Pharmaceutical Research, Vol. 15, No. 1, 1998

Review

The How and Why of Protein-carbohydrate Interaction: A Primer to the Theoretical Concept and a Guide to Application in Drug Design

Hans-Joachim Gabius[1,2]

Received June 26, 1997; accepted September 26, 1997

The common principles of molecular recognition with cooperative or bidentate hydrogen bonds, dispersion forces and hydrophobic packing govern the specificity of protein-carbohydrate interaction. Enthalpy/entropy-compensation is also valid, maintaining K_D-values in the range of 30 mM to 200 nM. The individual contributions of the enthalpic and entropic factors which originate from the receptor, the ligand and/or the solvent to the overall free energy change can at least be estimated by a combination of computer-assisted molecular modeling, NMR spectroscopy of the reactants before and after complex formation and thermodynamic measurements. The delineation of adaptable parameters such as ligand or receptor side chain flexibility points to a route to practicable guidelines for a rational design of mimetics in glycosciences.

KEY WORDS: lectin; antibody; glycoconjugate; glycoprotein; X-ray crystallography; NMR spectroscopy; molecular modeling; drug design; thermodynamics.

INTRODUCTION

The establishment of a biological code system requires the availability of a set of monomers. These modules can be covalently linked to convey oligomer-based signals (code units). In the cases of amino acids and nucleotides, sequence permutations govern the amount of information which can in principle be stored in oligomeric code words. Their total number can be readily calculated, yielding the assessment of the upper limit of the coding capacity for any class of biomolecules. For example, a set of six amino acids will generate 46,656 different peptide bond-linked structures (or 6.4×10^7 structures, if composed of any of the twenty monomers). When taking oligosaccharides into consideration, several additional factors are available to enhance the coding capacity besides the sequence. For example, various hydroxyl groups as potential participants in the glycosidic linkage are available in each unit, making a series of isomers accessible already at the level of disaccharides, e.g. α/β1-2, 1-3, 1-4 or 1-6. As a further consequence, branched structures can be readily established by monosaccharides, as graphically depicted in Fig. 1. To be able to quantitatively compare coding capacities, the same calculations have been performed to determine the total number of hexasaccharides relative to that of hexapeptides. Notably, the oligomer generation with six D-hexoses will theoretically yield the staggering group size of 1.05×10^{12} (or 1.44×10^{15} for branched structures from a set of twenty hexoses) isomers (1). Variability of attachment points for the glycosidic linkage, the α/β-anomerity, the occurrence of pyranose/furanose ring geometry and the inherent potential for branching thus opens the way to a biologically unsurpassed coding capacity.

Although knowledge accrued over the last decade indicates that the enzymatic machinery for glycan chain assembly restricts this theoretically attainable value (2), the compensating factor should not be neglected. The introduction of site-specific substitutions such as sulfate, phosphate or acetyl groups into a glycan leads to the presentation of unique epitopes especially at easily accessible non-reducing termini of the sugar antennae (1,3–5). Following the work on nucleic acids and proteins which founded the deciphering of biological code systems, the concept of a saccharide-based language has already been adequately

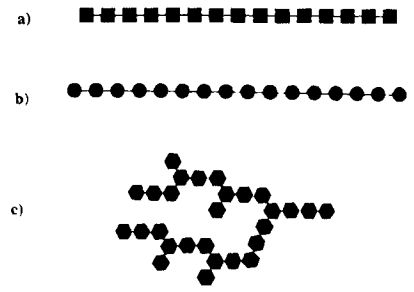

Fig. 1. Schematical representation of oligomer formation with nucleotides (a) and amino acids (b), which will generally lead to linear structures. In contrast, oligosaccharides have the inherent potential to be enzymatically synthesized as branched structures with enhanced coding capacity (c).

[1] Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität, Veterinärstr. 13, D-80539 München, Germany.

[2] To whom correspondence should be addressed.

0724-8741/98/0100-0023$15.00/0

supported, ascribing ligand properties to various sugar determinants in the course of respective investigations (6–8). To establish a recognitive interplay, sugar epitopes of cellular glycoconjugates bind to a target molecule whose molecular features are suitable for complex formation. In addition to immunoglobulins and enzymes such as the mentioned glycosyltransferases, lectins serve as receptors for glycoligands (7,8). They were first identified over a century ago as plant products with agglutinating potency. The analysis of mammalian tissues over the past decades has unambiguously revealed the presence of such activities in any organ studied (7). Presently, this work has already reached the level to delineate distinct lectin categories based on common molecular characteristics of the carbohydrate recognition domain. As shown in Table 1, five lectin families are currently distinguished in mammals, which are comprised of two to over twenty individual group members (7).

Having herewith underscored the coding potential of oligosaccharides and the actual expression of receptor molecules, the next step is to ask which intermolecular forces will govern the interaction of the two partners. The non-covalent interplay can be regarded as transient association of two at least partially complementary shapes irrespective of their biochemical nature. The relative spatial orientations of functional groups is visualized with diffraction data of a crystal. Such a snapshot of a complex in the solid state by X-ray crystallography will provide insight into the atomic interactions between the protein's carbohydrate recognition domain and the carbohydrate structure (7,9–15). It is therefore instructive to examine more closely the information in this respect, as gathered from electron density maps.

PROTEIN-CARBOHYDRATE INTERACTION IN THE CRYSTAL

The spatial orientations of the resolvable molecule groups in the crystal permit one to infer the involvement of various accommodating atomic forces. The chemical character of carbohydrates entails that hydroxyl groups and aliphatic C-H units at the epimeric centers and the exocyclic position can form axially/equatorially alternating patches of different degrees of hydrophilicity. Therefore, the entire surface area of any carbohydrate should be considered to be heterogeneous. Its amphiphilic character favors disparate contact types in the course of complex formation. Since lectins display a high degree of selectivity for distinct epimers such as galactose/mannose, it is expected that directional forces which distinguish the equatorial or axial position will be inferred from the structural depictions.

Providing selectivity and a certain degree of stability without impeding transient contact formation, hydrogen bonds serve as important factors in receptor-ligand interplay. For their formation, the sp^3-hybridized oxygen atoms of the hydroxyl groups and the ring offer two lone electron pairs as acceptors. Remarkably, the hydroxyl groups can simultaneously engage in a third contact with its proton as donor, which will—if active—completely arrest the respective group. In this situation, cooperative hydrogen bonds preferentially to donors from main- or side-chain amide groups, less frequently to hydroxyl groups of side chains and to a carbonyl/carboxyl-acceptor are established. Besides cooperative bonding, the spacing between two hydroxyl groups or the axial 4-OH group and the ring oxygen atom, which is about 2.8 Å for equatorial/equatorial and equatorial/axial configurations, allows the geometrical probing whether a suitably located planar side chain might simultaneously interact with the two neighbouring epitopes on the ligand to yield bidentate hydrogen bonds. The required complementary fit of the two modes of hydrogen bonding adds up to create an intricate network as a specificity control especially for receptor molecules with deep clefts such as bacterial periplasmic transport proteins (13). Unquestionably, the inherent spatial constraints are considerable. Theoretically, it would be desirable to provide some degree of adaptability to this binding mode to improve its versatility. The strategically favorable placement of a mediator of hydrogen bonds as acceptor/donor would spatially broaden the scope of this concept. As an integral part of a ligand-receptor network of interactions water molecules can fulfill this task.

Inclusion of water molecules is a general theme in intermolecular association (9–13,15–17). Although the precise delineation of their positions in time-averaged density maps is not a trivial task, even at a resolution of 2 Å or better (18), it is comforting to note that the locations of ordered water molecules appear to be conserved in related proteins, as documented for legume lectins (19). As an integral part of the protein's interacting shape, the presence of solvent molecules can optimize the number of hydrogen bonds which can be formed with the ligand. This physiological function of the versatile donor/acceptor of hydrogen bonds has encouraged the suggestion to deliberately engineer ligand mimetics with well-positioned water-binding sites embedded in the interacting surface (17).

Table 1. Current Categories for Classification of Various Mammalian Lectins

Family	Structural Motif	Carbohydrate Ligand	Modular Arrangement
C-type	conserved CRD	variable (mannose, galactose, fucose, heparin tetrasaccharide)	yes
I-type	immunoglobulin-like CRD	variable ($man_6glcNAc_2$, HNK-1 epitope, hyaluronic acid, α2,3/α2,6-sialyllactose)	yes
galectins (S-type)	conserved CRD	β-galactosides	variable
pentraxins	pentameric subunit arrangement	4,6-cyclic acetal of β-galactose, galactose, sulfated and phosphorylated monosaccharides	yes
P-type	homologous, not yet strictly defined CRD	mannose-6-phosphate-containing glycoproteins	yes

CRD: carbohydrate recognition domain; from [7].

The level of affinity of the synthetic products to a receptor should profit from the enhanced hydrogen-bonding potential which is introduced as a consequence of drug design. Interestingly, rearrangements of ordered water molecules may contribute to enable differential adaptation of the conformation of a ligand which has not completely lost its mobility (15). Thus, water should definitely not only be regarded as ubiquitous solvent, but as flexible building block within ligand association.

A similar role is played by the presence of a Ca^{2+}-ion in the C-type mannan-binding lectin (10,12). A lone electron shell of each of the vicinal equatorial 3′- and 4′-hydroxyl groups of D-mannose forms a direct coordination bond to the Ca^{2+}-ion. Moreover, it additionally positions four further side-chain amide or carboxyl groups to entirely occupy the hydrogen-bonding capacity of these vicinal donor/acceptor pairs (12). A Ca^{2+}-ion is not only an integral part of the binding site of this type of lectin. In the case of the pentraxin serum amyloid P component, the Ca^{2+}-ion interacts directly with the carboxyl group of methyl 4,6-O-(1-carboxyethylidene)-β-D-galactopyranoside or the phosphate group of the anionic ligand phosphoethanolamine (20). However, a direct coordination bond contact with the sugar is not made. Similar to legume lectins, its presence in the pentraxin seems to be a beneficial structural factor of the architecture of the binding site. It encompasses positioning of carboxylate/amide side-chain groups for hydrogen bonding to the ligand, thus operating to immobilize crucial side chains and to guide their positioning for acquiring the indispensable geometry of this highly directional polar interaction. In thermodynamic terms to be discussed in more detail later, its presence brings about a frame for ligand entry at a rather low entropic cost on the side of the receptor.

In contrast to the versatile hydrogen bond as a glue for transient complex building, dispersion forces lack an obvious directionality. As already alluded to, sugar units present an amphiphilic surface to the ligand-binding cleft of different degrees of depth. Nature's design has facilitated an array of structural concepts, ranging from completely burying the ligand to leaving a part of its surface exposed to the solvent. Patches of hydrophobic character such as that defined by the C3-C6-part of galactose (B face) can be a target for van der Waals interactions. It is intriguing to note that despite a lack of sequence homology, members of various classes of galactose-binding proteins, namely a bacterial enterotoxin, ricin, several leguminous lectins and galectins (11), all employ an aromatic side chain as part of the binding site for stacking to the hexose. The two rings are not perfectly aligned in parallel, but can show a positional distortion, which nonetheless will not impair the ensuing effect to remove the two non-polar surface patches from solvent contact (11,12). Aliphatic side chains are apparently infrequently used for this purpose. This observation lends credence to the assumption that the delocalized π-electrons are preferable interaction partners for the non-polar parts of the sugar ring whose aliphatic protons bear a net positive partial charge (12). If the configuration of any hydroxyl group in this part of the molecule is changed, the resulting epimer will exhibit a reduced affinity due to steric hindrance and/or the disturbing introduction of a polar group into a hydrophobic environment. Although the hydrophobic forces are not strictly dependent on directionality, they nonetheless contribute to determine the possible orientations of a ligand in the binding pocket. In this sense, dispersion forces cooperate with hydrogen bonds for accurate ligand accommodation.

When lectin and antibody structures are compared before and after complex formation, it is evident that binding of the ligand is not accompanied by a pronounced conformational change of the protein (12,15). However, it has to be registered that any conclusion should be judged in view of the inherent limitations of the technique. Concerning crystallography, the frequently necessary use of unphysiological conditions for initiation and growth of crystals, the inevitable interactions in the ordered crystal which will restrict mobility of surface residues, the time-averaged character of the coordinates and the demanding problem to ascribe adequate coordinates to solvent molecules argue strongly for cautious interpretation of electron density maps (15,21,22). Whereas these maps unquestionably reveal evidence for the bonding pattern in the crystal, they will fail to let us comprehend completely, which forces will drive the complex formation in any case. Therefore, it is necessary to supplement this information with structural data in solution by NMR and with thermodynamic measurements. Most notably, these approaches are valuable sources of information on the properties of the solvent and the ligand. Thermodynamic calculations, for example, can readily explain the involvement of a water molecule in the hydrogen bond network. Its positioning is connected with an unfavorable entropic term between 0–2 kcal mol^{-1} and a compensating enthalpic term of up to 4 kcal mol^{-1} (17,23). These calculations emphasize the necessity to discuss binding in terms of the—often still quite elusive—individual contributions to the free energy change (24–28). In order to prevent the reader from being confused by the intriguing complexity of this topic which currently is hotly debated, it is instructive to formalize the presentation by using guidelines derived from basic thermodynamics and to discuss special examples to illustrate the validity of certain predictions. When the reader keeps in mind that the outlined mechanisms are not mutually exclusive and can be operative in harmonious combination to varying extents, he can anticipate encountering fewer problems.

PROTEIN-CARBOHYDRATE INTERACTION IN SOLUTION

Basic Thermodynamic Considerations

The free energy change upon a non-covalent association is a measure of the affinity, expressed in equation (1):

$$\Delta G^\circ = -RT \ln K_A = RT \ln K_D \qquad (1)$$

A rather small free energy change of $\Delta G^\circ = -2$ kcal mol^{-1} will thus translate into a K_D-value of 30 mM. Further increases to $\Delta G^\circ = -9$ kcal mol^{-1} or -15 kcal mol^{-1} will be linked to K_D-values of 200 nM and 10 pM, respectively. According to the basic thermodynamic principle enthalpic and entropic factors contribute to the overall change in free energy, as given by equation (2):

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \qquad (2)$$

To gain an idea of the relative importance of each of these two factors for a certain binding process, it is necessary to determine these parameters by calorimetry. Although these measurements still require non-physiologically high concentrations

to record the occurring alterations, they open the way to eventually understand the actual driving forces. Starting with lysozyme's binding of oligosaccharide inhibitors (29), the determination of parameters of the thermodynamics of protein-carbohydrate recognition provided another example of the generally occurring entropy/enthalpy compensation (24–28). This correlation is commonly visualized by plotting values of $-\Delta H^\circ$ and $-T\Delta S^\circ$ in a coordinate system, based on the rearrangement of equation (2) to

$$-\Delta H^\circ = -\Delta G^\circ - T\Delta S^\circ \qquad (3)$$

Compilation of the available data and their computation in this coordinate system yields a strip of points which is delimited by parallel lines. Their intercepts result in values of -2 kcal mol^{-1} (K_D = 30 mM) and -9 kcal mol^{-1} (K_D = 200 nM) (24–27). Although the enthalpy of the association process can vary widely in a diagrammatic computation, the free energy appears to be nearly fixed, since any favorable gain by an enthalpic contribution is apparently neutralized by an entropic cost and vice versa. This compensation is effective irrespective of the nature of the receptor-ligand pairs. Further analysis of various receptor types has shown that the two competing influences can cover the correlated intervals between $-22 \leq \Delta H^\circ \leq 34$ kcal mol^{-1} and $-35 \leq \Delta S^\circ \leq 143$ cal K^{-1} mol^{-1} (30). Consequently, ΔG°-values should not pass the limit of -15 kcal mol^{-1} (K_D = 10 pM).

Whereas the actual occurrence of this phenomenon is not questioned, the physicochemical origin for the compensation is presently not precisely defined. Having first been attributed to the aqueous medium, in which biological interaction proceeds (31), the suggestion that the observed compensation arises from any weak intermolecular interaction, hydrogen-bonding in water just being the most frequently encountered example, can be regarded as an extension of the initial concept (30,32). When the intermolecular bond energies increase, measured as change in enthalpy, degrees of freedom will necessarily be restricted, forming the ensuing entropic cost. However, it is noteworthy that the range of entropic alterations is not limited to reduce the affinity. In those cases the favorable entropic factor, yielding an entropy-driven reaction, is counterbalanced by an enthalpic expense. To further delineate the individual components of the molecular basis of the well-accepted compensation, it is helpful to examine each reactant separately and to describe its behavior. This physicochemically meaningful strategy automatically leads to the three basic ideas, which are the core of the prevailing models. In detail, the contact establishment is inherently coupled to a loss of conformational degrees of freedom of the ligand and/or receptor and/or to reorganization of the solvation (24,27,33). Therefore, it is warranted to focus attention on the behavior of each reactant, starting with the ligand. This inspection is accompanied by the intention to point to possibilities for chemical manipulations aimed at the design of ligands with improved affinity. Since adaptation of shape and flexibility can be mastered synthetically, solid knowledge in this field is the basis for a rational receptor-targeted drug design.

The Ligand

As already emphasized, the precise representation of the conformational behavior of the ligand is a key to understand his functionality. If the range of spatial oscillations in solution

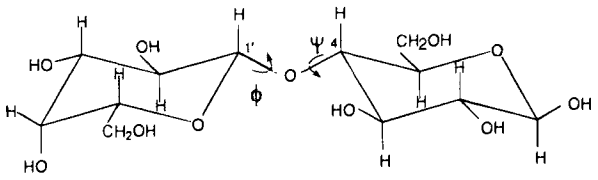

Fig. 2. Illustration of the two torsion angles ϕ and ψ of the glycosidic linkage in lactose. Changes in these two parameters describe the flexibility of the relative orientation of the two pyranose rings.

is rather small, the net entropic cost for a further immobilization in a binding pocket will not be considerable. The calculation of the impact of freezing of the intramolecular mobility requires having access to precise data on the number of conformers and the extent of population density for each conformational state. With respect to carbohydrates, molecular mechanics models define the relative spatial orientation of two sugar units by the two torsion angles about the glycosidic linkage (Fig. 2). With a consistent set of parameters for the description of the energy contents, referred to as a force field, iterative process steps yield a correlation of each devisable conformation to its relative energy content (25,28,34,35), as illustrated in Fig. 3. These procedures result in an energy-based category formation of the

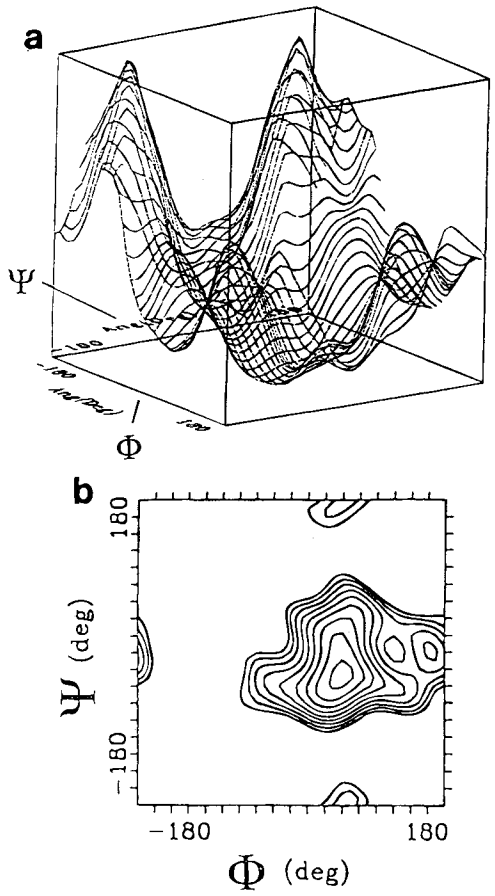

Fig. 3. Representation of the correlation between the values of the two torsion angles ϕ and ψ in a disaccharide (i.e. galβ1–3gal) and the energy level of any defined conformation in 3-dimensional (a) or 2-dimensional (b) schemes.

conformational space and are a step to infer to what extent ligands can be flexible under the given conditions. Remarkably, no generalization can be given for oligosaccharides. For example, as compiled for blood group-related epitopes recently (36), oligosaccharides can be very rigid (e.g. Le^x, Le^y), essentially rigid with at least one other conformation with 3–4% of the molecules forming this ensemble (H types 1,2 and 6) or rather flexible with distinct conformational families of almost equivalent importance (H types 3,4 and A types 1,2). Since the explicit inclusion of water molecules will drastically increase the number of atoms and consequently the number of equations expressing the motions, these simulations can probe the limits of the computer capacity. It is therefore conducive to gather experimental evidence on the atomic orientations in solution to corroborate the assumption of a significant, but limited motional flexibility.

NMR spectroscopy provides this indispensable input, albeit not to a degree to achieve absolute certainty about spatial coordinates. The analysis of the extent of proximity between interresidual protons in space serves to weight and limit the theoretically calculable conformational space. This dipolar interaction which involves magnetization transfer with a r^{-6}-dependence is termed nuclear Overhauser enhancement (NOE), effective up to an internuclear distance of 5 Å. With an abundance of these distance-dependent effects, the information can be computed into a molecular model, increasing in accuracy with the number of contacts. Unfortunately, the number of interresidue contacts is rather limited for oligosaccharides. Moreover, the time scale of the relaxation of the non-equilibrium state for nuclear spin energy levels relative to that of the dynamic fluctuations between conformers and the presence of distinct conformers itself are key factors which explain the caution that is exercized in the interpretation of such data sets. Conformational reorganization will lead to an averaging and to the depiction of a virtual conformation, to which a series of real conformations can contribute (24,37). Once the concept of interresidual motional flexibility is rightfully appreciated, the fact that a computable snapshot represents a time- and ensemble-averaged set of coordinates makes it readily explicable why computer-assisted calculations and NMR measurements establish an indispensable combination of mutually supplementing techniques.

This natural impediment may also shed light on the frustrating experience to crystallizing oligosaccharides. Thermodynamically, the entropic barrier built by the immobilization of torsion angles has been estimated to reach 1–2 kcal mol^{-1} at room temperature per affected angle in the carbohydrate ligand (24). When this barrier is purposefully lowered without risking loss of enthalpic stabilization, affinity enhancements are conceivable. Vice versa, strategies to improve the level of intermolecular interaction in terms of ΔH should avoid increasing the inherent flexibility of the ligand. Such target-directed approaches can affect the chemical structure and/or the presentation of the ligand. Intriguingly, the context of presentation of a distinct sugar sequence can well entail alterations in its conformational space and the dynamics of conformer transition. A glycan chain, e.g. at Asn297 on two IgG2a-type murine monoclonal antibodies (38), can be rendered more or less accessible by the context of its presentation, placing orientation on the same level as sequence as predictor of receptor selectivity. Although this attractive concept should presently not be oversold, its experimental elaboration may provide a (partial) answer to the tantalizing question why animal lectins with selectivity to abundantly expressed epitopes such as β-galactoside-containing termini or α2,6-sialylated derivatives thereof are capable of distinguishing between binding partners to remarkably different extents *in vitro* and *in vivo* (7). The consideration of motional flexibility is certainly not restricted to the ligand. It also applies to the receptor.

The Receptor

As already noted, the currently available evidence intimates that no major conformational changes in a carbohydrate-binding protein upon ligand binding. Although examples with a C-type lectin and hevein domain-containing plant agglutinins have revealed transitions of arginine/lysine or aromatic side chain presentation by ligand binding (39,40), it is nonetheless justified to focus further attention on the binding pocket. Its architecture is composed of various side chains, several of them participating in the polar and non-polar modes of interaction. Refined structural analysis, e.g., for the wheat germ agglutinin has demonstrated that limited ligand-dependent rearrangements of several residues such as Tyr64 and Ser114 take place (41). Since the resolution of the NMR spectrum of proteins with a molecular weight > 20 kDa is a formidable task, molecular dynamics simulations can be instrumental to estimate the entropic consequences of ligand binding for the receptor. In RNA-binding sites the importance of already restricted mobility of the contact sites prior to complex formation has been underscored (42). Alleviating spatial constraints and increasing their mobility yielded inactive variants, a proof for the impact of non-interacting side chains on affinity. The incorporation of Ca^{2+}-ions into the binding domain can thus be rationalized as an efficient way to reduce the overall entropic cost by preformation of an adapter surface for the ligand's shape. Disparate side-chain mobility in the vicinity of the ligand has also been attributed to the different binding behavior patterns of two legume lectins despite their conspicuous sequence similarities (43).

Not only side chains of the recognition domain architecture, but also other parts of a protein can have an influence on the binding process. The isosteric C2S-substitution in mammalian galectin-1, located approximately 20 Å from the binding domain, apparently exerts long-range effects with impact on the thermodynamics (44). Already subtle structural changes can evidently bring about compensating enthalpy/entropy alterations. Thus it is not surprising to observe that in the classes of lectins and antibodies, which preferentially bind to their ligands driven by dominant enthalpic and entropic contributions, respectively, exceptions occur, defying the general rule (15,26,45). This variability also holds true for agonist/antagonist binding, revealing an inverse compensating relationship for example in the cases of adenosine and β-adrenergic receptors (30) or for binding of drugs (daunomycin, ethidium bromide and netropsin) to synthetic DNA polymers (alternating copolymer and homopolymer duplexes) (46). Therefore, one has to acknowledge the remarkable problems in reliably predicting the molecular causes for affinity. Collection of data on the complex in solution can fill gaps in our understanding.

The Complex

One can probe the structural aspects of groups that are crucial to complex formation by synthesizing tailor-made ligand derivatives. For example, systematic substitutions of a single hydroxyl group by a C-H group in the ligand structure provides information on its actual hydrogen-bonding potential in the interaction and a means to experimentally verify the crystallographic results (27,47). If a ligand is not firmly associated with its receptor in solution and thus exchanges rapidly between free and bound states, transferred NOE effects can be detected (48). Similar to NOEs in free-state analysis, their presence as sharp signals again serves as an atomic ruler for the calculation of internuclear distances (28,48,49). An example of such a spectrum with signals indicative of intra- and interresidual contacts of ligand protons is shown in Fig. 4. Since magnetization transfer is not confined to dipolar through-space contacts of the ligand, it is essential that the participation of protein protons is experimentally excluded to avoid being led astray, e.g. by keeping experimental mixing times fairly short at low ligand concentrations and by monitoring the NOEs in the rotating frame under spin-locking conditions (28,50,51). The available evidence supports the view that a low-energy conformation of the ligand, being identical or close to the global energy minimum conformation, fits into the binding site (7,15,28,50,51). Conformer selection from different positions in the low-energy valley is possible, as shown for selectins or two galactoside-binding lectins (50,51).

Association to the lectin does not necessarily mean that the ligand abandons its flexibility about the glycosidic angles. Residual mobility of connected saccharide parts is still visible in several plant lectin complexes, namely *Aleuria aurantia* agglutinin/L-fuc-α1,6-glcNAcβ-1,R, lentil lectin/sucrose, ricin/methyl β-allolactoside and mistletoe lectin/gal-β1,3-galNAc-β1,R complexes (51–54). The ligand's mobility is thus not fully frozen to the benefit of the entropic contribution. On the contrary, the lack of a rigid association of the ligand over its entire length in such a case precludes strong intermolecular interactions as a source for a further enthalpic gain, providing a graphic illustration of enthalpy/entropy compensation. In cases where the affinity of the glycan is too high and prevents a rapid exchange, isotope-edited NMR-techniques are pertinent to probe structural aspects of the ligand, as documented recently for estrone-3-glucuronide in complex with a single-chain fragment of a specific antibody (55)

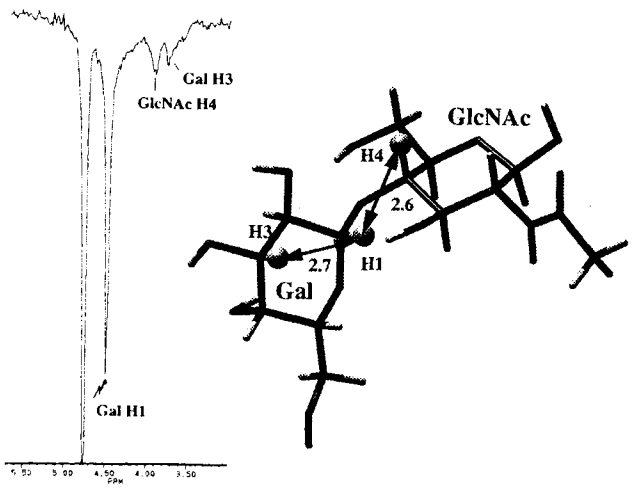

Fig. 4. Illustration of a 1-dimensional trNOE-spectrum of N-acetyllactosamine in the presence of the galactoside-binding mistletoe lectin and of the two contact points of the anomeric proton H1 of D-galactose at a distance of 2.7 Å or 2.6 Å in a molecular model of the disaccharide.

As already indicated in the description of the crystallographic view of the complex, solvent molecules can form an integral part of the interacting surface. Their role in modulating affinity and/or specificity of the overall interaction process extends beyond protein-carbohydrate recognition (17,56). But its participation in any biomolecular recognition is not the only reason to pay attention to the solvent. Conventional thermodynamic accounting requires to separate the overall enthalpy/entropy terms to separate the terms into individual fractions, which cannot neglect the solvent. Both partners of a recognition system are surrounded by water molecules. Notwithstanding the lack of a comprehensive physicochemical knowledge about the structure of bulk water and water in the vicinity of any surface part of biomolecules, the discussion of the solvent's properties reveals further important driving forces to attain populated positions in the equilibrium between association and dissociation. As a consequence, the discussion of the solvent, which is given in the next chapter, completes the crucial factors and at the same time highlights the possibility that multiple cooperative and opposing mechanisms are at disposition.

The Solvent

Placing emphasis on the solvent, any solution with reactants can be divided into normal bulk solvent, solvent molecules, which belong to the hydration shells of ligand/receptor, and firmly bound solvent molecules as part of the receptor/ligand. The boundaries between the categories should not be drawn too strictly, because water molecules at interfaces in the complex may well have residence times up to the msec-range, maintaining a partial disorder as protection from the high entropic penalty of forming a rigid surface (23,56). Maintaining a residual mobility while enthalpically contributing to the fluctuating network of hydrogen bonds forms a delicate balance. The answer to the question which factors will be dominant in the overall summation must take the structural features of the interacting solute surfaces into account. Release of water molecules from the vicinity of ionic/polar groups will be connected to an increase in entropy, because their motional freedom in bulk solvent is less restricted than in the hydration shell. Although the actual origin of the hydrophobic effect is still disputed (57,58), the entropically favorable release of ordered water molecules from strictly hydrophobic areas such as those of several sugar-binding monoclonal antibodies or the TATA box-binding protein upon complexation is considered as a further example of how solvent reorganization can influence the overall thermodynamics.

As already pointed out in a preceding paragraph, carbohydrate ligands harbor polar and non-polar epitopes, creating a polyamphiphilic surface with complementarity to the receptor. Again, it is not precisely clear how the arrangements within the hydration shells deviate from the bulk solvent. Monte Carlo simulations for a plant lectin have indicated that the perturbation

can lead to an enthalpically disadvantaged state with considerable mobility relative to bulk solvent and that it attenuates over a distance of three water molecules (59). The return of these water molecules to the bulk solvent in the course of complex formation can then provide an enthalpic driving force, as they leave the influence of the geometric constraints. They are assumed to be imposed on the water molecules by the polyamphiphilic surface that accounts for the increased disorder and the broken hydrogen bonds. Indeed, measurements of the thermodynamics of concanavalin A and a sugar ligand in D_2O confirm that solvent reorganization contributes to the measured net enthalpy (60). Notably, isotopic substitution of deuterium for protium is not without effect on the weak bonding force. The deuterium bonds are stronger and more localized than hydrogen bonds. Since NMR measurements are often performed in this solvent, the enthalpic gain and the entropic loss incurred by a change of medium document that the interpretation of spectroscopic data should not overlook this compensatory mechanism and the deviation from the situation in H_2O. Overall, the term "solvent reorganization" summarizes all types of solvent-solute and solvent-solvent interactions. The experimental evidence that the influence of the solvent will apparently have a bearing on the enthalpy of binding of various receptor types to varying extents underscores the interrelation of diverse parameters to be reckoned with in the understanding of the binding process (60). The ranking of each factor should be performed meticulously.

Although each calculation currently relies on estimates, an approximation of the individual components of the total entropy ΔS° can be attempted. Besides the term arising from solvation (ΔS°_{solv}) the complex formation is linked to changes in configurational, rotational and translational entropy (ΔS°_{config}, ΔS°_{rot}, ΔS°_{trans}). Using literature data on the cratic term which arises from the rotational and translational freezing of receptor and ligand by 1:1 complex formation (-8 cal mol^{-1} K^{-1}) and on ΔC°_p-values as an indicator of changes of solvent structure to estimate ΔS°_{solv} (26,61), the unitary part (which is composed of the first two terms) then presents only one unknown variable. Together with the measured total enthalpy change, ΔS°_{config} calculations provide a rough estimate on the extent of reduction of mobility of the protein (side chains and/or backbone) and/or the ligand (26).

Computer calculations together with measurements on a series of ligands are instrumental in relating structural details to thermodynamic data. Keeping the range of compensatory enthalpy/entropy changes in the respective plot and the given examples for obvious influences of subtle structural alterations on the thermodynamics in mind, it can be expected that cases will be found for each reasonable combination. As emphasized in the beginning of this section, the individual players of the recognition game cannot be viewed in isolation and the purposefully, albeit artificially separated discussion has been designed to underscore that each outlined factor can be involved in the overall process to a varying degree for different systems.

CONCLUSIONS

Acquisition of reliable information about structural and thermodynamic aspects of protein-carbohydrate interaction is the basis for any rational lectin-targeted drug design. The governing enthalpy/entropy-compensation is a gripping challenge, but still no reason to complain. The precise knowledge of the topology of a receptor's binding pocket can pave the way for procurement of a ligand structure which optimally exploits any favorable interaction. The hydrophobic space in the glycerol-binding subsite of influenza virus neuraminidase has thus been exploited to accommodate a bulky lipophilic group, namely a branched alkyl group, to gain an additional favorable hydrophobic interaction (62). However, engineering of the ligand for an assumed increased interaction potential may unfortunately be linked with loss or perturbation of those contacts which have been taken for granted, as work on glycogen phosphorylase inhibitors has elegantly proven (63). If such an alteration causes the sum of the individual contributions to fall short of the required threshold, the modified ligand is of no practical value.

On the level of the receptor, the introduction of mutations, as already mentioned for protein-RNA interaction by an aminoacyl-tRNA synthetase (42), will likewise modulate the overall affinity. The impact of such changes can even reach an enormous clinical importance. Emergence of certain protease-directed drug-resistant HIV-1 variants can be attributed to the key residues Val82 and Ile84. Mutational substitutions decrease the van der Waals contacts, thereby calling for larger or more flexible cyclic urea-type inhibitors to combat this complication (64). In view of the discussion in this review this task will not be trivial, as e.g. the entropic factor by immobilizing the flexible ligand may outweigh the enthalpic gain, again underscoring the delicate balance between the complex array of involved factors. When designing e.g. rigidified glycomimetics or lectin-like analogues to reduce the entropic cost, the elucidation of natural pathways how molecular association may still reach encounter-limited characteristics without suffering from kinetic impediments can be pertinent. The naturally occurring combination of two recognition sites with differing target selectivities has proven beneficial for solving this problem (7). The combination of carbohydrate recognition with ionic or hydrophobic interactions by compounds of suitable geometry for presentation of the crucial groups is already engendering intense activities by organic chemists. For example, the sugar part of a selectin ligand has been coupled synthetically at its reducing 1-position to a branched alkyl chain (2-tetradecylhexadecyl, mimicking the natural ceramide) to anchor the resulting conjugate to hydrophobic patches on the selectins' surfaces (65). This strategy may also aid to develop ligand analogues which can distinguish related receptor types with structural differences close to the binding sites. Due to the complicated question on the role of solvent molecules during the association process the educated guesses based on computer simulations or analogy require extensive experimental testing. In addition to the ligand clustering which affords an avidity effect for neoglycoconjugates, exploitable in diagnostic and therapeutic approaches (7,66–68), the list of chemical and physical factors whose tailoring can allow optimal presentation of an ideal shape for high-affinity binding can be derived from individual inspection for all compounds involved. Custom-made structural manipulations as a way of drug design as well as further insights on the role of context-specific ligand presentation are expected to form burgeoning branches of glycosciences in the next years.

ACKNOWLEDGMENTS

I am indebted to Dr. H.-C. Siebert and Dr. C.-W. von der Lieth for artful computer graphics.

REFERENCES

1. R. A. Laine. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives,* Chapman & Hall, Weinheim—London, 1997, pp. 1–14.
2. I. Brockhausen and H. Schachter. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives*, Chapman & Hall, Weinheim—London, 1997, pp. 79–112.
3. A. Varki. *Glycobiology* **3**:97–130 (1993).
4. E. F. Hounsell, M. J. Davies, and D. V. Renouf. *Glycoconjugate J.* **13**:19–26 (1996).
5. N. Sharon and H. Lis. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives,* Chapman & Hall, Weinheim—London, 1997, pp. 133–162.
6. J. Kopitz. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives,* Chapman & Hall, Weinheim, 1997, pp. 163–189.
7. H.-J. Gabius. *Eur J. Biochem.* **243**:543–576 (1997).
8. H. Rüdiger. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives,* Chapman & Hall, Weinheim—London, 1997, pp. 415–438.
9. A. Imberty, Y. Bourne, C. Cambillau, P. Rougé, and S. Pérez. *Adv. Biophys. Chem.* **3**:71–117 (1993).
10. N. Sharon. *Trends Biochem. Sci.* **18**:221–226 (1993).
11. J. M. Rini. *Annu. Rev. Biophys. Biomol. Struct.* **24**:551–577 (1995).
12. W. I. Weis and K. Drickamer. *Annu. Rev. Biochem.* **65**:441–473 (1996).
13. F. A. Quiocho. *Pure & Appl. Chem.* **61**:1293–1306 (1989).
14. G. Davies and B. Henrissat. *Chemistry & Biology* **3**:853–859 (1995).
15. D. R. Bundle. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives,* Chapman & Hall, Weinheim—London, 1997, pp. 311–331.
16. P. A. Karplus and C. Faerman. *Curr. Opinion Struct. Biol.* **4**:770–776 (1994).
17. J. E. Ladbury. *Chemistry & Biology* **3**:973–980 (1996).
18. M. Levitt and B. H. Park. *Structure* **1**:223–226 (1993).
19. R. Loris, P. P. G. Stas, and L. Wyns. *J. Biol. Chem.* **269**:26722–26733 (1994).
20. J. Emsley, H. E. White, B. P. O'Hara, G. Oliva, N. Srinivasan, I. J. Tickle, T. L. Blundell, M. B. Pepys, and S. P. Wood. *Nature* **367**:338–345 (1994).
21. G. Wagner, S. G. Hyberts, and T. F. Havel. *Annu. Rev. Biophys. Biomol. Struct.* **21**:167–198 (1992).
22. M. W. MacArthur, P. C. Driscoll, and J. M. Thornton. *Trends Biotechnol.* **12**:149–153 (1994).
23. J. D. Dunitz. *Science* **264**:670 (1994).
24. J. P. Carver. *Pure & Appl. Chem.* **65**:763–770 (1993).
25. S. Pérez, A. Imberty, and J. P. Carver. *Adv. Computat. Biol.* **1**:147–202 (1994).
26. E. J. Toone. *Curr. Opinion Struct. Biol.* **4**:719–728 (1994).
27. R. U. Lemieux. *Acc. Chem. Res.* **29**:373–380 (1996).
28. H.-C. Siebert, C.-W. von der Lieth, M. Gilleron, G. Reuter, J. Wittmann, J. F. G. Vliegenthart, and H.-J. Gabius. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives,* Chapman & Hall, Weinheim, 1997, pp. 291–310.
29. C. Bjurulf and I. Wadsö. *Eur. J. Biochem.* **31**:95–102 (1972).
30. P. Gilli, V. Ferretti, G. Gilli, and P. A. Borea. *J. Phys. Chem.* **98**:1515–1518 (1994).
31. R. Lumry and S. Rajender. *Biopolymers* **9**:1125–1227 (1970).
32. J. D. Dunitz. *Chemistry & Biology* **2**:709–712 (1995).
33. M. R. Eftink, A. C. Anusiem, and R. L. Biltonen. *Biochemistry* **22**:3884–3896 (1983).
34. S. W. Homans. In J. Montreuil, J. F. G. Vliegenthart, and H. Schachter (eds.) *Glycoproteins,* Elsevier Science B.V., Amsterdam, 1995, pp. 67–86.
35. C.-W. von der Lieth, T. Kozar, and W. E. Hull. *J. Mol. Struct.* **395–396**:225–244 (1997).
36. A. Imberty, E. Mikros, J. Koca, R. Mollicone, R. Oriol, and S. Pérez. *Glycoconjugate J.* **12**:331–349 (1995).
37. O. Jardetzky. *Biochim. Biophys. Acta* **621**:227–232 (1980).
38. K. D. White, R. D. Cummings, and F. J. Waxman. *J. Immunol.* **158**:426–435 (1997).
39. H. Komano, T. Kurama, Y. Nagasawa, and S. Natori. *Biochem. J.* **284**:227–230 (1992).
40. H.-C. Siebert, C.-W. von der Lieth, R. Kaptein, J. J. Beintema, K. Dijkstra, N. van Nuland, U. M. S. Soedjanaatmadja, A. Rice, J. F. G. Vliegenthart, C. S. Wright, and H.-J. Gabius. *PROTEINS* **28**:268–284 (1997).
41. C. S. Wright. *J. Mol. Biol.* **215**:635–651 (1990).
42. L. Ribas de Pouplana, D. S. Auld, S. Kim, and P. Schimmel. *Biochemistry* **35**:8095–8102 (1996).
43. M. C. Chervenak and E. J. Toone. *Biochemistry* **34**:5685–5695 (1995).
44. D. Gupta, M. Cho, R. D. Cummings, and C. F. Brewer. *Biochemistry* **35**:15236–15243 (1996).
45. E. M. Nashed and C. P. J. Glaudemans. *J. Biol. Chem.* **271**:8209–8214 (1996).
46. K. J. Breslauer, D. P. Remata, W. Y. Chou, R. Ferrante, J. Curry, D. Zaunczkowski, J. G. Snyder, and L. A. Marky. *Proc. Natl. Acad. Sci. USA* **84**:8922–8926 (1987).
47. D. Solís and T. Diaz-Mauriño. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives,* Chapman & Hall, Weinheim, 1997, pp. 345–354.
48. F. Ni. *Progr. Nucl. Magn. Reson. Spectrosc.* **26**:517–606 (1994).
49. T. Peters and B. M. Pinto. *Curr. Opinion Struct. Biol.* **6**:710–720 (1996).
50. L. Poppe, G. S. Brown, J. S. Philo, P. V. Nikrad, and B. H. Shah. *J. Am. Chem. Soc.* **119**:1727–1736 (1997).
51. M. Gilleron, H.-C. Siebert, H. Kaltner, C.-W. von der Lieth, T. Kozar, K. M. Halkes, E. Y. Korchagina, N. V. Bovin, H.-J. Gabius, and J. F. G. Vliegenthart. (in press).
52. T. Weimar and T. Peters. *Angew. Chem. Int. Ed. Engl.* **33**:88–91 (1994).
53. J. L. Asensio, F. J. Cañada, and J. Jiménez-Barbero. *Eur. J. Biochem.* **233**:618–630 (1995).
54. F. Casset, T. Hamelryck, R. Loris, J. R. Brisson, C. Tellier, M. H. Dao-Thi, L. Wyns, F. Poortmans, S. Pérez, and A. Imberty. *J. Biol. Chem.* **270**:25619–25628 (1995).
55. D. G. Low, M. A. Probert, G. Embleton, K. Seshadri, R. A. Field, S. W. Homans, J. Windust, and P. J. Davis. *Glycobiology* **7**:373–381 (1997).
56. J. W. R. Schwabe. *Curr. Opinion Struct. Biol.* **7**:126–134 (1997).
57. W. Blokzijl and J. B. F. N. Engberts. *Angew. Chem.* **105**:1610–1648 (1993).
58. B. Lee. *Meth. Enzymol.* **259**:555–576 (1995).
59. H. Beierbeck, L. T. J. Delbaere, M. Vandonselaar, and R. Lemieux. *Can. J. Chem.* **72**:463–470 (1994).
60. M. C. Chervenak and E. J. Toone. *J. Am. Chem. Soc.* **116**:10533–10539 (1994).
61. J. M. Sturtevant. *Proc. Natl. Acad. Sci. USA* **74**:2236–2240 (1977).
62. 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. *J. Am. Chem. Soc.* **119**:681–690 (1997).
63. K. A. Watson, E. P. Mitchell, L. N. Johnson, J. C. Son, C. J. F. Bichard, M. G. Orchard, G. W. J. Fleet, N. G. Oikonomakos, D. D. Leonidas, M. Kontou, and A. Papageorgioui. *Biochemistry* **33**:5745–5758 (1994).
64. P. J. Ala, E. E. Huston, R. M. Klabe, D. D. McCabe, J. L. Duke, C. J. Rizzo, B. D. Korant, R. J. DeLoskey, P. Y. S. Lam, C. N. Hodge, and C.-H. Chang. *Biochemistry* **36**:1573–1580 (1997).
65. H. Tsujishita, Y. Hiramatsu, N. Kondo, H. Ohmoto, H. Kondo, M. Kiso, and A. Hasegawa. *J. Med. Chem.* **40**:362–369 (1997).
66. H.-J. Gabius, K. Kayser, and S. Gabius. *Naturwissenschaften* **82**:533–543 (1995).
67. S. Gabius, K. Kayser, N. V. Bovin, N. Yamazaki, N. Kojima, H. Kaltner, and H.-J. Gabius. *Eur. J. Pharm. Biopharm.* **42**:250–261 (1996).
68. R. T. Lee and Y. C. Lee. In H.-J. Gabius and S. Gabius (eds.) *Glycosciences: Status and Perspectives,* Chapman & Hall, Weinheim, 1997, pp. 55–77.