods rely on earlier experimental studies for interpreting the structural and dynamics characteristics in the broad field of glycoscience. Significant steps have been made, among which the developments and implementations of force fields capable of taking into account the specificity of carbohydrates (stereoelectronic effects, *gauche* effects, and others), and their compatibility with the computational tools that have been developed earlier for proteins. Recently, methods for handling the many rotatable bonds in the flexible docking of conformationally mobile carbohydrates have been established. It has been recognized that the surfaces of carbohydrates and their derivatives present both hydrophobic and hydrophilic patches, and these remain a source of difficulty in modeling. Nevertheless, a balance between hydrophobic and hydrophilic patches is essential for the solubility of the carbohydrate and for molecular recognition. The existence of such a feature, combined with the enhanced conformational flexibility, is a unique aspect that explains how complex oligosaccharides can be transported through transmembrane proteins in a capillary-like and yet selective manner.

The force-field calculations of binding free energies and enthalpies with the desired level of accuracy remain to be improved and verified against well-characterized experimental data. Certainly, calculation of free-energy perturbations is a promising approach for the prediction of carbohydrate–receptor binding affinity. Such calculations cannot be performed without a full understanding of solvation. Progresses in this area require a better handling of hydration and the significant role played by solvation and desolvation of both carbohydrates and proteins, in their isolated states and also during the course of their interactions.

Computational modeling of protein–carbohydrate complexes faces multiple challenges that arise from the extreme flexibility of the ligand, the occurrence of CH–π interactions, the formation of an extensive network of hydrogen bonding (often involving water molecules), and the lack of specialized scoring functions. Nevertheless, these computational tools are achieving recognition as being as valuable as the other methods of structural investigation, and they have started to produce reliable and insightful results. They can help in reconciling the experimental results gathered from independent experiments under different conditions and environments, and in extrapolating the results. The wealth of successful applications in many different examples of protein interactions with carbohydrates is a testimony to the maturity of the modeling methods and protocols that have been developed. These instances of

success deal nevertheless almost exclusively with situations where proteins interact with carbohydrates, without any further catalytic action. Even in these instances, very few studies have been implemented that properly account for receptor flexibility in the carbohydrate-binding site.

Complementary computational methods need to be developed and/or integrated to allow the study of enzymatic reactions and the subsequent optimization of such biocatalyts. These methods, based on molecular robotics algorithms, would be used for an efficient virtual screening of high-dimensional configurational and conformational space. The ongoing developments of robotics algorithms are expected to provide efficient tools for exploring the dynamic functionality of enzymes. These are based on effective path-planning algorithms and fast geometric operators designed for complex articulated chains. Their aim is to reduce, in a significant but relevant way, the exploration of the combinatorial space of the enzyme sequences, based on the geometric feasibility for a ligand either to access or to leave the catalytic site in a "productive" way. The enhancement of the predictive performances of such algorithms will require the use of simplified energy functions to prefilter those conformations that are nonviable, and build the network of concerted motions while the enzyme is interacting with the ligand.

Investigation of the catalytic mechanism of inverting and retaining carbohydrate-processing enzymes requires high-level, DFT, and hybrid QM/MM calculations. The modeling studies have provided valuable information about the reaction mechanism of glycosidases and GTs. However, despite considerable progress in our understanding of structure–function relationships of these enzymes, there are challenges remaining. For example, the structure and action mechanism of GTs that are involved in the biosynthesis of various polysaccharides need to be determined. Studies on the catalytic reaction and the dynamic motions undergone by the enzymes are currently being addressed only separately. Many GTs exhibit conformational changes upon binding to the ligand and in the catalytic stage. For most enzymes, the nature and extent of these changes have not been well defined. To understand fully the catalytic mechanism of GTs, some questions therefore need to be answered: (i) Is the rate-limiting step chemical bond cleavage/formation, or is it conformational motion? (ii) Do the conformational fluctuations occur only when substrates are present? (iii) Is the loop closing a one-step process or are intermediates involved, and is there only one path to this conformation? (iv) Do motions from fully equilibrated conformational changes, as well as atomic vibration modes, contribute to and are they part of the complex reaction coordinate? Obtaining the entire picture of the reaction mechanism of GTs remains one of the major challenges for the rational design of specific and potent inhibitors/drugs for this class of enzymes. Consequently, developments are

required to set up "hybrid methods" of MD based on the principles of QM, with the purpose of studying the dynamics of electronic effects and charge transfer within the catalytic site. Such hybrid methods would include *ab initio* dynamics as developed by Car and Parrinello (CPMD) and a "classical" MD force field. Applications of these computational methods should allow further exploitation of the protein–carbohydrate interactions, especially for therapeutic purposes. Design of inhibitors of GHs and GTs that are transition-state analogues requires knowledge of the mechanism of the enzymatic reaction, along with the geometry and charge distribution of the transition state.

Biologically and technologically relevant questions are being identified and present major challenges. Since most carbohydrate-binding proteins, particularly lectins and adhesins, display rather low affinity and generally have narrow carbohydrate-recognition domains (involving less than a tetrasaccharide residue), their intrinsic specificities often reside in their valences, together with their various topologies. The simultaneous presentation of several appropriate and identical glycoside units converts relatively weak interactions into strong and specific recognition effects. This specificity is frequently referred to as "Sugar Code," deciphering of which may require the full characterization of not only a single carbohydrate–protein interaction but the interactions of several binding sites with multivalent carbohydrate ligands being presented. This factor may be the impetus for designing multivalent inhibitors in an anti-adhesion therapy and to understand how soluble lectins expressed as virulence factors by opportunistic bacteria may bind to epithelial cells of the host prior to invagination. Therefore, some physicochemical principles that underline their associations need to be considered in modeling such systems, for example, patches of glycolipids and glyco-surfaces. As the concept of a "glyco-landscape" becomes recognized, new modeling protocols need to be developed. They require novel computational tools capable of constructing the landscape resulting from the side-by-side arrangements of such glycoconjugates as glycolipids. A new paradigm will emerge, where attention will be given not only to the interaction of a protein with a single carbohydrate unit (the so-called tree vision), but rather the interaction with glyco-surfaces (the so-called glyco-canopy, a term coined by analogy to the "crown canopy," namely the uppermost layer in a forest formed by the crown of trees).

This concept is likely to progress as enhanced research on the solid-state degradation by enzymes of the polysaccharides of plant cell-walls offers major challenges and the promise of important applications. Plant biomass is an alternative natural source for chemical feedstocks, with a replacement cycle short enough to meet the demand of the world fuel market. The enzymatic hydrolysis of cellulose is still considered as the main limiting step in the production of biofuels from lignocellulosic biomass. This

step involves the action of three types of cellulose-degrading enzymes acting synergistically. For the design of a functional kinetic model, integrating the various properties of each enzyme along with its synergies, much can be learned by conducting carefully designed computer simulations. These would include the binding of the cellulose-binding domains and the catalytic domains of cellulases with various substrates, along with solvation models and thermodynamics protocols. Such an enormous computational challenge is delineating the new frontiers in the area of protein–carbohydrate interactions.

ACKNOWLEDGMENTS

Aspects of the work described here were supported by the Scientific Grant Agency of the Slovak Academy of Sciences (grants VEGA-02/0101/11), SAS-NSC JPR 2012/8 project, and Research & Development Operational Programs funded by ERDF (CEGreenI, Contract No. 26240120001 and CEGreenII, Contract No. 26240120025). Part of the research leading to this article has received funding from the European Commission's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 215536. The authors are indebted to Drs. Anne Imberty, Martin Franck, and Anita Sarkar for their contributions to scientific illustrations.

REFERENCES

1. R. J. Woods and M. B. Tessier, Computational glycoscience: Characterizing the spatial and temporal properties of glycans and glycan-protein complexes, *Curr. Opin. Struct. Biol.*, 20 (2010) 575–583.
2. S. Pérez, C. Gautier, and A. Imberty, Oligosaccharide conformations by diffraction methods, In: B. Ernst, G. W. Hart, P. Sinaÿ, (Eds.), *Oligosaccharides in Chemistry and Biology,* Wiley-VCH, Weinheim, 2000, pp. 969–1001.
3. S. Pérez, Oligosaccharides and polysaccharide conformations by diffraction methods, In: J. P. Kamerling, (Ed.), *Comprehensive Glycosciences B. Analytical Aspects (Scope and Limitations),* Elsevier, Amsterdam, 2007, pp. 347–387.
4. K. G. Rice, W. Pengguang, L. Brand, and Y. C. Lee, Experimental determination of oligosaccharide three-dimensional structure, *Curr. Opin. Struct. Biol.*, 3 (1993) 669–674.
5. T. Peters and B. M. Pinto, Structure and dynamics of oligosaccharides: NMR and modeling studies, *Curr. Opin. Struct. Biol.*, 6 (1996) 710–720.
6. A. Sarkar and S. Pérez, PolySac3DB: An annotated data base of 3 dimensional structures of polysaccharides, *BMC Bioinf.*, 13 (2012) 302.
7. S. Pérez, Molecular modeling in glycosciences, In: J. P. Kamerling, (Ed.), *Comprehensive Glycosciences, B. Analytical Aspects (Scope and Limitations),* Elsevier, Oxford, 2007, pp. 193–220.
8. S. Pérez and M. Kouwijzer, Shapes and interactions of polysaccharide chains, In: P. Finch, (Ed.), *Carbohydrates,* Springer Verlag, Berlin, 1999, pp. 258–293.
9. R. A. Laine, Invited commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: The isomer barrier to development of single-method saccharide sequencing or synthesis systems, *Glycobiology*, 4 (1994) 759–767.
10. A. Bohne-Lang, E. Lang, T. Forster, and C. W. von der Lieth, LINUCS: Linear notation for unique description of carbohydrate sequences, *Carbohydr. Res.*, 336 (2001) 1–11.

11. E. Banin, Y. Neuberger, Y. Altshuler, A. Halevi, O. Inbar, and A. Dekler, A novel linear code nomenclature for complex carbohydrates, *Trends Glycosci. Glycotechnol.*, 14 (2002) 127–137.
12. N. Kikuchi, A. Kameyama, S. Nakaya, S. Ito, T. Shikanai, Y. Takahasshi, and H. Narimatsu, The carbohydrate sequence markup language (CabosML): An XML description of carbohydrate structures, *Bioinformatics*, 21 (2005) 1717–1718.
13. S. Herget, R. Ranzinger, K. Maass, and C. W. von der Lieth, GlycoCT—A unifying sequence format for carbohydrates, *Carbohydr. Res.*, 343 (2008) 2162–2171.
14. IUPAC, Tentative rules for carbohydrate nomenclature, *Eur. J. Biochem.*, 21 (1971) 455–477.
15. IUPAC-IUB, Nomenclature of carbohydrates, *Pure Appl. Chem.*, 68 (1996) 1919–2008.
16. A. Varki, R. Kannagi, and B. P. Toole, Glycosylation changes in cancer, In: A. Varki, R. D. Cummings, J. D. Esko, H. D. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. E. Etzler, (Eds.), *Essentials of Glycobiology,* Laboratory Press, Cold Spring Harbor, NY, 2009. Chapter 44, available from http://www.ncbi.nlm.nih.gov/books/NBK1963.
17. A. Varki, Essentials of Glycobiology, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2009.
18. J. T. Edward, Stability of glycosides to acid hydrolysis, *Chem. Ind.*, 36 (1955) 1102–1104.
19. A. J. DeHoog, H. R. Buys, C. Altona, and E. Havinga, Conformation of non-aromatic ring compounds-LII: NMR spectra and dipole moments of 2-alkoxytetrahydropyrans, *Tetrahedron*, 25 (1969) 3365–3375.
20. R. U. Lemieux, A. A. Pavia, and J. C. Martin, Solvation effects on conformational equilibria. Studies related to the conformational properties of 2-methoxytetrahydropyran and related methyl glycopyranosides, *Can. J. Chem.*, 47 (1969) 4427–4439.
21. R. U. Lemieux and A. R. Morgan, The abnormal conformations of pyridinium α-glycopyranosides, *Can. J. Chem.*, 43 (1965) 2205–2213.
22. S. Wolfe, Gauche effect. Stereochemical consequences of adjacent electron pairs and polar bonds, *Acc. Chem. Res.*, 5 (1972) 102–111.
23. G. R. J. Thatcher (Ed.), The Anomeric Effect and Associated Stereoelectronic Effects, *ACS Symp. Ser.*, 539 (1993) 1–305.
24. I. Tvaroška and T. Bleha, Anomeric and exoanomeric effects in carbohydrate chemistry, *Adv. Carbohydr. Chem. Biochem.*, 47 (1989) 45–123.
25. I. Tvaroška and T. Bleha, Theoretical stereochemistry of molecules with heteroatoms linked to the tetrahedral center and the anomeric effect, *Chem. Pap.*, 39 (1985) 805–847.
26. A. J. Kirby, The Anomeric Effect and Related Stereoelectronic Effects at Oxygen, Springer-Verlag, Berlin, 1983.
27. E. C. Juaristi and G. Cuevas, The Anomeric Effect, CRC Press, Inc., Bocas Raton, Florida, 1995.
28. P. R. Graczyk and M. Mikolajczyk, Anomeric effect—Origin and consequences, In: E. L. W. Eliel and H. Samuel, (Eds.), *Topics in Stereochemistry,* Vol. 21, John Wiley & Sons, New York, 1994, pp. 159–349.
29. C. B. Romers, C. Altona, H. R. Buys, and E. Havinga, Geometry and conformational properties of some five- and six-membered heterocyclic compounds containing oxygen or sulfur, In: E. L. Eliel and N. L. Allinger, (Eds.), *Topics in Stereochemistry, Vol. 4*, John Wiley & Son, New York, 1969, pp. 39–97.
30. A. J. de Hoog and E. Havinga, Conformation of non-aromatic ring compounds. Part 64. NMR spectra and dipole moments of 2-alkylthiotetrahydropyrans, *Recl. Trav. Chim. Pays-Bas.*, 89 (1970) 972–979.
31. I. V. Alabugin, K. M. Gilmore, and P. W. Peterson, Hyperconjugation, *WIREs Comput. Mol. Sci.*, 1 (2011) 109–141.
32. I. Tvaroška and T. Bleha, Lone pair interactions in dimethoxymethane and anomeric effect, *Can. J. Chem.*, 57 (1979) 424–435.
33. I. Tvaroška and F. R. Taravel, Carbon-proton coupling constant in the conformational analysis in the conformational analysis of sugar molecules, *Adv. Carbohydr. Chem. Biochem.*, 51 (1995) 15–61.

34. M. Hricovini and I. Tvaroška, Conformational dependence of the one-bond carbon proton coupling-constants in oligosaccharides, *Magn. Reson. Chem.*, 28 (1990) 862–866.
35. V. G. S. Box, The role of lone pair interactions in the chemistry of the monosaccharides—Stereoelectronic effects in unsaturated monosaccharides, *Heterocycles*, 32 (1991) 795–807.
36. E. J. Cocinero, P. Carcabal, T. D. Vaden, J. P. Simons, and B. G. Davis, Sensing the anomeric effect in a solvent-free environment, *Nature*, 469 (2011) 76–79.
37. I. Tvaroška and T. Bleha, Different nature of the anomeric effect for methoxy and chlorine substituents, *Tetrahedron Lett.*, 16 (1975) 249–252.
38. G. F. Bauerfeldt, T. M. Cardozo, M. S. Pereira, and C. O. da Silva, The anomeric effect: The dominance of exchange effects in closed-shell systems, *Org. Biomol. Chem.*, 11 (2013) 299–308.
39. R. J. Gillespie, E. A. Robinson, and J. Pilme, Ligand close packing, molecular compactness, the methyl tilt, molecular conformations, and a new model for the anomeric effect, *Chem. Eur. J.*, 16 (2010) 3663–3675.
40. C. W. Wang, Z. H. Chen, W. Wu, and Y. R. Mo, How the generalized anomeric effect influences the conformational preference, *Chem. Eur. J.*, 19 (2013) 1436–1444.
41. O. Takahashi, K. Yamasaki, Y. Kohno, R. Ohtaki, K. Ueda, H. Suezawa, Y. Umezawa, and M. Nishio, The anomeric effect revisited. A possible role of the CH/*n* hydrogen bond, *Carbohydr. Res.*, 342 (2007) 1202–1209.
42. G. A. Jeffrey, J. A. Pople, and L. Radom, The application of *ab initio* molecular orbital theory to structural moieties of carbohydrates, *Carbohydr. Res.*, 38 (1974) 81–95.
43. G. A. Jeffrey, J. A. Pople, J. S. Binkley, and S. Vishveshwara, The application of *ab initio* molecular orbital theory to structural moieties of carbohydrates. Part III, *J. Am. Chem. Soc.*, 100 (1978) 373.
44. G. A. Jeffrey, J. A. Pople, and L. Radom, The application of *ab initio* molecular orbital theory to the anomeric effect. A comparison of theoretical predictions and experimental data, *Carbohydr. Res.*, 25 (1972) 117–131.
45. I. Tvaroška and J. P. Carver, *Ab initio* molecular orbital calculation on carbohydrate model compounds. 1. The anomeric effect in fluoro and chloro derivatives of tetrahydropyran, *J. Phys. Chem.*, 98 (1994) 6452–6458.
46. I. Tvaroška and J. P. Carver, *Ab-Initio* molecular orbital calculation of carbohydrate model compounds. 2. conformational analysis of axial and equatorial 2-methoxytetrahydropyrans, *J. Phys. Chem.*, 98 (1994) 9477–9485.
47. I. Tvaroška and J. P. Carver, *Ab initio* molecular orbital calculation of carbohydrate model compounds. 5. Anomeric, *exo*-anomeric, and reverse anomeric effects in *C*-, *N*-, and *S*-glycosyl compounds, *J. Phys. Chem.*, 100 (1996) 11305–11313.
48. I. Tvaroška and J. P. Carver, *Ab initio* molecular orbital calculation of carbohydrate model compounds. 6. The gauche effect and conformations of the hydroxymethyl and methoxymethyl groups, *J. Phys. Chem. B*, 101 (1997) 2992–2999.
49. I. Tvaroška and J. P. Carver, The anomeric and exo-anomeric effects of a hydroxyl group and the stereochemistry of the hemiacetal linkage, *Carbohydr. Res.*, 309 (1998) 1–9.
50. I. Tvaroška and J. P. Carver, *Ab initio* molecular orbital calculation of carbohydrate model compounds. 4. Flexibility of psi-type glycosidic bonds in carbohydrates, *J. Mol. Struct. (THEOCHEM)*, 395 (1997) 1–13.
51. H. L. Woodcock, D. Moran, R. W. Pastor, A. D. MacKerell, and B. R. Brooks, *Ab initio* modeling of glycosyl torsions and anomeric effects in a model carbohydrate: 2-Ethoxy tetrahydropyran, *Biophys. J.*, 93 (2007) 1–10.
52. A. J. Weldon, T. L. Vickrey, and G. S. Tschumper, Intrinsic conformational preferences of substituted cyclohexanes and tetrahydropyrans evaluated at the CCSD(T) complete basis set limit: Implications for the anomeric effect, *J. Phys. Chem.*, 109 (2005) 11073–11079.

53. C. J. Cramer, Anomeric and reverse anomeric effects in the gas phase and aqueous solution, *J. Org. Chem.*, 57 (1992) 7034–7043.
54. C. J. Cramer, D. G. Truhlar, and A. D. French, Exo-anomeric effects on energies and geometries of different conformations of glucose and related systems in the gas phase and aqueous solution, *Carbohydr. Res.*, 298 (1997) 1–14.
55. U. Salzner and P. v. R. Schleyer, *Ab initio* examination of anomeric effects in tetrahydropyrans 1,3-dioxane and glucose, *J. Org. Chem.*, 5 (1995) 2138–2155.
56. I. Tvaroška and T. Kožár, Theoretical studies on the conformation of saccharides. 3. Conformational properties of the glycosidic linkage in solution and their relation to the anomeric effect, *J. Am. Chem. Soc.*, 102 (1980) 6929–6936.
57. I. Tvaroška and T. Bleha, Calculation of solvent effect on conformation stability and anomeric effect in dimethoxymethane, *Collect. Czech. Chem. Commun.*, 45 (1980) 1883–1895.
58. A. Imberty and S. Pérez, Structure, conformation, and dynamics of bioactive oligosaccharides: Theoretical approaches and experimental validations, *Chem. Rev.*, 100 (2000) 4567–4588.
59. B. L. Foley, M. B. Tessier, and R. J. Woods, Carbohydrate force fields, *WIREs Comput. Mol. Sci.*, 2 (2012) 652–697.
60. E. Fadda and R. J. Woods, Molecular simulations of carbohydrates and protein-carbohydrate interactions: Motivation, issues and prospects, *Drug Discov. Today*, 15 (2010) 596–609.
61. I. Tvaroška and S. Pérez, Conformational-energy calculations for oligosaccharides: A comparison of methods and a strategy of calculation, *Carbohydr. Res.*, 149 (1986) 389–410.
62. R. H. Marchessault and S. Pérez, Conformations of the hydroxymethyl group in crystalline aldeohexopyranoses, *Biopolymers*, 18 (1979) 2369–2374.
63. R. Apweiler, H. Hermjakob, and N. Sharon, On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochim. Biophys. Acta*, 1473 (1999) 4–8.
64. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, The Protein Data Bank, *Nucleic Acids Res.*, 28 (2000) 235–242.
65. T. Lütteke and C. W. von der Lieth, Data mining the PDB for glyco-related data, *Methods Mol. Biol.*, 534 (2009) 293–310.
66. A. Imberty, H. Lortat-Jacob, and S. Pérez, Structural views of glycosaminoglycans proteins interactions, *Carbohydr. Res.*, 342 (2007) 430–439.
67. L. L. Kiessling, J. E. Gestwicki, and L. E. Strong, Synthetic multivalent ligands in the exploration of cell-surface interactions, *Curr. Opin. Chem. Biol.*, 4 (2000) 696–703.
68. M. Reynolds and S. Pérez, Thermodynamics and chemical characterization of protein-carbohydrate interactions: The multivalency issue, *C. R. Chim.*, 14 (2011) 74–95.
69. J. J. Lundquist and E. J. Toone, The cluster glycoside effect, *Chem. Rev.*, 102 (2002) 555–578.
70. E. R. Zuiderweg, Mapping protein-protein interactions in solution by NMR spectroscopy, *Biochemistry*, 41 (2002) 1–7.
71. J. Clarkson and I. D. Campbell, Studies of protein-ligand interactions by NMR, *Biochem. Soc. Trans.*, 31 (2003) 1006–1009.
72. J. Jiménez Barbero and T. Peters, TR-NOE experiments to study carbohydrate-protein interactions, In: J. Jiménez Barbero and T. Peters, (Eds.), *NMR Spectroscopy of Glycoconjugates,* Wiley-VCH Verlag GmbH & co, Weinheim, FRG, 2003, pp. 289–310.
73. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, J. D. Marth, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, Symbol nomenclature for glycan representation, *Proteomics*, 9 (2009) 5398–5399.
74. H. C. Siebert, J. Jiménez Barbero, S. Andre, H. Kaltner, and H. J. Gabius, Describing topology of bound ligand by transferred nuclear Overhauser effect spectroscopy and molecular modeling, *Methods Enzymol.*, 362 (2003) 417–434.

75. M. J. Burgering, R. Boelens, M. Caffrey, J. N. Breg, and R. Kaptein, Observation of inter-subunit nuclear Overhauser effects in a dimeric protein. Application to the Arc repressor, *FEBS Lett*., 330 (1993) 105–109.
76. P. J. M. Folkers, R. H. A. Folmer, R. N. H. Konings, and C. W. Hilbers, Overcoming the ambiguity problem encountered in the analysis of the nuclear Overhauser magnetic resonance spectra of symmetrical dimer proteins, *J. Am. Chem. Soc*., 115 (1993) 3798–3799.
77. C. Zwahlen, P. Legault, and S. J. Vincent, Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex, *J. Am. Vet. Med. Assoc*., 119 (1997) 6711–6721.
78. L. M. I. Koharudin and A. M. Gronenborn, Nuclear magnetic resonance studies of carbohydrate-protein interactions, In: E. Yuriev and P. A. Ramsland, (Eds.), *Structural Glycobiology,* CRC Press, 2013, pp. 29–45.
79. K. Landsteiner, The Specificity of Serological Reactions, Dover, New York, 1962.
80. J. P. McCoy, Jr., J. Varani, and I. J. Goldstein, Enzyme-linked lectin assay (ELLA): Detection of carbohydrate groups on the surface of unfixed cells, *Exp. Cell Res*., 151 (1984) 96–103.
81. J. F. Espinosa, J. L. Asensio, J. L. Garcia, J. Laynez, M. Bruix, C. Wright, H. C. Siebert, H. J. Gabius, F. J. Cañada, and J. Jiménez Barbero, NMR investigations of protein-carbohydrate interactions binding studies and refined three-dimensional solution structure of the complex between the B domain of wheat germ agglutinin and N, N', N"-triacetylchitotriose, *Eur. J. Biochem*., 267 (2000) 3965–3978.
82. K. J. Neurohr, N. M. Young, and H. H. Mantsch, Determination of the carbohydrate-binding properties of peanut agglutinin by ultraviolet difference spectroscopy, *J. Biol. Chem*., 255 (1980) 9205–9209.
83. R. V. Weatherman and L. L. Kiessling, Fluorescence anisotropy assays reveal affinities of C- and O-glycosides for concanavalin A(1), *J. Org. Chem*., 61 (1996) 534–538.
84. R. Seethala, Fluorescence polarization competition immunoassay for tyrosine kinases, *Methods*, 22 (2000) 61–70.
85. Y. M. Chabre and R. Roy, Design and creativity in synthesis of multivalent neoglycoconjugates, *Adv. Carbohydr. Chem. Biochem*., 63 (2010) 165–393.
86. M. Hartmann and T. Lindhorst, The bacterial lectin FimH, a target for drug discovery: Carbohydrate inhibitors of type 1 fimbriae-mediated bacterial adhesion, *Eur. J. Org. Chem*., 20–21 (2011) 3583–3609.
87. F. Sansone, L. Baldini, A. Casnati, and D. Ungaro, Calixarenes: From biomimetic receptors to multivalent ligands for biomolecular recognition, *New J. Chem*., 34 (2010) 2715–2728.
88. Y. Chevolot, C. Bouillon, S. Vidal, F. Morvan, A. Meyer, J. P. Cloarec, A. Jochum, J. P. Praly, J. J. Vasseur, and E. Souteyrand, DNA-based carbohydrate biochips: A platform for surface glyco-engineering, *Angew. Chem. Int. Ed*., 46 (2007) 2398–2402.
89. T. Darbre and J. L. Reymond, Glycopeptide dendrimers for biomedical applications, *Curr. Top. Med. Chem*., 8 (2008) 1286–1293.
90. M. Marradi, M. Martin-Lomas, and S. Penades, Glyconanoparticles polyvalent tools to study carbohydrate-based interactions, *Adv. Carbohydr. Chem. Biochem*., 64 (2010) 211–290.
91. M. Reynolds, M. Marradi, A. Imberty, S. Penades, and S. Pérez, Multivalent gold glycoclusters: High affinity molecular recognition by bacterial lectin PA-IL, *Chemistry*, 18 (2012) 4264–4273.
92. O. Renaudet, Recent advances on cyclopeptide-based glycoclusters, *Mini-Rev. Org. Chem*., 5 (2008) 274–286.
93. V. Spiwok, P. Lipovova, T. Skalova, E. Buchtelova, J. Hasek, and B. Kralova, Role of CH/pi interactions in substrate binding by *Escherichia coli* beta-galactosidase, *Carbohydr. Res*., 339 (2004) 2275–2280.

94. M. Wimmerova, S. Kozmon, I. Necasova, S. K. Mishra, J. Komarek, and J. Koca, Stacking interactions between carbohydrate and protein quantified by combination of theoretical and experimental methods, *PLoS One*, 7 (2012) e46032.
95. A. Poveda, J. L. Asensio, J. F. Espinosa, M. Martin-Pastor, J. Cañada, and J. Jiménez Barbero, Applications of nuclear magnetic resonance spectroscopy and molecular modeling to the study of protein-carbohydrate interactions, *J. Mol. Graph. Model.*, 15 (1997) 9–17.
96. M. C. Chervenak and E. J. Toone, A direct measure of the contribution of solvent reorganization to the enthalpy of binding, *J. Am. Chem. Soc.*, 116 (1994) 10533–10539.
97. S. M. Tschampel and R. J. Woods, Quantifying the role of water in protein-carbohydrate interactions, *J. Phys. Chem.*, 107 (2003) 9175–9181.
98. E. J. Sorin and V. S. Pande, Empirical force-field assessment: The interplay between backbone torsions and noncovalent term scaling, *J. Comput. Chem.*, 26 (2005) 682–690.
99. M. K. Dowd, A. D. French, and P. J. Reilly, Modeling of aldopyranosyl ring puckering with MM3 (92), *Carbohydr. Res.*, 264 (1994) 1–19.
100. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley, and R. J. Woods, GLYCAM06: A generalizable biomolecular force field. Carbohydrates, *J. Comput. Chem.*, 29 (2008) 622–655.
101. F. A. Momany and J. L. Willett, Computational studies on carbohydrates: *In vacuo* studies using a revised AMBER force field, AMB99C, designed for α-(1→4) linkages, *Carbohydr. Res.*, 326 (2000) 194–209.
102. T. M. Glennon and K. M. Merz, Jr., A carbohydrate force field for AMBER and its application to the study of saccharide to surface adsorption, *J. Mol. Struct. (THEOCHEM)*, 395–396 (1997) 157–171.
103. R. J. Woods, R. A. Dwek, C. J. Edge, and B. Fraser-Reid, Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM-93 parameter development, *J. Phys. Chem.*, 99 (1995) 3832–3846.
104. S. W. Homans, A molecular mechanical force field for the conformational analysis of oligosaccharides: Comparison of theoretical and crystal structures of Man.alpha.1-3Man.beta.1-4GlcNAc, *Biochemistry*, 29 (1990) 9110–9118.
105. W. Tempel, S. Tschampel, and R. J. Woods, The xenograft antigen bound to *Griffonia simplicifolia* lectin 1-B4, *J. Biol. Chem.*, 277 (2002) 6615–6621.
106. G. M. Bradbrook, J. R. Forshaw, and S. Pérez, Structure/thermodynamics relationships of lectin–saccharide complexes, *Eur. J. Biochem.*, 267 (2000) 4545–4555.
107. C. Clarke, R. J. Woods, J. Gluska, A. Cooper, M. A. Nutley, and G.-J. Boons, Involvement of water in carbohydrate-protein binding, *J. Am. Chem. Soc.*, 123 (2001) 12238–12247.
108. R. A. Bryce, I. H. Hillier, and J. H. Naismith, Carbohydrate-protein recognition: Molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A, *Biophys. J.*, 81 (2001) 1373–1388.
109. A. Pathiaseril and R. J. Woods, Relative energies of binding for antibody-carbohydrate-antigen complexes computed from free-energy simulations, *J. Am. Chem. Soc.*, 122 (2000) 331–338.
110. G. Liang, R. K. Schmidt, H. A. Yu, D. A. Cumming, and J. W. Brady, Free energy simulation studies of the binding specificity of mannose-binding protein, *J. Phys. Chem. B*, 100 (1996) 2528–2534.
111. K. Ramirez-Gualito, R. Alonso-Rios, B. Quiroz-Garcia, A. Rojas-Aguilar, D. Diaz, J. Jiménez Barbero, and G. Cuevas, Enthalpic nature of the CH/pi interaction involved in the recognition of carbohydrates by aromatic compounds, confirmed by a novel interplay of NMR, calorimetry, and theoretical calculations, *J. Am. Chem. Soc.*, 131 (2009) 18129–18138.
112. Z. R. Laughrey, S. E. Kiehna, A. J. Riemen, and M. L. Waters, Carbohydrate-pi interactions: What are they worth? *J. Am. Chem. Soc.*, 130 (2008) 14625–14633.

113. V. Spiwok, P. Lipovová, T. Skálová, E. Vondráčková, J. Dohnálek, J. Hašek, and B. Králová, Modelling of carbohydrate–aromatic interactions: *Ab initio* energetics and force field performance, *J. Comput. Aided Mol. Des.*, 19 (2005) 887–901.
114. S. Vandenbussche, D. Díaz, M. C. Fernández-Alonso, W. Pan, S. P. Vincent, G. Cuevas, F. J. Cañada, J. Jiménez Barbero, and K. Bartik, Aromatic–carbohydrate interactions: An NMR and computational study of model systems, *Chem. Eur. J.*, 14 (2008) 7570–7578.
115. M. B. Tessier, M. L. DeMarco, A. B. Yongye, and R. J. Woods, Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids, *Mol. Simul.*, 34 (2008) 349–364.
116. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.*, 117 (1995) 5179–5197.
117. K. N. Kirschner and R. J. Woods, Solvent interactions determine carbohydrate conformation, *Proc. Natl. Acad. Sci. U.S.A.*, 98 (2001) 10541–10545.
118. X. Biarnes, A. Ardevol, A. Planas, and C. Rovira, Substrate conformational changes in glycoside hydrolase catalysis. A first-principles molecular dynamics study, *Biocatal. Biotransfor.*, 28 (2010) 33–40.
119. M. Basma, S. Sundara, D. Calgan, T. Vernali, and R. J. Woods, Solvated ensemble averaging in the calculation of partial atomic charges, *J. Comput. Chem.*, 22 (2001) 1125–1137.
120. R. J. Woods, M. Khalil, W. Pell, S. H. Moffat, and V. H. Smith, Derivation of net atomic charges from molecular electrostatic potentials, *J. Comput. Chem.*, 11 (1990) 297–310.
121. R. J. Woods and R. Chappelle, Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates, *J. Mol. Struct. (THEOCHEM)*, 527 (2000) 149–156.
122. S. Pérez, A. Imberty, S. B. Engelsen, J. Gruza, K. Mazeau, J. Jiménez Barbero, A. Poveda, J.-F. Espinosa, B. P. van Eyck, G. Johnson, A. D. French, M. L. C. E. Kouwijzer, P. D. J. Grootenuis, A. Bernardi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten, and K. Rasmussen, A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates, *Carbohydr. Res.*, 314 (1998) 141–155.
123. R. D. Lins and P. H. Hunenberger, A new GROMOS force field for hexopyranose-based carbohydrates, *J. Comput. Chem.*, 26 (2005) 1400–1412.
124. L. D. Schuler, X. Daura, and W. F. van Gunsteren, An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase, *J. Comput. Chem.*, 22 (2001) 1205–1218.
125. L. D. Schuler and W. F. Van Gunsteren, On the choice of dihedral angle potential energy functions for n-alkanes, *Mol. Simul.*, 25 (2000) 301–319.
126. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, Interaction models for water in relation to protein hydration, In: B. Pullman, (Ed.), *Intermolecular Forces,* Reidel, Dordrecht, 1981, pp. 331–342.
127. N. S. Gandhi and R. L. Mancera, Free energy calculations of glycosaminoglycan-protein interactions, *Glycobiology*, 19 (2009) 1103–1115.
128. O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor, and A. D. Mackerell, Jr., Additive empirical force field for hexopyranose monosaccharides, *J. Comput. Chem.*, 29 (2008) 2543–2564.
129. O. Guvench, E. R. Hatcher, R. M. Venable, R. W. Pastor, and A. D. Mackerell, CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses, *J. Chem. Theory Comput.*, 5 (2009) 2353–2370.
130. E. Hatcher, O. Guvench, and A. D. Mackerell, CHARMM additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates and inositol, *J. Chem. Theory Comput.*, 5 (2009) 1315–1327.
131. A. D. MacKerell, N. Banavali, and N. Foloppe, Development and current status of the CHARMM force field for nucleic acids, *Biopolymers*, 56 (2000) 257–265.

132. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B*, 102 (1998) 3586–3616.
133. D. Kony, W. Damm, S. Stoll, and W. F. van Gunsteren, An improved OPLS-AA force field for carbohydrates, *J. Comput. Chem.*, 23 (2002) 1416–1429.
134. W. Damm, A. Frontera, J. Tirado–Rives, and W. L. Jorgensen, OPLS all-atom force field for carbohydrates, *J. Comput. Chem.*, 18 (1997) 1955–1970.
135. W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.*, 118 (1996) 11225–11236.
136. N. L. Allinger, F. Li, L. Yan, and J. C. Tai, Molecular mechanics (MM3) calculations on conjugated hydrocarbons, *J. Comput. Chem.*, 11 (1990) 868–895.
137. J.-H. Lii and N. L. Allinger, The MM3 force field for amides, polypeptides and proteins, *J. Comput. Chem.*, 12 (1991) 186–199.
138. M. Clark, R. D. Cramer, and N. Van Opdenbosch, Validation of the general purpose Tripos 5.2 force field, *J. Comput. Chem.*, 10 (1989) 982–1012.
139. A. Imberty, K. D. Hardman, J. P. Carver, and S. Pérez, Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A, *Glycobiology*, 1 (1991) 631–642.
140. S. Pérez, C. Meyer, and A. Imberty, Practical tools for molecular modeling of complex carbohydrates and their interactions with proteins, *Mol. Eng.*, 5 (1995) 271–300.
141. TRIPOS, SYBYL-X 1.3, Tripos, Tripos International, St. Louis, Missouri, USA, 1991–2011.
142. P. J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, *J. Med. Chem.*, 28 (1985) 849–857.
143. A. Bohne, E. Lang, and C. W. von der Lieth, SWEET—WWW based rapid 3D construction of oligo and polysaccharides, *Bioinformatics*, 15 (1999) 767–768.
144. J. Rosen, L. Miguet, and S. Pérez, Shape: Automatic conformation prediction of carbohydrates using a genetic algorithm, *J. Cheminform.*, 1 (2010) 16.
145. C. J. Cramer, Essentials of Computational Chemistry: Theories and Models, Wiley, Chichester, 2004.
146. D. Young, A Practical Guide for Applying Techniques to Real World Problems, Wiley, New York, 2001.
147. A. R. Leach, Molecular Modelling: Principles and Applications, 2nd ed. Prentice Hall, Harlow, England; New York, 2001.
148. I. N. Levine, Quantum Chemistry, 6th ed. Pearson Prentice Hall, Upper Saddle River, NJ, 2009.
149. I. Tvaroška and T. Bleha, Molecular orbital studies of the conformations of dimethoxyethane, *J. Mol. Struct.*, 24 (1975) 249–259.
150. I. Tvaroška and J. P. Carver, Theoretical studies on the conformation of saccharides. 15. AM1 calculation of relative stabilities and geometries of conformers, *J. Chem. Res. Synop.* (1991) 6–7.
151. V. Spiwok and A. D. French, Modelling the effect of solvents on carbohydrates, *Mini-Rev. Org. Chem.*, 8 (2011) 249–255.
152. J. P. McNamara, A. Muslim, H. Abdel-Aal, H. Wang, M. Mohr, I. H. Hillier, and R. A. Bryce, Towards a quantum mechanical force field for carbohydrate reparametrized semi-empirical MO approach, *Chem. Phys. Lett.*, 394 (2004) 429–436.
153. Y. Zhong, B. A. Bauer, and S. Patel, Solvation properties of N-acetyl-β-glucosamine: Molecular dynamics study incorporating electrostatic polarization, *J. Comput. Chem.*, 32 (2011) 3339–3353.

154. A. D. French, A. M. Kelterer, G. P. Johnson, M. K. Dowd, and C. J. Cramer, HF/6-31G*energy surfaces for disaccharide analogs, *J. Comput. Chem.*, 22 (2001) 65–78.
155. S. E. Barrows, F. J. Dulles, C. J. Cramer, A. D. French, and D. G. Truhlar, Relative stability of alternative chair forms and hydroxymethyl conformations of β-D-glucopyranose, *Carbohydr. Res.*, 276 (1995) 219–251.
156. A. D. French, Combining computational chemistry and crystallography for a better understanding of the structure of cellulose, *Adv. Carbohydr. Chem. Biochem.*, 67 (2012) 19–93.
157. R. R. Andrade and C. O. da Silva, On the route of the determination of monosaccharides conformations, *Mini-Rev. Org. Chem.*, 8 (2011) 239–248.
158. A. D. French and G. P. Johnson, Quantum mechanics studies of cellobiose conformations, *Can. J. Chem.*, 84 (2006) 603–612.
159. A. D. French and G. P. Johnson, Role of starting geometries in quantum mechanics studies of cellobiose, *Mol. Simul.*, 34 (2008) 365–372.
160. B. J. Smith, A conformational study of 2-oxanol: Insight into the role of ring distortion on enzyme-catalyzed glycosidic bond cleavage, *J. Am. Chem. Soc.*, 119 (1997) 2699–2706.
161. B. J. Smith, P. M. Colman, M. Von Itzstein, B. Danylec, and J. N. Varghese, Analysis of inhibitor binding in influenza virus neuraminidase, *Protein Sci.*, 10 (2001) 689–696.
162. R. G. Parr and W. Yang, Density-Functional Theory of Atoms and Molecules, Oxford University Press/Clarendon Press, New York; Oxford, England, 1989.
163. G. E. Scuseria and V. N. Staroverov, Progress in the development of exchange-correlation functionals, In: C. E. Dykstra, G. Frenking, K. S. Kim, G. E. Scuseria, (Eds.), *Theory and Applications of Computational Chemistry: The First Forty Years,* Elsevier, Amsterdam, 2005, pp. 669–724.
164. Y. Zhao, N. E. Schultz, and D. G. Truhlar, Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions, *J. Chem. Theory Comput.*, 2 (2006) 364–382.
165. Y. Zhao and D. G. Truhlar, The M06 suite of functionals for main density group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.*, 120 (2008) 215–241.
166. Y. Zhao and D. G. Truhlar, Density functionals with broad applicability in chemistry, *Acc. Chem. Res.*, 41 (2008) 157–167.
167. A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.*, 98 (1993) 5648–5652.
168. C. T. Lee, W. T. Yang, and R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density, *Phys. Rev. B*, 37 (1988) 785–789.
169. M. Appell, G. Strati, J. L. Willett, and F. A. Momany, B3LYP/6-311++G study of α- and β-D-glucopyranose and 1,5-anhydro-D-glucitol: $^{4}C_{1}$ and $^{1}C_{4}$ chairs, $^{3,O}B$ and $B_{3,O}$ boats, and skew-boat conformations, *Carbohydr. Res.*, 339 (2004) 537–551.
170. M. Appell, J. L. Willett, and F. A. Momany, DFT study of α- and β-D-mannopyranose at the B3LYP/6-311++G** level, *Carbohydr. Res.*, 340 (2005) 459–468.
171. F. Momany, U. Schnupf, and J. L. Willett, DFT-MD studies of glucose and epimers: Anomeric ratios, rotamer populations, and hydration energies, *Carbohydr. Res.*, 345 (2010) 503–511.
172. F. A. Momany, M. Appell, G. Strati, and J. L. Willett, B3LYP/6-311++G study of monohydrates of α- and β-D-glucopyranose: Hydrogen bonding, stress energies, and effect of hydration on internal coordinates, *Carbohydr. Res.*, 339 (2004) 553–567.
173. F. A. Momany, M. Appell, J. L. Willett, and W. B. Bosma, B3LYP/6-311++G** geometry-optimization study of pentahydrates of α- and β-D-glucopyranose, *Carbohydr. Res.*, 340 (2005) 1638–1655.
174. F. A. Momany, M. Appell, J. L. Willett, U. Schnupf, and W. B. Bosma, DFT study of alpha- and beta-D-galactopyranose at the B3LYP/6-311++G** level of theory, *Carbohydr. Res.*, 341 (2006) 525–537.

175. U. Schnupf, J. L. Willett, W. B. Bosma, and F. A. Momany, DFT study of α and β-D-allopyranose at the B3LYP/6-311++G** level of theory, *Carbohydr. Res.*, 342 (2007) 196–216.
176. G. L. Strati, J. L. Willet, and F. A. Momany, *Ab initio* computational study of β-cellobiose conformers using B3LYP/6-311++G, *Carbohydr. Res.*, 337 (2002) 1833–1849.
177. A. D. French, G. P. Johnson, C. J. Cramer, and G. I. Csonka, Conformational analysis of cellobiose by electronic structure theories, *Carbohydr. Res.*, 350 (2012) 68–76.
178. D. A. Navarro and C. A. Stortz, DFT/MM modeling of the five-membered ring in 3,6-anhydrogalactose derivatives and its influence on disaccharide adiabatic maps, *Carbohydr. Res.*, 343 (2008) 2292–2298.
179. C. A. Stortz, Conformational pathways of simple six-membered rings, *J. Phys. Org. Chem.*, 23 (2010) 1173–1186.
180. M. Hricovini, B3LYP/6-311++G** study of structure and spin-spin coupling constant in methyl 2-*O*-sulfo-α-L-iduronate, *Carbohydr. Res.*, 341 (2006) 2575–2580.
181. M. Hricovini, E. Scholtzova, and F. Bizik, B3LYP/6-31 1++G** study of structure and spin-spin coupling constant in heparin disaccharide, *Carbohydr. Res.*, 342 (2007) 1350–1356.
182. F. Momany and U. Schnupf, DFT optimization and DFT/MD studies of glucose, ten explicit water molecules enclosed by implicit solvent, COSMO, *Comput. Theor. Chem.*, 1029 (2014) 57–67.
183. G. I. Csonka, Proper basis set for quantum mechanical studies of potential energy surfaces of carbohydrates, *J. Mol. Struct. (THEOCHEM)*, 584 (2002) 1–4.
184. G. I. Csonka and A. D. French, Evaluation of density functionals and basis sets for carbohydrates, *J. Chem. Theory Comput.*, 5 (2009) 679–692.
185. A. Berces, G. Enright, T. Nukada, and M. Whitfield, The conformational origin of the barrier to the formation of neighboring group assistance in glycosylation reactions: A dynamical density functional theory study, *J. Am. Chem. Soc.*, 123 (2001) 5460–5464.
186. T. Nukada, A. Berces, and D. M. Whitfield, Can the stereochemical outcome of glycosylation reactions be controlled by the conformational preferences of the glycosyl donor? *Carbohydr. Res.*, 337 (2002) 765–774.
187. D. M. Whitfield and T. Nukada, DFT studies of the role of C-2-O-2 bond rotation in neighboring-group glycosylation reactions, *Carbohydr. Res.*, 342 (2007) 1291–1304.
188. D. M. Whitfield, Complications of modeling glycosylation reactions: Can the anomeric conformation of a donor determine the glycopyranosyl oxacarbenium ring conformation? *Carbohydr. Res.*, 356 (2012) 191–195.
189. D. M. Whitfield, Plausible transition states for glycosylation reactions, *Carbohydr. Res.*, 356 (2012) 180–190.
190. D. M. Whitfield, DFT studies of the ionization of α and β glycopyranosyl donors, *Carbohydr. Res.*, 342 (2007) 1726–1740.
191. A. R. Ionescu, D. M. Whitfield, and M. Z. ZgIerski, O-2 Substituted pyranosyl oxacarbenium ions are C-2-O-2 2-fold rotors with a strong *syn* preference, *Carbohydr. Res.*, 342 (2007) 2793–2800.
192. A. R. Ionescu, D. M. Whitfield, M. Z. Zgierski, and T. Nukada, Investigations into the role of oxacarbenium ions in glycosylation reactions by *ab initio* molecular dynamics, *Carbohydr. Res.*, 341 (2006) 2912–2920.
193. T. Nukada, A. Berces, M. Z. Zgierski, and D. M. Whitfield, Exploring the mechanism of neighboring group assisted glycosilation reactions, *J. Am. Chem. Soc.*, 120 (1998) 13291–13295.
194. Y. Zhao, N. Gonzalez-Garcia, and D. G. Truhlar, Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods, *J. Phys. Chem.*, 109 (2005) 2012–2018.
195. X. F. Xu, I. M. Alecu, and D. G. Truhlar, How well can modern density functionals predict internuclear distances at transition states? *J. Chem. Theory Comput.*, 7 (2011) 1667–1676.

196. Y. Zhao and D. G. Truhlar, How well can new-generation density functional methods describe stacking interactions in biological systems? *Phys. Chem. Chem. Phys.*, 7 (2005) 2701–2705.
197. Y. Zhao and D. G. Truhlar, Benchmark databases for nonbonded interactions and their use to test density functional theory, *J. Chem. Theory Comput.*, 1 (2005) 415–432.
198. S. Grimme, Density functional theory with London dispersion corrections, *WIREs Comput. Mol. Sci.*, 1 (2011) 211–228.
199. S. T. Schneebeli, A. D. Bochevarov, and R. A. Friesner, Parameterization of a B3LYP specific correction for noncovalent interactions and basis set superposition error on a gigantic data set of CCSD(T) quality noncovalent interaction energies, *J. Chem. Theory Comput.*, 7 (2011) 658–668.
200. D. G. Truhlar and Y. Zhao, Exploring the limit of accuracy of the global hybrid meta density functional for main-group thermochemistry, kinetics, and noncovalent interactions, *J. Chem. Theory. Comput.*, 4 (2008) 1849–1868.
201. D. G. Truhlar and Y. Zhao, Density functionals with broad applicability in chemistry, *Acc. Chem. Res.*, 41 (2008) 157–167.
202. P. Hobza, J. Sponer, and T. Reschel, Density-functional theory and molecular clusters, *J. Comput. Chem.*, 16 (1995) 1315–1325.
203. S. Kozmon, R. Matuska, V. Spiwok, and J. Koca, Three-dimensional potential energy surface of selected carbohydrates CH/pi dispersion interactions calculated by high-level quantum mechanical methods, *Chem. Eur. J.*, 17 (2011) 5680–5690.
204. S. Kozmon, R. Matuska, V. Spiwok, and J. Koca, Dispersion interactions of carbohydrates with condensate aromatic moieties: Theoretical study on the CH/pi interaction additive properties, *Phys. Chem. Chem. Phys.*, 13 (2011) 14215–14222.
205. B. J. Lynch, P. L. Fast, M. Harris, and D. G. Truhlar, Adiabatic connection for kinetics, *J. Phys. Chem. A*, 104 (2000) 4811–4815.
206. Y. Zhao and D. G. Truhlar, Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions, *J. Phys. Chem. A*, 109 (2005) 5656–5667.
207. J. Kona and I. Tvaroška, Comparative DFT study on the alpha-glycosidic bond in reactive species of galactosyl diphosphates, *Chem. Pap.*, 63 (2009) 598–607.
208. L. Simon and J. M. Goodman, How reliable are DFT transition structures? Comparison of GGA, hybrid-meta-GGA and meta-GGA functionals, *Org. Biomol. Chem.*, 9 (2011) 689–700.
209. P. E. M. Siegbahn and F. Himo, The quantum chemical cluster approach for modeling enzyme reactions, *WIREs Comput. Mol. Sci.*, 1 (2011) 323–336.
210. L. H. Hu, P. Soderhjelm, and U. Ryde, On the convergence of QM/MM energies, *J. Chem. Theory Comput.*, 7 (2011) 761–777.
211. C. Sumowski, C. Vanessa, and C. Ochsenfeld, A convergence study of QM/MM isomerization energies with the selected size of the QM region for peptidic systems, *J. Phys. Chem. A*, 113 (2009) 11734–11741.
212. L. H. Hu, J. Eliasson, J. Heimdal, and U. Ryde, Do quantum mechanical energies calculated for small models of protein-active sites converge? *J. Phys. Chem. A*, 113 (2009) 11793–11800.
213. A. Warshel and M. Levitt, Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *J. Mol. Biol.*, 103 (1976) 227–249.
214. I. Solt, P. Kulhanek, I. Simon, S. Winfield, M. C. Payne, G. Csanyi, and M. Fuxreiter, Evaluating boundary dependent errors in QM/MM simulations, *J. Phys. Chem. B*, 113 (2009) 5728–5735.
215. H. M. Senn and W. Thiel, QM/MM studies of enzymes, *Curr. Opin. Chem. Biol.*, 11 (2007) 182–187.
216. H. M. Senn and W. Thiel, QM/MM methods for biological systems, In: M. Reiher, (Ed.), *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations, Topics in Current Chemistry*, Vol. 268, 2007, pp. 173–290.
217. H. M. Senn and W. Thiel, QM/MM methods for biomolecular systems, *Angew. Chem. Int. Ed.*, 48 (2009) 1198–1229.

218. O. Acevedo and W. L. Jorgensen, Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions, *Acc. Chem. Res.*, 43 (2010) 142–151.
219. K. E. Ranaghan and A. J. Mulholland, Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods, *Int. Rev. Phys. Chem.*, 29 (2010) 65–133.
220. H. Lin and D. G. Truhlar, QM/MM: What have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.*, 117 (2007) 185–199.
221. S. C. Kamerlin, M. Haranczyk, and A. Warshel, Progress in *ab initio* QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pKa, redox reactions and solvation free energies, *J. Phys. Chem. B*, 113 (2009) 1253–1272.
222. Y. Zhong, B. A. Bauer, and S. Patel, Solvation properties of N-Acetyl-β-glucosamine: Molecular dynamics study incorporating electrostatic polarization, *J. Comput. Chem.*, 32 (2011) 3339–3353.
223. U. C. Singh and P. A. Kollman, A combined *ab initio* quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular systems—Applications to the $CH_3Cl + Cl^-$ exchange reaction and gas phase protonation of polyethers, *J. Comput. Chem.*, 7 (1986) 718–730.
224. M. J. Field, P. A. Bash, and M. Karplus, A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations, *J. Comput. Chem.*, 11 (1990) 700–733.
225. R. B. Murphy, D. M. Philipp, and R. A. Friesner, A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments, *J. Comput. Chem.*, 21 (2000) 1442–1457.
226. R. B. Murphy, D. M. Philipp, and R. A. Friesner, Frozen orbital QM/MM methods for density functional theory, *Chem. Phys. Lett.*, 321 (2000) 113–120.
227. D. M. Philipp and R. A. Friesner, Mixed *ab initio* QM/MM Modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide, *J. Comput. Chem.*, 20 (1999) 1468–1494.
228. A. E. Torda, R. M. Scheek, and W. F. van Gunsteren, Time-dependent distance restraints in molecular dynamics simulations, *Chem. Phys. Lett.*, 157 (1989) 289–294.
229. A. E. Torda, R. M. Scheek, and W. F. van Gunsteren, Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat, *J. Mol. Biol.*, 214 (1990) 223–235.
230. A. E. Torda, R. M. Brunne, T. Huber, H. Kessler, and W. F. van Gunsteren, Structure refinement using time-averaged J-coupling constant restraints, *J. Biomol. NMR*, 3 (1993) 55–66.
231. A. P. Nanzer, W. F. van Gunsteren, and A. E. Torda, Parametrization of time-averaged distance restraints in MD simulations, *J. Biomol. NMR*, 6 (1995) 313–320.
232. D. A. Pearlman and P. A. Kollman, Are time-averaged restraints necessary for nuclear-magnetic-resonance refinement—A model study for DNA, *J. Mol. Biol.*, 220 (1991) 457–479.
233. D. A. Pearlman, How well do time-averaged J-coupling restraints work? *J. Biomol. NMR*, 4 (1994) 279–299.
234. P. M. S. Hendrickx, F. Corzana, S. Depraetere, D. A. Tourwe, K. Augustyns, and J. C. Martins, The use of time-averaged $^3J_{HH}$ restrained molecular dynamics (tar-MD) simulations for the conformational analysis of five-membered ring systems: Methodology and applications, *J. Comput. Chem.*, 31 (2010) 561–572.
235. Z. Q. Zhang, S. A. McCallum, J. Xie, L. Nieto, F. Corzana, J. Jiménez Barbero, M. Chen, J. Liu, and R. J. Linhardtt, Solution structures of chemoenzymatically synthesized heparin and its precursors, *J. Am. Chem. Soc.*, 130 (2008) 12998–13007.
236. M. K. Gilson and B. Honig, Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis, *Proteins*, 4 (1988) 7–18.
237. D. Sitkoff, K. A. Sharp, and B. Honig, Accurate calculation of hydration free energies using macroscopic solvent models, *J. Phys. Chem. B*, 98 (1994) 1978–1988.

238. N. Basdevant, H. Weinstein, and M. Ceruso, Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study, *J. Am. Chem. Soc*., 128 (2006) 12766–12777.
239. J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case, Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate DNA helices, *J. Am. Chem. Soc*., 120 (1998) 9401–9409.
240. H. Gohlke and D. A. Case, Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex RAS–RAF, *J. Comput. Chem*., 25 (2004) 238–250.
241. P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Acc. Chem. Res*., 33 (2000) 889–897.
242. M. L. DeMarco and R. J. Woods, Structural glycobiology: A game of snakes and ladders, *Glycobiology*, 18 (2008) 426–440.
243. C. Chipot, and A. Pohorille, (Eds.), In: Free Energy Calculations: Theory and Applications in Chemistry and Biology, Springer, Berlin/Heidelberg, 2007.
244. C. Dellago and P. G. Bolhuis, Transition path sampling and other advances simulation techniques for rare events, *Adv. Polym. Sci*., 221 (2009) 167–233.
245. A. Laio, J. VandeVondele, and U. Rothlisberger, A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations, *J. Chem. Phys*., 116 (2002) 6941–6947.
246. A. Laio and F. L. Gervasio, Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics and material sciences, *Rep. Prog. Phys*., 71 (2008) 126601.
247. A. Barducci, M. Bonomi, and M. Parrinello, Metadynamics, *WIREs Comput. Mol. Sci*., 1 (2011) 826–843.
248. B. Ensing, M. De Vivo, Z. Liu, P. Moore, and M. L. Klein, Metadynamics as a tool for exploring free energy landscapes of chemical reactions, *Acc. Chem. Res*., 39 (2006) 73–81.
249. L. Sutto, S. Marsili, and G. F. Luigi, New advances in metadynamics, *WIREs Comput. Mol. Sci*., 2 (2012) 771–779.
250. M. Iannuzzi, A. Laio, and M. Parrinello, Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics, *Phys. Rev. Lett*., 90 (2003) 238302.
251. A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello, Assessing the accuracy of metadynamics, *J. Phys. Chem. B*, 109 (2005) 6714–6721.
252. F. L. Gervasio, A. Laio, and M. Parrinello, Flexible docking in solution using metadynamics, *J. Am. Chem. Soc*., 127 (2005) 2600–2607.
253. X. Biarnes, A. Ardevol, A. Planas, C. Rovira, A. Laio, and M. Parrinello, The conformational free energy landscape of beta-D-glucopyranose. Implications for substrate preactivation in beta-glucoside hydrolases, *J. Am. Chem. Soc*., 129 (2007) 10686–10693.
254. V. Spiwok, B. Kralova, and I. Tvaroška, Modelling of β-D-glucopyranose ring distortion in different force fields: A metadynamics study, *Carbohydr. Res*., 345 (2010) 530–537.
255. R. Car and M. Parrinello, Unified approach for molecular-dynamics and density-functional theory, *Phys. Rev. Lett*., 55 (1985) 2471–2474.
256. V. Spiwok and I. Tvaroška, Metadynamics modelling of the solvent effect on primary hydroxyl rotamer equilibria in hexopyranosides, *Carbohydr. Res*., 344 (2009) 1575–1581.
257. V. Spiwok and I. Tvaroška, Conformational free energy surface of alpha-N-acetylneuraminic acid: An interplay between hydrogen bonding and solvation, *J. Phys. Chem. B*, 113 (2009) 9589–9594.
258. P. Oborsky, I. Tvaroška, B. Kralova, and V. Spiwok, Toward an accurate conformational modeling of iduronic acid, *J. Phys. Chem. B*, 117 (2013) 1003–1009.

259. X. Biarnes, J. Nieto, A. Planas, and C. Rovira, Substrate distortion in the Michaelis complex of Bacillus 1,3-1,4-beta-glucanase—Insight from first principles molecular dynamics simulations, *J. Biol. Chem.*, 281 (2006) 1432–1441.
260. A. L. van Bueren, A. Ardevol, J. Fayers-Kerr, B. Luo, Y. M. Zhang, M. Sollogoub, Y. Bleriot, C. Rovira, and G. J. Davies, Analysis of the reaction coordinate of alpha-L-fucosidases: A combined structural and quantum mechanical approach, *J. Am. Chem. Soc.*, 132 (2010) 1804–1806.
261. A. J. Thompson, J. Dabin, J. Iglesias-Fernandez, A. Ardevol, Z. Dinev, S. J. Williams, O. Bande, A. Siriwardena, C. Moreland, T. C. Hu, D. K. Smith, H. J. Gilbert, C. Rovira, and G. J. Davies, The reaction coordinate of a bacterial GH47 alpha-mannosidase: A combined quantum mechanical and structural approach, *Angew. Chem. Int. Ed.*, 51 (2012) 10997–11001.
262. A. Ardevol, X. Biarnes, A. Planas, and C. Rovira, The conformational free-energy landscape of beta-D-mannopyranose: Evidence for a S-1(5) ->B-2, B-5 ->S-O(2) catalytic itinerary in beta-mannosidases, *J. Am. Chem. Soc.*, 132 (2010) 16058–16065.
263. I. J. Barker, L. Petersen, and P. J. Reilly, Mechanism of xylobiose hydrolysis by GH43 beta-xylosidase, *J. Phys. Chem. B*, 114 (2010) 15389–15393.
264. A. Ardevol and C. Rovira, The molecular mechanism of enzymatic glycosyl transfer with retention of configuration: Evidence for a short-lived oxocarbenium-like species, *Angew. Chem. Int. Ed.*, 50 (2011) 10897–10901.
265. S. Kirillova, J. Cortes, A. Stefaniu, and T. Simeon, An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins, *Proteins*, 70 (2008) 131–143.
266. M.M.P.L. Mouawad, C. Breton, A. Imberty, A. Thomas, The conformational transition mechanism of ABO histo-blood group glycosyltransferases: Insights from normal modes analysis, personal communication.
267. N. Haspel, M. Moll, M. Baker, W. Chiu, and L. Kavraki, Tracing conformational changes in proteins, *BMC Struct. Biol.*, 10 (2010) S1.
268. J. Cortes, S. Barbe, M. Erard, and T. Simeon, Encoding molecular motions in voxel maps, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 8 (2011) 557–563.
269. S. Barbe, J. Cortes, T. Simeon, P. Monsan, M. Renaud-Simeon, and I. Andre, A mixed molecular modeling-robotics approach to investigate lipase large molecular motions, *Proteins*, 79 (2011) 2517–2529.
270. J. Cortés, D. Thanh Le, R. Iehl, and T. Simeon, Simulating ligand-induced conformational changes in proteins using a mechanical disassembly method, *Phys. Chem. Chem. Phys.*, 12 (2010) 8268–8276.
271. J. Cortes, L. Jaillet, and T. Simeon, Disassembly path planning for complex articulated objects, *IEEE Trans. Robot. Autom.*, 24 (2), (2008) 475–481.
272. V. Lafaquiere, S. Barbe, S. Puech-Guenot, D. Guieysse, J. Cortes, P. Monsan, T. Simeon, I. Andre, and M. Remaud-Simeon, Control of lipase enantioselectivity by engineering the substrate binding site and access channel (ChemBioChem 17/2009), *Chem. Biol. Chem.*, 10 (2009) 2677.
273. D. Guieysse, J. Cortés, S. Puech-Guenot, S. Barbe, V. Lafaquière, P. Monsan, T. Simeon, I. Andre, and M. Remaud-Siméon, A structure-controlled investigation of lipase enantioselectivity by a path-planning approach, *Chem. Biol. Chem.*, 9 (2008) 1308–1317.
274. I. Al-Bluwi, M. Vaisset, T. Simeon, and J. Cortes, Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods, *BMC Struct. Biol.*, 13 (2013) S2.
275. T. Lengauer and M. Rarey, Computational methods for biomolecular docking, *Curr. Opin. Struct. Biol.*, 6 (1996) 402–406.
276. P. Bultinck, H. D. Winter, W. Langenaecker, and J. P. Tollenare, Computational Medicinal Chemistry for Drug Discovery, Marcel Dekker, Inc., New York, 2004.
277. R.C. Walker, The development of a QM/MM based linear response method and its application to proteins, Department of Chemistry, Doctor of Philosophy, Imperial College, London, 2003, p. 274.

278. T. J. A. Ewing, S. Makino, A. G. Skillman, and I. D. Kuntz, DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases, *J. Comput. Aided Mol. Des.*, 15 (2001) 411–428.
279. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.*, 19 (1998) 1639–1662.
280. 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, and P. S. Shenkin, Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.*, 47 (2004) 1739–1749.
281. I. Muegge and Y. C. Martin, A general and fast scoring function for protein-ligand interactions: A simplified potential approach, *J. Med. Chem.*, 42 (1999) 791–804.
282. G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head, A critical assessment of docking programs and scoring functions, *J. Med. Chem.*, 49 (2006) 5912–5931.
283. A. Nurisso, B. Blanchard, A. Audfray, L. Rydner, S. Oscarson, A. Varrot, and A. Imberty, Role of water molecules in structure and energetics of *Pseudomonas aeruginosa* lectin I interacting with disaccharides, *J. Biol. Chem.*, 285 (2010) 20316–20327.
284. E. Yuriev and M. Agostino, Challenges and advances in computational docking: 2009 in review, *J. Mol. Recognit.*, 24 (2011) 149–164.
285. A. D. Hill and P. J. Reilly, A Gibbs free energy correlation for automated docking of carbohydrates, *J. Comput. Chem.*, 29 (2008) 1131–1141.
286. A. Kerzmann, D. Neumann, and O. Kohlbacher, SLICK—Scoring and energy functions for protein-carbohydrate interactions, *J. Chem. Inf. Model.*, 46 (2006) 1635–1642.
287. J. S. Taylor and R. M. Burnett, DARWIN: A program for docking flexible molecules, *Proteins*, 41 (2000) 173–191.
288. A. Kerzmann, J. Fuhrmann, O. Kohlbacher, and D. Neumann, BALLDock/SLOC: A new method for protein carbohydrate interactions, *J. Chem. Inf. Model.*, 48 (2008) 1616–1625.
289. M. Agostino, M. S. Sandrin, P. E. Thompson, E. Yuriev, and P. A. Ramsland, In silico analysis of antibody-carbohydrate interactions and its application to xenoreactive antibodies, *Mol. Immunol.*, 47 (2009) 233–246.
290. T. A. Beyer, J. E. Sadler, J. I. Rearick, J. C. Paulson, and R. L. Hill, Glycosyltransferases and their use in assessing oligosaccharide structure and structure-function relationships, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 52 (1981) 23–175.
291. H. Schachter, Enzymes associated with glycosylation, *Curr. Opin. Struct. Biol.*, 1 (1991) 755–765.
292. R. Kleene and E. G. Berger, The molecular and cell biology of glycosyltransferases, *Biochim. Biophys. Acta*, 1154 (1993) 283–325.
293. J. Montreuil, J. F. G. Vliegenthart, and H. Schachter, New comprehensive biochemistry, In: A. Neuberger and L. L. M. van Deenen, (Eds.), *Glycoproteins*, Vol. 29a, Elsevier Science B.V., Amsterdam, 1995, p. 644.
294. J. A. Campbell, G. J. Davies, V. Bulone, and B. Henrissat, A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities, *Biochem. J.*, 326 (1997) 929–942. [letter] [published erratum appears in Biochem. J. 329(Pt. 3) (1998) 719].
295. P. M. Coutinho, E. Deleury, G. J. Davies, and B. Henrissat, An evolving hierarchical family classification for glycosyltransferases, *J. Mol. Biol.*, 328 (2003) 307–317.
296. B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics, *Nucleic Acids Res.*, 37 (2009) D233–D238.

297. J. Liu and A. Mushegian, Three monophyletic superfamilies account for the majority of the known glycosyltransferases, *Protein Sci*., 12 (2003) 1418–1431.
298. M. Igura, N. Maita, J. Kamishikiryo, M. Yamada, T. Obita, K. Maenaka, and D. Kohda, Structure-guided identification of a new catalytic motif of oligosaccharyltransferase, *EMBO J*., 27 (2008) 234–243.
299. M. L. Sinnott, Catalytic mechanisms of enzymic glycosyl transfer, *Chem. Rev*., 90 (1990) 1171–1202.
300. D. L. Zechel and S. G. Withers, Glycosyl transferase mechanism, In: C. D. Poulter, (Ed.), *Comprehensive Natural Products Chemistry*, Vol. 5, Elsevier, New York, 1999, pp. 279–314.
301. G. Davies, M. L. Sinnott, and S. G. Withers, Glycosyl transfer, In: M. L. Sinott, (Ed.), *Comprehensive Biological Catalysis,* Academic Press Limited, New York, 1998, pp. 119–208.
302. D. L. Zechel and S. G. Withers, Glycosidase mechanisms: Anatomy of a finely tuned catalyst, *Acc. Chem. Res*., 33 (2000) 11–18.
303. C. Breton, J. Mucha, and C. Jeanneau, Structural and functional features of glycosyltransferases, *Biochimie*, 83 (2001) 713–718.
304. I. Tvaroška, Structural insights into the catalytic mechanism and transition state of glycosyltransferases using *ab initio* molecular modeling, *Trends Glycosci. Glycotechnol*., 17 (2005) 177–190.
305. I. Tvaroška, Molecular modeling of retaining glycosyltransferases, *ACS Symp. Ser*., 930 (2006) 285–301.
306. C. Breton, L. Snajdrova, C. Jeanneau, J. Koca, and A. Imberty, Structures and mechanisms of glycosyltransferases, *Glycobiology*, 16 (2006) 29r–37r.
307. L. L. Lairson and S. G. Withers, Mechanistic analogies amongst carbohydrate modifying enzymes, *Chem. Commun.* (2004) 2243–2248.
308. I. B. H. Wilson, C. Breton, A. Imberty, and I. Tvaroška, Molecular basis for the biosynthesis of oligo- and polysaccharides, In: B. O. T. Fraser-Reid, J. Thiem, G. L. Côté, S. Flitsch, Y. Ito, H. Kondo, S.-I. Nishimura, B. Yu, (Eds.), *Glycoscience Chemistry and Chemical Biology,* 2008, pp. 2267–2323.
309. B. Henrissat, G. Sulzenbacher, and Y. Bourne, Glycosyltransferases, glycoside hydrolases: Surprise, surprise!, *Curr. Opin. Struct. Biol*., 18 (2008) 527–533.
310. V. L. Y. Y. Ran Zhang and S. G. Withers, Mechanisms of enzymatic glycosyl transfer, In: M. Lew and L. Hung-Wen, (Eds.), *Comprehensive Natural Products II. Chemistry and Biology, Enzymes and Enzyme Mechanisms*, Vol. 8, Elsevier, 2010, pp. 385–422.
311. R. Zhang, V. L. Y. Yip, and S. G. Withers, Mechanisms of enzymatic glycosyl transfer, In: M. Lew and L. Hung-Wen, (Eds.), *Comprehensive Natural Products II,* Elsevier, Oxford, 2010, pp. 385–422.
312. I. Tvaroška, I. Andre, and J. P. Carver, *Ab initio* molecular orbital study of the catalytic mechanism of glycosyltransferases: Description of reaction pathways and determination of transition state structures for inverting N-acetylglucosaminyltransferases, *J. Am. Chem. Soc*., 122 (2000) 8762–8776.
313. I. Andre, I. Tvaroška, and J. P. Carver, On the reaction pathways and determination of transition-state structures for retaining α-galactosyltransferases, *Carbohydr. Res*., 338 (2003) 865–877.
314. I. Tvaroška, I. Andre, and J. P. Carver, Catalytic mechanism of the inverting N-acetylglucosaminyltransferase I: DFT quantum mechanical model of the reaction pathway and determination of the transition state structure, *Glycobiology*, 13 (2003) 559–566.
315. I. Tvaroška, Molecular modeling insights into the catalytic mechanism of the retaining galactosyltransferase LgtC, *Carbohydr. Res*., 339 (2004) 1007–1014.
316. U. M. Unligil and J. M. Rini, Glycosyltransferase structure and mechanism, *Curr. Opin. Struct. Biol*., 10 (2000) 510–517.
317. S. Kozmon and I. Tvaroška, Catalytic mechanism of glycosyltransferases: Hybrid quantum mechanical/molecular mechanical study of the inverting N-acetylglucosaminyltransferase I, *J. Am. Chem. Soc*., 128 (2006) 16921–16927.
318. H. Schachter, F. Reck, and H. Paulsen, Use of synthetic oligosaccharide substrate analogs to map the active sites of N-acetylglucosaminyltransferases I and II, *Methods Enzymol*., 363 (2003) 459–475.

319. B. Ramakrishnan, V. Ramasamy, and P. K. Qasba, Structural snapshots of beta-1,4-galactosyltransferase-1 along the kinetic pathway, *J. Mol. Biol.*, 357 (2006) 1619–1633.
320. B. Ramakrishnan and P. K. Qasba, Structure-based design of β1,4-galactosyltransferase I (β4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity, *J. Biol. Chem.*, 277 (2002) 20833–20839.
321. B. Ramakrishnan, E. Boeggeman, V. Ramasamy, and P. K. Qasba, Structure and catalytic cycle of beta-1,4-galactosyltransferase, *Curr. Opin. Struct. Biol.*, 14 (2004) 593–600.
322. P. K. Qasba, B. Ramakrishnan, and E. Boeggeman, Substrate-induced conformational changes in glycosyltransferases, *Trends Biochem. Sci.*, 30 (2005) 53–62.
323. B. Ramakrishnan and P. Qasba, Crystal structure of lactose synthase reveals a large conformational change in the catalytic component, the β1,4-galactosyltransferase-I, *J. Mol. Biol.*, 310 (2001) 205–218.
324. B. Ramakrishan, P. V. Balaji, and P. K. Qasba, Crystal structure of β1,4-Galactosyltransferase complex with UDP-Gal reveals an oligosaccharide acceptor binding site, *J. Mol. Biol.*, 318 (2002) 491–502.
325. M. Krupicka and I. Tvaroška, Hybrid quantum mechanical/molecular mechanical investigation of the beta-1,4-Galactosyltransferase-I mechanism, *J. Phys. Chem. B*, 113 (2009) 11314–11319.
326. C. R. Torres and G. W. Hart, Topography and polypeptide distribution of terminal *N*-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc, *J. Biol. Chem.*, 259 (1984) 3308–3317.
327. G. W. Hart, M. P. Housley, and C. Slawson, Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins, *Nature*, 446 (2007) 1017–1022.
328. G. W. Hart and R. J. Copeland, Glycomics hits the big time, *Cell*, 143 (2010) 672–676.
329. I. Tvaroška, S. Kozmon, M. Wimmerova, and J. Koca, Substrate-assisted catalytic mechanism of O-GIcNAc transferase discovered by quantum mechanics/molecular mechanics investigation, *J. Am. Chem. Soc.*, 134 (2012) 15563–15571.
330. M. B. Lazarus, Y. S. Nam, J. Y. Jiang, P. Sliz, and S. Walker, Structure of human O-GlcNAc transferase and its complex with a peptide substrate, *Nature*, 469 (2011) 564–567.
331. M. B. Lazarus, J. Y. Jiang, T. M. Gloster, W. F. Zandberg, G. E. Whitworth, D. J. Vocadlo, and S. Walker, Structural snapshots of the reaction coordinate for O-GlcNAc transferase, *Nat. Chem. Biol.*, 8 (2012) 966–968.
332. M. S. Macauley, G. E. Whitworth, A. W. Debowski, D. Chin, and D. J. Vocadlo, O-GlcNAcase uses substrate-assisted catalysis—Kinetic analysis and development of highly selective mechanism-inspired inhibitors, *J. Biol. Chem.*, 280 (2005) 25313–25322.
333. M. Schimpl, X. W. Zheng, V. S. Borodkin, D. E. Blair, A. T. Ferenbach, A. W. Schuttelkopf, I. Navratilova, T. Aristotelous, O. Albarbarawi, D. A. Robinson, M. A. Macnaughtan, and D. M. F. van Aalten, O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis, *Nat. Chem. Biol.*, 8 (2012) 969–974.
334. D. Adlercreutz, K. Mannerstedt, W. Wakarchuk, N. Dovichi, O. Hindsgaul, and M. M. Palcic, UDP-5'-thio galactose: A valuable tool for mechanistic studies on glycosyltransferases, *Glycobiology*, 20 (2010) 1517.
335. I. Tvaroška, S. Kozmon, M. Wimmerova, and J. Koca, QM/MM investigation of the catalytic mechanism of metal-ion-independent core 2 β1,6-N-acetylglucosaminyltransferase, *Chem. Eur. J.*, 19 (2013) 8153–8162.
336. A. Monegal and A. Planas, Chemical rescue of alpha 3-galactosyltransferase. Implications in the mechanism of retaining glycosyltransferases, *J. Am. Chem. Soc.*, 128 (2006) 16030–16031.
337. N. Soya, Y. Fang, M. M. Palcic, and J. S. Klassen, Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases, *Glycobiology*, 21 (2011) 547–552.

338. K. Persson, H. D. Ly, M. Dieckelmann, W. W. Wakarchuk, S. G. Withers, and N. C. J. Strynadka, Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs: Towards a mechanism, *Nat. Struct. Biol.*, 8 (2001) 166–175.
339. J. C. Errey, S. S. Lee, R. P. Gibson, C. M. Fleites, C. S. Barry, P. M. J. Jung, A. C. O'Sullivan, B. G. Davis, and G. J. Davies, Mechanistic insight into enzymatic glycosyl transfer with retention of configuration through analysis of glycomimetic inhibitors, *Angew. Chem. Int. Ed.*, 49 (2010) 1234–1237.
340. S. S. Lee, S. Y. Hong, J. C. Errey, A. Izumi, G. J. Davies, and B. G. Davis, Mechanistic evidence for a front-side, S(N)i-type reaction in a retaining glycosyltransferase, *Nat. Chem. Biol.*, 7 (2011) 631–638.
341. H. Gomez, I. Polyak, W. Thiel, J. M. Lluch, and L. Masgrau, Retaining glycosyltransferase mechanism studied by QM/MM methods: Lipopolysaccharyl-alpha-1,4-galactosyltransferase C transfers alpha-galactose via an oxocarbenium ion-like transition state, *J. Am. Chem. Soc.*, 134 (2012) 4743–4752.
342. H. Gomez, J. M. Lluch, and L. Masgrau, Essential role of glutamate 317 in galactosyl transfer by alpha 3GalT: A computational study, *Carbohydr. Res.*, 356 (2012) 204–208.
343. H. Gomez, R. Rojas, D. Patel, L. A. Tabak, J. M. Lluch, and L. Masgrau, A computational and experimental study of O-glycosylation. Catalysis by human UDP-GalNAc polypeptide:GalNAc transferase-T2, *Org. Biomol. Chem.*, 12 (2014) 2645–2655.
344. T. Trnka, S. Kozmon, I. Tvaroška, T. Bleha, J. Koca, Stepwise catalytic mechanism via short-lived intermediate inferred from combined QM/MM MERP and PES calculations on retaining glycosyltransferase ppGalNAcT2, personal communication.
345. A. Bobovska, J. Kona, and I. Tvaroška, Theoretical study on the catalytic mechanism of the retaining alfa-1,2-mannosyltransferase Kre2p/Mnt1p: The impact of different metal ions on catalysis, *Org. Biomol. Chem.*, 12 (2014) 4201–4210.
346. L. N. Gastinel, C. Bignon, A. K. Misra, O. Hindsgaul, J. H. Shaper, and D. H. Joziasse, Bovine alpha1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases, *EMBO J.*, 20 (2001) 638–649.
347. H. Jamaluddin, P. Tumbale, S. G. Withers, K. R. Acharya, and K. Brew, Conformational changes induced by binding UDP-2F-galactose to alpha-1,3-galactosyltransferase—Implications for catalysis, *J. Mol. Biol.*, 369 (2007) 1270–1281.
348. H. D. Ly, B. Lougheed, W. W. Wakarchuk, and S. G. Withers, Mechanistic studies of a retaining α-galactosyltransferase from *Neisseria meningitidis*, *Biochemistry*, 41 (2002) 5075–5085.
349. L. Snajdrova, P. Kulhanek, A. Imberty, and J. Koca, Molecular dynamics simulations of glycosyltransferase LgtC, *Carbohydr. Res.*, 339 (2004) 995–1006.
350. L. L. Lairson, B. Henrissat, G. J. Davies, and S. G. Withers, Glycosyltransferases: Structures, functions, and mechanisms, *Annu. Rev. Biochem.*, 77 (2008) 521–555.
351. B. Waldscheck, M. Streiff, W. Notz, W. Kinzy, and R. R. Schmidt, Alpha(1-3)-galactosyltransferase inhibition based on a new type of disubstrate analogue, *Angew. Chem. Int. Ed.*, 40 (2001) 4007–4011.
352. P. Compain and O. R. Martin, Carbohydrate mimetics-based glycosyltransferase inhibitors, *Bioorg. Med. Chem.*, 9 (2001) 3077–3092.
353. M. Raab, S. Kozmon, and I. Tvaroška, Potential transition-state analogs for glycosyltransferases. Design and DFT calculations of conformational behavior, *Carbohydr. Res.*, 340 (2005) 1051–1057.
354. L. Sihelniková, S. Kozmon, and I. Tvaroška, DFT and docking study of potential transition state analogue inhibitors of glycosyltransferases, *Collect. Czech. Chem. Commun.*, 73 (2008) 591–607.
355. U. Krengel and A. Imberty, Crystallography and lectin structure database, In: L. N. Carol, (Ed.), *Lectins,* Elsevier Science B.V., Amsterdam, 2007, pp. 15–50.
356. O. Sulak, E. Lameignere, M. Wimmerova, and A. Imberty, Specificity and affinity studies in lectin/carbohydrate interactions, *Carbohydr. Chem.*, 35 (2009) 357–372.
357. D. Neumann, C. M. Lehr, H. P. Lenhof, and O. Kohlbacher, Computational modeling of the sugar-lectin interaction, *Adv. Drug Deliv. Rev.*, 56 (2004) 437–457.

358. A. Nurisso, S. Kozmon, and A. Imberty, Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III, *Mol. Simul.*, 34 (2008) 469–479.
359. J. Adam, Z. Kriz, M. P. Prokop, M. Wimmerova, and J. Koca, *In silico* mutagenesis and docking studies of *Pseudomonas aeruginosa* PA-IIL lectin predicting binding modes and energies, *J. Chem. Inf. Model.*, 48 (2008) 2234–2242.
360. B. Blanchard, A. Nurisso, E. Hollville, C. Tetaud, J. Wiels, M. Pokorna, M. Wimmerova, A. Varrot, and A. Imberty, Structural basis of the preferential binding for globo-series glycosphingolipids displayed by *Pseudomonas aeruginosa* lectin I, *J. Mol. Biol.*, 383 (2008) 837–853.
361. G. Cioci, E. P. Mitchell, C. Gautier, M. Wimmerova, D. Sudakevitz, S. Pérez, N. Gilboa-Garber, and A. Imberty, Structural basis of calcium and galactose recognition by the lectin PA-IL of *Pseudomonas aeruginosa*, *FEBS Lett.*, 555 (2003) 297–301.
362. S. K. Mishra, J. Sund, J. Aqvist, and J. Koca, Computational prediction of monosaccharide binding free energies to lectins with linear interaction energy models, *J. Comput. Chem.*, 33 (2012) 2340–2350.
363. S. K. Mishra, J. Adam, M. Wimmerova, and J. Koca, *In silico* mutagenesis and docking study of *Ralstonia solanacearum* RSL lectin: Performance of docking software to predict saccharide binding, *J. Chem. Inf. Model.*, 52 (2012) 1250–1261.
364. H. C. Siebert, S. Y. Lu, R. Wechselberger, K. Born, T. Eckert, S. Liang, C. W. von der Lieth, J. Jiménez Barbero, R. Schauer, J. F. Vliegenthart, T. Lütteke, S. Andre, H. Kaltner, H. J. Gabius, and T. Kozar, A lectin from the Chinese bird-hunting spider binds sialic acids, *Carbohydr. Res.*, 344 (2009) 1515–1525.
365. M. Agostino, E. Yuriev, and B. Ramakrishan, A computational approach for exploring carbohydrate recognition by lectins in innate immunity, *Front. Immunol.*, 2 (2011) 23.
366. O. Sperling, A. Fuchs, and T. K. Lindhorst, Evaluation of the carbohydrate recognition domain of the bacterial adhesin FimH: Design, synthesis and binding properties of mannoside ligands, *Org. Biomol. Chem.*, 4 (2006) 3913–3922.
367. T. K. Lindhorst, K. Bruegge, A. Fuchs, and O. Sperling, A bivalent glycopeptide to target two putative carbohydrate binding sites on FimH, *Beilstein J. Org. Chem.*, 6 (2010) 801–809.
368. C. Grabosch, M. Hartmann, J. Schmidt-Lassen, and T. K. Lindhorst, Squaric acid monoamide mannosides as ligands for the bacterial lectin FimH: Covalent inhibition or not? *Chembiochem*, 12 (2011) 1066–1074.
369. J. J. Reina, S. Sattin, D. Invernizzi, S. Mari, L. Martinez-Prats, G. Tabarani, F. Fieschi, R. Delgado, P. M. Nieto, J. Rojo, and A. Bernardi, 1,2-Mannobioside mimic: Synthesis, DC-SIGN interaction by NMR and docking, and antiviral activity, *ChemMedChem*, 2 (2007) 1030–1036.
370. J. J. Reina, I. Diaz, P. M. Nieto, N. E. Campillo, J. A. Paez, G. Tabarani, F. Fieschi, and J. Rojo, Docking, synthesis, and NMR studies of mannosyl trisaccharide ligands for DC-SIGN lectin, *Org. Biomol. Chem.*, 6 (2008) 2743–2754.
371. N. Obermajer, S. Sattin, C. Colombo, M. Bruno, U. Svajger, M. Anderluh, and A. Bernardi, Design, synthesis and activity evaluation of mannose-based DC-SIGN antagonists, *Mol. Divers.*, 15 (2011) 347–360.
372. A. Imberty, R. Mollicone, E. Mikros, P.-A. Carrupt, S. Pérez, and R. Oriol, How do antibodies and lectins recognize histo-blood group antigens? A 3D-QSAR study by comparative molecular field analysis (CoMFA), *Bioorg. Med. Chem.*, 4 (1996) 1979–1988.
373. F. Strino, J. H. Lii, C. A. Koppisetty, P. G. Nyholm, and H. J. Gabius, Selenoglycosides in silico: *Ab initio*-derived reparameterization of MM4, conformational analysis using histo-blood group ABH antigens and lectin docking as indication for potential of bioactivity, *J. Comput. Aided Mol. Des.*, 24 (2010) 1009–1021.
374. M. Agostino, M. S. Sandrin, P. E. Thompson, E. Yuriev, and P. A. Ramsland, Identification of preferred carbohydrate binding modes in xenoreactive antibodies by combining conformational filters and binding site maps, *Glycobiology*, 20 (2010) 724–735.

375. C. A. Koppisetty, W. Nasir, F. Strino, G. E. Rydell, G. Larson, and P. G. Nyholm, Computational studies on the interaction of ABO-active saccharides with the norovirus VA387 capsid protein can explain experimental binding data, *J. Comput. Aided Mol. Des.*, 24 (2010) 423–431.
376. W. Nasir, M. Frank, C. A. Koppisetty, G. Larson, and P. G. Nyholm, Lewis histo-blood group alpha1,3/alpha1,4 fucose residues may both mediate binding to GII.4 noroviruses, *Glycobiology*, 22 (2012) 1163–1172.
377. J. Topin, J. Arnaud, A. Sarkar, M. Audry, E. Gillon, S. Pérez, H. Jamet, A. Varrot, A. Imberty, and A. Thomas, Deciphering the glycan preference of bacterial lectins by glycan array and molecular docking with validation by microcalorimetry and crystallography, *PLos One*, 8 (2013) e71149.
378. R. D. Astronomo and D. R. Burton, Carbohydrate vaccines: Developing sweet solutions to sticky situations? *Nat. Rev. Drug Discov.*, 9 (2010) 308–324.
379. H. Clausen and S.-i. Hakomori, ABH and related histo-blood group antigens; immunochemical differences in carrier isotypes and their distribution. 1, *Vox Sang.*, 56 (1989) 1–20.
380. P. A. Ramsland, W. Farrugia, T. M. Bradford, P. Mark Hogarth, and A. M. Scott, Structural convergence of antibody binding of carbohydrate determinants in Lewis Y tumor antigens, *J. Mol. Biol.*, 340 (2004) 809–818.
381. A. M. van Roon, N. S. Pannu, J. P. de Vrind, G. A. van der Marel, J. H. van Boom, C. H. Hokke, A. M. Deelder, and J. P. Abrahams, Structure of an anti-Lewis X Fab fragment in complex with its Lewis X antigen, *Structure*, 12 (2004) 1227–1236.
382. D. C. de Geus, A. M. van Roon, E. A. Thomassen, C. H. Hokke, A. M. Deelder, and J. P. Abrahams, Characterization of a diagnostic Fab fragment binding trimeric Lewis X, *Proteins*, 76 (2009) 439–447.
383. A. Imberty, E. Mikros, J. Koca, R. Mollicone, R. Oriol, and S. Pérez, Computer simulation of histo-blood group oligosaccharides: Energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens, *Glycoconj. J.*, 12 (1995) 331–349.
384. M. Agostino, C. Jene, T. Boyle, P. A. Ramsland, and E. Yuriev, Molecular docking of carbohydrate ligands to antibodies: Structural validation against crystal structures, *J. Chem. Inf. Model.*, 49 (2009) 2749–2760.
385. J. Milland, E. Yuriev, P. X. Xing, I. F. McKenzie, P. A. Ramsland, and M. S. Sandrin, Carbohydrate residues downstream of the terminal Gal alpha (1,3) Gal epitope modulate the specificity of xenoreactive antibodies, *Immunol. Cell Biol.*, 85 (2007) 623–632.
386. E. Yuriev, M. S. Sandrin, and P. A. Ramsland, Antibody-ligand docking: Insights into peptide-carbohydrate mimicry, *Mol. Simul.*, 34 (2008) 461–468.
387. M. Kearns-Jonker, N. Barteneva, R. Mencel, N. Hussain, I. Shulkin, A. Xu, M. Yew, and D. V. Cramer, Use of molecular modeling and site-directed mutagenesis to define the structural basis for the immune response to carbohydrate xenoantigens, *BMC Immunol.*, 8 (2007) 3.
388. R. P. Oomen, N. M. Young, and D. R. Bundle, Molecular modeling of antibody-antigen complexes between the *Brucella abortus* O-chain polysaccharide and a specific monoclonal antibody, *Protein Eng.*, 4 (1991) 427–433.
389. R. Kadirvelraj, J. Gonzalez-Outeirino, B. L. Foley, M. L. Beckham, H. J. Jennings, S. Foote, M. G. Ford, and R. J. Woods, Understanding the bacterial polysaccharide antigenicity of *Streptococcus agalactiae* versus *Streptococcus pneumoniae*, *Proc. Natl. Acad. Sci. U.S.A.*, 103 (2006) 8149–8154.
390. B. Vulliez-Le Normand, F. A. Saul, A. Phalipon, F. Belot, C. Guerreiro, L. A. Mulard, and G. A. Bentley, Structures of synthetic O-antigen fragments from serotype 2a *Shigella flexneri* in complex with a protective monoclonal antibody, *Proc. Natl. Acad. Sci. U.S.A.*, 105 (2008) 9976–9981.
391. F. X. Theillet, M. Frank, B. Vulliez-Le Normand, C. Simenel, S. Hoos, A. Chaffotte, F. Bélot, C. Guerreiro, F. Nato, A. Phalipon, L. A. Mulard, and M. Delepierre, Dynamic aspects of antibody: oligosaccharide complexes characterized by molecular dynamics simulations and saturation transfer difference nuclear magnetic resonance, *Glycobiology*, 21 (2011) 1570–1579.

392. J. R. Brisson, H. Baumann, A. Imberty, S. Pérez, and H. J. Jennings, Helical epitope of the group B meningococcal alpha(2-8)-linked sialic acid polysaccharide, *Biochemistry*, 31 (1992) 4996–5004.
393. L. Kjellen and U. Lindahl, Proteoglycans: Structures and interactions, *Annu. Rev. Biochem.*, 60 (1991) 443–475.
394. D. J. D. Johnson, W. Li, T. E. Adams, and J. A. Huntington, Antithrombin-S195A factor Xa-heparin structure reveals the allosteric mechanism of antithrombin activation, *EMBO J.*, 25 (2006) 2029–2037.
395. R. Skinner, J.-P. Abrahams, J. C. Whisstock, A. M. Lesk, R. W. Carrell, and M. R. Wardell, The 2.6 Å structure of antithrombin indicates a conformational change at the heparin binding site, *J. Mol. Biol.*, 266 (1997) 601–609.
396. P. D. J. Grootenhuis and C. A. A. Vanboeckel, Constructing a molecular-model of the interaction between antithrombin-III and a potent heparin analog, *J. Am. Chem. Soc.*, 113 (1991) 2743–2747.
397. W. Bitomsky and R. C. Wade, Docking of glycosaminoglycans to heparin-binding proteins: Validation for αFGF, βFGF, and antithrombin and application to IL-8, *J. Am. Chem. Soc.*, 121 (1999) 3004–3013.
398. L. Jin, J. P. Abrahams, R. Skinner, M. Petitou, R. N. Pike, and R. W. Carrell, The anticoagulant activation of antithrombin by heparin, *Proc. Natl. Acad. Sci. U.S.A.*, 94 (1997) 14683–14688.
399. M. Hricovini, M. Guerrini, A. Bisio, G. Torri, M. Petitou, and B. Casu, Conformation of heparin pentasaccharide bound to antithrombin III, *Biochem. J.*, 359 (2001) 265–272.
400. L. Pellegrini, D. F. Burke, F. von Delft, B. Mulloy, and T. L. Blundell, Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin, *Nature*, 407 (2000) 1029–1034.
401. R. Raman, G. Venkataraman, S. Ernst, V. Sasisekharan, and R. Sasisekharan, Structural specificity of heparin binding in the fibroblast growth factor family of proteins, *Proc. Natl. Acad. Sci. U.S.A.*, 100 (2003) 2357–2362.
402. D. Rossi and A. Zlotnik, The biology of chemokines and their receptors, *Annu. Rev. Immunol.*, 18 (2000) 217–243.
403. T. M. Handel, Z. Johnson, S. E. Crown, E. K. Lau, M. Sweeney, and A. E. Proudfoot, Regulation of protein function by glycosaminoglycans—As exemplified by chemokines, *Annu. Rev. Biochem.*, 74 (2005) 385–410.
404. H. Lortat-Jacob, A. Grosdidier, and A. Imberty, Structural diversity of heparan sulfate binding domains in chemokines, *Proc. Natl. Acad. Sci. U.S.A.*, 99 (2002) 1229–1234.
405. M. G. Costa, P. R. Batista, C. S. Shida, C. H. Robert, P. M. Bisch, and P. G. Pascutti, How does heparin prevent the pH inactivation of cathepsin B? Allosteric mechanism elucidated by docking and molecular dynamics, *BMC Genomics*, 11(Suppl. 5), (2010) S5.
406. C. J. Rogers, P. M. Clark, S. E. Tully, R. Abrol, K. C. Garcia, W. A. Goddard, IIIrd, and L. C. Hsieh-Wilson, Elucidating glycosaminoglycan-protein-protein interactions using carbohydrate microarray and computational approaches, *Proc. Natl. Acad. Sci. U.S.A.*, 108 (2011) 9747–9752.
407. N. Sapay, E. Cabannes, M. Petitou, and A. Imberty, Molecular modeling of the interaction between heparan sulfate and cellular growth factors: Bringing pieces together, *Glycobiology*, 21 (2011) 1181–1193.
408. S. A. Samsonov, J. Teyra, and M. T. Pisabarro, Docking glycosaminoglycans to proteins: Analysis of solvent inclusion, *J. Comput. Aided Mol. Des.*, 25 (2011) 477–489.
409. S. A. Samsonov, J.-P. Gehrcke, and M. T. Pisabarro, Flexibility and explicit solvent in molecular dynamics based docking of protein–glycosaminoglycan systems, *J. Chem. Inf. Model.*, 54 (2014) 582–592.
410. T. Schirmer, T. Keller, Y. Wang, and J. Rosenbusch, Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution, *Science*, 267 (1995) 512–514.

411. R. Dutzler, T. Schirmer, M. Karplus, and S. Fischer, Translocation mechanism of long sugar chains across the maltoporin membrane channel, *Structure*, 10 (2002) 1273–1284.
412. P. Cunningham, I. Afzal-Ahmed, and R. J. Naftalin, Docking studies show that D-glucose and quercetin slide through the transporter GLUT1, *J. Biol. Chem.*, 281 (2006) 5797–5803.
413. S. Dang, L. Sun, Y. Huang, F. Lu, Y. Liu, H. Gong, J. Wang, and N. Yan, Structure of a fucose transporter in an outward-open conformation, *Nature*, 467 (2010) 734–738.
414. D. E. Koshland, Stereochemistry and mechanism of enzymatic reactions, *Biol. Rev.*, 28 (1953) 416–436.
415. A. C. T. Vanscheltinga, S. Armand, K. H. Kalk, A. Isogai, B. Henrissat, and B. W. Dijkstra, Stereochemistry of chitin hydrolysis by a plant chitinase lysozyme and X-Ray structure of a complex with allosamidin—Evidence for substrate assisted catalysis, *Biochemistry*, 34 (1995) 15619–15623.
416. A. Vasella, G. J. Davies, and M. Bohm, Glycosidase mechanisms, *Curr. Opin. Chem. Biol.*, 6 (2002) 619–629.
417. D. J. Vocadlo and G. J. Davies, Mechanistic insights into glycosidase chemistry, *Curr. Opin. Chem. Biol.*, 12 (2008) 539–555.
418. G. Legler, Glycoside hydrolases: Mechanistic information from studies with reversible and irreversible inhibitors, *Adv. Carbohydr. Chem. Biochem.*, 48 (1990) 319–384.
419. T. D. Heightman and A. T. Vasella, Recent insights into inhibition, structure, and mechanism of configuration-retaining glycosidases, *Angew. Chem. Int. Ed.*, 38 (1999) 750–770.
420. C. S. Rye and S. G. Withers, Glycosidase mechanisms, *Curr. Opin. Chem. Biol.*, 4 (2000) 573–580.
421. C. C. Blake, L. N. Johnson, G. A. Mair, A. C. North, D. C. Phillips, and V. R. Sarma, Crystallographic studies of the activity of hen egg-white lysozyme, *Proc. R. Soc. Lond. B Biol. Sci.*, 167 (1967) 378–388.
422. C. C. Blake, D. F. Koenig, G. A. Mair, A. C. North, D. C. Phillips, and V. R. Sarma, Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution, *Nature*, 206 (1965) 757–761.
423. C. C. Blake, G. A. Mair, A. C. North, D. C. Phillips, and V. R. Sarma, On the conformation of the hen egg-white lysozyme molecule, *Proc. R. Soc. Lond. B Biol. Sci.*, 167 (1967) 365–377.
424. G. J. Davies, A. Planas, and C. Rovira, Conformational analyses of the reaction coordinate of glycosidases, *Acc. Chem. Res.*, 45 (2012) 308–316.
425. L. E. Tailford, W. A. Offen, N. L. Smith, C. Dumon, C. Morland, J. Gratien, M. P. Heck, R. V. Stick, Y. Bleriot, A. Vasella, H. J. Gilbert, and G. J. Davies, Structural and biochemical evidence for a boat-like transition state in beta-mannosidases, *Nat. Chem. Biol.*, 4 (2008) 306–312.
426. V. M.-A. Ducros, D. L. Zechel, G. N. Murshudov, H. J. Gilbert, L. Szabo, D. Stoll, S. G. Withers, and G. J. Davies, Substrate distortion by a β-Mannanase: Snapshots of the Michaelis and covalent-intermediate complexes suggest a $B_{2,5}$ conformation for the transition state, *Angew. Chem. Int. Ed.*, 41 (2002) 2824–2827.
427. I. Tews, A. Perrakis, A. Oppenheim, Z. Dauter, K. S. Wilson, and C. E. Vorgias, Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease, *Nat. Struct. Biol.*, 3 (1996) 638–648.
428. G. Sulzenbacher, H. Driguez, B. Henrissat, M. Schulein, and G. J. Davies, Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: Substrate distortion gives rise to the preferred axial orientation for the leaving group, *Biochemistry*, 35 (1996) 15280–15287.
429. G. J. Davies, L. Mackenzie, A. Varrot, M. Dauter, A. M. Brzozowski, M. Schulein, and S. G. Withers, Snapshots along an enzymatic reaction coordinate: Analysis of a retaining beta-glycoside hydrolase, *Biochemistry*, 37 (1998) 11707–11713.
430. X. Biarnes, A. Ardevol, J. Iglesias-Fernandez, A. Planas, and C. Rovira, Catalytic itinerary in 1,3-1,4-beta-glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state, *J. Am. Chem. Soc.*, 133 (2011) 20301–20309.

431. D. Cremer and J. A. Pople, A general definition of ring puckering coordinates, *J. Am. Chem. Soc*., 97 (1975) 1354–1358.
432. L. Petersen, A. Ardevol, C. Rovira, and P. J. Reilly, Mechanism of cellulose hydrolysis by inverting GH8 endoglucanases: A QM/MM metadynamics study, *J. Phys. Chem. B*, 113 (2009) 7331–7339.
433. L. Petersen, A. Ardevol, C. Rovira, and P. J. Reilly, Molecular mechanism of the glycosylation step catalyzed by Golgi alpha-mannosidase II: A QM/MM metadynamics investigation, *J. Am. Chem. Soc*., 132 (2010) 8291–8300.
434. N. F. Bras, M. J. Ramos, and P. A. Fernandes, DFT studies on the beta-glycosidase catalytic mechanism: The deglycosylation step, *J. Mol. Struct. (THEOCHEM)*, 946 (2010) 125–133.
435. N. F. Bras, S. A. Moura-Tamames, P. A. Fernandes, and M. J. Ramos, Mechanistic studies on the formation of glycosidase-substrate and glycosidase-inhibitor covalent intermediates, *J. Comput. Chem*., 29 (2008) 2565–2574.
436. N. F. Bras, P. A. Fernandes, and M. J. Ramos, QM/MM studies on the beta-galactosidase catalytic mechanism: Hydrolysis and transglycosylation reactions, *J. Chem. Theory Comput*., 6 (2010) 421–433.
437. D. H. Juers, T. D. Heightman, A. Vasella, J. D. McCarter, L. Mackenzie, S. G. Withers, and B. W. Matthews, A structural view of the action of *Escherichia coli* (lacZ) beta-galactosidase, *Biochemistry*, 40 (2001) 14781–14794.
438. M. J. T. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian 03, Revision D.01/D.02, Gaussian, Inc., Wallingford, CT, 2004.
439. S. M. Tschampel, M. R. Kennerty, and R. J. Woods, TIP5P-consistent treatment of electrostatics for biomolecular simulations, *J. Chem. Theory Comput*., 3 (2007) 1721–1733.
440. K. A. Brameld, W. D. Shrader, B. Imperiali, and W. A. Goddard, Substrate assistance in the mechanism of family 18 chitinases: Theoretical studies of potential intermediates and inhibitors, *J. Mol. Biol*., 280 (1998) 913–923.
441. O. Passos, P. A. Fernandes, and M. J. Ramos, Theoretical insights into the catalytic mechanism of beta-hexosaminidase, *Theor. Chem. Acc*., 129 (2011) 119–129.
442. A. Bottoni, G. P. Miscione, and M. De Vivo, A theoretical DFT investigation of the lysozyme mechanism: Computational evidence for a covalent intermediate pathway, *Proteins*, 59 (2005) 118–130.
443. A. L. Bowman, I. M. Grant, and A. J. Mulholland, QM/MM simulations predict a covalent intermediate in the hen egg white lysozyme reaction with its natural substrate, *Chem. Commun*. (2008) 4425–4427.
444. J. Jitonnom, V. S. Lee, P. Nimmanpipug, H. A. Rowlands, and A. J. Mulholland, Quantum mechanics/molecular mechanics modeling of substrate-assisted catalysis in family 18 chitinases: Conformational changes and the role of Asp142 in catalysis in chitinase B, *Biochemistry*, 50 (2011) 4697–4711.
445. J. Jitonnom, A. J. Mulholland, P. Nimmanpipug, and V. S. Lee, Hybrid QM/MM study on the deglycosylation step of chitin hydrolysis catalysed by chitinase B from *Serratia marcescens*, *Maejo Int. J. Sci. Technol*., 5 (2011) 47–57.

446. A. Bottoni, G. P. Miscione, and M. Calvaresi, Computational evidence for the substrate-assisted catalytic mechanism of O-GlcNAcase. A DFT investigation, *Phys. Chem. Chem. Phys*., 13 (2011) 9568–9577.
447. D. J. Vocadlo, O-GlcNAc processing enzymes: Catalytic mechanisms, substrate specificity, and enzyme regulation, *Curr. Opin. Chem. Biol*., 16 (2012) 488–497.
448. B. L. Mark, D. J. Vocadlo, D. Zhao, S. Knapp, S. G. Withers, and M. N. James, Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 beta-N-acetylhexosaminidase inhibitor, *J. Biol. Chem*., 276 (2001) 42131–42137.
449. G. E. Whitworth, M. S. Macauley, K. A. Stubbs, R. J. Dennis, E. J. Taylor, G. J. Davies, I. R. Greig, and D. J. Vocadlo, Analysis of PUGNAc and NAG-thiazoline as transition state analogues for human O-GlcNAcase: Mechanistic and structural insights into inhibitor selectivity and transition state poise, *J. Am. Chem. Soc*., 129 (2007) 635–644.
450. K. A. Brameld and W. A. Goddard, III, Substance distortion to a boat conformation at subsite-1 is critical in the mechanism of family, 18 chitinases, *J. Am. Chem. Soc*., 120 (1998) 3571–3580.
451. B. L. Mark, D. J. Vocadlo, S. Knapp, B. L. Triggs-Raine, S. G. Withers, and M. N. G. James, Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase, *J. Biol. Chem*., 276 (2001) 10330–10337.
452. D. R. Rose, Structure, mechanism and inhibition of Golgi alpha-mannosidase II, *Curr. Opin. Struct. Biol*., 22 (2012) 558–562.
453. A. Varki, Functions of sialic acids in intact mammals, *FASEB J*., 11 (1997) A1443.
454. N. R. Taylor and M. von Itzstein, Molecular modeling studies on ligand-binding to sialidase from influenza virus and the mechanism of catalysis, *J. Med. Chem*., 37 (1994) 616–624.
455. J. A. Barnes and I. H. Williams, Quantum mechanical molecular mechanical approaches to transition state structure: Mechanism of sialidase action, *Biochem. Soc. Trans*., 24 (1996) 263–268.
456. M. E. S. Soliman, J. J. R. Pernia, I. R. Greig, and I. H. Williams, Mechanism of glycoside hydrolysis: A comparative QM/MM molecular dynamics analysis for wild type and Y69F mutant retaining xylanases, *Org. Biomol. Chem*., 7 (2009) 5236–5244.
457. M. E. S. Soliman, G. D. Ruggiero, J. J. R. Pernia, I. R. Greig, and I. H. Williams, Computational mutagenesis reveals the role of active-site tyrosine in stabilising a boat conformation for the substrate: QM/MM molecular dynamics studies of wild-type and mutant xylanases, *Org. Biomol. Chem*., 7 (2009) 460–468.
458. D. M. A. Guerin, M. B. Lascombe, M. Costabel, H. Souchon, V. Lamzin, P. Beguin, and P. M. Alzari, Atomic (0.94 angstrom) resolution structure of an inverting glycosidase in complex with substrate, *J. Mol. Biol*., 316 (2002) 1061–1069.
459. J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett*., 77 (1996) 3865–3868.
460. L. Ballut, N. Sapay, E. Chautard, A. Imberty, and S. Ricard-Blum, Mapping of heparin/heparan sulfate binding sites on alpha-beta3 integrin by molecular docking, *J. Mol. Recognit*., 26 (2013) 76–85.
461. N. S. Gandhi, C. Freeman, C. R. Parish, and R. L. Mancera, Computational analyses of the catalytic and heparin-binding sites and their interactions with glycosaminoglycans in glycoside hydrolase family 79 endo-β-D-glucuronidase (heparanase), *Glycobiology*, 22 (2012) 35–55.
462. N. Sapay, E. Cabannes, M. Petitou, and A. Imberty, Molecular model of human heparanase with proposed binding mode of a heparan sulfate oligosaccharide and catalytic amino acids, *Biopolymers*, 97 (2012) 21–34.
463. J. N. Varghese, W. G. Laver, and P. M. Colman, Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution, *Nature*, 303 (1983) 35–40.
464. J. N. Varghese, J. L. McKimm-Breschkin, J. B. Caldwell, A. A. Kortt, and P. M. Colman, The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor, *Proteins*, 14 (1992) 327–332.

465. M. von Itzstein, W.-Y. Wu, G. B. Kok, M. S. Pegg, J. C. Dyason, B. Jin, T. V. Phan, M. L. Smythe, H. F. White, S. W. Oliver, P. M. Colman, J. N. Varghese, D. M. Ryan, J. M. Woods, R. C. Bethell, V. J. Hotham, J. M. Cameron, and C. R. Penn, Rational design of potent sialidase-based inhibitors of influenza virus replication, *Nature*, 363 (1993) 418–423.
466. C. U. Kim, W. Lew, M. A. Williams, H. Liu, L. Zhang, S. Swaminathan, N. Bischofberger, M. S. Chen, D. B. Mendel, C. Y. Tai, W. G. Laver, and R. C. Stevens, Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: Design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity, *J. Am. Chem. Soc.*, 119 (1997) 681–690.
467. J. A. McCullers, E. Hoffmann, V. C. Huber, and A. D. Nickerson, A single amino acid change in the C-terminal domain of the matrix protein M1 of influenza B virus confers mouse adaptation and virulence, *Virology*, 336 (2005) 318–326.
468. R. Chachra and R. C. Rizzo, Origins of resistance conferred by the R292K neuraminidase mutation via molecular dynamics and free energy calculations, *J. Chem. Theory Comput.*, 4 (2008) 1526–1540.
469. M. Raab and I. Tvaroška, The binding properties of the H5N1 influenza virus neuraminidase as inferred from molecular modeling, *J. Mol. Model.*, 17 (2011) 1445–1456.
470. J. Lehtio, J. Sugiyama, M. Gustavsson, L. Fransson, M. Linder, and T. T. Teeri, The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules, *Proc. Natl. Acad. Sci. U.S.A.*, 100 (2003) 484–489.
471. A. L. Demain, M. Newcomb, and J. H. D. Wu, Cellulase, clostridia, and ethanol, *Microbiol. Mol. Biol. Rev.*, 69 (2005) 124–154.
472. B. Mertz, A. D. Hill, C. Mulakala, and P. J. Reilly, Automated docking to explore subsite binding by glycoside hydrolase family 6 cellobiohydrolases and endoglucanases, *Biopolymers*, 87 (2007) 249–260.
473. O. V. Oliveira, L. C. Freitas, T. P. Straatsma, and R. D. Lins, Interaction between the CBM of Cel9A from *Thermobifida fusca* and cellulose fibers, *J. Mol. Recognit.*, 22 (2009) 38–45.
474. T. Yui, H. Shiiba, Y. Tsutsumi, S. Hayashi, T. Miyata, and F. Hirata, Systematic docking study of the carbohydrate binding module protein of Cel7A with the cellulose Iα crystal model, *J. Phys. Chem. B*, 114 (2010) 49–58.
475. T. Imai, C. Boisset, M. Samejima, K. Igarashi, and J. Sugiyama, Unidirectional processive action of cellobiohydrolase Cel7A on *Valonia* cellulose microcrystals, *FEBS Lett.*, 432 (1998) 113–116.
476. H. Hashimoto, Recent structural studies of carbohydrate-binding modules, *Cell. Mol. Life Sci.*, 63 (2006) 2954–2967.
477. M. F. Crowley, E. C. Uberbacher, C. L. Brooks, III, R. C. Walker, M. R. Nimlos, and M. E. Himmel, Developing improved MD codes for understanding processive cellulases, *J. Phys. Conf. Ser.*, 125 (2008) 12049–12056. Institute of Physics Publishing.
478. S. Parthasarathy, G. Ravindra, H. Balaram, P. Balaram, and M. R. Murthy, Structure of the *Plasmodium falciparum* triosephosphate isomerase-phosphoglycolate complex in two crystal forms: Characterization of catalytic loop open and closed conformations in the ligand-bound state, *Biochemistry*, 41 (2002) 13178–13188.
479. A. L. Jansma, J. P. Kirkpatrick, A. R. Hsu, T. M. Handel, and D. Nietlispach, NMR analysis of the structure, dynamics, and unique oligomerization properties of the chemokine CCL27, *J. Biol. Chem.*, 285 (2010) 14424–14437.
480. P. J. Collins, L. F. Haire, Y. P. Lin, J. Liu, R. J. Russell, P. A. Walker, J. J. Skehel, S. R. Martin, A. J. Hay, and S. J. Gamblin, Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants, *Nature*, 453 (2008) 1258–1261.
481. G. Parsiegla, C. Reverbel-Leroy, C. Tardif, J. P. Belaich, H. Driguez, and R. Haser, Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action, *Biochemistry*, 39 (2000) 11238–11246.

482. E. Brun, F. Moriaud, P. Gans, M. J. Blackledge, F. Barras, and D. Marion, Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by *Erwinia chrysanthemi*, *Biochemistry*, 36 (1997) 16074–16086.
483. J. Kraulis, G. M. Clore, M. Nilges, T. A. Jones, G. Pettersson, J. Knowles, and A. M. Gronenborn, Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing, *Biochemistry*, 28 (1989) 7241–7257.
484. S. Raghothama, P. J. Simpson, L. Szabo, T. Nagy, H. J. Gilbert, and M. P. Williamson, Solution structure of the CBM10 cellulose binding module from *Pseudomonas* xylanase A, *Biochemistry*, 39 (2000) 978–984.
485. J. Tormo, R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham, and T. A. Steitz, Crystal structure of a bacterial family-III cellulose-binding domain: A general mechanism for attachment to cellulose, *EMBO J*., 15 (1996) 5739–5751.
486. G. Y. Xu, E. Ong, N. R. Gilkes, D. G. Kilburn, D. R. Muhandiram, M. Harris-Brandts, J. P. Carver, L. E. Kay, and T. S. Harvey, Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy, *Biochemistry*, 34 (1995) 6993–7009.
487. M. J. Cuneo, A. Changela, L. S. Beese, and H. W. Hellinga, Structural adaptations that modulate monosaccharide, disaccharide, and trisaccharide specificities in periplasmic maltose-binding proteins, *J. Mol. Biol*., 389 (2009) 157–166.
488. O. Sulak, G. Cioci, M. Delia, M. Lahmann, A. Varrot, A. Imberty, and M. Wimmerova, A TNF-like trimeric lectin domain from *Burkholderia cenocepacia* with specificity for fucosylated human histo-blood group antigens, *Structure*, 18 (2010) 59–72.
489. T. Murase, R. B. Zheng, M. Joe, Y. Bai, S. L. Marcus, T. L. Lowary, and K. K. Ng, Structural insights into antibody recognition of mycobacterial polysaccharides, *J. Mol. Biol*., 392 (2009) 381–392.
490. E. C. Schulz, A. K. Bergfeld, R. Ficner, and M. Muhlenhoff, Crystal structure analysis of the polysialic acid specific O-acetyltransferase NeuO, *PLoS One*, 6 (2011) e17403.
491. A. Sarkar and S. Pérez, Protein-carbohydrate interactions—Computational aspects, In: E. Yuriev and P. A. Ramsland, (Eds.), *Structural Glycobiology*, CRC Press, Boca Raton, Fl, 2013, pp. 71–111.
492. Y. Bourne, P. Rouge, and C. Cambillau, X-ray structure of a biantennary octasaccharide-lectin complex refined at 2.3-A resolution, *J. Biol. Chem*., 267 (1992) 197–203.
493. E. A. Merritt, S. Sarfaty, M. G. Jobling, T. Chang, R. K. Holmes, T. R. Hirst, and W. G. Hol, Structural studies of receptor binding by cholera toxin mutants, *Protein Sci*., 6 (1997) 1516–1528.
494. T. Lütteke, M. Frank, and C. W. von der Lieth, Carbohydrate Structure Suite (CSS): Analysis of carbohydrate 3D structures derived from the PDB, *Nucleic Acids Res*., 33 (2005) D242–D246.
495. W. S. Somers, J. Tang, G. D. Shaw, and R. T. Camphausen, Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1, *Cell*, 103 (2000) 467–479.
496. E. Lameignere, T. C. Shiao, R. Roy, M. Wimmerova, F. Dubreuil, A. Varrot, and A. Imberty, Structural basis of the affinity for oligomannosides and analogs displayed by BC2L-A, a *Burkholderia cenocepacia* soluble lectin, *Glycobiology*, 20 (2010) 87–98.
497. S. Elgavish and B. Shaanan, Structures of the *Erythrina corallodendron* lectin and of its complexes with mono- and disaccharides, *J. Mol. Biol*., 277 (1998) 917–932.
498. E. A. Merritt, P. Kuhn, S. Sarfaty, J. L. Erbe, R. K. Holmes, and W. G. Hol, The 1.25 A resolution refinement of the cholera toxin B-pentamer: Evidence of peptide backbone strain at the receptor-binding site, *J. Mol. Biol*., 282 (1998) 1043–1059.
499. A. Audfray, J. Claudinon, S. Abounit, N. Ruvoen-Clouet, G. Larson, D. F. Smith, M. Wimmerova, J. Le Pendu, W. Romer, A. Varrot, and A. Imberty, Fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes, *J. Biol. Chem*., 287 (2012) 4335–4347.

500. J. R. Walker, B. Nagar, N. M. Young, T. Hirama, and J. M. Rini, X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure, *Biochemistry*, 43 (2004) 3783–3792.
501. M. Franck and S. Schloissnig, Bioinformatics and molecular modeling in glycobiology, *Cell. Mol. Life Sci.*, 67 (2010) 2749–2772.
502. B. Chen, E. M. Vogan, H. Gong, J. J. Skehel, D. C. Wiley, and S. C. Harrison, Structure of an unliganded simian immunodeficiency virus gp120 core, *Nature*, 433 (2005) 834–841.
503. T. Blundell, D. Carney, S. Gardner, F. Hayes, B. Howlin, T. Hubbard, J. Overington, D. A. Singh, B. L. Sibanda, and M. Sutcliffe, 18th Sir Hans Krebs lecture. Knowledge-based protein modelling and design, *Eur. J. Biochem.*, 172 (1988) 513–520.
504. R. Loris, P. P. Stas, and L. Wyns, Conserved waters in legume lectin crystal structures. The importance of bound water for the sequence-structure relationship within the legume lectin family, *J. Biol. Chem.*, 269 (1994) 26722–26733.
505. R. A. Laskowski, D. S. Moss, and J. M. Thornton, Main-chain bond lengths and bond angles in protein structures, *J. Mol. Biol.*, 231 (1993) 1049–1067.
506. D. A. Simmons and E. Romanowska, Structure and biology of *Shigella flexneri* O antigens, *J. Med. Microbiol.*, 23 (1987) 289–302.
507. A. Imberty and S. Pérez, Molecular modeling of glycosaminoglycans and interactions with protein receptors—Methods and progress, In: M. Delehedde and H. Lortat-Jacob, (Eds.), *New Developments in Therapeutic Glycomics,* Research Signpost, Kerala, 2006, pp. 185–201.
508. R. Dutzler, Y. F. Wang, P. Rizkallah, J. P. Rosenbusch, and T. Schirmer, Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway, *Structure*, 4 (1996) 127–134.
509. J. N. Varghese, P. M. Colman, A. van Donkelaar, T. J. Blick, A. Sahasrabudhe, and J. L. McKimm-Breschkin, Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases, *Proc. Natl. Acad. Sci. U.S.A.*, 94 (1997) 11808–11812.
510. J. N. Varghese, V. C. Epa, and P. M. Colman, Three-dimensional structure of the complex of 4-guanidino-Neu5Ac2en and influenza virus neuraminidase, *Protein Sci.*, 4 (1995) 1081–1087.
511. J. N. Varghese, P. W. Smith, S. L. Sollis, T. J. Blick, A. Sahasrabudhe, J. L. McKimm-Breschkin, and P. M. Colman, Drug design against a shifting target: A structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase, *Structure*, 6 (1998) 735–746.
512. S. Y. Ding, Q. Xu, M. Crowley, Y. Zeng, M. Nimlos, R. Lamed, E. A. Bayer, and M. E. Himmel, A biophysical perspective on the cellulosome: New opportunities for biomass conversion, *Curr. Opin. Biotechnol.*, 19 (2008) 218–227.
513. S. Pérez and D. Samain, Structure and engineering of celluloses, *Adv. Carbohydr. Chem. Biochem.*, 65 (2010) 25–116.