JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Exploring Carbohydrate−Peptide Interactions in the Gas Phase: Structure and Selectivity in Complexes of Pyranosides with *N*-Acetylphenylalanine Methylamide

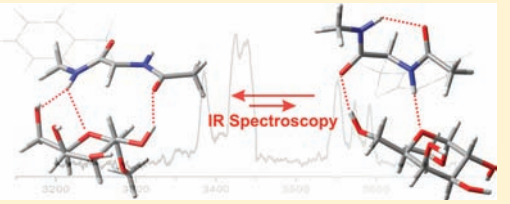
Emilio J. Cocinero,[†,§] Pierre Çarçabal,[†,∥] Timothy D. Vaden,[†,⊥] Benjamin G. Davis,[*,‡] and John P. Simons[*,†]

[†]Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, U.K.

[‡]Chemistry Research Laboratory, Department of Chemistry, University of Oxford, Mansfield Road, Oxford OX1 3TA, U.K.

Supporting Information

ABSTRACT: The physical basis of carbohydrate−peptide interactions has been explored by probing the structures of a series of complexes generated in a solvent-free environment under molecular beam conditions. A combination of double-resonance IR−UV spectroscopy and quantum-chemical calculations has established the structures of complexes of the model, *N*-acetyl-L-phenylalanine methylamide, bound to the α and β anomers of methyl D-gluco- and D-galactopyranoside as guests. In all cases, the carbohydrates are bound through hydrogen bonding to the dipeptide chain, although with some differing patterns. The amino acid host "engages" with the most suitable pair of neighboring conjugate sites on each carbohydrate; the specific choice depends on the conformation of the peptide backbone *and* the configuration and conformation of the carbohydrate ligand. None of the structures is supported by "stacking" interactions with the aromatic ring, despite their common occurrence in bound carbohydrate−protein structures.



INTRODUCTION

Interactions between carbohydrates and proteins are intimately involved in the biological machinery of a vast number of biological processes: mammalian infection, inflammation, fertilization, and cellular recognition provide just a few examples.[1] Selective molecular recognition by specific protein carbohydrate binding modules[2] is controlled by the disposition, conformation, and configuration of the locally interacting sites and the subtle interplay of the multivalent interactions between the protein and the ligand, which are often mediated by neighboring water molecules and ions, typically Ca^{2+}.

Crystallographic investigations of the structures of protein-bound carbohydrates in the solid state and NMR spectroscopic structural investigations conducted in solution have revealed the operation of several different types of interaction.[3] Hydrogen bonds readily form between carbohydrates and proteins, predominantly to the backbone and side groups such as amide or carboxylate groups on acidic amino acids (asparagine and glutamine or aspartate and glutamate) or protonated side groups on basic amino acids (e.g., lysine or arginine). Further impetus can be provided by hydrophobic interactions between the apolar regions of the carbohydrate and aromatic residues (tyrosine, tryptophan, and, less commonly, phenylalanine)[2a] promoted perhaps (in aqueous environments) by the exclusion of surrounding water molecules from the apolar interface,[4] but binding to aromatic residues could also involve a contribution from dispersive so-called CH−π interactions.[5] Indeed, stacked structures supported solely by weak, dispersive CH−π interactions between pyranose and aromatic rings have been identified recently through vibrational spectroscopic interrogation of carbohydrate−arene complexes isolated in the gas phase and stabilized at low temperatures in a molecular beam.[6,7]

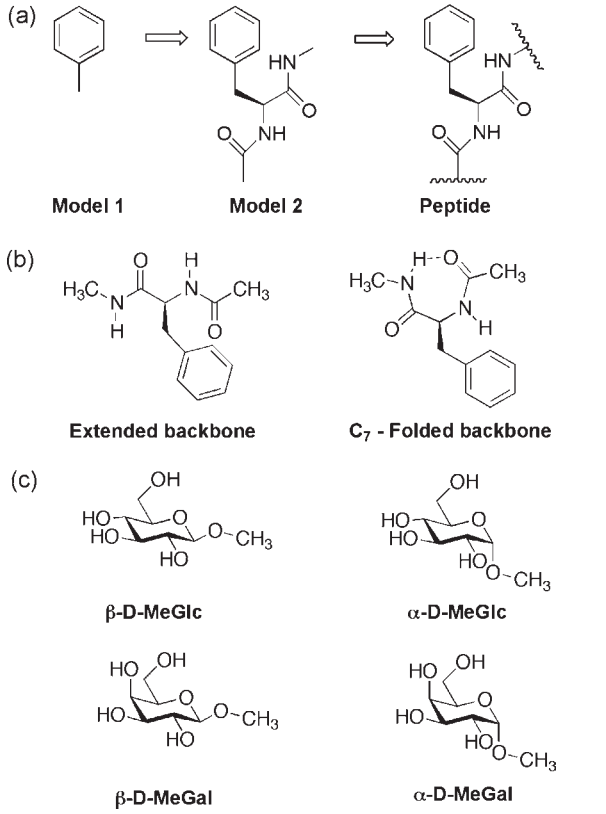
These spectroscopic investigations, using toluene as a truncated model for phenylalanine, focused solely on the carbohydrate−aromatic interactions and were designed to exclude any competition from hydrogen-bonded interactions involving peptide or neighboring water molecules. If a peptide chain *had* been present and polar interactions were found to be dominant within the isolated complexes, then the relative prevalence of "stacked" aromatic−carbohydrate structures in the architecture of natural protein−carbohydrate complexes in aqueous environments could be assessed. It might implicate exclusion of water from the hydrophobic interface rather than the weak CH−π interactions identified in isolated arene complexes. In crystalline environments, stacked carbohydrate−protein structures could be "frozen in", supported by a network of hydrogen bonds between the protein and the trapped carbohydrate.

To address these issues,[8] test the "hydrophobic binding" hypothesis,[4] and also probe potential origins of conformational selectivity (and molecular recognition) in carbohydrate−peptide interactions, we developed a modular approach in which a truncated peptide segment was attached to the aromatic unit (model 1) to create a suitable motif, *N*-acetyl-L-phenylalanine methylamide, a dipeptide identified as model 2 in Scheme 1a. The isolated motif adopts two alternative conformations,[9] either extended, providing a template for a dimeric β-sheet structure, or folded, incorporating a C_7

Received: November 25, 2010
Published: March 04, 2011

dx.doi.org/10.1021/ja109664k | J. Am. Chem. Soc. 2011, 133, 4548–4557

Scheme 1. Schematic Structural Representations of (a) the Development from a Truncated Aromatic (Model 1)[6] to a Truncated Peptide (Model 2) to a Polypeptide; (b) the Major Backbone Conformations of the Peptide Model, *N*-Acetyl-L-phenylalanine Methylamide; and (c) the Carbohydrate Partners, the α and β Anomers of Methyl D-Gluco- and D-Galactopyranoside (α/β-D-MeGlc and α/β-D-MeGal)



γ-turn linking the "outer" methylaminyl (MeNH) and acetyl (MeC═O) groups (Scheme 1b). Its carbohydrate guests, the α and β anomers of methyl D-gluco- and D-galactopyranoside (α/β-MeGlc and α/β-MeGal) (Scheme 1c), provide a structurally varied set of ligands. Substituting MeGlc for MeGal changes the configuration at OH-4 from equatorial (MeGlc) to axial (MeGal), and changing the anomeric structure from α to β switches the configuration at O-1 from axial (α) to equatorial (β). Their choice recognizes (and explores) the importance of two of the general factors controlling selective bonding and chiral recognition, identified by Zehnacker and Suhm[10] as molecular flexibility (provided here by the exocyclic hydroxymethyl group) and cooperativity (provided by the potential for generating chains of inter- and intramolecular hydrogen bonds). The architectures of the resulting carbohydrate−amino acid complexes were characterized in the gas phase as before[6] through a combination of mass-selected electronic and vibrational spectroscopy, using tunable IR laser radiation to probe the OH (carbohydrate) and NH (peptide) vibrational bands, and an iterative sequence of force-field and density functional theory (DFT) calculations to analyze and interpret the experimental data.

METHODS

Vaporization, Complex Formation, and Spectroscopy. *N*-acetyl-L-phenylalanine methylamide and the α and β anomers of methyl D-gluco- and D-galactopyranoside were synthesized or obtained as commercial samples. The host−carbohydrate complexes were generated in the gas phase using a combination of pulsed laser desorption and molecular beam procedures similar to those described earlier.[6b,c] Ground powder samples of the carbohydrate, α/β-MeGlc or α/β-MeGal, and the capped amino acid in a 5:1 molar ratio were thoroughly mixed with graphite powder [~20:80 (w/w) graphite/organic], and this mixture was deposited as a thin uniform surface layer on a graphite substrate and placed in a vacuum chamber close to and just below the exit of a pulsed cylindrical-nozzle expansion valve (0.8 mm diameter). The species desorbed from the surface were entrained and cooled in an expanding argon jet (~4 bar backing pressure) before passing into the detection chamber through a 2 mm diameter skimmer to create a collimated molecular beam. This was crossed by one, two, or three tunable laser beams in the extraction region of a linear Wiley−McClaren time-of-flight mass spectrometer (R. M. Jordan). Mass-resolved one- or two-color resonant two-photon ionization (R2PI) spectra of selected molecular complexes were recorded using one or two frequency-doubled pulsed Nd:YAG-pumped dye lasers(Lambda Physik and Sirah). In single-laser experiments, these were operated at 10 Hz, but in double-resonance experiments, the first laser (first UV or IR photon) was pulsed at 5 Hz and the second one at 10 Hz to allow subtraction of the background signals. The conformer-specific vibrational spectra were subsequently recorded in the OH and NH stretch regions by IR ion-dip (IRID) double-resonance spectroscopy[11] using UV radiation tuned onto selected R2PI absorption bands and tunable IR radiation in the range 3200−3800 cm^{-1}. The IR radiation (line width 2−3 cm^{-1}, 5 mJ/pulse) was provided by the idler output of an OPO/OPA laser system (LaserVision) pumped by a pulsed Nd/YAG laser (Continuum, Surelite II). Several spectra, typically ≥5, were recorded and averaged to achieve acceptable signal-to-noise (S/N) ratios.

Computational Strategy. The structural search followed an iterative approach that began with a series of unrestricted surveys of the many possible complex structures until no additional new structures were obtained. These surveys, which generated 3000−4000 structures, were conducted using the Monte Carlo multiple-minimization procedure and the large-scale low-frequency-mode torsional sampling procedures implemented in the MacroModel software (MacroModel version 8.5.207, Schrödinger, LLC).[12] The initial sets of structures were grouped into families distinguished by their association with extended or folded peptide backbones and by the number of strong hydrogen-bonded interactions they presented. Subsequent geometry optimization of the lower-energy members of each family (typically ~50 structures with relative energies ≤15 kJ mol^{-1}) using DFT with the "dispersion-sensitive" M05-2X functional[13] and a 6-31+G** basis set, as implemented in the Gaussian 03 suite of programs,[14] led to a new set of relative energies (corrected for zero-point energy), molecular structures, and vibrational spectra that could then be compared with experiment. The predicted harmonic wavenumbers of the NH and OH stretch modes were scaled by the recommended[15] factor of 0.9419 to accommodate anharmonicity and aid comparisons with the experiment. Depending on the level of agreement between the computed and experimental spectra, the structures that provided the closest levels of agreement, judged first by the degree of correspondence between the experimental and computed wavenumbers of each vibrational band and second by their calculated relative energies, were further optimized, guided by the experimental spectra. The process was repeated iteratively to minimize the calculated relative energies and achieve the best accord with the experimental spectra.

RESULTS AND DISCUSSION

The Isolated Partners. The capped amino acid and the carbohydrates, α/β-MeGlc and α/β-MeGal, are all flexible molecules, and the conformations they favor when isolated in

dx.doi.org/10.1021/ja109664k | J. Am. Chem. Soc. 2011, 133, 4548–4557

the gas phase need not be the same as those they adopt when they are bound within a molecular complex. The backbone of the isolated capped amino acid can adopt extended and folded conformations (see Scheme 1b), and in principle, both the "outer" and "inner" pairs of NH and C═O groups could provide binding sites for the carbohydrates. In the folded structure, however, the choice is constrained by the *intra*molecular hydrogen bond linking the two "outer" groups, MeNH and MeC═O, which blocks one of the pairs of potential *inter*molecular hydrogen-bonding sites (see Scheme 1). This opens up the possible role of peptide conformational selectivity within carbohydrate−peptide complexes—an aspect of molecular recognition. Similarly, the isolated carbohydrates may populate a number of alternative rotameric conformations, and peptide complexes selecting conformations that enhance and optimize the network of inter- and intramolecular hydrogen bonding may emerge. This type of behavior has been observed previously in gas-phase hydrated carbohydrate complexes, where conformational optimization, for example, through rotation of the exocyclic hydroxymethyl group from a G+ to a G− conformation coupled with a switch in the cyclic hydroxyl orientation from counterclockwise (OH-4 → OH-3 → OH-2 → O-1) to clockwise (OH-2 → OH-3 → OH-4),[16b,c] can create deep, localized water-binding pockets and strong regioselectivity for water insertion.[16]

Electronic Spectroscopy: Qualitative Considerations. Figure 1 compares the UV R2PI spectrum of isolated *N*-acetyl-L-phenylalanine methylamide with the corresponding spectra recorded (in the parent ion channels) for the peptide bound to the α and β anomers of MeGal and MeGlc. [Since the one-color R2PI spectrum of the α-D-MeGlc complex was heavily saturated (see the Supporting Information), it was recorded under two-color conditions using a low-intensity tunable laser to excite the initial resonant transition and a much more intense single-frequency laser pulse (at 37 244 cm^{-1}) delayed by 50 ns to create the ion]. A weak band in the one-color R2PI spectrum that appeared at ∼37 450 cm^{-1} (shown in the Supporting Information) also appeared in the hydrated ion channel and was associated with a fragmenting ternary hydrated complex, [β-D-MeGlc−*N*-acetyl-L-phenylalanine methylamide−H_2O]$^+$.

Since the aromatic side chain provides the UV chromophore, the R2PI spectra largely reflect the environment of the amino acid host. The peaks labeled I, II, and III in the isolated host (following Gerhard's notation[9]) have been associated, through subsequent interrogation of their IRID vibrational spectra coupled with ab initio calculations,[9] with the band origins of its extended (I) and folded (II and III) conformations; conformers II and III differ only in the orientation of the aromatic side group. The spectrum associated with conformer I includes a short vibronic progression, and the weak feature at ∼37 415 cm^{-1}, to the red of its band origin, can be assigned as a "hot band".

Qualitative comparisons between the R2PI spectrum of *N*-acetyl-L-phenylalanine methylamide and the succeeding spectra associated with its carbohydrate complexes reveal two striking patterns: those associated with the β anomers present very similar UV spectral signatures that are in close correspondence with the bands associated with the extended conformer (I), but no features appear in the regions associated with the folded conformers (II and III). This suggests the exclusive selection of the extended peptide backbone by both β anomers, β-D-MeGlc and β-D-MeGal, despite their different configurations at OH-4.

The spectra associated with the α anomers are more complicated. Each displays a feature at low wavenumber, close to the band origin of conformer III. The α-D-MeGal complex also displays a band close to that for conformer I, suggesting the population of extended *and* folded peptide structures; the bands appearing at intermediate wavenumbers perhaps suggest the population of additional structures. Since the intense band generated by the α-D-MeGlc complex at ∼37 480 cm^{-1}, which might also be related to conformer I, is significantly displaced from the corresponding bands in the other complexes, further discussion of it is deferred until additional information becomes available.

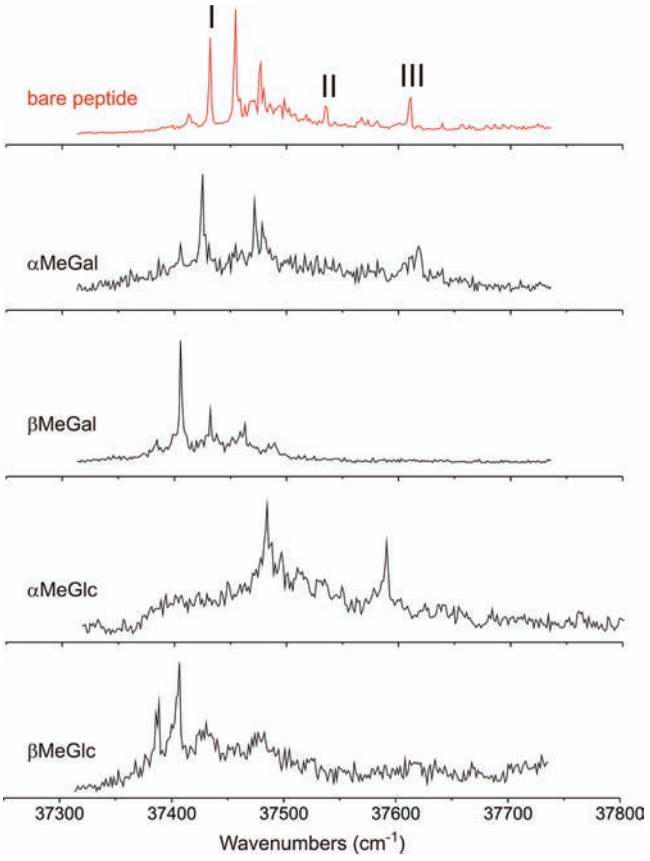


Figure 1. R2PI spectra of *N*-acetyl-L-phenylalanine methylamide and its complexes with the α and β anomers of the carbohydrates MeGlc and MeGal, recorded in their respective parent ion channels.

Vibrational Spectroscopy: Quantitative Considerations. *The β Anomers.* Figures 2 and 3 show the IRID spectra recorded in the NH and OH stretch regions of the β-D-MeGlc and β-D-MeGal complexes, respectively, together with "stick spectra" associated with their calculated low-energy structures. In each case, the spectrum, which was recorded with the UV probe laser tuned to the band origin of the corresponding R2PI spectrum, was unchanged when probed on the other vibronic bands except for a decrease in the S/N ratio, indicating significant population of one structure only. The structures that provide the best agreement with the experimental data both correspond to the minimum-energy configurations and are both associated with an extended peptide backbone conformation, consistent with the initial inference based on the R2PI spectral comparisons. The two complexes incorporate quite different carbohydrate binding geometries, however, and are associated with different hydrogen-bonding networks.

The β-D-MeGlc complex (Figure 2) is bound through two intermolecular hydrogen bonds linking the "outer" groups on the

dx.doi.org/10.1021/ja109664k | *J. Am. Chem. Soc.* 2011, 133, 4548–4557

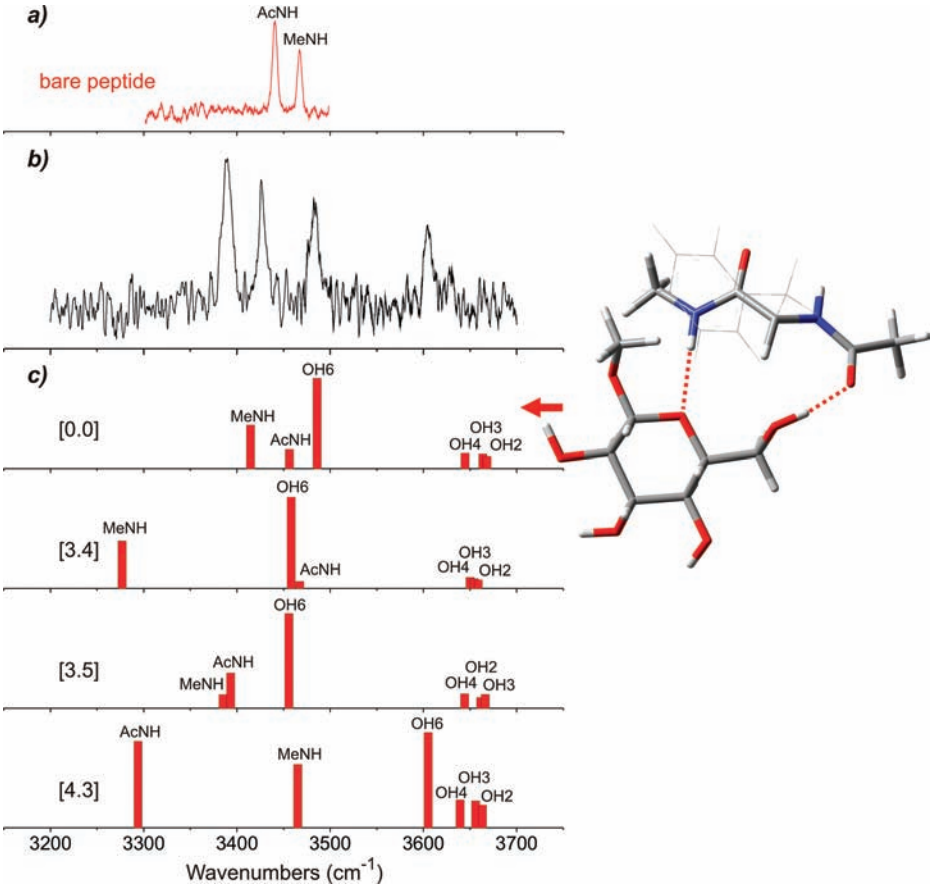


Figure 2. Experimental IRID spectrum of (a) isolated *N*-acetyl-L-phenylalanine methylamide conformer I, (b) its complex with β-D-MeGlc, and (c) the corresponding computed vibrational spectra associated with its calculated lowest-energy structures (whose relative energies in kJ mol^{-1} are shown in square brackets). In the structure at the right, the benzyl side group is shown as a thin-wire representation to aid clarity and emphasize the peptide backbone conformation.

peptide backbone, MeNH and MeC═O, to oxygen O-5 on the pyranose ring and OH-6 on the hydroxymethyl group; the latter is rotated away from its equilibrium G+ orientation to optimize its interaction with the carbonyl group. This structure does *not* incorporate a stacked interaction between the pyranose and aromatic rings, though it does allow the anomeric methoxy group to face the aromatic ring, perhaps introducing a weak dispersive $OCH_3 \rightarrow \pi$ interaction. The two intermolecular hydrogen bonds displace the associated MeNH and OH-6 vibrational modes toward lower wavenumbers by ~70 and ~150 cm^{-1}, respectively, but the wavenumber of the "spectator" AcNH mode is virtually unchanged. The same is true for the "spectator" OH-2, -3, and -4 modes, which continue to be linked through a chain of weak intramolecular hydrogen bonds, OH-4 → OH-3 → OH-2 → O-1; they are associated with the bands at higher wavenumbers lying between ~3600 and 3640 cm^{-1}, as they are in the isolated carbohydrate.[17]

The vibrational spectrum associated with the β-D-MeGal complex (Figure 3) appears at first sight to be very similar to that of the β-D-MeGlc complex, but a closer comparison reveals important quantitative differences. The two modes MeNH and OH-2, which are strongly displaced by ~80 and ~210 cm^{-1}, respectively, through intermolecular hydrogen bonding, are located at markedly lower wavenumbers; the same is also true for *all* of the remaining OH bands, which are now located between ~3550 and 3600 cm^{-1} (displaced by ~50 cm^{-1}). The spectral shifts suggest the incorporation of an extensive hydrogen-bonded network, generating a high level of cooperativity.[18] The structure that provides the best fit for the β-D-MeGlc complex has a relative energy of 10.1 kJ mol^{-1} when MeGal is substituted for MeGlc, and in addition, it does not reproduce these increased displacements to lower wavenumber (Figure 3).

In the light of these experimental data, the initial computational structural survey was extended, with particular attention focused on the erstwhile "minimum-energy structure" (labeled [8.3] in Figure 3c). Although this structure did present a cooperatively hydrogen-bonded chain, OH-4 → OH-3 → OH-2 → O═C, the G+ orientation of the hydroxymethyl group prevented its further extension; a rotation of the hydroxymethyl group about the C5−C6 axis into a G− orientation, however, would bring the OH-6 group much closer to the "outer" MeNH site. This could provide a better, more fitting structure, so following this lead, a further series of iterative searches was conducted. It did indeed reveal a new cooperatively hydrogen-bonded structure (see Figure 3) with an energy 8.3 kJ mol^{-1} below that of the previous "minimum-energy structure". Like the β-D-MeGlc complex, it was still bound to the extended host backbone by intermolecular hydrogen bonds at the "outer" MeNH and MeC═O sites, but these were now connected to O-6, O-5, and OH-2. The new structure was supported by an enhanced cooperatively bonded chain, MeNH → OH-6 → OH-4 → OH-3 → OH-2 → O═CMe, and the calculated OH modes were

dx.doi.org/10.1021/ja109664k |*J. Am. Chem. Soc.* 2011, 133, 4548–4557

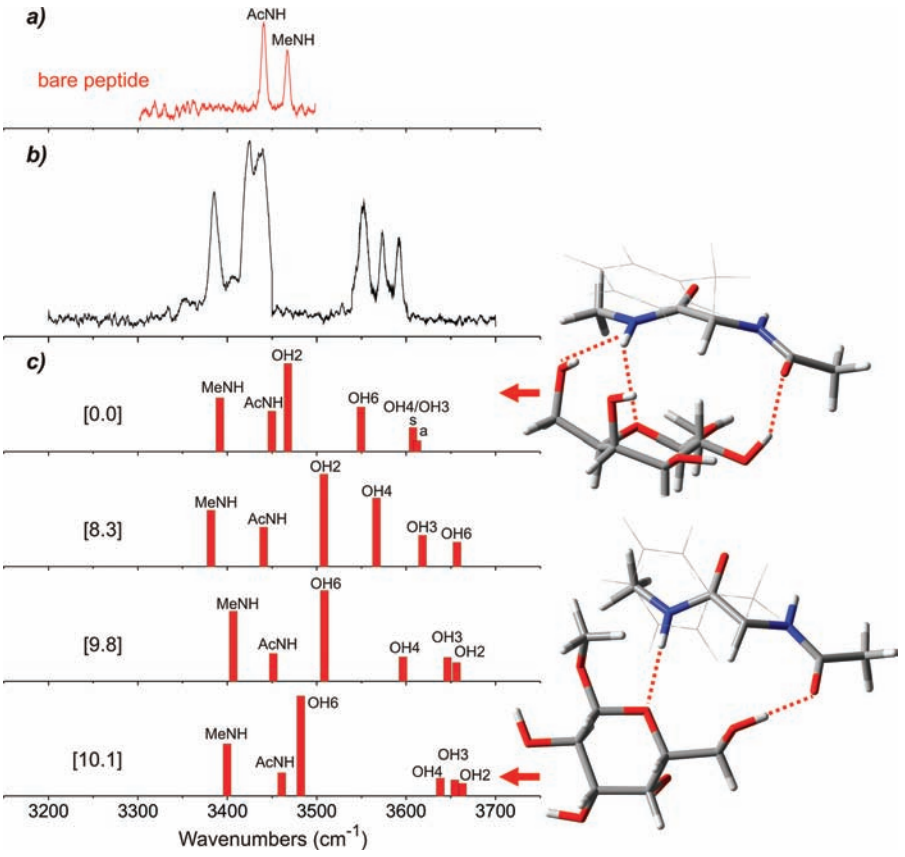


Figure 3. Experimental IRID spectrum of (a) the extended *N*-acetyl-L-phenylalanine methylamide conformer I, (b) its complex with β-D-MeGal, and (c) the corresponding computed vibrational spectra associated with its calculated lowest-energy structures (whose relative energies in kJ mol^{-1} are shown in square brackets).

all displaced further toward lower wavenumbers. The predicted vibrational spectrum was evidently in much closer correspondence with experiment, where each of the six bands in its IRID spectrum was clearly resolved. The high S/N ratio perhaps reflects the particular stability conferred on the β-MeGal complex by the cooperative network.

The difference between the structures of the β-D-MeGlc and β-D-MeGal complexes is caused primarily by the altered configuration of OH-4, which changes from equatorial in Glc to axial in Gal. This, coupled with the rotation of the hydroxymethyl group from a G+ to a G− orientation, facilitates the formation of an intramolecular hydrogen bond between OH-6 and OH-4 and the consequent development of the extended cooperative chain. As a result, the change from Glc to Gal alters the relative orientation of the pyranoside ring to create contacts between the O-6 and OH-2 sites on the carbohydrate and the MeNH and MeC═O sites on the peptide backbone. In the β-D-MeGlc complex, the equatorial orientation at OH-4 prevents generation of the extended cooperative chain, and the corresponding contacts involve O-5 and OH-6.

The α *Anomers.* IRID spectra associated with the α-D-MeGal complex and probed via the UV R2PI features that lie close to the band origins of the extended (I) and folded (III) conformations of the peptide backbone in free *N*-acetyl-L-phenylalanine methylamide are presented in Figures 4 and 5, respectively. The two figures also show the computed spectra associated with the lowest-energy members of each conformational family. Reassuringly, the computed spectrum associated with an extended backbone structure provides by far the best correspondence with the experimental spectrum shown in Figure 4; it also has the lowest relative energy. The overall structure of the extended complex is very similar to the one assigned to the corresponding complex with β-D-MeGal: the carbohydrate engages with the peptide through its most distal groups, OH-6 and OH-2, bound by hydrogen bonds linking MeNH to O-6 (and weakly to O-5) and OH-2 to MeC═O. The structure is again supported by the cooperatively hydrogen-bonded chain MeNH → OH-6 → OH-4 → OH-3 → OH-2 → O═CMe, which is facilitated in turn by the axial orientation of OH-4 and the G− orientation of the hydroxymethyl group.

There are some striking and precise differences between the two anomers, however; the IRID spectra of the extended, α-D-MeGal and β-D-MeGal complexes are *not* identical. The "free" AcNH mode in the α complex still lies at ∼3415 cm^{-1}, but the hydrogen-bonded MeNH mode at ∼3340 cm^{-1}, which principally engages O-6, is displaced by a further 50 cm^{-1} to lower wavenumbers. Similarly, while the displacements of the bands associated with the "spectator" groups (OH-6, -4, and -3) remain very similar, the displacement of the hydrogen-bonded OH-2 mode (at ∼3510 cm^{-1} in the extended α complex) is only ∼140 cm^{-1}, which is ∼70 cm^{-1} *less* than before. Together, these suggest increased H-bonding at O-6 at the expense of engagement at OH-2. Since the magnitudes of the spectral shifts reflect the relative hydrogen-bond distances and angles, the hydrogen bond linking MeNH to OH-6 in the extended α complex must be shorter (stronger) than that in the β complex and the one linking OH-2 to O═CMe must be longer (weaker). The calculations provide following hydrogen-bond distances: 1.99 Å for MeNH···O-6 in the

dx.doi.org/10.1021/ja109664k |*J. Am. Chem. Soc.* 2011, 133, 4548–4557

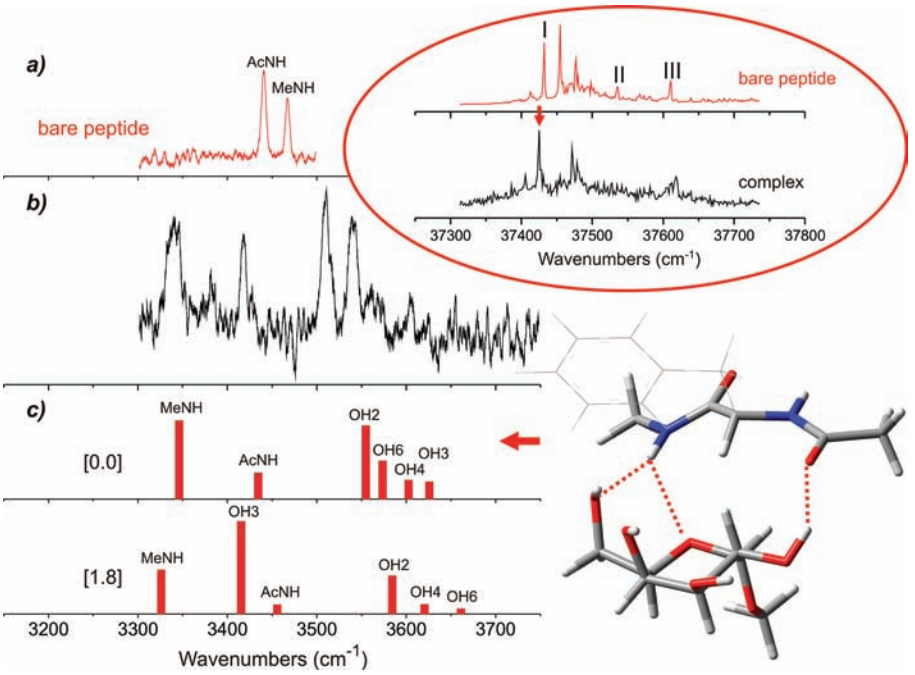


Figure 4. (a) Experimental IRID spectrum associated with the extended peptide backbone (conformer I) in *N*-acetyl-L-phenylalanine methylamide, (b) the corresponding spectrum associated with its α-D-MeGal complex (recorded with the UV probe laser centered on the band identified by an arrow; see the inset), and (c) computed vibrational spectra associated with the lowest-energy members of the family of extended peptide backbone structures. Relative energies in kJ mol^{-1} are shown in square brackets.

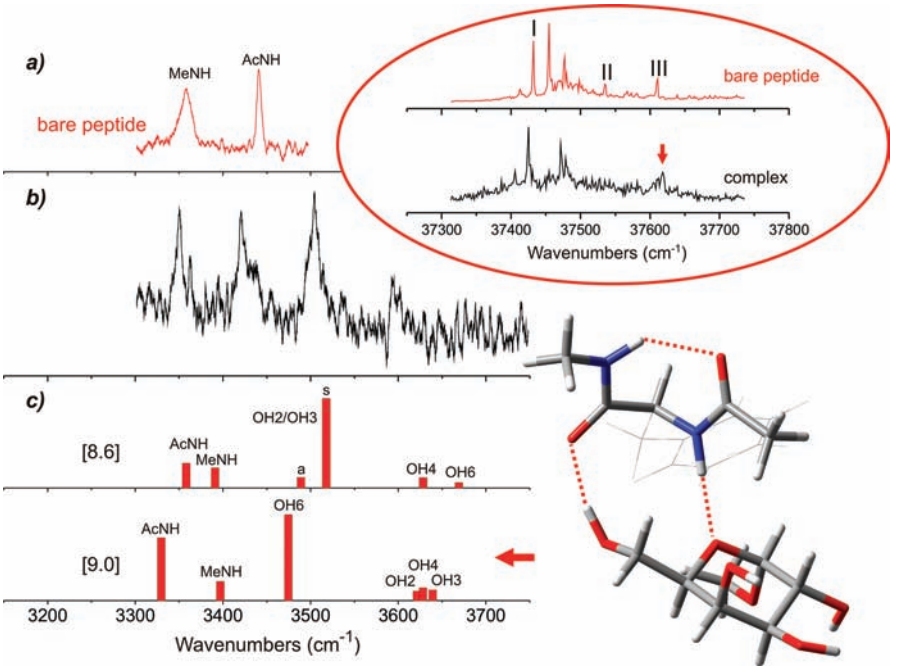


Figure 5. (a) Experimental IRID spectrum associated with the folded peptide backbone in *N*-acetyl-L-phenylalanine methylamide (conformer III), (b) the corresponding spectrum associated with its α-MeGal complex (recorded with the UV probe laser centered on the band identified by an arrow; see the inset), and (c) computed vibrational spectra associated with the lowest-energy members of the family of folded peptide backbone structures. The energies in kJ mol^{-1} shown in square brackets are relative to the global minimum structure shown in Figure 4.

following α complex (cf. ∼2.14 Å for the β complex); ∼1.95 Å for OH-2···O=CMe in the α complex (cf. ∼1.90 Å for the β complex). On the other hand, in the β complex, the stronger bifurcated hydrogen bond linkage to O-5 compensates for the weaker bond linking MeNH to O-6, since the MeNH···O-5 distance shortens from ∼2.61 Å (α) to ∼2.39 Å (β), implying an increased electron density or polarization on O-5. Evidently, the intermolecular bonding and geometry are subtly influenced by

dx.doi.org/10.1021/ja109664k |*J. Am. Chem. Soc.* 2011, 133, 4548–4557

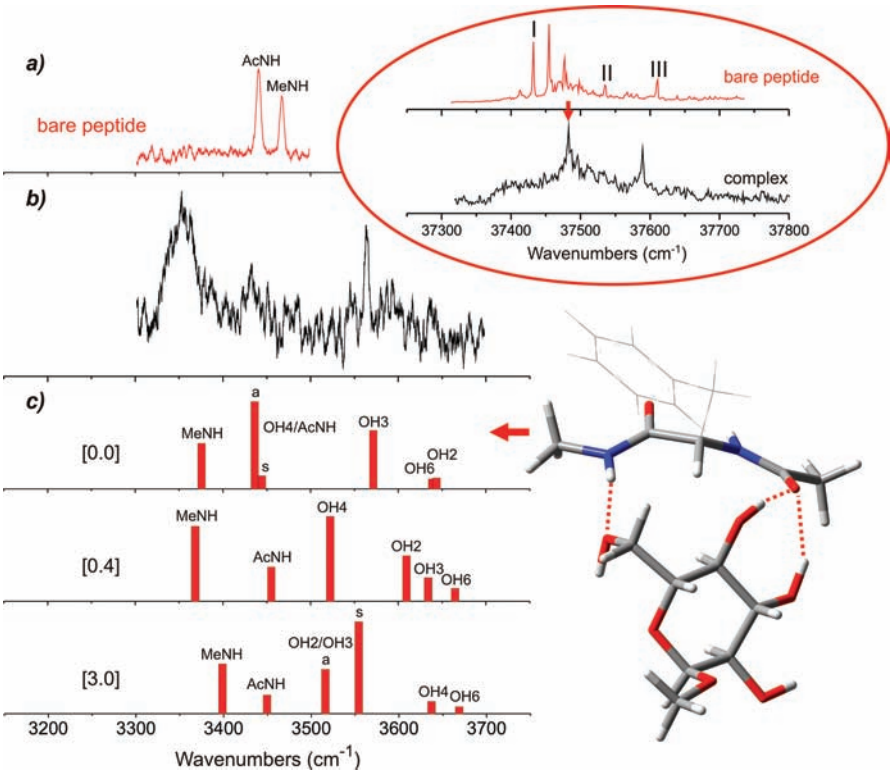


Figure 6. (a) Experimental IRID spectrum associated with the extended peptide backbone in *N*-acetyl-L-phenylalanine methylamide (conformer I), (b) the corresponding spectrum associated with its α-D-MeGlc complex (recorded with the UV probe laser centered on the band identified by the arrow; see the inset), and (c) computed vibrational spectra associated with the lowest-energy members of the family of extended peptide backbone structures. Relative energies in kJ mol^{-1} are shown in square brackets.

the orientation of the anomeric OCH_3 group; the striking stereoelectronic implications of this anomeric sensitivity are discussed elsewhere.[19]

Analysis of the IRID spectrum shown in Figure 5 confirms its provisional association with a complex incorporating the folded peptide backbone conformation III. In this conformation, hydrogen bonding between the "outer" binding sites, MeNH and MeC═O, provides the scaffold supporting the cyclic backbone structure, and they are no longer available for *inter*molecular binding. Instead, the binding is now located at the "inner" sites, AcNH and C═O(Phe), and the structure providing the closest correspondence with experiment attaches the carbohydrate to these sites through hydrogen bonds to O-5 and OH-6. It is one of the two lowest-energy members of the family of "folded" complexes, calculated to lie ∼9 kJ mol^{-1} above the global minimum, consistent with its low R2PI spectral intensity and relative population.

The two strong IR bands located at ∼3340 and ∼3500 cm^{-1} are assigned to the strongly hydrogen-bonded AcNH and OH-6 modes; the intermediate band located at ∼3420 cm^{-1} is assigned to the MeNH mode and the weaker blended feature at ∼3600 cm^{-1} to the cluster of "spectator" OH groups (OH-2, -3, and -4). The smaller displacement of the MeNH mode relative to the isolated peptide indicates some weakening of the *intra*molecular peptide hydrogen bond within the complex. In contrast to the extended α-D-MeGal complex, the intermolecular bonding between OH-6 and C═O(Phe) prevents formation of the OH-6 → OH-4(axial) intramolecular hydrogen bond and in so doing removes a key linkage in the formation of an extended cooperative chain. Consequently, this minor, folded peptide conformer is both disruptive of the preferred α-D-MeGal interactions and insensitive to the carbohydrate configuration. The IRID spectra associated with the intermediate bands in the R2PI spectrum of the α-D-MeGal complex, centered at ∼37 470 and ∼37 480 cm^{-1}, present overlapping band contours (see the Supporting Information) that suggest contributions from additional structure(s). Unfortunately, the spectra were too complex to allow their unambiguous assignment.

Figure 6 compares the IRID spectrum of the extended conformer I of isolated *N*-acetyl-L-phenylalanine methylamide with that of its complex with α-D-MeGlc, recorded with the UV probe laser centered on the band at ∼37 480 cm^{-1}, and with the calculated vibrational spectra associated with the lowest-energy members of the family of extended peptide complexes. The spectrum associated with the global minimum structure again provides the best correspondence with experiment. The carbohydrate is bound at the "outer" sites of an extended peptide structure, but in contrast to the complex with the *β*-anomer, the MeNH group is now linked to O-6 rather than O-5, and the MeC═O group is strongly bonded to OH-4 and more weakly to OH-3. There is little displacement of the AcNH vibration at ∼3430 cm^{-1}, which is now coupled to the strongly displaced OH-4 mode, but the MeNH band at ∼3350 cm^{-1} is displaced by ∼120 cm^{-1}, reflecting its strong intermolecular bonding to O-6. In the corresponding complex of α-D-MeGal (Figure 4), the axial orientation of OH-4 created a different intermolecular structure that was promoted by its *intra*molecular bonding to OH-3 and the extended cooperative hydrogen-bonded network terminated by links to O-6 and OH-2. Here the hydrogen-bonding network of α-D-MeGlc is noncooperatively bridged from OH-4 to OH-3 via the C═O group.

The results shown in Figure 7, which complete the set, compare the calculated and experimental IRID spectra of the α-D-MeGlc

dx.doi.org/10.1021/ja109664k | *J. Am. Chem. Soc.* 2011, 133, 4548–4557

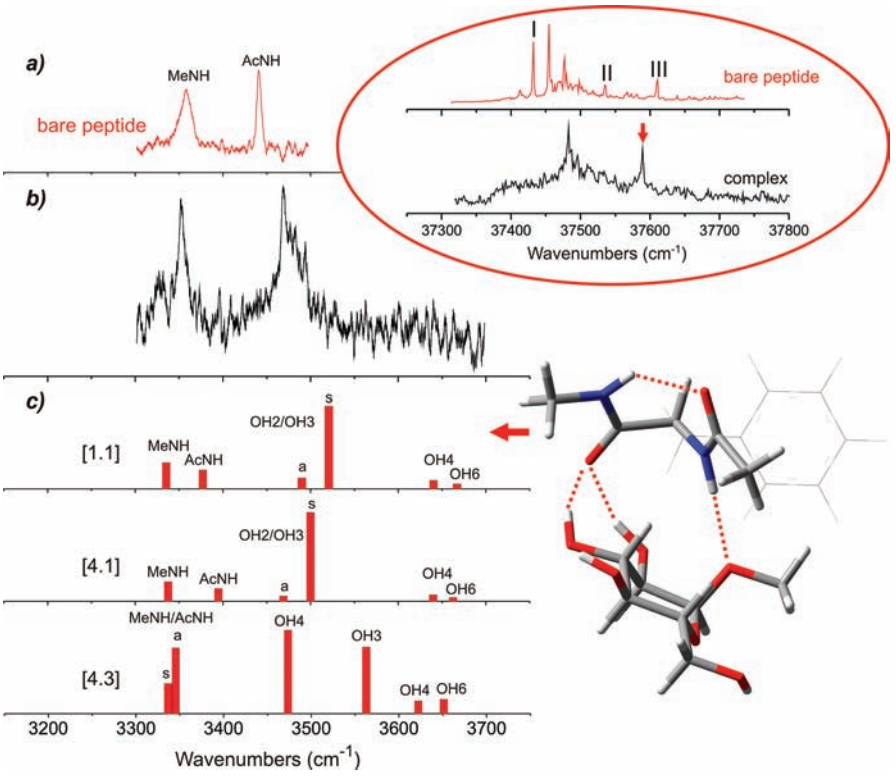


Figure 7. (a) Experimental IRID spectrum associated with the folded peptide backbone in *N*-acetyl-L-phenylalanine methylamide (conformer III), (b) the corresponding spectrum associated with its α-D-MeGlc complex (recorded with the UV probe laser centered on the band identified by an arrow; see the inset), and (c) computed vibrational spectra associated with the lowest-energy members of the family of folded peptide backbone structures. Energies in kJ mol^{-1} shown in square brackets are relative to the global minimum structure shown in Figure 6.

complex probed on the UV band at ~37 590 cm^{-1}, close to the band origin of folded conformer III in the isolated peptide. The spectrum is best reproduced by the lowest-energy (or perhaps a combination of the two lowest-energy) members of the family of folded peptide complexes; apart from differing orientations of the aromatic side group, their structures are identical. Unlike the α-D-MeGal complexes (see Figures 4 and 5), the relative energies of the folded and extended peptide backbone structures are quite close, consistent with their comparable relative R2PI spectral intensities, which suggest similar populations. The carbohydrate is linked at the "inner" binding sites, AcNH and C═O(Phe), which are now bonded to O-1 and to both OH-2 and OH-3, respectively. The vibrational band associated with the "outer" MeNH group (which provides the intramolecular hydrogen bond supporting the folded peptide backbone) is still located at ~3350 cm^{-1}, but it is blended with the "inner" AcNH band, which is displaced by ~100 cm^{-1}. Similarly, the coupled vibrational bands associated with OH-2 and OH-3, which lie in the region 3470−3490 cm^{-1}, are displaced by 130−150 cm^{-1}. Like the extended peptide complex of α-D-MeGlc, the hydrogen-bonding network in the folded peptide complex is again non-cooperatively bridged via the C═O group, but this time from OH-3 to OH-2.

■ FURTHER IMPLICATIONS AND CONCLUDING REMARKS

The full set of carbohydrate−peptide complex structures that provide the best agreement with the experimental IR spectra are summarized in Figure 8. They are largely distinct, and their binding motifs are selective, highlighting the exquisite sensitivity of the host−carbohydrate interaction in the gas phase. A number of striking conclusions emerge:

(1) The designed modular addition of the truncated peptide backbone to the tolyl motif prevents the exclusive formation of stacked carbohydrate−aromatic structures supported by apolar CH−π interactions between the pyranose ring and the aromatic side chain, even though these can be formed in the gas phase in the absence of the backbone.[6]

(2) The sensitive conformational equilibrium between an "extended" and a "folded" (C_7 γ-turn) state, which is established by the conformational flexibility of the truncated peptide, is remarkably and selectively displaced through interactions with carbohydrates.

(3) The intermolecular structures are sensitive to epimeric changes (from equatorial to axial at C-4) and to anomeric changes ("the tail can wag the dog"). In two structurally similar "extended peptide" complexes (and in the *absence* of solvent), the change from β to α in MeGal greatly weakens the bond linking MeNH to O-5, and in the corresponding complexes with MeGlc, it eliminates the bond altogether, suggesting an important role associated with stereoelectronic effects.[19,20]

Stacking Interactions. Even when the aromatic and pyranose rings are nearly coplanar, as in the β-D-MeGal complex, the curved apolar face of the carbohydrate (which presents axially oriented CH-1, -3, and -5 and equatorial CH-4 groups) is actually

dx.doi.org/10.1021/ja109664k |*J. Am. Chem. Soc.* 2011, 133, 4548–4557

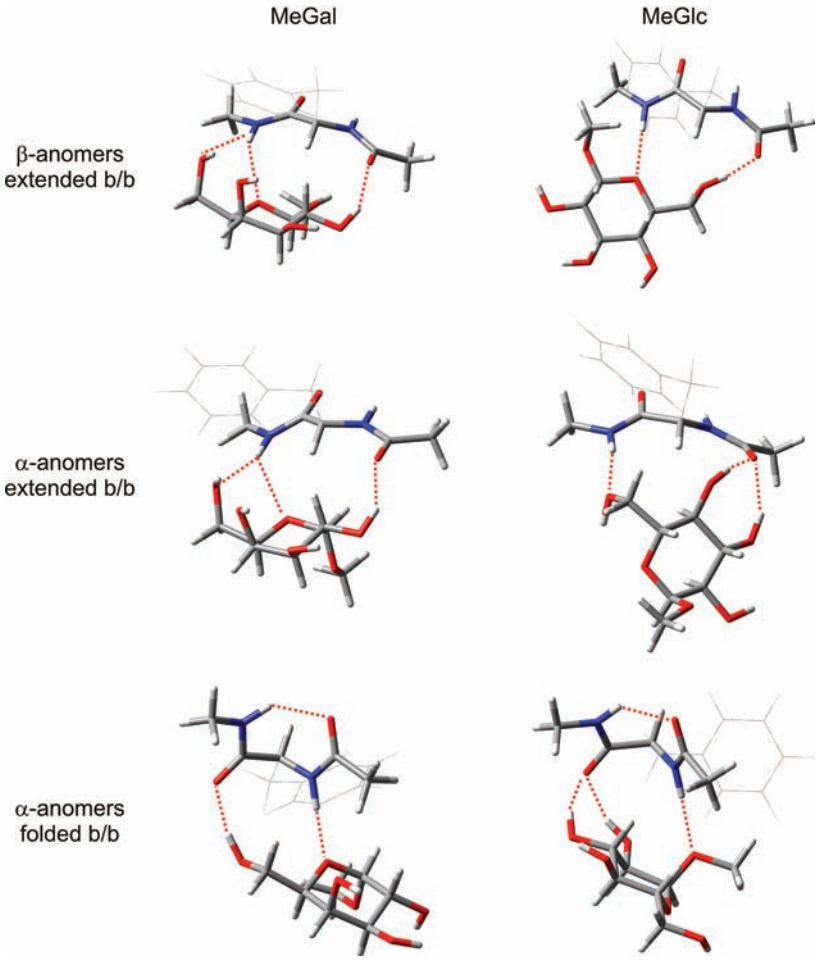


Figure 8. Computed structures of carbohydrate−*N*-acetyl-L-phenylalanine methylamide complexes that provide the best comparisons with their experimental IRID spectra.

directed *away* from the aromatic π-electron system. Thus, addition of a peptide-like motif equipped with NH and C=O hydrogen-bond donors and acceptors dominates a competent π-interacting group (CH_2Ph); CH−π interactions are evidently too weak to compete against multivalent hydrogen bonding in the gas phase.

One of the motivations for investigating the machinery of carbohydrate−peptide interactions in the gas phase was to provide a benchmark reference, free of interference from interactions with the environment, particularly water. The results obtained though modular partitioning of the interactions of carbohydrates with aromatic peptide motifs in the gas phase indirectly provide strong and rare support for the role of water in the architecture of natural protein−carbohydrate complexes in aqueous environments. In "real life", the involvement of water is crucial, although its role in providing a driving force for carbohydrate binding to proteins is still debated.[8] The new results suggest that the prevalence of "stacked" carbohydrate−aromatic side chain structures in such proteins may be driven by the exclusion of water from the hydrophobic interface[4] rather than binding through the intrinsic CH−π dispersive interactions that, although possible, are clearly weaker than hydrogen bonding (or electrostatic interactions at protonated or anionic sites). In protein structures, of course, the more rigid evolved pockets are likely to dictate the creation of stacked structures, but our observations provide an indication of energetic contributions in possible evolutionary pathways.

Carbohydrate Selectivity. When the host backbone is extended, the carbohydrates bind to the "external" sites, MeNH and MeC=O, but when these are engaged in supporting a C_7 γ-turn, the carbohydrates select the alternative pair of "internal" sites, AcNH and C=O(Phe). Particular stability is conferred by the development of cooperative hydrogen-bonded networks. The host "engages" with the most suitable pair of neighboring conjugate sites on each carbohydrate, a little like interacting "teeth" on a rotating gear wheel. The specific choice depends on the conformation of the peptide backbone *and* the configuration and conformation of the carbohydrate ligand. The peptide "selects" the most strongly bound structures. In the extended α- and β-D-MeGal complexes, the choice depends upon the axial orientation of OH-4 and the flexibility of the hydroxymethyl group, which rotates from a preferred G+ orientation to a G− orientation; together they facilitate the development of a cooperative chain of hydrogen bonds incorporated into similar structures. In contrast, in the extended β-D-MeGlc complex, the development of a cooperative network is prevented by the equatorial orientation of OH-4, which accounts for the *differing* structures. This sensitivity recalls a remark made by Gabius et al.[22] in an essay on the chemical biology of the sugar code. Commenting on the selective binding of carbohydrates to lectins,

dx.doi.org/10.1021/ja109664k |*J. Am. Chem. Soc.* 2011, 133, 4548–4557

they wrote, "The selectivity of the molecular rendezvous ... explains the exquisite way in which individual code letters, for example D-glucose and its 4-epimer D-galactose, are distinguished. Thus, a change of only one hydroxy group from the equatorial to the axial orientation ... is sufficient to establish distinct letters."

The (unexpected) preference for binding β anomers to the extended peptide backbone conformation also opens the possibility of investigating the strength and selectivity of carbohydrate binding to small amyloidogenic peptides, which also present extended β-sheet backbone conformations. Inhibition of their dimerization[21] might conceivably have some therapeutic potential; as in the recent reports of the modulation of amyloid-β aggregation and toxicity by carbohydrates.[23]

■ ASSOCIATED CONTENT

Supporting Information. R2PI spectra of α-D-MeGlc−peptide and α-D-MeGal−peptide complexes, additional IRID spectra of α-D-MeGal−peptide complexes, Cartesian coordinates and relative energies of the six complex structures shown in Figure 8, and complete ref 14. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

john.simons@chem.ox.ac.uk; ben.davis@chem.ox.ac.uk

Present Addresses

[§]Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco, (UPV−EHU), Apartado 644, E-48940 Bilbao, Spain.

[∥]Institut des Sciences Moléculaire d'Orsay-CNRS, Université Paris Sud, 91405 Orsay Cedex, France.

[⊥]Department of Chemistry and Biochemistry, Rowan University, 201 Mullica Hill Road, Glassboro, NJ 08028, USA.

■ ACKNOWLEDGMENT

We are grateful to the following for their support: the Leverhulme Trust for the award of an Emeritus Fellowship (J.P.S.); the Royal Society and the Wolfson Foundation for a Research Merit Award (B.G.D.); the EPSRC (B.G.D.); the Spanish Ministry (MICINN) for a "Ramón y Cajal" Contract (E.J.C.); the Royal Society for a USA/Canada Research Fellowship and Linacre College for a Junior Research Fellowship (T.D. V.); the STFC for the provision of equipment from the Laser Loan Pool; and the Oxford Supercomputing Centre. Dr. Zheng Su and Dr. Conor Barry contributed to some of the experimental and preparative work.

■ REFERENCES

(1) (a) Varki, A. *Glycobiology* **1993**, *3*, 97–130. (b) Weiss, W. I.; Drickamer, K. *Annu. Rev. Biochem.* **1996**, *65*, 441–473. (c) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720. (d) Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357–2364. (e) Dwek, R. A.; Butters, T. D. *Chem. Rev.* **2002**, *102*, 283–284. (f) Gabius, H. J.; Siebert, H. C.; Andre, S.; Jiménez-Barbero, J.; Rudiger, H. *ChemBioChem* **2004**, *5*, 740–764.

(2) (a) Boraston, A. D.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769–781. (b) Ambrosi, M.; Cameron, N. R.; Davis, B. G. *Org. Biomol. Chem.* **2005**, *3*, 1593–1608. (c) Hashimoto, H. *Cell. Mol. Life Sci.* **2006**, *63*, 2954–2967. (d) Mazik, M. *Chem. Soc. Rev.* **2009**, *38*, 935–956.

(3) (a) Lemieux, R. U. *Chem. Soc. Rev.* **1989**, *18*, 347–374. (b) Quiocho, F. A. *Pure Appl. Chem.* **1989**, *61*, 1293–1306. (c) Lis, H.; Sharon, N. *Chem. Rev.* **1998**, *98*, 637–674. (d) Jiménez-Barbero, J.; Asensio, J. L.; Cañada, F. J.; Poveda, A. *Curr. Opin. Struct. Biol.* **1999**, *9*, 549–555.

(4) Lemieux, R. U. *Acc. Chem. Res.* **1996**, *29*, 373–380.

(5) Fernández-Alonso, M. C.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G. *J. Am. Chem. Soc.* **2005**, *127*, 7379–7386.

(6) (a) Screen, J.; Stanca-Kaposta, E. C.; Gamblin, D. P.; Liu, B.; Snoek, L. C.; Davis, B. G.; Simons, J. P. *Angew. Chem., Int. Ed.* **2007**, *46*, 3644–3648. (b) Stanca-Kaposta, E. C.; Gamblin, D. P.; Screen, J.; Liu, B.; Snoek, L. C.; Davis, B. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4444–4451. (c) Su, Z.; Stanca-Kaposta, E. C.; Cocinero, E. J.; Davis, B. G.; Simons, J. P. *Chem. Phys. Lett.* **2009**, *471*, 17–22.

(7) (a) Sharma, R.; McNamara, J. P.; Raju, R. K.; Vincent, M. A.; Hillier, I. H.; Morgado, C. A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2767–2774. (b) Raju, R. K.; Ramraj, A.; Vincent, M. A.; Hillier, I. H.; Burton, N. A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6500–6508. (c) Raju, R. K.; Ramraj, A.; Hillier, I. H.; Vincent, M. A.; Burton, N. A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3411–3416.

(8) Klein, E.; Ferrand, Y.; Barwell, N. P.; Davis, A. P. *Angew. Chem., Int. Ed.* **2008**, *47*, 2693–2696.

(9) Gerhards, M.; Unterberg, C.; Gerlach, A.; Jansen, A. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2682–2690.

(10) Zehnacker, A.; Suhm, M. *Angew. Chem., Int. Ed.* **2008**, *47*, 6970–6992.

(11) Robertson, E. G.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1–18.

(12) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrikson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440–467.

(13) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289–300.

(14) Frisch, M. J.; et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Pittsburgh, PA, 2003.

(15) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.

(16) (a) Stanca-Kaposta, E. C.; Gamblin, D. P.; Cocinero, E J.; Frey, J.; Kroemer, R. T.; Davis, B. G.; Fairbanks, A. J.; Simons, J. P. *J. Am. Chem. Soc.* **2008**, *130*, 10691–10696. (b) Cocinero, E. J.; Stanca-Kaposta, E. C.; Scanlan, E. M.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Chem.—Eur. J.* **2008**, *14*, 8947–8955. (c) Cocinero, E. J.; Stanca-Kaposta, E. C.; Dethlefsen, M.; Liu, B.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Chem.—Eur. J.* **2009**, *15*, 13427–13434.

(17) Simons, J. P.; Çarçabal, P.; Davis, B. G.; Gamblin, D. P.; Hünig, I.; Jockusch, R. A.; Kroemer, R. T.; Marzluff, E. M.; Snoek, L. C. *Int. Rev. Phys. Chem.* **2005**, *24*, 489–532.

(18) (a) Dashnau, J. L.; Sharp, K. A.; Vanderkooi, J. M. *J. Phys. Chem. B* **2005**, *109*, 24152–24159. (b) Deshmuhk, M. M.; Bartolotti, L. J.; Gadre, S. R. *J. Phys. Chem. A* **2008**, *112*, 312–321.

(19) Cocinero, E. J.; Çarçabal, P.; Vaden, T. D.; Simons, J. P.; Davis, B. G. *Nature* **2011**, *469*, 76–79.

(20) (a) *The Anomeric Effect and Associated Stereoelectronic Effects*; Thatcher, G. R. J., Ed.; ACS Symposium Series 539; American Chemical Society: Washington, DC, 1993. (b) Salzner, U.; Schleyer, P. v. R. *J. Org. Chem.* **1994**, *59*, 2138–2155.

(21) Fricke, H.; Gerlach, A.; Unterberg, C.; Wehner, M.; Schrader, T.; Gerhards, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 900–904.

(22) Gabius, H.-J.; Siebert, H.-C.; André, S.; Jiménez-Barbero, J.; Rüdiger, H. *ChemBioChem* **2004**, *5*, 740–764.

(23) (a) McLaurin, J.; Golomb, R.; Jurewicz, A.; Antel, J. P.; Fraser, P. E. *J. Biol. Chem.* **2000**, *24*, 18495–18502. (b) Liu, R.; Barkhordarian, H.; Emadi, S.; Park, C. B.; Sierks, M. R. *Neurobiol. Dis.* **2005**, *20*, 74–78. (c) Nitz, M.; Fenili, D.; Darabie, A. A.; Wu, L.; Cousins, J. E.; McLaurin, J. *FEBS J.* **2008**, *275*, 1663–1674.