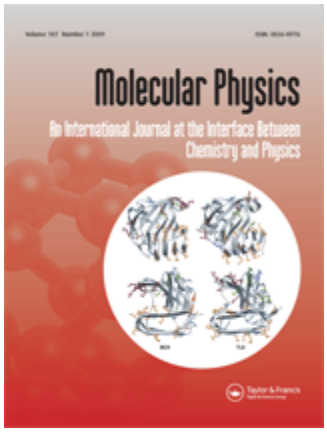
Good Vibrations: Probing Biomolecular Structure and Interactions through Spectroscopy in the Gas Phase

J.P. Simons



Good Vibrations: Probing Biomolecular Structure and Interactions through Spectroscopy in the Gas Phase

Journal:	*Molecular Physics*
Manuscript ID:	TMPH-2009-0277.R1
Manuscript Type:	Invited Article
Date Submitted by the Author:	05-Oct-2009
Complete List of Authors:	Simons, J.P.; University of Oxford
Keywords:	biomolecules, vibrational spectroscopy, conformation, peptides, carbohydrates

Good Vibrations: Probing Biomolecular Structure and Interactions through Spectroscopy in the Gas Phase

J. P. Simons

Chemistry Department, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ UK

Key Words: biomolecules; vibrational spectroscopy; conformation; peptides; carbohydrates.

Abstract.

Spectroscopic investigations of biomolecular structure and interactions in the gas phase, free from environmental 'disturbance', have become a growth industry over the last decade. The great majority however, have taken their 'bio-pedigrees' as a given rather than an issue to be addressed. This review attempts to move towards aspects of their biological, or perhaps more accurately, biochemical relevance, by reviewing some recent and current investigations of peptide and carbohydrate systems in the gas phase, chosen not just because they are 'do-able' but because of the biochemical questions they address. Hopefully, this will help to foster better communication between the molecular physics, biochemistry and molecular biology communities.

1. Prologue

The functioning of biological molecules depends fundamentally on their shape and flexibility and on their interactions with local environments – interactions which may also provide a feed-back influence on their preferred structures and bio-active conformations: living organisms are complex and also crowded systems. Molecular biologists can pay due respect to this by pursuing a holistic approach, "molecular systems biology", which aims to explain the behaviour and the machinery of complex biological systems in terms of all their molecular components and interactions. Physical scientists favour the alternative reductionist approach, diluting the system and taking it to bits to characterize its separated components and, after that has been achieved, to explore the interactions and consequences that may follow the initial steps in its controlled reassembly or its disassembly. Following this approach, the reductionist begins by isolating the key components of the machine in the gas phase, either individual molecules or their molecular building blocks, in order to characterize their *un*perturbed conformational landscapes and the non-covalent molecular and ionic interactions which control them. Subsequently, the structural and sometimes chemical consequences of *inter*molecular interactions are explored in a controlled way by probing the structure and bonding in their molecular (often hydrated) and ionic complexes, beginning with 1:1 complexes and then moving on to larger clusters of increasing size or complexity.

Spectroscopic and computational investigations of isolated biomolecular structures conducted by physical chemists, following the bottom-up approach, have focused on small molecules such as neurotransmitters, or drugs, or the molecular building blocks of proteins, nucleic acids and carbohydrates, selected on the basis of their

perceived "biological relevance" but also, to be honest, their experimental and computational accessibility. Their biological selection through evolution has seldom been a primary concern (though investigations of the ultraviolet photo-stability of nucleobases and nucleosides for example, have provided a notable exception). Clearly, there is an enormous gap between the bottom-up reductionist approach, which seeks an understanding of the conformational landscapes, accessible structures and interactions of its individual molecular components and the top-down systems approach, which seeks an understanding of the machinery of biological systems as a whole. Living biological systems depend upon cooperative intermolecular interactions; isolated molecules are inanimate. Without conversations between the component biomolecules the system is dead. The same is also true in a sociological sense: isolated scientists and scientific communities are closed systems while living systems are open and depend upon communications beyond their local confines. The following article attempts to move a little further towards the aspiration of biological relevance, or perhaps more accurately, biochemical relevance, by reviewing some recent and current spectroscopic investigations of small biomolecular systems in the gas phase, chosen not just because they are 'do-able' but because of the biochemical questions they address. Hopefully, this will help to foster better communication between the molecular physics, biochemistry and molecular biology communities.

2. Asking 'relevant' questions

The key issues addressed by structural biology are structure, interaction and function; they are intimately connected of course. Function lies in the province of molecular biology but biomolecular structure and interactions lie primarily in the province of

molecular physics. One of the broad strategies through which these qualities can be probed, to provide links to the working components and to the machinery of function, is based upon a combination of molecular spectroscopy and computation. In solution this generally involves varieties of NMR spectroscopy and molecular dynamics simulations. In the gas phase it generally involves combinations of electronic and vibrational spectroscopy, mass spectrometry, and electronic structure and force field calculations using density functional theory or ab initio methods for systems held at low temperatures or molecular dynamics simulations at more elevated temperatures. Potential connections between structure and function can be illustrated by considering some of the questions that are currently being addressed through gas phase spectroscopic investigations of peptides, glyopeptides and carbohydrates.

Peptides. The peptide chains selected by evolutionary pressures in proteins fold into specific conformations to achieve the 'bioactive' secondary, tertiary and quaternary structures associated with their native states, and to allow specific kinds of interaction within their clusters or during their molecular encounters. What are their intrinsic secondary structures and how do they develop as the peptide chain-lengths increase? How do their intrinsic structures and the selectivity of their intermolecular interactions depend upon the composition and sequence of their amino acid residues? To what extent are their folded structures controlled by non-covalent *intra*molecular interactions, by environmental, usually aqueous *inter*molecular interactions, or by a balance of both? Since proteins have evolved in an aqueous world, are their native structures conserved in the absence of an aqueous environment, and how unique or robust are these structures?

Glycopeptides. Many functional proteins are glycosylated - decorated with oligosaccharides that are covalently linked to the peptide chains via *O*- or *N*-atom linkages during or following their translation from RNA in the ribosome. *N*-linked glycans are bound exclusively to asparagine residues, which are invariably located in the conserved peptide sequon, Asn-Xxx-Ser/Thr - as long as the neighbouring residue, Xxx, is not proline. The glycans are also invariably constructed around a core conserved pentasaccharide unit to provide *motifs*, the so-called glycocode, which can be interpreted for example, by a variety of lectins (proteins that recognize specific carbohydrates) and by the enzymes which assist in the construction and trafficking of cellular glycoproteins. Why has nature settled on these particular *motifs* in *N*-linked glycoproteins? How do they influence the folding of the peptide chain? Are the intrinsic conformations populated by glycans, or indeed carbohydrates and oligosaccharides in general, in the gas phase the same as those populated in aqueous environments?

Carbohydrates. Recent spectroscopic studies of carbohydrates in the absence of solvent have highlighted clearly delineated conformational landscapes which display well-defined minima, [1] but comparisons between the intrinsic, gas-phase structures of carbohydrates and their structures in aqueous solution have led to the conclusion that their inherent conformational preferences and consequently their secondary structures are indeed modified by solvent interactions.[2-4] Molecular dynamics investigations of naturally occurring *N*-glycans in aqueous solution have identified enduring structural *motifs* supported by explicitly bound water molecules. [5] In such a scenario, questions arise as to the mechanism through which “bare” glycans acquire each solvent molecule. Is their acquisition specific and selective, and is there a midpoint of partial solvation that

is a tipping point in the balance between inherent and solvated structural preferences? How does hydration influence their preferred structures: do the glycan structures that are selected by nature provide water-binding pockets which help to sustain their bio-active conformations in aqueous environments? Does hydration influence their molecular recognition by other proteins?

To begin to answer questions like these, the first essential requirement is knowledge of their intrinsic conformational structures, those they can access when their biological environment is stripped away and they are transferred from the condensed phase into the gas phase. Subsequent spectroscopic interrogation through pulsed tunable infrared laser excitation, coupled with force field and DFT or ab initio electronic structure calculations, can lead to their structural assignment and more importantly, identify the *primary* interactions within the isolated molecules that govern their intrinsic structural landscapes. Similar questions can then be asked about their *secondary* interactions with isolated elements of their biological environment, through the interrogation of 'custom-made' biomolecular complexes, created and stabilized at low temperature under controlled conditions, for example those incorporating bound water molecules or formed through self-association, or through binding to key components of protein molecular recognition sites.

This review provides a snapshot of some of the progress that is being made towards answering these questions, illustrated in part but by no means exclusively, by recent spectroscopic investigations conducted in the author's laboratory. To gain a more comprehensive spectroscopic view of biological molecules in the gas phase the reader is encouraged to read Schermann's excellent monograph, 'Spectroscopy and Modelling of

Biomolecular Building Blocks' [6]. There is also a rapidly increasing number of more specialist reviews, notably by de Vries and Hobza [7] dealing with biomolecular building blocks (primarily, nucleobases and peptides); Mons *et al*. [8], dealing with secondary structure in small, capped oligopeptides; Grégoire *et al*. [9] dealing with 'warm' protonated oligopeptides; Rizzo *et al.* [10,11] addressing the growth of helical structure in cold protonated peptides with increasing chain length; Polfer and Oomens [12], dealing with a wide range of bare and solvated ions 'of biological relevance'; and Simons *et al.* [1], dealing with bare and hydrated carbohydrates. There are also many excellent reviews of the shapes and interactions of the much larger, and more realistic, biomolecular ions which can be accessed through ion mobility/mass spectrometry (IMS/MS) investigations in the gas phase, including 'real' proteins and protein complexes [13-17]. These exciting and powerful strategies complement those provided by optical spectroscopy but their detailed discussion lies outside the scope of the present article, as do investigations of biomolecular building blocks isolated at very low temperature in helium nanodroplets [18].

3. Practical considerations

3.1. *Experimental*

Tunable infrared (IR) laser excitation of small biomolecules entrained in an expanding supersonic jet, or of biomolecular ions held in a cooled ion-trap, provides a powerful means of isolating them at low temperatures in the gas phase and probing their vibrational spectroscopy at high resolution. Many alternative experimental strategies, recently reviewed in references [6, 19], have been devised to achieve this. Transferring neutral biomolecules, or their ions into the gas phase most commonly involves pulsed

laser desorption from the solid phase into an expanding supersonic jet (typically He or Ar), or electrospray ionization (ESI) from solution; both methods manage to keep the often fragile, volatilized biomolecules intact. The most widely used means of recording their vibrational spectra in a jet-cooled environment typically involves double resonance excitation. In the simplest and most direct strategy, a resonant IR laser pulse depletes the selected absorbing molecular conformational population and the transient depletion is detected mass spectrometrically, following ionization of the selected molecular species by resonant two-photon ultraviolet (UV) laser excitation. Vibrational action ("ion-dip") spectra of the cold populations of each of the selected molecular conformers, tautomers, or molecular complexes isolated or stabilized in the jet, are generated by scanning the IR laser frequency. Vibrational spectra of trapped, conformationally selected, cold biomolecular ions can be recorded using a similar strategy; protonated oligopeptides emerging from an ESI source, mass-selected by quadrupole filtering and stored in a cold multipole ion trap have been detected via their daughter ions, generated by UV laser excited predissociation [11]. The predissociation yield can be greatly enhanced by combining UV photo-excitation of the peptide ions with (subsequent) IR (CO_2) laser excitation and the collisional cooling in the cold ion trap allows very high spectral resolution. [11,20]

In both examples, the use of a double resonance IR-UV scheme necessarily requires the presence of an UV chromophore, provided in almost all cases by an aromatic 'tag' or by an aromatic amino acid residue. One way of circumventing this limitation is to replace detection via resonant two photon ionization with direct, single photon ionization in the VUV although this may compromise the advantage of individual conformer

selection. [21] Mass selected vibrational action spectra of biomolecular ions can also be generated directly, without the need for an aromatic UV chromophore, through resonant IR multiphoton dissociation (IRMPD) combined with mass spectrometric detection of the resulting daughter ions (or less sensitively, the depletion of the parent ion) to provide lower quality but extremely valuable IR action spectra. They reflect the absorption of resonant IR photons coupled with rapid intramolecular vibrational redistribution (IVR) of the absorbed energy and eventual decomposition [12, 22]. The majority of investigations of biomolecular ions, typically protonated or cationised amino acids and peptides held in an ion trap, have been conducted in the so-called 'finger-print' region, between ca. 1000-2000 cm^{-1}, using the high power radiation provided by free electron lasers [12], but the strategy has also been successfully employed under supersonic expansion conditions to probe the (often more informative) O-H, N-H and C-H regions using table-top IR lasers [23,24]. In both situations, the biomolecular ions were either trapped or generated at 'warm' temperatures, typically in the range 300 – 400K. The elevated temperature and the dynamical population of several conformers can, and generally does increase the spectral complexity and the spectral resolution is necessarily poorer. On the other hand, since these conditions are somewhat closer to 'real life', the relative conformational populations are likely to be governed by entropic as well as energetic factors.

3.2 *Computational*

The analysis and interpretation of biomolecular vibrational spectra recorded at low temperatures are absolutely dependent upon reliable DFT and *ab initio* calculations. These provide a theoretical view of the spectra to compare with experiment, the structural landscapes which they reflect, and the non-covalent interactions that control them.

Spectral, conformational and structural assignments can be based upon a combination of several factors which collectively home in either on unique assignments or, when this is not possible, a limited set of alternative assignments that best fit observation (though the "best fit" may sometimes be thought subjective by skeptics). These include the overall qualitative match between the patterns of observed and calculated vibrational spectra; the correlation between the relative conformer populations (usually estimated from relative spectral intensities) and the ordering and magnitude of their calculated relative zero point energies, or for highly flexible molecules, their calculated pre-expansion free energies. In some cases, unambiguous assignments have been aided by 'chemical' means, for example the selective ^{15}N isotopic labeling of protonated peptides [11].

A typical assignment strategy begins with the generation of the large set of feasible conformers or cluster structures that may be possible in theory (the à la carte menu), although in practice, only a small number of these will be populated (the actual choice). This initial 'brute force' procedure can be done by performing an extensive and iterative Monte Carlo search of the conformational landscape generated by a molecular mechanics force field, followed by their statistical analysis and sorting into distinct conformational families. With increasing biomolecular size and flexibility, these procedures become more and more complicated and a full search of the potential structures becomes increasingly arduous. In these cases the 'brute force' approach can be modified in favour of an intelligent feed-back strategy in which the experimental data are used to guide the calculations. Specific spectral features can signal the characteristics of probable structures, for example groups involved in specific hydrogen bonding can be obtained from the positions of their associated vibrational bands in the experimental IR

spectrum. Structural families which do not provide an acceptably close match (at that level) to the signature of the experimental spectrum can be eliminated from further calculations.

Following the initial survey, a broad set of relevant lowest energy structures (typically with relative energies <50 kJ mol^{-1}) are submitted to calculations at the Hartree Fock level using a moderately sized basis set, and the most stable of these are re-optimized using DFT methods, generally using the B3LYP hybrid correlation functional, which gives reliable results for hydrogen bonded biomolecules; for larger molecules, the most and least sensitive regions of the molecule can be separated, using the ONIOM approximation [25, 26] to treat the two regions at higher and lower levels of theory. The relative energies of the lower energy conformers are then re-calculated, typically using Møller-Plesset second order perturbation (MP2) theory to include electron correlation, and a larger basis set and the lowest-lying DFT structures, i.e. those which might conceivably be populated under the experimental conditions employed, are submitted for frequency calculations. These provide the harmonic vibrational frequencies and IR intensities and the zero point and free energy corrections; experimentally calibrated "anharmonic" scaling factors [27] are used to compare the calculated spectra with the experimental ones. Alternatively, in small systems accurate anharmonicities can be determined *ab initio* using vibrational SCF methods. [28]

Although these strategies have been widely and successfully used to interpret the vibrational spectral signatures of many cold, hydrogen-bonded biomolecules and their molecular complexes they seldom provide absolute agreement with all the observed frequencies, particularly those associated with strongly coupled vibrations or those which

have been strongly displaced through H-bonded interactions. The principal source of error is associated with the vibrational anharmonicity.[27,28] In addition, computed molecular conformations, and in consequence predicted vibrational spectra, can become seriously inadequate when dispersive interactions are important.[29] Calculations conducted using MP2 theory, the least expensive ab initio method which takes them into account, are known to overestimate dispersive interactions. They also become very expensive for molecules larger than one building block unit, although the 'resolution of identity' (RI) approximation, which reduces the demands on computer time without seriously compromising the degree of approximation, alleviates this problem. DFT calculations using the B3LYP functional are notoriously unable to reflect the influence of dispersion. This has encouraged, and continues to encourage the development of new or improved, computationally efficient methods, [31] for example DFT calculations which have been modified by adding a weighted empirical dispersion term to the total DFT energy, D-DFT, [32] or calculations which employ a functional that takes proper account of electron exchange and correlation. [33] Other more sophisticated methods include spin-component-scaled MP2 (SCS-MP2) [34,35], and a combination of density fitting and local MP2 (DF-LMP2) which facilitates calculations for larger molecules (~40 non-hydrogen atoms) using medium sized basis sets (at least triple-zeta). [36]

The methods so far described, while appropriate at low temperatures, where only a small number of 'frozen' structures are likely to be populated, are inappropriate for flexible biomolecular populations probed at elevated temperatures. Under these conditions, barriers to large amplitude side-chain motions and also to back-bone conformational isomerization may be surmounted, to create a dynamical population of

conformers; furthermore entropic differences between them can alter their (free) energy surfaces. Although a 'zero order' view may still be obtained through comparisons between experimental spectra and calculated composite spectra associated with a judicious selection of individual 'frozen' back-bone conformational structures, a higher level understanding of the relation between the experimentally observed biomolecular vibrational spectrum and its underlying structural and dynamical properties can be sought through a Born-Oppenheimer molecular dynamical (BOMD) approach. [9,37,38] In this approach the dynamical IR spectrum is related only to the time-dependent dipole moment and, provided the forces acting on the nuclei are accurately determined and the dipole moment time correlation functions are accurately calculated along the trajectory, its Fourier transform will provide an accurate simulation of the experimental vibrational spectrum [9]. In addition, since the finite temperature dynamics take place on all accessible parts of the potential energy surface and not just the minima, anharmonicities are automatically taken into account.

4. Scenes from life

4.1. The evolution of secondary structure in isolated peptides

The H-bonded interactions in peptide chains, and their structural environments, are reflected very directly in the spectral shifts of their localized NH stretch (amide A) modes. They are not coupled to other amide vibrational modes and are generally only weakly coupled to each other but most importantly of all, they are extremely sensitive to the strength of the local H-bonded interactions and any cooperative effects. The interactions generate dispersed spectral patterns to create distinct signatures, located typically in the wave number range *ca.* 3200-3450 cm^{-1}, which can be fully resolved for

short oligopeptides, typically with up to ~5 residues, in the gas phase at low temperatures, or partially resolved at higher temperatures or for larger peptides. NH→O=C interactions can also be reflected, less dramatically, by shifts in the bands associated with the C=O stretch (amide I, *ca.* 1700-1800 cm^{-1}) and NH bending (amide II, *ca.* 1400-1500 cm^{-1}) modes, though the latter may also be coupled to C-N stretch modes. These can be used to aid the conformational assignment of frozen peptide conformations at low temperature through comparisons with harmonic DFT calculations but molecular dynamics simulations do not support their use at temperatures ~300K. [9]

Following the reductionist 'building block' approach, the initial target of most of the spectroscopic investigations of peptide structure conducted in the gas phase has been to identify the key structural interactions – principally, intramolecular hydrogen-bonded NH→O=C attractions along the peptide backbone and sometimes, hydrogen-bonded or dispersive attractions involving interactions with the side chain(s). These, together with the opposing steric repulsions, determine their twists and turns. The intrinsic locally folded structures within short synthetic peptide sequences have been explored to identify the basic *motifs* which can control the extended secondary structures of much longer peptide chains (in the absence of environmental interactions), either through their periodic repetition, to create for example helices or pleated sheets, or through the formation of reverse turns which help to create tertiary protein structures; basic types of cyclic H-bonded *motifs* are shown schematically in Figure 1. The shorter, C_5 and C_7 *motifs* create local β_L and γ turns; the longer C_{10} and C_{13} *motifs*, when periodically repeated, generate the β-3_{10} and the more extended α-helical secondary structures.

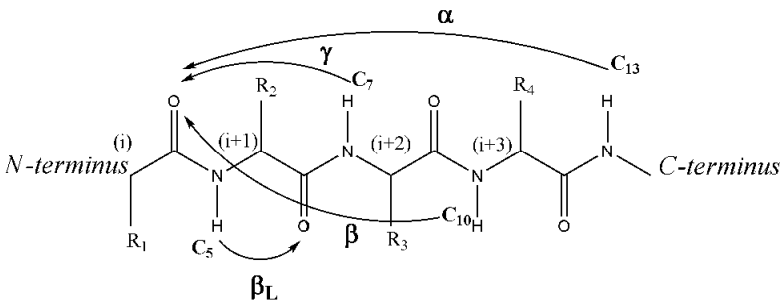


Figure 1. Peptide backbone turns created by hydrogen bonding.

Many groups [6-8] have explored the development of these *motifs* in small, neutral peptides - the appearance of β_L, γ, β and α-turns in the bench-mark series of capped oligo-peptides, Ac-Phe.$(Xxx)_{n=0\text{-}2}$-NH_2, systematically characterized by Mons and his co-workers, [8] provide an excellent illustration of the approach. The acetyl cap, located at the N-terminus, creates model di- (n=0), tri- (n=1) or tetra- (n=2) peptide sequences; the non-polar amino acids such as glycine, alanine, valine or proline, provide the chain of Xxx residues; and the phenylalanine residue provides the necessary UV chromophore required by the IR-UV detection strategy. Peptide conformers, stabilized through collisional cooling in a free jet expansion, have been individually resolved and characterized through analysis of their NH vibrational spectra coupled with DFT calculations.

The shortest peptides (n=0) adopted conformations presenting local β_L-C_5 and γ-C_7 turns. Peptides with n=1, adopted conformations displaying sequences of β_L-γ or γ-γ turns but, in addition, these model tripeptides could also adopt conformations (calculated to be their global minimum structures) which incorporated the more extended C_{10} turn, linked through a hydrogen bond from the amide cap at the C-terminus to the acetyl cap at the N-terminus. [8] The conformational structures of peptides with n=2 were

dependent on the sequence of amino acid residues. When the Phe residue was located at the N-terminus (Ac-FAA-NH_2), the principal structure incorporated a γ-γ sequence (the building block of a 2_7 ribbon – a helical structure based on a periodic two residue C_7-C_7 sequence); a centrally located Phe residue (Ac-AFA-NH_2) generated two C_{10} turns (the building block of a 3_{10} helix); its location at the C-terminal (Ac-AAF-NH_2) generated a C_{10}-C_7 sequence. In all three cases, the aromatic residue promoted an additional side-chain-backbone interaction, through local NH→π bonding. The structures are shown together with their associated experimental and calculated vibrational spectra in Figure 2. [8].

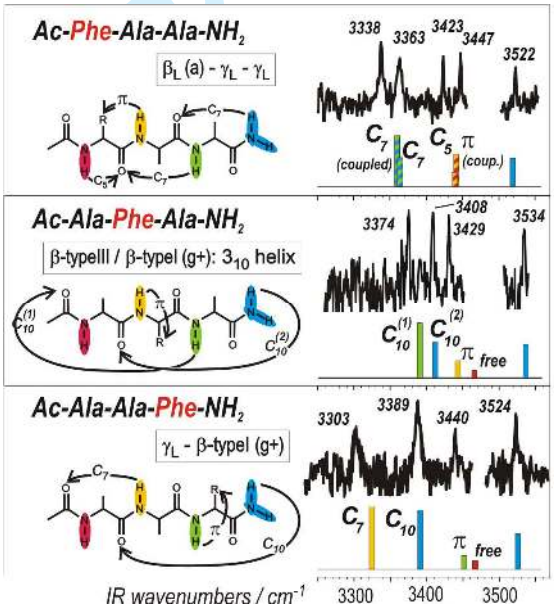


Figure 2. The experimental IR ion depletion spectra and the DFT computed vibrational spectra and conformational structures of three capped tripeptides, Ac-FAA-NH_2, Ac-AFA-NH_2 and Ac-AAF-NH_2. Note their sensitivity to the amino acid sequence. [8]

The first intimation of a C_{13} α-turn in an isolated gas phase oligopeptide was provided by de Vries's group through a pioneering IR-UV investigation [39] of the

uncapped penta-peptide, FDASV (Phe.Asp.Ala.Ser.Val). Remarkably, only a single conformer (or perhaps a small family of similar conformers) could be detected in the free jet expansion. The best fit between the experimental and DFT computed vibrational spectra was provided by a structure incorporating a C_{13} α-turn, linked through hydrogen bonding between the terminal residues, $NH_V \rightarrow O{=}C_F$, (see Figure 3). The band just below 3600 cm^{-1}, signaling a 'free' carboxylic OH, was assigned to overlapping contributions from the OH groups in the terminal valine residue and the aspartic acid side-chain. Their involvement in local OH→X interactions would have been expected to compromise the extended C_{13}-α turn. Despite this achievement, the limited resolution of the vibrational spectrum and the questionable accuracy of DFT calculations conducted using the B3LYP functional and a rather small basis set, 6-31G**, constrained their precision and their unambiguous assignment, a limitation recognized by practitioners in the field but not always explicitly acknowledged.

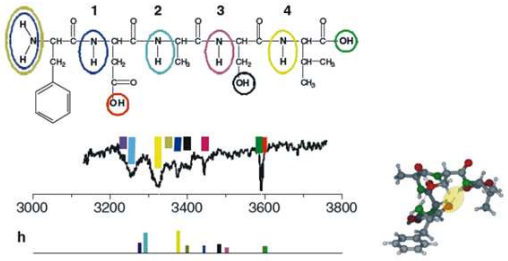
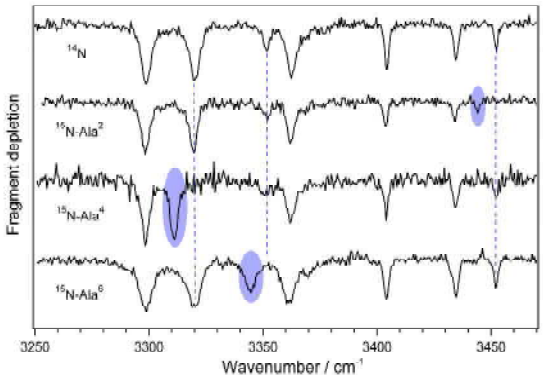


Figure 3. The IR depletion spectrum of the pentapeptide, FDASV: experiment vs. DFT theory. [39]

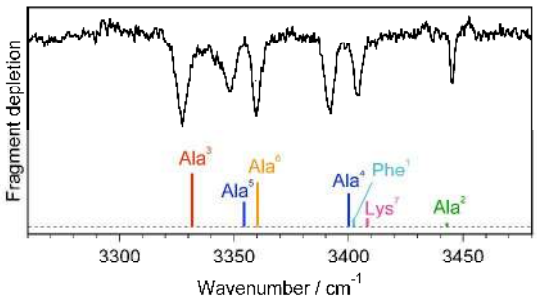
The most precise spectroscopic investigations of the beginnings of secondary structure in oligopeptides of increasing chain length have been provided by Rizzo's

group, through measurements of high resolution NH vibrational spectra in the series of systematically varied protonated model oligopeptides, Ac-FA_5KH^+, Ac-$FA_{10}KH^+$ and Ac-KH^+FA_{10} (Ac-Phe.Ala_5.$LysH^+$, Ac-Phe.Ala_{10}.$LysH^+$ and Ac-$LysH^+$.Phe.Ala_{10}), isolated in the gas phase and stored in a cold ion trap [11,12] (see section 3.1). Phenylalanine again provides the essential UV chromophore. The presence of the positive charge introduces much stronger intramolecular interactions and locating the lysine residue at the C-terminus, rather than the N-terminus, encourages formation of an α-helical peptide structure, stabilised through electrostatic interaction with the large dipole of the helix. [15] This is not the case when the proton is sited on a lysine residue at the N-terminus.

The series of vibrational spectra shown in Figure 4(a), where each of the peptide NH bands in the lowest energy conformer of Ac-FA_5KH^+ are clearly resolved, demonstrates the state of the art. The spectra also provide a clear resolution of their ^{14}N-^{15}N isotopic shifts to allow unambiguous vibrational and structural assignments – and a stringent assessment of the accuracy of DFT calculations, see Figure 4(b). The scaled DFT frequencies for oligopeptides of this size are not perfect but they are not that bad either! The spectral signatures of the larger, A_{10} peptides were, not surprisingly, more congested and in consequence, not quite fully resolved.



(a)



(b)

Figure 4. (a) 'Natural' and isotopically labeled IR depletion spectra of the lowest energy conformer of the protonated peptide, Ac-FA$_5$KH$^+$; (b) comparison with a DFT calculation (B3LYP/6-31++G**) for the lowest energy conformation of the natural peptide. [11]

What about their conformational structures? Do their spectral signatures reflect the development of α-helical turns? Figure 5 shows results obtained for the two lowest energy populated conformers of Ac-FA$_5$KH$^+$: both display C_{13} helical α-turns and a series of C_{10} turns as well as more local, C_5 and NH-π, hydrogen-bonded interactions. Strikingly, the 'free' carboxylic OH band at ~3570 cm^{-1} which appears when the lysine residue is located at the N-terminus, disappears when it is at the C-terminus, indicating its involvement in hydrogen-bonding and lending support to the structural conclusions based on ion mobility measurements. The first appearance of a fully developed, secondary α-

helical structure very likely occurs in the extended peptide, Ac-$FA_{10}KH^+$, but this has not yet been rigorously established.

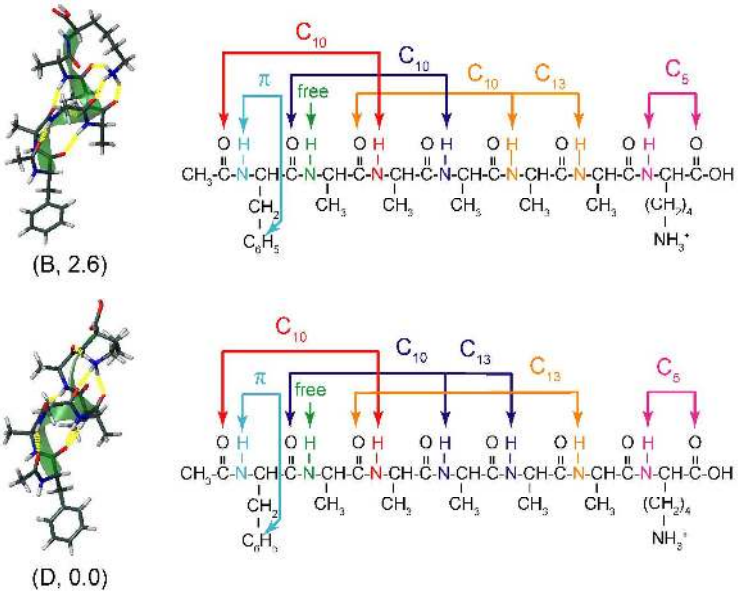


Figure 5. Conformational structures of the two lowest energy conformers of Ac-FA_5KH^+. [11]

We turn now to the pleated β-sheet *motif*, which links two peptide strands in a parallel or antiparallel orientation through an alternating series of NH→O=C and C=O←HN bonds to provide a key protein structural element. Its 'spontaneous' appearance in small model oligopeptide dimers, isolated in the gas phase and independent of any environmental control, has been demonstrated by Gerhards' group [40,41] using the capped model tripeptide, Ac-Val.Tyr(Me)-NHMe, to simulate a minimalist protein segment. (Replacement of the phenolic OH in the tyrosine residue by a methoxy group prevented it acting as a hydrogen bond donor). The IR ion depletion spectrum of the monomer recorded in the amide A, I and II regions, when coupled with DFT and *ab initio*

calculations, identified a stretched tripeptide backbone conformation which could serve as a template for β-sheet formation. This was duly confirmed by the spectral signature of the dimer which could be assigned uniquely, to the antiparallel β-sheet structure shown schematically, in figure 6.

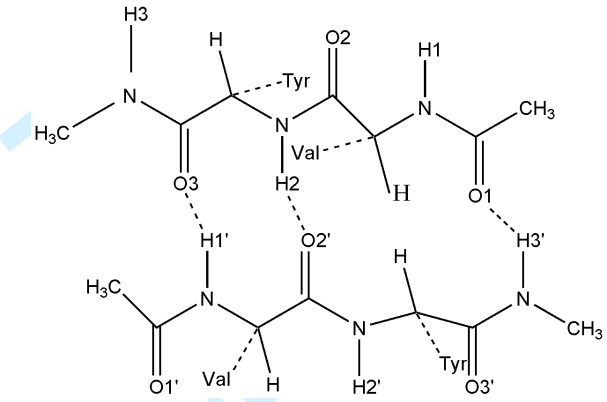


Figure 6. Schematic representation of the β-sheet *motif* in the model dimer, (Ac-ValTyr(Me)-OMe)$_2$. [41]

The formation of β-sheet structures also has a more sinister, pathological significance since they are major components of the insoluble amyloid fibrils that are associated with the formation of intra- or intercellular plaques and protein misfolding diseases such as Alzheimer's and Huntington's disease; fibrils are constructed through the non-covalent aggregation of β-sheet units. Several short amino acid sequences that form common β-sheet building block *motifs* have been identified in amyloid proteins. [42] One of them, the hexa-peptide sequence, ^{306}VQIVYK^{311} (Val.Glu.iso-Leu.Val.Tyr.Lys) which occurs in the so-called tau-peptide amyloids, belongs to a family of 'steric zipper' *motifs*, which contribute to fibril aggregation through interdigital

packing of the neighbouring β-strand side chains. Can the strategy employed by Gerhards to characterize the simplest model β-sheet *motif*, be successfully extended to characterize

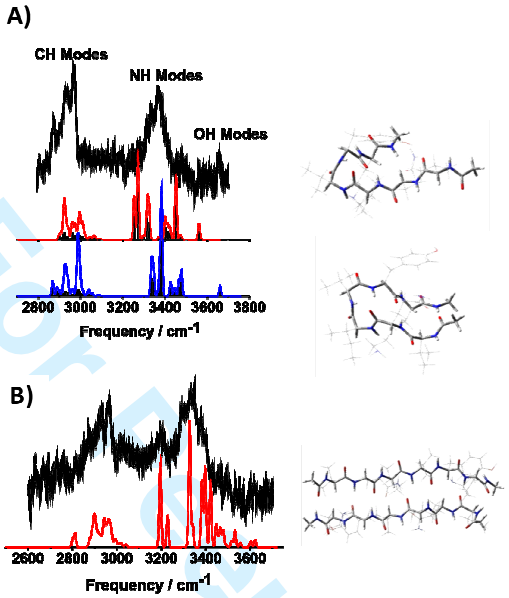


Figure 7. IR depletion spectrum of (A) the zipper peptide Ac-VQIVYK-NHMe and (B) its dimer isolated at low temperature in the gas phase, together with DFT (B3LYP) calculations of the best fit spectra and the corresponding conformational structures. [43,44]

the *inherent* conformation of the much larger, and biochemically more relevant, steric zipper *motif*? Figure 7 provides an affirmative answer. It shows the NH vibrational spectra of the capped peptide, Ac-VQIVYK-NHMe and its dimer, isolated in the gas phase at low temperature, recorded by Vaden *et al.*, [43,44] and the associated structures they reflect.

The cluster of bands lying between 3500 cm^{-1} and 3200 cm^{-1} is associated with a series of increasingly strongly hydrogen-bonded NH modes; those lying below 3000 cm^{-1} are associated predominantly with the (very large number of) CH modes. Despite the inevitable spectral congestion, which could also be due in part to contributions from more

Deleted: were

Deleted: were

than one conformer, comparisons between the experimental and computed vibrational signatures indicate a preference for a β-strand conformation even in the isolated monomer. This is facilitated by a β-hairpin turn, which provides two, loosely extended peptide chains linked by successive hydrogen bonds.

The spectrum of the peptide dimer, shown in Figure 7(b), is in remarkably good, qualitative agreement with the computed vibrational signatures associated with β-sheet structures, for example the antiparallel β-sheet shown alongside it; although alternative parallel structures could not be excluded, *non* β-sheet structures could. [44] Their binding energies were more than sufficient to straighten out the β-hairpin conformation of the monomeric peptide. The β-sheet *motif* is generated spontaneously, in the gas phase, without the intervention of the environment, driven by thermodynamic rather than kinetic control, to create a structure in which the linked peptide backbone forms a central stem and the side chain groups lie on the outside - precisely the *motif* displayed by the 'zipper' structural units which aggregate together within an amyloid fibril.

These results raise interesting issues in regard to the role of an aqueous environment. As Wyttenbach and Bowers have noted [13], "With life processes taking place in a water environment…there is a constant competition between hydration and interaction between biomolecules carrying out a biochemical process." The behaviour of the Alzheimer's amyloid-β (Aβ) protein fragment provides an instructive example. In the gas phase molecular dynamics calculations [45] predict a structure with a hydrophilic core and the hydrophobic residues on the outside, *cf.* the gas phase structure of the zipper peptide, but in an aqueous environment its conformational structure is predicted to turn inside out. On the other hand many protein interactions occur in hydrophobic

Deleted: d
Deleted: was
Deleted: was
Deleted: M
Deleted: s of the Alzheimer's amyloid-β (Aβ) protein fragment *'in vacuo'*
Deleted: for example,
Deleted: ed
Deleted: wa
Deleted: by immersion in an aqueous environment

environments, in cell membranes for example, where water molecules are largely excluded. These issues will be revisited in section 4.3 dealing with carbohydrates and their molecular recognition. For the moment we simply note the remarkable absence of any reported spectroscopic (as opposed to computational) investigations of hydrated neutral peptides in the gas phase save for one exception, the singly hydrated, model β-sheet segment, Ac-Val.Tyr(Me)-NHMe [40] (in marked contrast to the many ion mobility-mass spectrometric (IMS-MS) investigations of protonated peptide hydration, [46,47]). The anti-climactic result was a structure in which the water molecule simply acted as a spectator, bound as a hydrogen bond acceptor to the terminal methyl amino group and exerting no influence on the extended peptide backbone conformation.

Deleted: (

Deleted: s

Deleted:), save for one single exception reported by Gerhards *et al.*, [40] who established the structure of the singly hydrated, model β-sheet segment, Ac-Val.Tyr(Me)-NHMe.

All of the structural investigations discussed so far, have been undertaken at low temperatures. What happens when the temperature is closer to that of a living organism? What structural *motifs* do protonated peptides, for example, access when they are produced through electrospray ionization and held in an *un*cooled ion trap at ambient temperature (~300K), or when they are created at higher temperatures (~350K) through protonation within a photo-excited donor-acceptor complex. [23] The set of protonated alanine oligopeptides, $(Ala)_{n=2-8}H^+$, created through the latter strategy provide an instructive series [48]; their resonant IR-MPD spectra are shown in Figure 8. Despite the increased spectral complexity and the poor resolution it is still possible to discern the development of characteristic vibrational signatures as the peptide chain length increases – note, for example, the sudden appearance of the strongly displaced feature when n≥7. Can they be converted into corresponding conformational structures or structural

Deleted: ose

Deleted: s

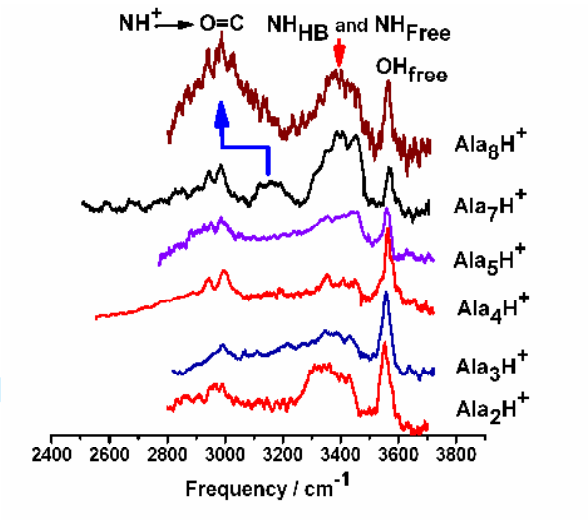


Figure 8. Resonant IRMPD spectra of protonated alanine oligopeptides, generated in the gas phase at T ~ 350K [48]

families? As discussed in section 3.2, there are two ways to approach this question, and both have been applied. The 'poor man's' approach, qualitatively matches the broad structural features against a series of computed composite vibrational spectra, containing contributions from each of the lowest lying, thermally accessible (in terms of ΔG rather ΔE), frozen conformational structures [48]. The more rigorous approach recognizes their dynamical character, and quantitatively matches the detailed contours of the vibrational spectra with those computed through Born-Oppenheimer molecular dynamics calculations which take into account the dynamics of the accessible populations of inter-converting conformers and the vibrational anharmonicities. [9,37,38]

The poor man's approach has identified a series of conformational transitions with increasing chain length, from linear (protonated on a carbonyl backbone) to folded or cyclic (protonated at the N-terminus) when n=3, through to cyclic (n≥4) and eventually, globular structures (all protonated at the N-terminus) when the chain length has increased

to n≥7. The globular structure corresponds to full charge solvation of the protonated terminal amino-group in sharp contrast to the α-helical structures adopted by alanine oligopeptides, when terminated by a protonated lysine residue. These results mirror similar conclusions based upon ion mobility spectroscopy. [14] A successful application [38] of the more realistic, DFT-based molecular dynamics approach is shown in figure 9. The averaged spectrum shown in figure 9C includes all the structural families explored during the MD trajectories, *i.e.*, the extended and cyclic families shown alongside figures 9A and B. The figure also includes, for comparison (9D), vibrational spectra computed for the corresponding frozen conformations.

Deleted: , folded

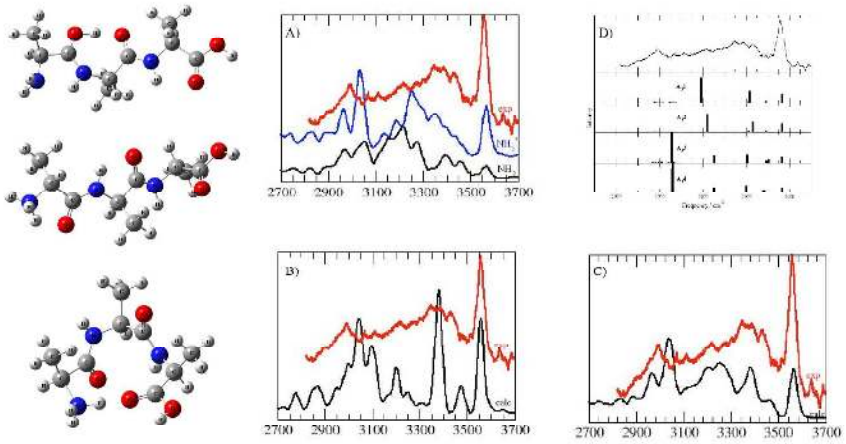
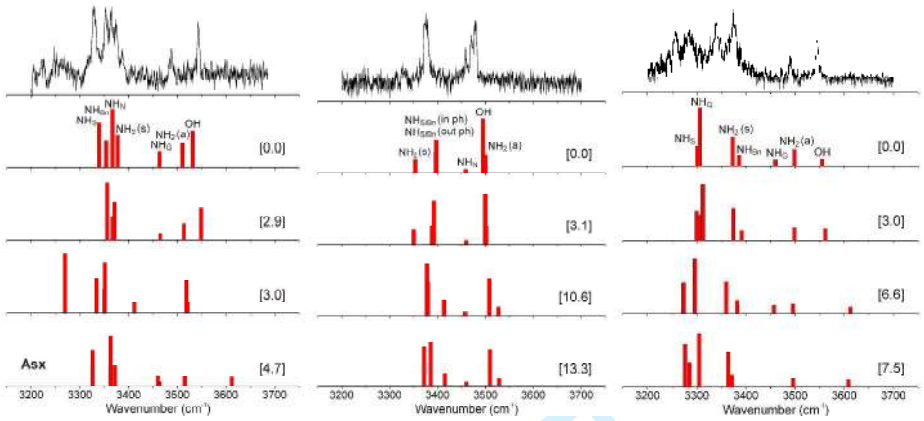


Figure 9. The molecular dynamics approach: experimental and computed vibrational spectra of $(Ala)_3H^+$ reflecting trajectories exploring three conformational families. (A), extended, with protonation on a carbonyl group (NH_2) or on the N-terminus (NH_3^+) and (B), cyclic, linked by hydrogen bonding between the C-terminus and the protonated N-terminus; (C), the average over all conformations; (D), DFT computed spectra of the corresponding frozen conformations (extended and folded at the top, cyclic at the bottom). [38]

Deleted: folded with protonation

As a final illustration, we turn to a recent double resonance, IR-UV spectroscopic and computational study [49] of the consensus *motif*, Asn.Xxx.Ser/Thr (NXS/T; X≠Pro), present in all *N*-linked glycoproteins. It was designed to address some of the 'biologically relevant' questions posed in section 2, particularly the possibility of a structural basis for the evolutionary choice of the NXS/T *motif*. Figure 10 presents the vibrational spectra, recorded at low temperature in a supersonic expansion, of three capped peptides: one



(a) (b) (c)

Figure 10. Experimental IR ion depletion spectra of (a), Ac-NGS-NHBn and two 'chemical' mutants, (b) Ac-NPS-NHBn and (c) Ac-QGS-NHBn, together with the computed vibrational spectra of their four lowest energy conformations: relative energies in kJ mol^{-1}. [49]

containing an 'allowed' sequence, NGS, and two containing 'forbidden' sequences, the 'chemical' mutants, NPS and QGS. The mutations were introduced to identify the intrinsic structural consequences of replacing the central glycine (G) residue with proline (P), or replacing the asparagine residue (N), where the acetamido side chain provides the unique glycosylation site, with glutamine (Q) where the side chain is slightly extended by the presence of an extra methylene group.

Each peptide presents a distinct vibrational spectrum indicating differences in the pattern of hydrogen-bonded interactions and consequently, their associated secondary structures. Mutations make a difference! Figure 10 also includes DFT/B3LYP calculations of the scaled vibrational spectra associated with each of their four, lowest energy conformers; they include backbone NH (amide A) modes, Asn or Gln side-chain

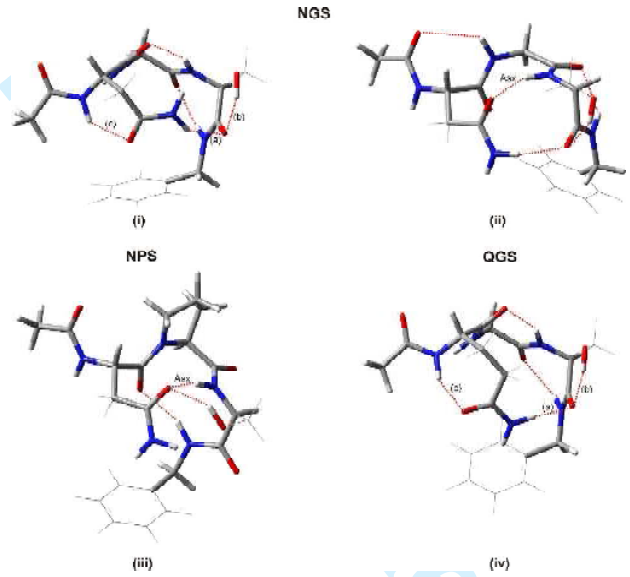


Figure 11. Lowest energy conformations of (i) Ac-NGS-NHBn, (iii) Ac-QGS-NHBn, and (iv) Ac-NPS-NHBn, and (ii) the 'Asx' turn conformation of Ac-NGS-NHBn, lying at relative energy 4.7 kJ mol^{-1}. [49]

NH_2 modes (symmetric and antisymmetric), and the Ser OH mode. In each case, the experimental spectra match quite well, the computed vibrational spectra associated with the lowest energy structures, shown in Figure 11 (i) NGS, (iii) NPS and (iv) QGS.

The intrinsic structures adopted by their peptide backbones represent a compromise between conflicting demands: those imposed by hydrogen-bonded interactions with the carboxamide side chain, their own internal interactions and in NPS, the conformational constraint associated with the proline residue, which promotes the

formation of a C_{10} β-turn. How do these structures compare with those displayed in the condensed phase or in 'real-life'? In aqueous (or DMSO) solutions, the preferred conformation of NGS, identified through 2D-NMR measurements coupled with MD simulations [50], presents an 'Asx' turn, created by hydrogen bonding between a backbone amide and the carbonyl group on the acetamido side chain, $(NH)_S \rightarrow O=C$. This does not correspond to the intrinsic global minimum energy structure populated in the gas phase, but to the excited structure shown in Figure 11 (ii), a few kJ mol^{-1} higher in energy. However it is also blessed with a much larger electric dipole, 7.85D *cf.* 2.52D and in the high dielectric environment of an aqueous or DMSO solution it becomes the lowest energy conformation. [49]

In 'real life', the great majority of N-linked glycoproteins present extended NXS/T conformations, with most probable $(OH)_{S/T}\ldots(N)_{side\text{-}chain}$ distances ~7.3Å, [51] more than double the distance in NGS, both in its global minimum gas phase conformation and in the 'Asx' conformation it adopts in solution. Evidently, glycosylation of the side chain during protein biosynthesis results in a large conformational change which extends the local conserved segment of the peptide backbone. Such a change cannot occur when the central residue is proline and furthermore, the hydrogen bonded interactions which sustain the folded peptide backbone in NGS are weaker than those in QGS: food for thought.

4.2. Spectral signatures and structural *motifs* in isolated and selectively hydrated carbohydrates

In contrast to amino acids, which can only form linear peptide chains, carbohydrate units may link together to form branched chains, for example *N*-linked

glycans or the Lewis antigens, or linear chains, for example cellulose polysaccharides. The Lewis antigens are key blood group determinants and they can also be major immuno-therapeutic targets – for example, the trisaccharide, Lewis^x, which is expressed by many epithelial cancers. Its intrinsic gas phase structure is not the same as its structure in the condensed phase or when bound as a protein ligand. [52]

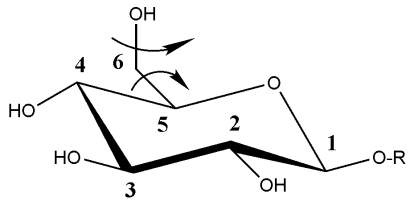
Carbohydrate molecular recognition by proteins is a key factor in many biological processes but the factors dictating selective carbohydrate–protein interactions are not easily discerned. Selectivity must depend in part upon the conformational structures of the interacting carbohydrate ligand and its receptor but because of their flexibility and the universality of hydrogen-bonded interactions in aqueous environments, their structures may also be influenced by explicit hydration, dehydration, or the direct involvement of water at the receptor site. Modeling and understanding carbohydrate structures, their interactions with water, with proteins, or with both, presents a notoriously complex challenge. Characterizing their *intrinsic* three dimensional structures marks the key starting point on the road towards meeting it – the first step on the way to investigating and understanding their interactions with other molecules, not least water.

It should come as no surprise, given the theme of this article, to learn that mass-selected double resonance vibrational spectroscopy, coupled with DFT and *ab initio* computation, is providing a powerful means of meeting the challenge. [1] Their spectral signatures are extremely sensitive to their local H-bonded environments, and they provide an excellent diagnostic when compared with DFT and *ab initio* calculations, facilitating their spectroscopic and conformational assignment. [1] The strategy is being exploited to resolve and characterize the intrinsic gas phase conformations of a series of 'biologically

relevant' mono- and oligo-saccharides, including glycan building blocks and Lewis antigens; the structures of some of their hydrated and molecular complexes, created at low temperature in a supersonic expansion; the hydrogen-bonded networks they support (or which support them); and the factors which determine their conformational and structural preferences.

4.2.1. Monosaccharides

Hexoses such as the glucopyranoside shown schematically in figure 12 in its $^4C_1$ chair configuration, can in principle adopt many conformations. The exocyclic hydroxymethyl group, and its terminal OH6 group can adopt *gauche* or *trans* orientations, denoted by the labels, G+, G-, T and g+, g-, t, respectively. The peripheral OH groups often form extended hydrogen-bonded chains which may present counter-clockwise (OH4→OH3→OH2→O1) or clockwise (O1→OH2→OH3→OH4) orientations, which can be denoted by the labels, cc or c.



Dihedral angles		
(O5-C5-C6-O6)	-60°	G-
	+60°	G+
	180°	T
(C5-C6-O6-H6)	-60°	g-
	+60°	g+
	180°	t

Figure 12. Schematic representation of the glucose hexapyranoside; the stereochemistry at C1 can generate equatorial (β, as shown) or axial (α) anomers.

The lowest energy conformers of the "tagged" hexapyranose, phenyl β-D-glucopyranoside (βpGlc), have been individually resolved through double resonance UV hole burning experiments. [1] Their distinct IR depletion spectra in the mid IR, where the four OH bands can be partially or fully resolved, and in the 'fingerprint' region at lower wave numbers, see figure 13, allow their assignment to the three rotamers of the

exocyclic hydroxymethyl group, oriented to the "right" (G+), or "upwards" (G-) or to the "left" (T), see figure 13(b). An illustration of the spectral sensitivity to changes in the local H-bonded environment is provided by the shift to low wave number of the OH6 vibrational band, labeled σ6, in the ccTg+ conformer which reflects the hydrogen bond OH6→OH4, facilitated by the Tg+ orientation of the exocyclic CH_2OH group.

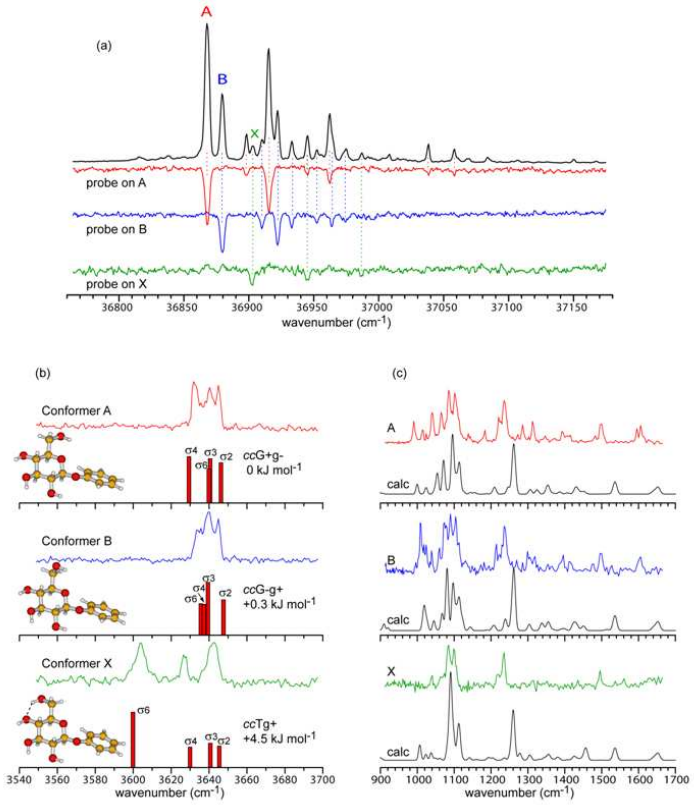


Figure 13. (a) UV hole burn and (b,c) IR depletion and DFT computed vibrational spectra of phenyl β-D-glucopyranoside. [1]

The sensitivity is displayed rather more dramatically in the series of IR spectra shown in figure 14, which compares the OH "vibrational signature" of the β anomer of galactose (where OH4 is axially oriented), with the corresponding signatures of α and β

fucose (6-deoxy galactose, where the CH_2OH group is replaced by CH_3) and xylose (where there is no exocyclic group and OH4 is equatorially oriented). Each of the four OH bands in galactose, σ2-6, is clearly resolved but σ6 is "missing" from the vibrational spectra of fucose and xylose, which both lack the exocyclic CH_2OH group.

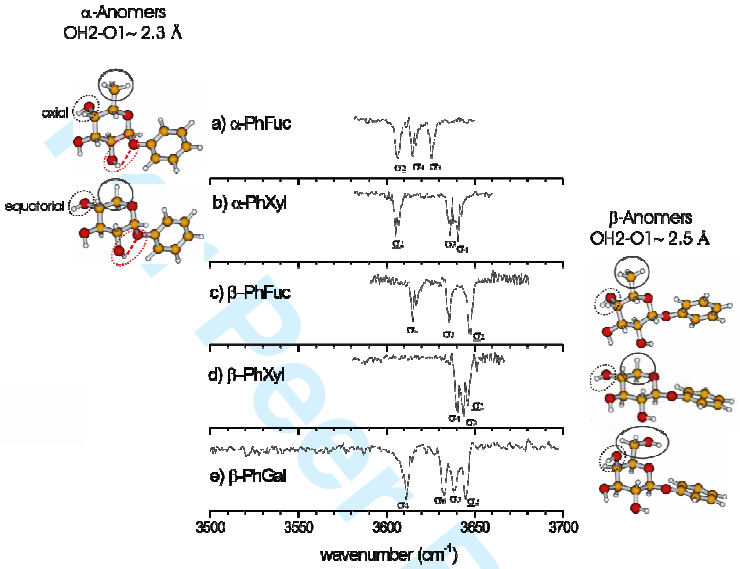


Figure 14. "Vibrational signatures" of galactose (with OH4 axially oriented and containing an exocyclic CH_2OH group), fucose, (6-deoxy galactose, containing a CH_3 group), and xylose (OH4 equatorially oriented, no exocyclic group).

More strikingly, the α and β anomers display a characteristic alternating displacement of σ2 from higher (in the β-anomers) to lower (α-anomers) wave numbers, reflecting the decrease in their respective OH2-O1 distances. There is a similar alternating shift in the position of σ4, between galactose and fucose on the one hand and xylose on the other, which reflects the change in orientation of OH4 from axial (Gal, Fuc) to equatorial (Xyl) and the consequent change in the OH4-O3 hydrogen-bond distance.

4.2.2. Hydration of monosaccharides: conformation, selectivity and the rules of the game.

The conformational preferences in hydrated glucose (Glc), galactose, (Gal) and mannose (Man), which has an axial OH2 group, are controlled primarily by the flexibility of their exocyclic hydroxymethyl groups. Further controls are exerted by their anomeric configuration and by the relative orientations (axial *vs.* equatorial) of their hydroxyl groups - in fucose and xylose, where there is no hydroxymethyl group these are the only controls available. All three factors can operate separately or collectively, and may adapt the intrinsic carbohydrate conformations in order to optimize the cooperative sequence of intra- and inter-molecular hydrogen-bonded interactions in their hydrated complexes and to create binding site selectivity. [2-4] In effect, the bound water molecules act like 'spies' surveying the electrostatic topography of the carbohydrate to seek out, or create, the most favoured binding site(s).

The first (and second) bound water molecules in Glc, Gal and Man invariably occupy sites near the hydroxymethyl group, inserting between OH6 and its nearest neighbour, OH4, or between OH6 and O5, the (4,6) or (6,5) sites: figures 15 and 16 provide two clear illustrations of this rule. In singly and doubly hydrated phenyl β-D-glucopyranoside, the preferred conformation of the hydroxymethyl group changes from

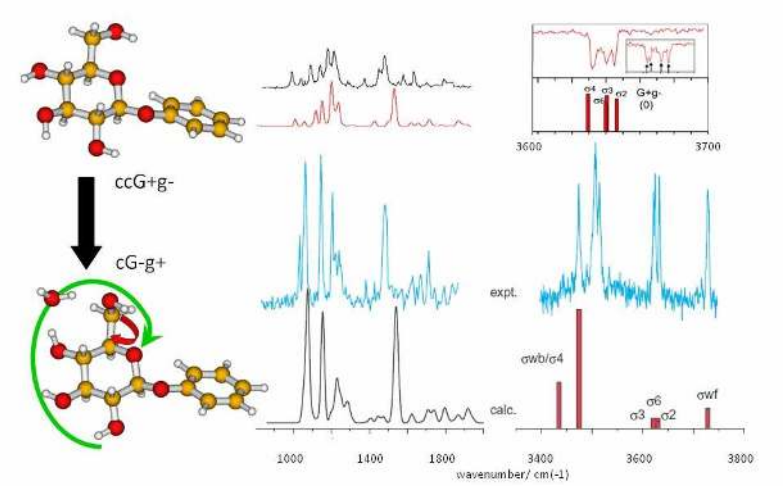


Figure 15. Experimental (upper) and computed (lower) IR spectra of the most populated conformers of phenyl β-D-glucopyranoside and its monohydrate. Note the change in conformation promoted by explicit hydration, which allows selective insertion of the bound water molecule at the (4,6) site.

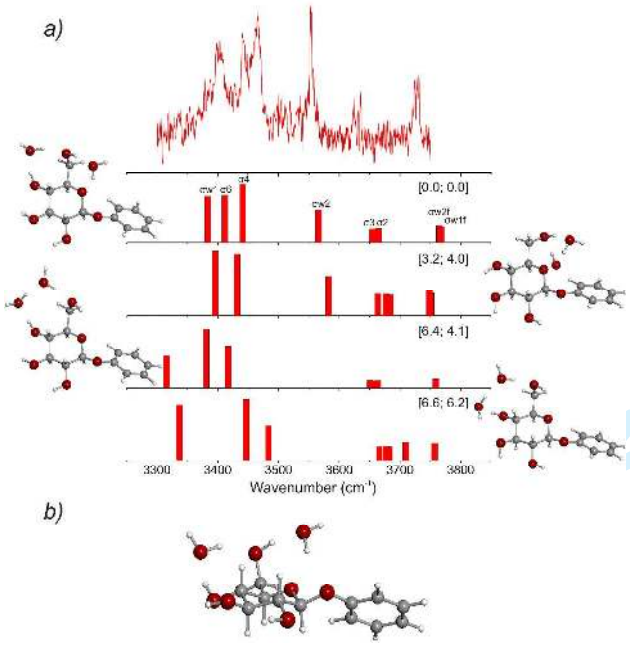


Figure 16. (a) IR spectrum of doubly hydrated phenyl β-D-glucopyranoside and the computed vibrational spectra, structures, relative energies and free energies (kJ mol^{-1}) of its four lowest energy structures. (b) side-on view of the populated, minimum energy structure. [3]

G+g- to G-g+ while the orientation of the peripheral OH groups switches from counter-clockwise, 'cc' to clockwise, 'c'. These changes provide attractive 'pockets' at the 4,6 site, for insertion of the first bound water molecule, (Figure 15) and the (6,5) site for insertion of the second (Figure 16) – signalled by the appearance of the strongly displaced OH bands lying below 3600 cm^{-1}. Their vibrational signatures are in good accord with those predicted for their global minimum energy structures.

Counter-intuitively, despite the equatorial orientation of all its hydrophilic groups, the two (separately) bound water molecules in the doubly hydrated sugar are both located above the ring rather than around its edge, directed by the "perpendicular" G-g+ orientation of the hydroxymethyl group. The IR spectra of triply hydrated hydrated phenyl β-D-mannose, where OH2 is oriented axially, revealed similar behaviour, with the first water molecule occupying the (4,6) site and the second and third inserted, as a water dimer, between OH6 and the axial OH2 group [2]. Strikingly, all three water molecules were located on the hydrophilic, polar face, organized through cooperative hydrogen bonding into a cyclic, clockwise orientation, dictated by the (perturbed) conformation of the carbohydrate to which they were attached; its hydrophobic, apolar face remained 'dry'.

In contrast to mannose (and glucose and galactose), the IR spectra of the amino carbohydrate, 2-*N*-acetyl glucosamine (GlcNAc), and its hydrated complexes reveal structures where the bound water molecules are marshaled into place not by the hydroxymethyl group, but by the acetamido group, which rotates out of the pseudo-plane of the pyranose ring. Indeed its propensity to act as a hydration focus is so great that in the trihydrate, all three water molecules are focused around the acetamido group, [3]

linked to the C=O and NH groups above and below the (2,3) site, behaviour that might help to explain the differing and widespread roles of GlcNAc, and perhaps GalNAc, in nature.

4.2.2. Oligosaccharides: probing the glycosidic linkage

The 3D structures of linear and branched oligosaccharides depend upon the nature of the glycosidic linkages which connect their individual monosaccharide units, in particular their torsional angles and anomeric configurations. Although the heavy atom, chair conformations of the linked (pyranoside) units are generally retained, those of their peripheral OH or hydroxymethyl (or other substituent) groups may not be. If their local structural *motifs are* wholly, or at least partially retained, the vibrational spectra of the isolated monosaccharide 'building blocks' provide an 'alphabet' of vibrational signatures, which can be used to refine the assignment of the larger oligosaccharide structures in which they are incorporated. When this is not the case, the structural modifications, for example the disruption of structural *motifs* promoted by inter-residue hydrogen bonding, will be encoded in their altered IR spectra. The first type of behaviour, nicely illustrated by the vibrational spectrum of the mannose disaccharide, Manα(1,3)Man, one of the terminal arms of the core branched penta-saccharide unit in *N*-linked glycoproteins, has been discussed by Çarçabal *et al.* [53] Striking examples of the second type are provided by recent investigations of the (phenyl tagged) cellulose building block, cellobiose (Glcβ(1,4)Glc), its mono-hydrate and its C4' epimer, lactose (Galβ(1,4)Glc); [54,55] the two disaccharide structures are shown schematically in Figure 17.

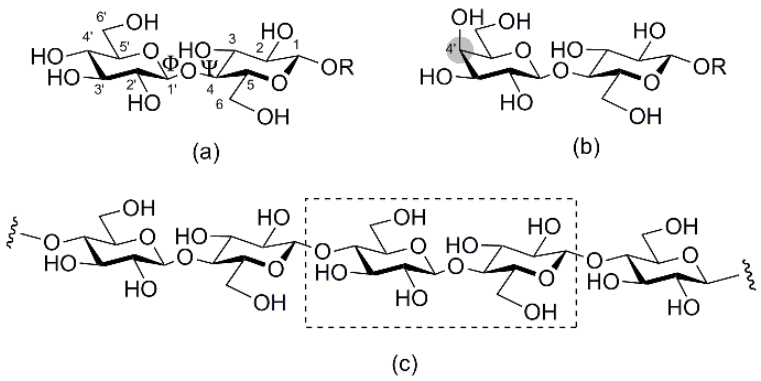


Figure 17. Schematic molecular structures of (a) cellobiose, Glcβ(1,4)Glc, (b) its C4' epimer, lactose, Galβ(1,4)Glc, and (c) cellulose. The glycosidic dihedral angles, Φ and Ψ, can be defined in terms of the heavy atom sequences, Φ: O5'-C1'-O-C4 and Ψ: C1'-O-C4-C3.

Comparisons between their experimental and computed vibrational spectra, shown in Figure 18, reveal their intrinsic conformational structures. Each one corresponds to the DFT calculated global minima, all associated with a *cis* glycosidic configuration supported by two inter-ring hydrogen-bonds. In the cellulose disaccharide, cellobiose, these link the neighbouring hydroxymethyl groups, OH6'→OH6, and two of the OH groups, OH2'→OH3, (one of the links in the cooperative clockwise $(OH{\rightarrow}O)_n$ chain, OH4′→OH3′→OH2′→OH3→OH2). The modest displacement of the vibrational modes, σ6' and σ2', to ~3545 cm^{-1}, reflects the moderate strength of the inter-ring hydrogen bonding. In lactose, the changed orientation of the OH4′ group from equatorial to axial, promotes a strong OH4′→OH6′ interaction which dramatically alters the conformational landscape, reversing the orientation of the cooperative $(OH\text{-}O)_n$ chain and extending it. [55] The enhanced cooperativity creates much stronger inter-ring binding through the OH6′→OH6 and OH3→OH2′ hydrogen bonds, reflected in its vibrational signature which shifts to substantially lower wavenumbers in comparison with cellobiose; the

vibrational spectrum now displays three intense bands associated with the modes, $\sigma 4'$ and $\sigma 6'$ (coupled through strong hydrogen bonding) and $\sigma 3$.

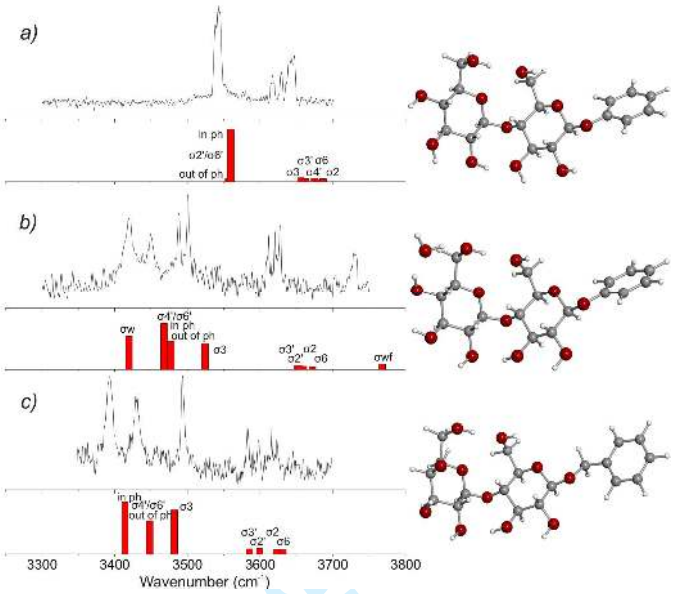


Figure 18. Experimental IR depletion spectra and DFT computed vibrational spectra and structures of phenyl-tagged cellobiose (a), its singly hydrated complex (b) and lactose (c). [54,55]

When cellobiose is hydrated, the water molecule inserts at the favoured glucose binding site, between OH4′ and the neighbouring hydroxyl group OH6′. This promotes a major structural reorganization, creating a configuration and an associated vibrational signature that parallels and now closely resembles that of its *un*hydrated epimer, lactose. The bound water molecule takes on, in effect, the role of the axial OH4', acting as a structural element and greatly strengthening the inter-ring hydrogen bonding.

Do these results have anything to do with 'real life' since the cellobiose building blocks in cellulose polymers (and also in the solid state and in aqueous solution) typically adopt a *trans* configuration, associated with the alternating structure shown in Figure 17(c). This could be created directly at each enzyme catalyzed polymerization step, now

generally accepted to occur at the end of the growing β1,4-linked chain; alternatively it might be accessed subsequently by relaxation from an initial *cis* orientation. A direct mechanism is problematic, since each polymerization step, adding one glucose residue at a time, could be expected to create a uniform, non-alternating *cis* configuration. In the growing cellulose polymer, however, the successive OH4′ groups are capped (in a glycosidic bond) which breaks the cooperative $(OH\text{-}O)_n$ chain in the free disaccharide. This might well facilitate a change in conformation about the glycosidic linkages from *cis* to *trans* in the growing polymer, allowing formation of the sequence of OH3→O5′ inter-ring hydrogen-bonds *along* the extended polymer chain. It also removes the favored (4′,6′) water binding site, which might contribute to the intrinsic resistance of cellulose polymer to solvation. The typical *trans* conformation in cellulose may not be a result of inherent inter-residue interaction but instead, arise through a combination of its extended structure and micro-hydration.

The key branched trimannose *motif*, Manα(1,3)Manα(1,6)Man, found in the core of most hydrated *N*-linked glycans, might also be strengthened by acceptance of a water molecule, this time into its branch "fork". The provision of a stiffened structural scaffolding unit by the conserved glycan core could facilitate the recognition processes, 'molecular discussions' that take place at the glycan termini, and also protect the peptide chain from enzymic attack. Indeed, the recurrence of this site in many fully solvated glycan structures could suggest the existence of a "water pocket" in *N*-glycans that conserves a bisecting water molecule as a beneficial structural element - the specific glycosidic linkages may have been selected *because* they can be 'stiffened' by direct or

solvent mediated hydrogen-bonded interactions across them, thereby influencing the shape of the glycan, its structural integrity and biological specificity.

Despite the complexity of its conformational landscape and 'a fortiori', the vast number of potential water binding sites, both (phenyl tagged) trimannoside (triman) and its singly hydrated complex (triman·H_2O), when cooled to low temperature in the gas phase, populate single structures only. Their IR depletion spectra are shown in figure 19, together with their DFT computed vibrational spectra and the corresponding structures of their lowest energy conformers. [56] The bare mannose trisaccharide adopts

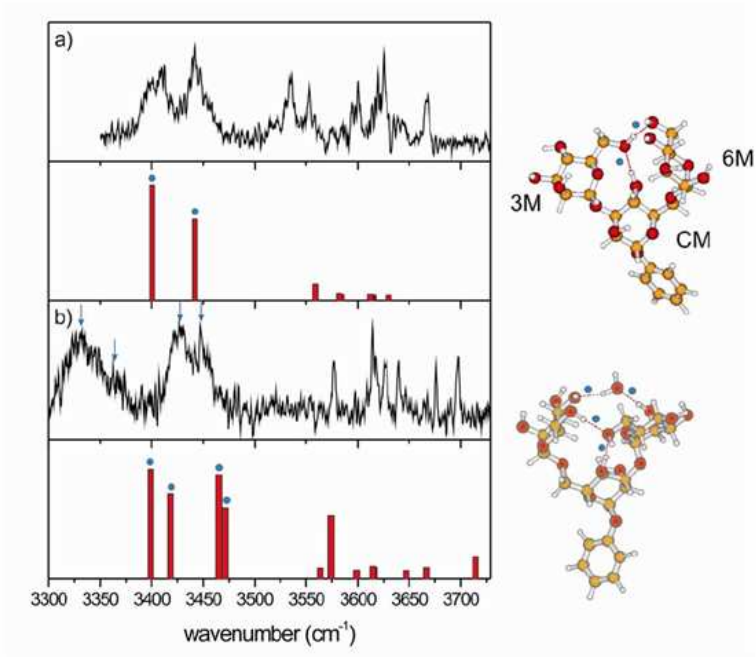


Figure 19. Experimental and computed IR spectra of (a) the trimannoside (phenyl)α-Manα(1,3)Manα(1,6)Man and (b) its singly hydrated complex. The dots indicate the strongly H-bonded OH groups in the structures and the corresponding vibrational bands in the computed spectra. The arrows identify these bands in the experimental spectra. [56]

a 'closed' conformation, held together by strong intramolecular hydrogen bonds. These connect the α(1,3) and α(1,6) arms of the branched trimannoside unit, labeled 3M and 6M in Figure 19(a), *via* the hydroxymethyl group on 3M, which is linked to the hydroxymethyl group on 6M, as well as the OH group at the bisecting site on the central mannose ring, labeled CM. The pair of strongly displaced bands at ~3400 and ~3440 cm^{-1} are associated with their symmetric and antisymmetic vibrational modes. The IR spectrum of the hydrated complex, shown in figure 19(b), presents *two* pairs of strongly displaced bands reflecting a more tightly bound structure. It incorporates a water molecule bound between the 3M and 6M arms; these remain linked to the central site, but *via* a modified H-bonding *motif*, now mediated by the hydroxymethyl group on the 6M arm.

The 'closed' conformational structure of the bare trimannoside, and its singly hydrated complex displayed at low temperature in the gas phase, contrasts with the 'open' (dynamical) structure determined in aqueous solution at 298K, using NMR techniques coupled with MD calculations. One of the set of computed singly hydrated structures however, the centrally hydrated structure shown in Figure 20(a), presents an (entropically favoured) open conformation which is in remarkably close correspondence with that determined in aqueous solution. The structure can be generated by a simple rotation about the α(1,3) linkage in the bare molecule. An even more striking comparison can be made between the calculated structure of the open trimannoside unit, Figure 20(a) and the average computed structure of the core pentasaccharide segment, $(Man)_3(GlcNAc)_2$, shown in Figure 20(b), based upon NMR and MD simulations of the 'water-bridged' *N*-glycan $(Man)_9(GlcNAc)_2$ in an aqueous environment at 298K. [5] This

is highlighted in Figure 20(c), which compares the 'frozen' trimannoside structure shown in Figure 20(a) with the structure of the core pentasaccharide unit in $(Man)_9(GlcNAc)_2$, where the most persistent explicit hydration site incorporates a 'keystone' water molecule H-bonded to the central OH4 site. The similarity between the open structure and the relative orientations of the α(1,3) and α(1,6) 'wings' in the trimannoside and in the corresponding unit in the core pentasaccharide is remarkable. Perhaps the conformation of the key branched (trimannoside) segment is indeed driven, in aqueous solution by the creation of a water 'pocket' in the branch 'fork'; it would not be difficult for the water molecule linked to the central OH4 site to slip into the neighbouring water pocket.

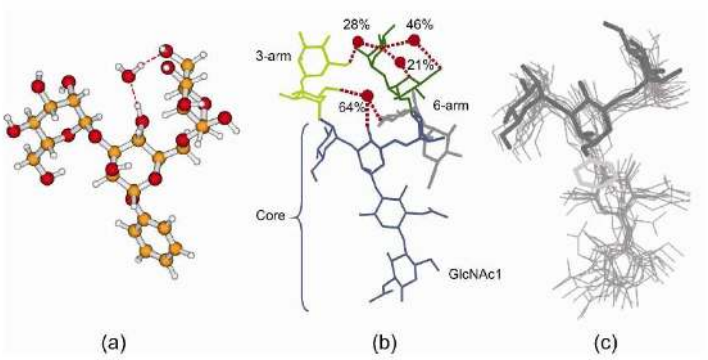


Figure 20. Comparison between (a) the computed open conformational structure of the monohydrate, trinan·H_2O, (b) the computed average conformation of the *N*-glycan, $(Man)_9(GlcNAc)_2$, in aqueous solution at 298K [5], and (c) its core pentasaccharide segment, $(Man)_3(GlcNAc)_2$. Black rods indicate the trisaccharide framework, light grey rods, the phenyl tag and grey thin wires, the pentasaccharide framework.

4.3. Carbohydrate molecular recognition

Protein–carbohydrate interactions are fundamentally important in a wide array of biological processes: infection, fertilization, inflammation, and cellular recognition provide just a few examples. X-ray crystal structures of simple protein–carbohydrate complexes often identify the hydroxymethyl group as one of the favoured recognition

points, with monosaccharide ligands bound at their 4,6 or 6,5 sites – interestingly, also the favoured binding sites in their hydrated complexes. The recognition sites on the proteins are often located adjacent to aromatic amino acid residues, tryptophan, tyrosine or phenylalanine, where the bound carbohydrates can adopt a stacking *motif*, with the aromatic side group providing a platform for the ligand. This has been associated with 'apolar CH–π bonding' which can supplement the hydrogen bonded interactions involving polar OH groups on the carbohydrate and neighbouring polar groups in the local protein environment or bonding that may be mediated through electrostatic interactions with nearby ions, such as Ca^{2+}.

The identification of local structures at carbohydrate–protein binding sites, primarily through X-ray crystallographic or NMR spectroscopic investigations, does not establish the detailed nature of the interactions which create them. Binding to aromatic residues for example, could include hydrogen bonding as well as dispersion interactions and there are arguments which suggest they might actually be driven, or assisted, by entropic effects associated with the hydrophobic exclusion of surrounding water molecules, which provide part of their natural environment. Explicitly bound, conserved water molecules, are commonly involved in the architecture of protein–carbohydrate complexes. This might be because they provide a stronger overall protein–carbohydrate interaction but hydration might also maintain a favoured ligand conformation. The question then arises as to whether the recognition *motif* is that of the carbohydrate or its hydration shell. Consider again, the hydrophilic and hydrophobic faces displayed in multiply hydrated phenyl β-D-mannose. The 'dry' hydrophobic face is ready for stacking interactions with aromatic residues, see Figure 21 but the hydrophilic face is crowned by

an ordered array of water molecules with a structure directed by the conformation of the underlying carbohydrate—which may itself be altered by hydration! Which conformation does the protein recognize?

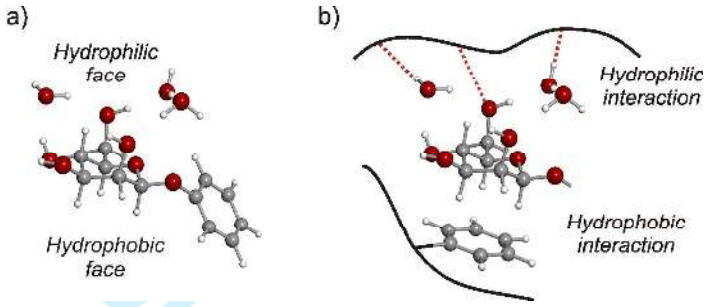


Figure 21. (a) The local structure of triply hydrated (phenyl) β-D-mannose and (b) a cartoon representation of its potential interactions at a protein recognition site.

The spectroscopic detection of weakly bound complexes between carbohydrates and toluene (a model for the aromatic side chain of phenylalanine free of its polar backbone) formed at low temperatures in the gas phase, has established their intrinsic stability in the absence of solvent water molecules and also the general existence of an intermolecular attraction. [57,58] Since the complexes are formed in the absence of water and at low temperature, entropic contributions which could be provided by the reorganization of surrounding water molecules, the essence of the proposed 'hydrophobic interaction' in aqueous environments, cannot operate – which is not to say that they do not contribute in the *condensed* phase but simply, that there is an intrinsic attractive intermolecular potential. In most of the cases explored, stacked structures, involving CH–π dispersion interactions predominated, exemplified by the spectra shown in Figure 22, although complexes bound through specific OH–π hydrogen-bonded interactions could also be identified. The OH vibrational signature of the complex between toluene and methyl α-L-fucopyranoside (the enantiomeric form found in nature), Figure 22, is very

similar to that of the bare carbohydrate and its lowest lying computed structures all present stacked conformations, with the aromatic molecule bound below, or above the pyranoside ring; their computed binding energies were ~20 kJ mol^{-1}. [59]

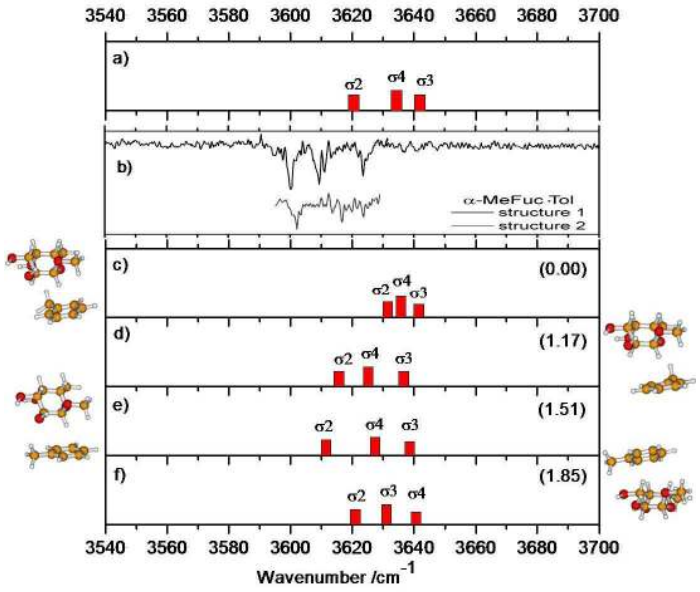


Figure 22. (a) Computed OH vibrational spectrum of methyl α-L-fucopyranoside; (b) Experimental IR depletion spectra of the two populated conformers of its complex with toluene; and (c)–(f) computed vibrational spectra, structures and relative energies (in kJ mol^{-1}) of its lowest lying computed structures, calculated using the MO5-2X functional to include the contribution made by dispersion interaction. (From ref. [59], reproduced with permission from Elsevier).

Having established the reality of bonding with the aromatic side chain, the next step was to move to a more realistic model which also includes the peptide backbone. The simplest way to do this was to replace toluene with the capped amino acid, Ac-Phe-NHMe which provides a model dipeptide receptor. Figure 23 shows the IR depletion spectrum of its complex with methyl β-D-glucopyranoside.

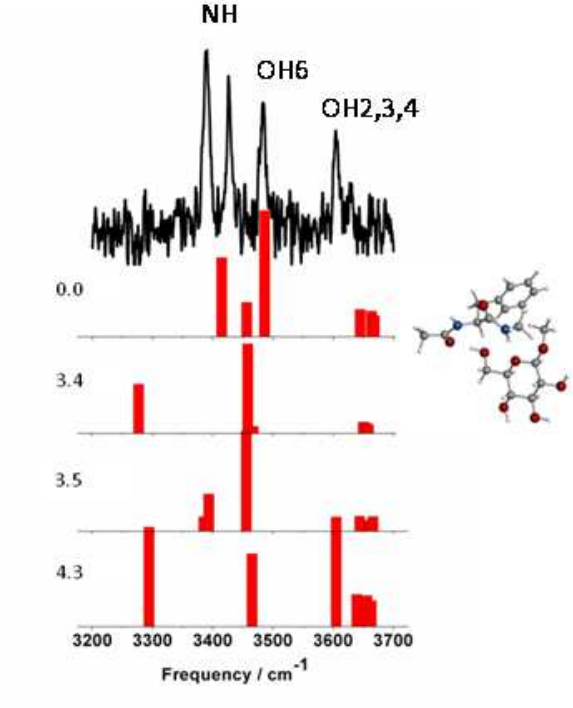


Figure 23. Experimental and DFT (MO5-2X/6-31+G**) computed vibrational spectra and lowest energy (kJ mol^{-1}) structure of a carbohydrate-peptide complex: methyl β-D glucopyranoside with capped phenylalanine, Ac-Phe-NHMe.

An initial qualitative comparison with the NH vibrational spectrum of the bare peptide [60] suggested retention of the extended peptide conformation. In addition, the strong displacement of one of the OH bands to lower wave number indicated strong hydrogen bonding. These speculations were supported by the results of DFT calculations, using either B3LYP or MO5-2X functionals (to assess the possible role of dispersive interactions). The 'best fit' spectrum shown in Figure 23, corresponds to a structure in which the carbohydrate is bound to the peptide chain through hydrogen bonding from the hydroxymethyl group (OH6) to the terminal carbonyl group and from the terminal NH(Me) group back to the carbohydrate. As anticipated, the peptide adopts its favoured

extended 'β-sheet' peptide backbone conformation but the orientation of its aromatic side chain, adjacent to the methyl group of the carbohydrate, suggested the possibility of a weak dispersive interaction with the methyl group, not the pyranoside ring. The peptide complex with methyl β-D-galactopyranoside presented similar behaviour.

Early experiments for the complexes of the peptide with the two corresponding α-anomers also indicated two similar structures but their vibrational signatures were not at all like those of the β-anomers – an example of chiral recognition in the gas phase [61] and an indication perhaps, of anomeric selectivity in carbohydrate-protein binding, even in so simple a model system – which is, perhaps, a good point on which to end the article.

5. Conclusions

Albert Einstein has been quoted as saying "If we knew what it was we were doing, it would not be called research"; this article has tried to address a slightly different issue, "If we did not know *why* we were doing it, perhaps it might not deserve to be called research." At first it was a sufficient challenge to characterize the intrinsic structures of biomolecules large enough to be of biochemical interest but now that challenge has been met it is time to take stock and take seriously, the next and more important challenge, to explore their molecular interactions and to interact with the biochemical world. There seems little point in spectroscopic and computational explorations of the structure and interactions of biomolecules *per se*, unless the explorations are conducted within a biochemically relevant context. This brief review has provided some very recent case histories to illustrate how this can be done.

Formatted: Indent: First line: 0 pt

Formatted: Font: Italic

Acknowledgments

Firstly, I thank Prof Benjamin Davis for his inspirational research collaboration and guidance in matters biochemical; secondly, I am grateful to Profs Marie-Pierre Gaigeot and Thomas Rizzo, and Drs Michel Mons and Timothy Vaden for their generous help in providing some of the illustrative figures; and lastly but certainly not least, I owe a great debt to Drs Emilio Cocinero, Cristina Stanca-Kaposta, Timothy Vaden and Pierre Çarçabal for being such good colleagues in recent years.

References

[1] J.P. Simons, P. Çarçabal, B.G. Davis, D.P. Gamblin, I. Hünig, R.A. Jockusch, R.T. Kroemer, E.M. Marzluff and L.C. Snoek, Intl. Rev. Phys. Chem., **24**, 489 (2005).

[2] E.J. Cocinero, E.C. Stanca-Kaposta, E.M. Scanlan, D.P. Gamblin, B.G. Davis and J.P. Simons, Chem. Eur. J., **14**, 8947 (2008).

[3] E.J. Cocinero, E.C. Stanca-Kaposta, M. Dethlefsen, B. Liu, D.P. Gamblin, B.G. Davis and J.P. Simons, Chem. Euro J., **15**, (2009), in press.

[4] J.P. Simons, B.G. Davis, E.J. Cocinero, D.P. Gamblin and E.C. Stanca-Kaposta, Tetrahedron Asymmetry, **20**, 718 (2009).

[5] R.J. Woods, A. Pathiaseril, M.R. Wormald, C.J. Edge and R.A. Dwek, Eur. J. Biochem., **258**, 372 (1998).

[6] J.-P. Schermann, *Spectroscopy and Modelling of Biomolecular Building Blocks*, Elsevier, Amsterdam, 2008.

[7] M.S. de Vries and P. Hobza, Ann. Rev. Phys. Chem., **58**, 585 (2007).

[8] W. Chin, F. Piuzzi, I. Dimicoli and M. Mons, Phys. Chem. Chem. Phys., **8**, 1033 (2006).

[9] G. Grégoire, M.-P. Gaigeot, D. C. Marinica, J. Lemaire, J.-P. Schermann and C. Desfrançois, Phys. Chem. Chem. Phys., **9**, 3082, (2007).

[10] J.A. Stearns, C. Seaiby, O.V. Boyarkin and T.R. Rizzo, Phys. Chem. Chem. Phys., **11**, 125 (2009).

[11] T.R. Rizzo, J.A. Stearns and O.V. Boyarkin, Intl. Rev. Phys. Chem. **28**, 481 (2009).

[12] N.C. Polfer and J. Oomens, Mass Spectrometry Rev., **28**, 468 (2009).

[13] T. Wyttenbach and M.T. Bowers, Ann. Rev. Phys. Chem., **58**, 511, (2007).

[14] M.F. Jarrold, Ann. Rev. Phys. Chem., **51**, 179 (2000).

[15] M.F. Jarrold, Phys. Chem. Chem. Phys., **9**, 1659 (2007).

[16] B.C. Bohrer, S.I. Mererbloom, S.L. Koeniger, A.E. Hilderbrand and D.E. Clemmer, Ann. Rev. Anal. Chem., **1**, 293 (2008).

[17] M. Sharon and C.V. Robinson, Ann. Rev. Biochem., **76**, 167 (2007).

[18] M.Y. Choi and R.E. Miller, J. Phys. Chem. A, **111**, 2475 (2007).

[19] E.C. Stanca-Kaposta and J.P. Simons, in *Handbook of High Resolution Spectroscopy,* Ed. F.Merkt and M.Quack, Wiley, New York, 2010.

[20] M. Guidi, U.J. Lorenz, G. Papadopoulos, O.V. Boyarkin and T.R. Rizzo, J. Phys. Chem. A, **113**, 797 (2009).

[21] Y. Hu and E.R. Bernstein, J. Chem. Phys., **128**, 164311 (2008).

[22] J. Oomens, B.G. Sartakov, G. Meiger and G. Von Helden, Intl. J. Mass Spectrometry, 254, **1** (2006).

[23] N.A. Macleod and J.P. Simons, Mol. Phys., **104**, 3317 (2006).

[24] T.D. Vaden, S.A.N. Gowers and L.C. Snoek, Phys. Chem. Chem. Phys., **11**, 5843 (2009).

[25] G.S. Tschumper and K. Morokuma, J. Mol. Struct. (THEOCHEM), **592**, 137 (2002).

[26] J.-C. Poully, G. Grégoire and J.-P. Schermann, J. Phys. Chem. A, **113**, 8020 (2009).

[27] Y. Bouteiller, J.-C. Gillet, G. Grégoire and J.-P. Schermann, J. Phys. Chem. A, **112**, 11656 (2008).

[28] A.A. Adesokan and R.B. Gerber, J. Phys. Chem.A, **113**, 1905 (2009).

[29] D. Reha, H. Valdes, J. Vondrasek, P. Hobza, A. Abu-Riziq, B. Crews and M.S. deVries, Chem. Eur. J., **11**, 6803 (2005).

[30] J. Černý, P. Jurečka, P. Hobza and H. Valdés, J. Phys. Chem. A, **111**, 1146 (2007); T. Kubař, P. Jurečka, J. Černý, J. Řezáč, M. Otyepka, H. Valdés and P. Hobza, J. Phys. Chem. A, **111**, 5642 (2007).

[31] P. Hobza, Phys. Chem. Chem. Phys., **10**, 2581 (2008).

[32] T. Schwabe and S. Grimme, Phys. Chem. Chem. Phys., **9**, 3397 (2007).

[33] Y. Zhao and D.G. Truhlar, J. Chem. Theo. Comp. **3**, 289 (2007).

[34] S. Grimme, J. Chem. Phys., **118**, 9095 (2003).

[35] J.G. Hill, J. Platts and H.-J. Werner, Phys. Chem. Chem. Phys. **8**, 4072 (2006).

[36] H.-J. Werner, F.R. Manby and P.J. Knowles, J. Chem. Phys., **118**, 8148 (2003).

[37] M.-P. Gaigeot, J. Phys. Chem. A, **112**, 13507 (2008).

[38] A. Cimas, T.D. Vaden, T.S.J.A. de Boer, L.C. Snoek and M.-P. Gaigeot, J. Chem. Theo. Comp., **5**, 1068 (2009).

[39] A. Abo-Riziq, J.E. Bushnell, B. Crews, M. Callahan, L. Grace and M.S. de Vries, Chem. Phys. Lett., **431**, 227 (2006).

[40] H. Fricke, A. Gerlach, C. Unterberg, P. Rzepecki, T. Schrader and M. Gerhards, Phys. Chem. Chem. Phys., **6**, 4636 (2004).

[41] H. Fricke, A. Funk, T. Schrader and M. Gerhards, J. Am. Chem. Soc. **130**, 4692 (2008).

[42] M.R. Sawaya, S. Sambashivan, R. Nelson, M.I. Ivanova, S.A. Sievers, M.I. Apostol, M.J. Thompson, M. Balbirnie, J.J.W. Wiltzius, H.T. McFarlane, A. Ø. Madsen, C. Riekel and D. Eisenberg, Nature, **447**, 453 (2007).

[43] T.D. Vaden, S.A.N. Gowers, T.S.J.A. de Boer, J.D. Steill, J. Oomens and L.C. Snoek, J. Am. Chem. Soc., **130**, 14640 (2008).

[44] T.D. Vaden, S.A.N. Gowers and L.C. Snoek, J. Am. Chem. Soc., **131**, 2472 (2009).

[45] A. Baumketner, S.L. Bernstein, T. Wyttenbach, G. Bitan D.B. Teplow, M.T. Bowers and J.E. Shea, Protein Sci., **15**, 420 (2006).

[46] M.F. Jarrold, Ann. Rev. Phys. Chem., **51**, 179 (2000).

[47] T. Wyttenbach, D. Liu and M.T. Bowers, Intl. J. Mass Spectrometry, **240**, 221 (2005).

[48] T.D. Vaden, T.S.J.A. de Boer, J.P. Simons, L.C. Snoek, G. Suhai and B. Paisz, J. Phys. Chem. A, **112**, 4608 (2008).

[49] E.J. Cocinero, E.C. Stanca-Kaposta, D.P. Gamblin, B.G. Davis and J.P. Simons, J. Am. Chem. Soc., **131**, 1282 (2009).

[50] B. Imperiali and K.L. Shannon, Biochemistry, **30**, 4374 (1991).

[51] A.-J. Petrescu, M.R. Wormald, R.A. Dwek, Curr. Opin. Struct. Biol., **16**, 600 (2006).

[52] Z. Su, B. Wagner, E.C. Cocinero, B. Ernst and J.P. Simons, Chem. Phys. Letters, **477**, 365 (2009).

[53] P. Çarçabal, I. Hünig, B. Liu, D.P. Gamblin, R.A. Jockusch, R.T. Kroemer, L.C.

Snoek, B.G. Davis, A.J. Fairbanks and J.P. Simons, J. Am. Chem. Soc., **128**, 1976 (2006).

[54] E.J. Cocinero, D.P.Gamblin, B.G. Davis and J.P. Simons, J. Am. Chem. Soc., **131**, 11117 (2009).

[55] R.A. Jockusch, R.T. Kroemer, F.O. Talbot, L.C. Snoek, P. Çarçabal, J.P. Simons, M. Havenith, J.M. Bakker, I. Compagnon, G. Meijer and G. von Helden, J. Am. Chem. Soc., **126**, 5709 (2004).

[56] E.C. Stanca-Kaposta, D.P. Gamblin, E.J. Cocinero, J. Frey, R.T. Kroemer, B.G. Davis, A.J. Fairbanks and J.P. Simons, J. Am. Chem. Soc., **130**, 10691 (2008).

[57] J. Screen, E.C. Stanca-Kaposta, D.P. Gamblin, B. Liu, L.C. Snoek, B.G. Davis and J.P. Simons, Angew. Chem., Int. Ed., **46**, 3644 (2007).

[58] E.C. Stanca-Kaposta, D. P. Gamblin, J. Screen, B. Liu, L.C. Snoek, B.G. Davis and J. P. Simons, Phys. Chem. Chem. Phys., **9**, 4444 (2007).

[59] Z. Su, E.C. Stanca-Kaposta, E.J. Cocinero; B.G. Davis and J.P. Simons, Chem. Phys. Letters, **471**, 17 (2009).

[60] M. Gerhards, C. Unterberg, A. Gerlach and A. Jansen, Phys. Chem. Chem. Phys., **6**, 2682 (2004).

[61] A. Zehnacker and M. Suhm, Angew. Chem., Int. Ed., **47**, 6970 (2008).