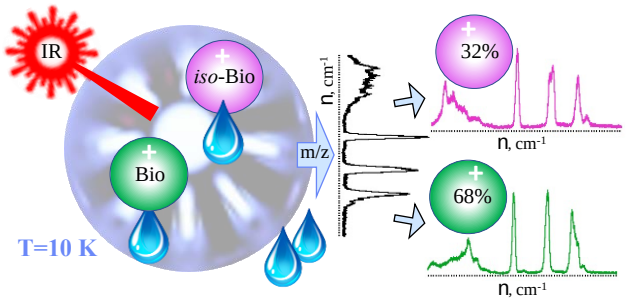

Identification of Isomeric Biomolecules by IR Spectroscopy of Solvent-Tagged Ions.

Erik Saparbaev, Andrei Zviagin, and Oleg V. Boyarkin*

Laboratoire de Chimie Physique Moléculaire, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

ABSTRACT: Difference in functionality of many isomeric biomolecules requires their analytical identification for life science studies. We present a universal approach for quantitative identification of different small to midsize isomeric biomolecules that can be brought to the gas phase from solution by electrospray ionization (ESI). The method involves IR fragment cold ion spectroscopy of analyte molecules that are incompletely desolvated by soft ESI. The use of solvent molecules as natural tags removes a need for adding to solutions any special compounds, which may interfere with LC or mass-spectrometric measurements. The tested peptides and especially monosaccharides and lipids exhibit highly isomer-specific IR fragment spectra of such non-covalent complexes, which were produced from water, methanol, acetonitrile and 2-butanol solutions. The relative concentrations in solution mixtures of, for instance, two isomeric dipeptides can be quantified with the accuracy of 1.6 % and 2.9 % for the acquisition time of 25 min and, potentially, 5 s, respectively; for three isomeric phospho-octapeptides the accuracy becomes 4.1% and 11% for 17 min and, potentially, 10 s measurement, respectively.

Isomeric identification of biomolecules remains one of the essential challenges in analytical chemistry. There is no single universal approach capable of reliable identification of isoforms for the large variety of biomolecules. X-ray diffraction, NMR, and cryo-electron microscopy may provide direct structural information for biomolecules, although each of these methods has certain limitations and practical drawbacks. Owing high resolution and fast data acquisition, mass spectrometry (MS) is the current tool of choice for high throughput identifications of biomolecules in numerous -omics applications, but it remains intrinsically low-capable of distinguishing the isomers. Liquid chromatography (LC) and ion mobility (IM) techniques, which can be conveniently combined with MS analysis, may separate isomers, tagging them with retention and arrival time, respectively. The resolution of IM instruments is rapidly increasing,[1] often allowing separation of isomers with similar collisional cross-sections.[2-5] The higher is the LC/IM resolution the more stringent are the requirements for stability of experimental conditions to reproduce the time tags from day-to-day and in different laboratories. This intrinsic issue shifts the focus from separation to well-reproducible identification of isomers.

Different from LC and IM, spectroscopy reflects molecular structure on a quantum mechanical level, which makes optical spectra highly reproducible fingerprints of analytes. The recently developed combination of UV photodissociation spectroscopy (PD) of cold ions and high-resolution MS (2D UV-MS)[6] demonstrates unprecedent capabilities in quantitative isomeric identification of many types of biomolecules, including peptides,[7,8] drugs,[9] carbohydrates[10,11] and lipids.[12] Along with UVPD, different types of IRPD spectroscopy of ionized biological molecules have been developed for identifications of their isomers.[13-18] A combination of IM and IRPD spectroscopy of the ions that are cryogenically tagged by low-polarized small molecules[19] (e.g., D_2, N_2) demonstrates a particularly promising alternative for isomeric identifications of carbohydrates.[20,21] Cryo-tagging requires, however, the conditions in a PD cold ion trap to be optimized simultaneously for tagging and cooling of the ions. The trap temperature, for instance, is to remain relatively high to supress

condensation of N_2 tag molecules onto RF electrodes. Higher temperature reduces spectral resolution and, ultimately, the accuracy of identifications. Moreover, because the tagging is not 100% efficient, the spectra are measured by detecting the reduction of the number of the complexes due to IRPD. Such depletion spectra are inherently non-baseline free, which limits their quality and therefore the accuracy of the method.

Non-covalent tagging directly in solution prior to a MS filtering of the desired ions solves these problems, but requires addition of foreign tag molecules (e.g., alkali atoms[13] for IR or aromatic molecules[10,22] for UV spectroscopy, respectively), which make the solution not always compatible with, for instance, LC separation or may interfere with MS. Here we present a universal and free from these drawbacks method of isomeric quantification with tagging in solution. The method employs IRPD spectroscopy of cold analyte ions that are non-covalently tagged directly by solvent molecules. We demonstrate this baseline-free approach for identification of isomeric dipeptides, phosphorylated oligo-peptides, monosaccharides and lipids dissolved in some common solvents. The accuracy of the method is evaluated by quantification of isomeric peptides in their solution mixtures and it is compared (for dipeptides) with the performance by depletion spectroscopy.

Our experimental setup has been described elsewhere (see SI for details).[11] Briefly, the protonated solvated com-plexes are produced from solution by a nano-ESI source configured for a gentle ionization.[23] The ions are mass-selected by a quadrupole mass-filter and guided into a cold (T=6K) octupole trap, where they are collisionally cooled and then undergo photofragmentation by a pulse of a tuneable IR OPO. The second quadrupole mass-filter detects the appearing charged fragments or the reduction of the number of parent complexes for the fragment or depletion IR spectroscopy, respectively.

The library-based quantitative identification of isomers in their mixtures includes preliminary measurements of IR spectra for all known isomers (library of standards) that are suspected to be present in a mixture. Then, the spectrum of the ions electrosprayed from an unknown analyte solution is measured too. This spectrum is numerically decomposed to a linear combination of the library spectra.[6] The coefficients of this decomposition give the relative concentrations of the mixed in solution isomers.

As a test case, we measured IR fragment spectra of doubly hydrated dipeptides AY and YA by detecting the appearance of the bare peptides as the PD products. The spectra of the two isomers appear visibly distinguishable (Fig. 1a-b) and can be used for quantification of the isomers in solution mixtures. To evaluate the accuracy of such identifications, we mixed both isomers in different proportions and measured IRPD spectra of the mixtures (e.g., Fig. 1c). Each of the spectra was numerically decomposed in the basis set of the spectra of the pure isomers. The results imply that, the relative concentration of these isomers in a mixture can be determined with the root-mean-square deviation (RMSD) of 1.6% (Fig. 1d). The accuracy dropped to 2.9%, however, when the IR depletion spectra of the same isomers and their mixtures (Fig. S1) were used in the decompositions (Fig. S2). This comparison quantifies the advantage of the baseline-free fragment over the depletion spectroscopy, which is widely used for interrogating cryogenically gas-tagged ions. The measurement of one spectrum in the whole range of 3380-3800 cm^{-1} takes 25 minutes. While this timescale is acceptable for offline measurements of the library standards, much faster acquisition time is highly desired, for instance, to make the method compatible with LC separation. We earlier suggested and demonstrated that the analysis of UV spectra measured for library standards may reveal the sets of isomer-characteristic wavelengths, at which the spectra are most different (e.g., peaks, band origins, etc.).[6,24] Once pre-determined, the spectra of analyte mixtures need to be measured at these few specific wavelengths only. Herein we performed the decompositions by taking PD intensities in the measured IR spectra of the standards and analyte mixtures only at five "specific" wavenumbers (Table S1). Compared with the whole spectral range, the accuracy of isomeric quantifications with such truncated set of data decreases, in average, from 1.6% to a still appreciable value of 2.9%. In return, the time of measurements at five wavelengths, potentially, can be reduced by 300 times to 5 seconds, which would make the method compatible with online LC separation. Ironically, the accuracy becomes the same as in the case of decomposition of depletion spectra in the entire spectral range. This comparison demonstrates the potential advantage of the baseline-free technique over the depletion approach from the point of view of data acquisition rate: for the same accuracy the former is 300 times faster than the latter.

The accuracy of isomeric quantification typically reduces with increasing the number and the size of molecules due to their more congested spectra and the accumulated errors of the decomposition.[9,25] Figure 2a-c shows IR fragment spectra of three doubly hydrated model phosphorylated octapeptides, $pTSA_3TSY$, TSA_3pTSY, TSA_3TSpY, which differ by the position of the phosphogroup. It is worth mentioning that the former two isomers could not be resolved using standard UPLC separation.[6] The position, intensities and widths of a few peaks, which are due to NH- and OH-stretch

transitions, make the three spectra visually distinguishable, particularly in the spectral range of 3550–3800 cm^{-1}. To assess the accuracy of the method with this larger set of larger peptides, we prepared four test mixtures of the isomers and measured their IR spectra in this range (e.g., Fig. 2d). The decomposition over the entire range results in RMSD of 4.1% (Fig. 2e). A truncated decomposition with 10 (visually most characteristic) data points (Table S1) reduced the accuracy to almost 11%. This can be compared with the earlier quantification of five isomeric phospho-octapeptides by the method of 2D UV-MS fingerprinting. It simultaneously measures several UVPD spectra by detecting all abundant photofragments with a broadband MS analyzer. The achieved RMSD was 6.5% and 8.1% for the decompositions with a large spectral range and with retaining 10 wavelengths, respectively.[6] Although more powerful and accurate, 2D UV-MS fingerprinting requires a high-resolution broadband MS detection (e.g., Orbitrap, TOF) and is limited to the species that absorb in UV. In this respect, IRPD of hydrated biomolecules is a technically simpler but universal alternative to the UV-MS fingerprinting with still very good ability to quantify isomers of small to midsize biomolecules.

The higher performance of the 2D UV-MS method roots from the fact that UVPD spectra reflect not only absorption transitions, but also a complex dissociation dynamic that involves the ground and excited electronic states of an ion. As a result of this complexity, the UVPD spectra, measured with detection of different photofragments often appear quite distinct. IRPD of bare or tagged ions is a statistical process occurring on the electronic ground surface only, such that, regardless of the vibration used for excitation (by pulses of ns duration), the weakest bond breaks first. Because the IR fragment MS of ions tagged by gases or, as herein, by solvent molecules is trivial, the bare ion only, there is no sense of performing 2D IR-MS measurements.

Along with water, a few other solvents are often used in LC and ESI-MS analysis of biomolecules. The choice of a particular solvent (mixture of solvents) is determined, in particular, by solubility of an analyte and the performance of LC and ESI. A high solubility implies an appreciable analyte-to-solvent binding energy. This energy may not be, however, sufficient for preserving solvated complexes of the analytes during ESI. In addition, larger solvent molecules may make the spectra of the complexes more congested. Regarding this, we tested the IRPD approach with isomeric monosaccharides, *D*-glucosamine (*D*-GlcN; m=179.171 Da) and *D*-galactosamine (*D*-GalN), dissolved in three most common for LC-MS solvents: water, methanol and acetonitrile (Fig.3a-f). Finally, we compared IR fragment spectra of two diastereomeric sphingolipids (m=299.5 Da) tagged by single molecules of 2-butanol (m=74.1 Da), which particularly readily dissolves these lipids (Fig. 3g-h). Regardless of the solvent, the spectra of all the isomeric pairs are visually quite distinct, particularly, when compared with the spectra of YA/AY (Fig.1). We therefore expect that the accuracy in isomeric quantification of the two amino sugars and the two stereometric sphingolipids by IRPD should be higher than that for the dipeptides.

To sum up, cold-ion IRPD spectroscopy of analyte–solvent non-covalent complexes enables library-based quantitative identifications of isomeric biomolecules of different types. The formation of such complexes in solution is natural, as far as the solvent is suitable for dissolving the analyte. Gentle ESI preserves the complexes during transmission to the gas phase, while cryogenic cooling conserves them and suppresses spectroscopic inhomogeneous broadening.

Isomeric YA/AY dipeptides can be identified with 1.6% accuracy during 25 min or, potentially, with 2.9% accuracy just in 5 s, required for spectroscopic sampling at 5 specific IR wavelengths. The accuracy drops to 4% for three phospho-octapeptides due to the higher number and larger size of the isomers. Compared with a commonly employed IR depletion (e.g., of gas-tagged ions), the background-free fragmentation IR spectroscopy of solvent-tagged ions exhibits much better performance in terms of accuracy and/or acquisition rate. The spectra of the tested isomeric monosaccharides (GlcN/GalN) appear well-distinguished with the three tested most common solvents: water, methanol and acetonitrile. A use of a much larger solvent molecules (2-butanol) still allows for excellent stereoisomer -specific spectra of sphingosine C-18 lipids.

Overall, IR spectroscopy of naturally solvent-tagged molecules is viewed as a simple and universal alternative approach for accurate quantitative identification of, essentially, any small to midsize isomers that can be electrosprayed. The method does not require any modifications of an analyte solution optimised for MS or MS-LC analysis and, potentially, can be combined with online LC separation. Switching from spectroscopy mode to a normal LC-MS operation can be done upon request within seconds by increasing collisional activation, for instance, in an ESI source to destroy non-covalent complexes.

Practical implementation of the analytical methods that are based on spectroscopy of cold ions requires tailored MS instruments that will be compact, affordable and simpler than our multipurpose research grade 6 K spectrometer. Such cooling requires a use of bulky (total of >130 kg) two-stage closed cycle refrigerators. We earlier demonstrated that even a ten-fold increase of the temperature still allows for a good accuracy in quantification of isomeric biomolecules.[9] This result enables the use of compact and light (~5 kg) pressure-wave generators that cool down to

~40 K. The development of an affordable spectroscopic module, which is be based on such a cooler and which can be safely added to commercial high-resolution MS instruments is currently underway in our laboratory.

AUTHOR INFORMATION

Corresponding Author

* E-mail: oleg.boiarkin@epfl.ch

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We thank the Swiss National Science Foundation (grant 200020_204072) for supporting this work.

REFERENCES

(1) Kirk, A. T.; Bohnhorst, A.; Raddatz, C. R.; Allers, M.; Zimmermann, S. Ultra-high-resolution ion mobility spectrometry-current instrumentation, limitations, and future developments, *Anal. Bioanal. Chem.* **2019**, *411*, 6229-6246.

(2) Hofmann, J.; Hahm, H. S.; Seeberger, P. H.; Pagel, K. Identification of carbohydrate anomers using ion mobility-mass spectrometry, *Nature* **2015**, *526*, 241-244.

(3) Shliaha, P. V.; Baird, M. A.; Nielsen, M. M.; Gorshkov, V.; Bowman, A. P.; Kaszycki, J. L.; Jensen, O. N.; Shvartsburg, A. A. Characterization of Complete Histone Tail Proteoforms Using Differential Ion Mobility Spectrometry, *Anal. Chem.* **2017**, *89*, 5461-5466.

(4) Leaptrot, K. L.; May, J. C.; Dodds, J. N.; McLean, J. A. Ion mobility conformational lipid atlas for high confidence lipidomics, *Nat. Commun.* **2019**, *10*, 985.

(5) Meier, F.; Kohler, N. D.; Brunner, A. D.; Wanka, J. H.; Voytik, E.; Strauss, M. T.; Theis, F. J.; Mann, M. Deep learning the collisional cross sections of the peptide universe from a million experimental values, *Nat. Commun.* **2021**, *12*, 1185.

(6) Kopysov, V.; Makarov, A.; Boyarkin, O. V. Colors for molecular masses: fusion of spectroscopy and mass spectrometry for identification of biomolecules, *Anal. Chem.* **2015**, *87*, 4607-4611.

(7) Kopysov, V.; Makarov, A.; Boyarkin, O. V. Nonstatistical UV Fragmentation of Gas-Phase Peptides Reveals Conformers and Their Structural Features, *J. Phys. Chem. Lett.* **2016**, *7*, 1067-1071.

(8) Kopysov, V.; Gorshkov, M. V.; Boyarkin, O. V. Identification of Isoforms of Aspartic Acid Residue in Peptides by 2D UV-MS Fingerprinting of Cold Ions, *Analyst* **2018**, *143*, 833-836.

(9) Kopysov, V.; Makarov, A.; Boyarkin, O. V. Identification of Isomeric Ephedrines by Cold Ion UV Spectroscopy: Toward Practical Implementation, *Anal. Chem.* **2017**, *89*, 544-547.

(10) Saparbaev, E.; Kopysov, V.; Yamaletdinov, R.; Pereverzev, A.; Boyarkin, O. V. Interplay of H-bonds with Aromatics in Isolated Complexes Identifies Isomeric Carbohydrates, *Angew. Chem. Int. Ed.* **2019**, *58*, 7346-7350.

(11) Saparbaev, E.; Kopysov, V.; Aladinskaia, V.; Ferrieres, V.; Legentil, L.; Boyarkin, O. V. Identification and Quantification of Any Isoforms of Carbohydrates by 2D UV-MS Fingerprinting of Cold Ions, *Anal. Chem.* **2020**, *92*, 14624-14632.

(12) Saparbaev, E.; Yamaletdinov, R.; Boyarkin, O. V. Identification of Isomeric Lipids by UV Spectroscopy of Noncovalent Complexes with Aromatic Molecules, *Anal. Chem.* **2021**, *93*, 12822-12826.

(13) Polfer, N. C.; Valle, J. J.; Moore, D. T.; Oomens, J.; Eyler, J. R.; Bendiak, B. Differentiation of isomers by wavelength-tunable infrared multiple-photon dissociation-mass spectrometry: Application to glucose-containing disaccharides, *Anal. Chem.* **2006**, *78*, 670-679.

(14) Mucha, E.; Florez, A. I. G.; Marianski, M.; Thomas, D. A.; Hoffmann, W.; Struwe, W. B.; Hahm, H. S.; Gewinner, S.; Schollkopf, W.; Seeberger, P. H.; von Helden, G.; Pagel, K. Glycan Fingerprinting via Cold-Ion Infrared Spectroscopy, *Angew. Chem. Int. Edit.* **2017**, *56*, 11248-11251.

(15) Schindler, B.; Laloy-Borgna, G.; Barnes, L.; Allouche, A. R.; Bouju, E.; Dugas, V.; Demesmay, C.; Compagnon, I. Online Separation and Identification of Isomers Using Infrared Multiple Photon Dissociation Ion Spectroscopy Coupled to Liquid Chromatography: Application to the Analysis of Disaccharides Regio-Isomers and Monosaccharide Anomers, *Anal. Chem.* **2018**, *90*, 11741-11745.

(16) Martens, J.; van Outersterp, R. E.; Vreeken, R. J.; Cuyckens, F.; Coene, K. L. M.; Engelke, U. F.; Kluijtmans, L. A. J.; Wevers, R. A.; Buydens, L. M. C.; Redlich, B.; Berden, G.; Oomens, J. Infrared ion spectroscopy: New opportunities for small-molecule identification in mass spectrometry - A tutorial perspective, *Anal. Chim. Acta.* **2020**, *1093*, 1-15.

(17) Kirschbaum, C.; Saied, E. M.; Greis, K.; Mucha, E.; Gewinner, S.; Schollkopf, W.; Meijer, G.; von Helden, G.; Poad, B. L. J.; Blanksby, S. J.; Arenz, C.; Pagel, K. Resolving Sphingolipid Isomers Using Cryogenic Infrared Spectroscopy, *Angew. Chem. Int. Ed.* **2020**, *59*, 13638-13642.

(18) Kirschbaum, C.; Greis, K.; Mucha, E.; Kain, L.; Deng, S.; Zappe, A.; Gewinner, S.; Schollkopf, W.; von Helden, G.; Meijer, G.; Savage, P. B.; Marianski, M.; Teyton, L.; Pagel, K. Unravelling the structural complexity of glycolipids with cryogenic infrared spectroscopy, *Nat. Commun.* **2021**, *12*, 1201.

(19) Kamrath, M. Z.; Garand, E.; Jordan, P. A.; Leavitt, C. M.; Wolk, A. B.; Van Stipdonk, M. J.; Miller, S. J.; Johnson, M. A. Vibrational Characterization of Simple Peptides Using Cryogenic Infrared Photodissociation of H-2-Tagged, Mass-Selected Ions, *J. Am. Chem. Soc.* **2011**, *133*, 6440-6448.

(20) Khanal, N.; Masellis, C.; Kamrath, M. Z.; Clemmer, D. E.; Rizzo, T. R. Glycosaminoglycan Analysis by Cryogenic Messenger-Tagging IR Spectroscopy Combined with IMS-MS, *Anal. Chem.* **2017**, *89*, 7601-7606.

(21) Ben Faleh, A.; Warnke, S.; Rizzo, T. R. Combining Ultrahigh-Resolution Ion-Mobility Spectrometry with Cryogenic Infrared Spectroscopy for the Analysis of Glycan Mixtures, *Anal. Chem.* **2019**, *91*, 4876-4882.

(22) Screen, J.; Stanca-Kaposta, E. C.; Gamblin, D. P.; Liu, B.; Macleod, N. A.; Snoek, L. C.; Davis, B. G.; Simons, J. P. IR-spectral signatures of aromatic-sugar complexes: probing carbohydrate-protein interactions, *Angew. Chem. Int. Ed.* **2007**, *46*, 3644-3648.

(23) Zviagin, A.; Kopysov, V.; Nagornova, N. S.; Boyarkin, O. V. Tracking local and global structural changes in a protein by cold ion spectroscopy, *Phys. Chem. Chem. Phys.* **2022**, *24*, 8158-8165.

(24) Lobas, A. A.; Solovyeva, E. M.; Saparbaev, E.; Gorshkov, M. V.; Boyarkin, O. V. Accelerating photofragmentation UV Spectroscopy–Mass spectrometry fingerprinting for quantification of isomeric peptides, *Talanta* **2021**, *232*.

(25) Kopysov, V.; Boyarkin, O. V. Resonance Energy Transfer Relates the Gas-Phase Structure and Pharmacological Activity of Opioid Peptides, *Angew. Chem. Int. Ed.* **2016**, *55*, 689-692.

Figures

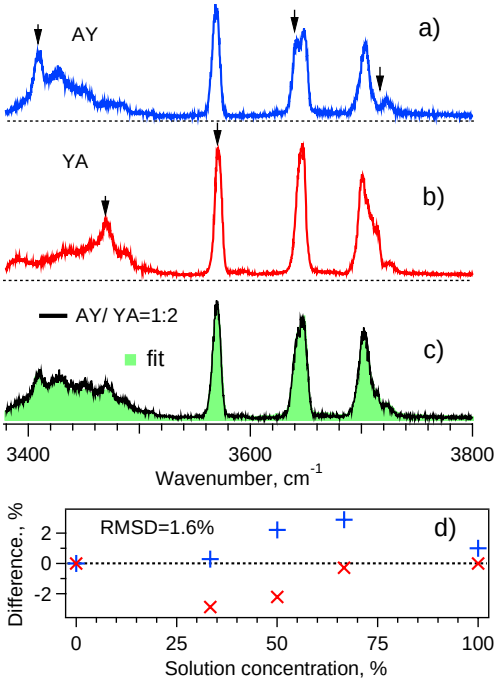

Figure 1. IR fragment spectra of (a) [AlaTyr-$(H_2O)_2$]-H^+, (b) [TyrAla-$(H_2O)_2$]-H^+ cold isomeric complexes, and of the complexes electrosprayed from (c) 1:2 solution mixture of the isomers and its fit spectrum (green space). The arrows point to the five wavenumbers used in the truncated decompositions. (d) The difference between the calculated and the prepared solution concentrations of the isomers (AY–blue and YA–red crosses) in five test mixtures.

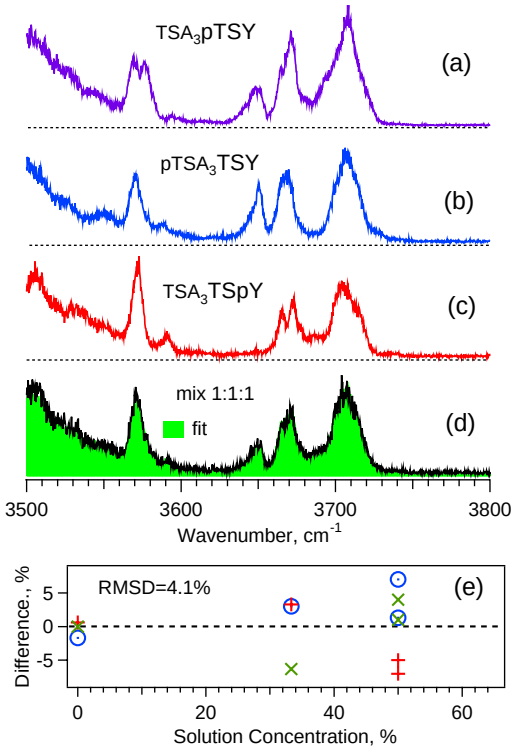

Figure 2. IR fragment spectra of (**a**) $[TSA_3pTSY-(H_2O)_2]-H^+$, (**b**) $[pTSA_3TSY--(H_2O)_2]-H^+$, (**c**) $[TSA_3TSpY--(H_2O)_2]-H^+$ and (**d**) of the equimolar solution mixture of the three isomeric complexes (black) and its fit (green). (**e**) The difference between the calculated and the actual solution concentrations of the isomers in four test mixtures (TSA_3pTSY–blue circles, $pTSA_3TSY$–red and TSA_3TSpY–green crosses).

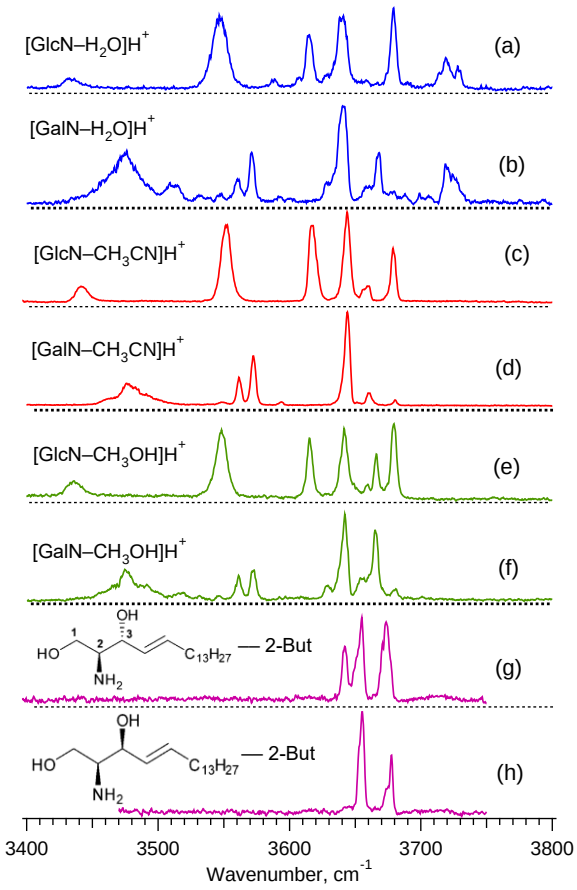

Figure 3. IR fragment spectra of protonated (**a**, **c**, **e**) GlcN and (**b**, **d**, **f**) GalN carbohydrates doubly solvated by (**a–b**) water, (**c–d**) singly solvated by acetonitrile and (**e–f**) methanol, and of (**g**) *D*-erythro-sphingosine and (**h**) *L*-threo-sphingosine singly solvated by 2-butanol.