Evaluation of the ONIOM Method for Interpretation of Infrared Spectra of Gas-Phase Molecules of Biological Interest

Jean-Christophe Poully, Gilles Grégoire, and Jean-Pierre Schermann*

Laboratoire de Physique des Lasers, UMR 7538 CNRS, Université Paris 13, 93430 Villetaneuse, France

Received: February 24, 2009; Revised Manuscript Received: May 5, 2009

The prediction accuracy of the ONIOM method for the interpretation of infrared spectra of gas-phase molecules of biological interest has been investigated. With the use of experimental results concerning amino acids, small peptides, and sugars taken from the literature, mode-specific local scaling factors have been determined for different high-layer/low-layer couples. A significant improvement is noticed when using local scaling factors with respect to global factors. The B3LYP/6-31G*:AM1 level turns out to offer the best trade-off between computational expense and accuracy. In the case of the RGD peptide, the B3LYP/6-31G*:AM1 and the B3LYP/3-21G levels require similar computational expense, but the former yields structures and predicted spectra comparable to those obtained from pure B3LYP/6-31G* calculations with a factor of 2 in time-saving gain. The experimental infrared spectrum of doubly charged gas-phase vancomycin ions has been recorded in the 1000−2000 cm^{-1} range and compared to predicted spectra of three different conformers at the B3LYP/6-31G*:AM1 level. This demonstrates the possible interpretation of IR spectra of relatively large systems (178 atoms) with the use of rather modest computational means.

1. Introduction

The combined quantum mechanical and molecular mechanical (QM/MM) method[1−3] is widely applied in structure and dynamics studies of molecular systems. A localized region is treated at a more or less elaborate quantum level (high layer), whereas its surrounding is treated at a much simpler and economical level (low layer). This method has been in particular used for the interpretation of infrared spectra. Accurate simulations of infrared spectra require rather computationally demanding quantum calculations and are generally conducted at the density functional theory (DFT) level. They are thus restricted to systems containing a number of atoms typically less than 200 atoms, even with very efficient treatments such as RI-DFT-D[4] or SCC-DFTB.[5] Less computer-time-demanding approaches using force-fields or semiempirical (SE) methods are applicable to considerably larger molecular sizes but usually lead to sizable prediction errors, even with the use of sophisticated improvements.[6,7] Through mixing of quantum and classical descriptions, the QM/MM approach allows interpretation of infrared spectra of large-size biomolecular systems[8−10] as well as small-size systems embedded in explicit solvents.[11−16]

During recent years, experimental gas-phase infrared spectroscopy studies of neutral[17−20] and ionic[21−24] species of biological interest have been conducted. The goal of these studies is the investigation of intrinsic structures in absence of any environment. Nearly ideal conditions, although somewhat far from biological reality, are then fulfilled and rigorously correspond to those encountered in quantum chemistry calculations then allowing direct comparison between experiments and simulations. The size of the gas-phase investigated systems has been steadily increasing from elementary building blocks such as nucleobases, amino acids, or sugars[25] up to much larger systems such as decapeptides,[24] gramicidin,[26] and cytochrome *c*.[27] Exploration of potential energy surfaces and simulation of infrared spectra thus require more and more computational means.

We here wish to evaluate the possibility of using the QM/MM approach for reasonably accurate predictions of infrared spectra compatible with the improved spectral resolution obtained in gas-phase studies conducted with lasers as compared, for example, to Fourier transform infrared (FT-IR) spectroscopy in condensed phase.[12,28] For that purpose, we evaluate the efficiency of the ONIOM method through comparison between experimental spectra and simulations conducted with different high- and low-level couples. In a first step, we determine specific scaling factors[29,30] that take anharmonicity into account and minimize prediction errors in harmonic calculations. For that purpose, we use available experimental infrared spectra obtained by different groups for small mass-selected and conformer-selected gas-phase molecular species. We further use those scaling factors to interpret experimental results obtained from infrared resonant multiphoton dissociation (IRMPD) spectroscopy of intermediate size molecular systems.

2. Determination of Mode-Specific Scaling Factors Appropriate to the ONIOM Method

The here presented mode-specific scaling factors have been obtained from simulations of infrared spectra of mass-selected and conformer-selected molecules.[30] Those simulations have been carried out using the ONIOM method[1,3,31] as implemented in the Gaussian 03 package.[32] Our goal is to achieve reasonably accurate predictions of vibrational frequencies of systems larger than those feasible by treating the whole system with pure DFT methods. We thus aim to set a minimum number of atoms in the high layer. In gas-phase experiments conducted with lasers, infrared spectra are generally not recorded over ranges as wide as in FT-IR experiments.[33] Instead a rather small number of vibrational modes providing crucial information about structures are particularly scrutinized. Modes involving groups such as

* To whom correspondence should be addressed. E-mail: jean-pierre.schermann@univ-paris13.fr.

10.1021/jp901696d CCC: $40.75
Published on Web 06/18/2009

TABLE 1: Scaling Factors $a_{scal}^{\nu,L}$ for the B3LYP/6-31G* ONIOM High-Layer Level with Three Different Low-Layer Levels and for Two Pure DFT B3LYP Levels[a]

vibrational mode/method	ν(OH)	ν(NH_2)as	ν(NH_2)s	ν(NH)	ν(CO)	γ(NH_2)	γ(NH)	global scale factor
B3LYP/6-31G*:AM1	0.977	0.958	0.958	0.958	0.957	0.964	0.964	0.962
B3LYP/6-31G*:HF/STO-3G	0.973	0.957	0.957	0.952	0.956	0.964	0.967	0.960
B3LYP/6-31G*:B3LYP/STO-3G	0.974	0.958	0.959	0.955	0.954	0.964	0.970	0.962
B3LYP/3-21G	1.061	0.974	0.983	0.978	0.995	0.943	0.991	1.001
B3LYP/6-31G*	0.974	0.961	0.962	0.96	0.961	0.963	0.963	0.965

[a] Those scaling factors are given for seven vibrational modes providing structural information.

TABLE 2: Mean Error Dispersion $\sigma_{\nu,L}$ for the B3LYP/6-31G* ONIOM High-Layer Level with Three Different Low-Layer Levels and for Two Pure DFT B3LYP Levels

vibrational mode	ν(OH)	ν(NH_2)as	ν(NH_2)s	ν(NH)	ν(CO)	γ(NH_2)	γ(NH)	single scale factor
B3LYP/6-31G*:AM1	16.9	7.2	13.4	23.2	6.8	5.9	8.2	28.8
B3LYP/6-31G*:HF/STO-3G	17.1	8.0	8.5	15.7	11.6	6.3	10.1	27.5
B3LYP/6-31G*:B3LYP/STO-3G	22.0	7.7	11.7	21.0	9.2	6.7	11.2	28.7
B3LYP/3-21G	47.8	22.4	46.5	17.9	8.1	17.7	14.5	131.6
B3LYP/6-31G*	9.1	11.1	12.6	9.0	4.9	7.4	2.7	24.2

N−H, N−H_2, O−H, and C═O that are very often engaged in hydrogen bonds[31] receive most of the attention.

When carrying out a two-layer frequency calculation with the ONIOM method, one can freely choose atoms that are included either in the high layer or in the low layer. However, even investigating localized vibrations such as C═O stretches,[34] we observe that a high layer only containing C and O atoms most often leads to poor results in terms of mean error dispersion. The here presented results have been obtained by adopting the following selection procedure. All C═O, N−H, N−H_2, and O−H groups as well as the atoms linking two or more of these groups and atoms H-bound to these groups were set in the high layer treated at the B3LYP/6-31G(d) level. Similarly, in the case of an aromatic ring including an N−H group such as the tryptophan side chain, setting only this N−H group in the high layer provides large errors for N−H stretching and bending mode frequencies. Atoms sharing delocalized electron density with the investigated groups must also be set as high-layer atoms.

Since vibrational frequency calculations of molecules of biological interest are usually carried out within the harmonic approximation, we first determine appropriate mode-specific scaling factors $a_{scal}^{\nu,L}$ for each given vibrational mode ν using different high-layer/low-layer levels L. This choice of mode-specific scaling factors aims to lower prediction errors as compared to global scaling factors covering the whole near-infrared range. Those scaling factors are obtained by minimizing errors between predicted and experimental values for molecules constituting a training set. The aim being to predict as accurately as possible IR spectra of relatively large systems, the molecules in the used training set contain more than 25 atoms. Their gas-phase IR spectra have been obtained by de Vries and Hobza,[19] Chin and co-workers,[35,36] and Simons and co-workers[37,38] with the conformer-selective IR−UV double-resonance technique,[39−41] leading to unambiguous assignment of experimental peaks. N−H, N−H_2, C═O (amide I) stretching modes and N−H_2, N−H (amide II) bending modes are here investigated for 24 peptide conformers (the list is provided in the Supporting Information). Seven sugar conformers are used for the O−H stretching mode. No extra scaling factor is introduced to take into account the effect of H-bonding strength on frequency shifts. One should thus use these factors cautiously when the studied molecules contain strongly H-bound N−H, C═O, or O−H groups with red-shifts larger than typically 150 cm^{-1}. The methodology followed to obtain the mean error dispersions $\sigma_{\nu,L}$ has been described in detail in a preceding paper[30] where transferable scaling factors were determined for the B3LYP and BPW91 functionals. In brief, for each high-level/low-level of theory L and each considered vibrational mode ν, individual scaling factors $a_{scal}^{i,j,\nu,L}$ are obtained by dividing the published experimental value $\nu_{exp}^{i,j,\nu}$ for conformer j of molecule i by the corresponding calculated value $\nu_{cal}^{i,j,\nu,L}$. The here proposed "gas-phase" transferable specific scaling factors $a_{scal}^{\nu,L}$ are then the arithmetic averages of the individual values deduced from the training set. The predicted frequency values are by definition equal to $\nu_{predict}^{i,j,\nu,L} = a_{scal}^{\nu,L}\nu_{cal}^{i,j,\nu,L}$. Those mode-specific scaling factors are given in Table 1. In order to evaluate their prediction capabilities and their transferability, we provide in Table 2 the standard deviations $\sigma_{\nu,L}^2$ given by

$$\sigma_{\nu,L}^{\ 2} = \frac{n\sum(\nu_{predict}^{i,j,\nu,L} - \nu_{exp}^{i,j,\nu})^2 - (\sum(\nu_{predict}^{i,j,\nu,L} - \nu_{exp}^{i,j,\nu}))^2}{n(n-1)}$$

for the here considered high-level/low-level couples. n is the number of considered frequency values.

We have chosen the widely used B3LYP functional[42,43] for the high layer. In a previous study,[30] we have shown that the performances of the 6-31G* basis set with respect to simulation of infrared spectra are quite comparable to those of much more computer-time-demanding basis sets such as 6-311++G**. The 6-31G* basis set can be considered, as far as prediction of IR spectra are concerned (and not for energy), as the most efficient "value for money". For the low layer, force-fields[44] or semiempirical methods can be chosen. Since most of the amount of computer time (typically 99%) is used for the high-level calculations, we here employ the semiempirical AM1 method that has previously been employed as well for the high layer[45] as for the low layer.[29,46−48] AM1 has been shown to provide very valuable results for the interpretation of structures[49] and IR spectra.[34] It is here tested as low-level method as well as the HF/STO-3G and B3LYP/STO-3G methods.

An example of partitioning between high and low layers in a conformer of [Ac-FGG] is displayed in the Supporting Information. The treatment of borders between the high and low layers is a delicate task in QM/SE or QM/MM methods.[50,51] We here simply use the basic link-atom solution set as default in ONIOM. It consists in cutting bonds between high-layer and low-layer atoms in a symmetric manner, leading to one unpaired

TABLE 3: QM/SE Partition, Number of Basis Functions, and Relative Computer Time Required for Computation of Vibrational Frequencies of Four Different Peptides at the ONIOM B3LYP/6-31G*:AM1, B3LYP/6-31G*, and B3LYP/3-21G Levels (Computer Time for Required for IR Spectrum Simulation of Ac-F at the ONIOM B3LYP/6-31G*:AM1 Level Taken As Unit Time)

studied peptides	Ac-F	Ac-FGG	Ac-FA	$[RGD + H]^+$
no. of atoms in the model system for ONIOM	11	25	17	38
total no. of atoms	29	43	39	47
no. of basis functions for the B3LYP/model system calculation with ONIOM	117	249	183	317
no. of basis functions for B3LYP/6-31G*	253	385	338	406
no. of basis functions for B3LYP/3-21G	163	247	218	262
computer time for ONIOM B3LYP/6-31G*:AM1	1	12	3.3	21.5
computer time for B3LYP/6-31G*	12	42	28	53
computer time for B3LYP/3-21G	3.6	13.3	8.3	16.6

electron for each border atom, and capping the dangling bonds with "dummy" H atoms. We optimize each structure at the chosen level of theory, calculate the corresponding IR spectrum, and check for absence of imaginary frequencies. We then assign experimental peaks to computed frequencies for O−H, asymmetric and symmetric $N-H_2$, N−H, and C═O stretches, $N-H_2$ and N−H bends. We also compare the performances of ONIOM methods to those of all-atoms B3LYP calculations with the 3-21G and 6-31G* basis sets. The results concerning the mean error dispersions $\sigma_{\nu,L}$ are presented in Table 2.

Surprisingly, increasing the level of the low layer does not lead to any significant improvement of performances as far as the prediction accuracy of the studied vibrational mode frequencies is concerned. Results obtained with the use of the DFT level B3LYP/STO-3G are less satisfying than with the semiempirical AM1 level both for mean errors and mean error deviations. The HF/STO-3G level performs better than AM1 only for ν(N−H) and $\nu(N-H_2)_s$ stretching modes. In this last case, the computer time required for structure optimization and IR spectrum calculation of, for example, the Ac-FGG protected peptide starting from the same nonoptimized structure becomes 50% larger. When using a parallel version of Gaussian 03, one should, however, prefer HF/STO-3G as the low layer since the calculation of frequencies is not run in parallel at the AM1 level. We thus conclude that the B3LYP/6-31G*:AM1 method represents the best trade-off between computational expense and prediction accuracy for QM/SE vibrational frequency calculations, and we employ it in the following part of this study. QM/SE methods save computer time and memory and thus allow, for a given available computing power, the simulation of IR spectra of systems that would otherwise require a too large number of basis functions for pure DFT calculations. In order to quantify this computer time saving, we compare in Table 3 the amount of computer time required to calculate vibrational frequencies with the same computer for four peptides. If N_{HL} and N_{tot} are, respectively, the number of atoms in the high layer and the total number of atoms, this table shows that the computer time saving scales approximately as $(N_{HL}/N_{tot})^3$.

The number of basis functions involved in the calculation performed at the B3LYP level governs the amount of time required for computing the frequency spectrum. The time required for simulating IR spectra at the AM1 level is much shorter (e.g., 60 times shorter than the pure B3LYP/6-31G* level for the calculation of $[RGD + H]^+$ peptide vibrational frequencies, see below). Therefore, as can be easily guessed, the number of atoms in the high-level system must be as small as possible.

3. Interpretation of Experimental IRMPD Spectra with the ONIOM Method

We compare IR spectra predicted by ONIOM to experimental spectra obtained at room temperature with the free electron laser (FEL) facility CLIO of the Orsay University.[52] The data are obtained by IRMPD spectroscopy, a quasi-universal technique allowing the recording of IR spectra of isolated ionized species. This method has been previously described into detail in refs 53 and 22. Briefly, ions are generated in gas phase by means of an electrospray source and mass-selected in a quadrupole ion trap. Following illumination by a free electron IR laser beam scanning the spectral region of interest, here 1000−2000 cm^{-1}, the ionic fragmentation rate induced by sequential resonant absorption of several photons is monitored and constitutes the experimental IR spectrum. The conformers of a molecular ion exhibit different intramolecular interaction patterns (H-bonds, NH−π, salt bridges, etc.) and thus possess different vibrational modes providing structural signatures. It may turn out that several conformers coexist at low temperature or interconvert at room temperature and can then contribute to the experimentally recorded spectrum.

3.1. Arginine−Glycine−Aspartic Acid (RGD) Peptide. The RGD amino acid sequence plays a major role in cell adhesion since it is specifically recognized by the transmembrane integrins.[54] In a previous study,[55] we recorded its IRMPD spectrum and interpreted the experimental results by means of pure DFT and Car−Parrinello quantum molecular dynamics. The protonated $[RGD + H]^+$ peptide possesses two low-lying charge-solvated conformers CS1 and CS2 populated at 300 K, both with the proton localized on the arginine side chain. In conformer CS1, this guanidinium group interacts with both carboxyl groups, i.e., the RGD C-terminal and the Asp side chain (no free C═O group), whereas the guanidinium group only interacts with the Asp side chain in conformer CS2.

In the present study, the corresponding structures are optimized and their vibrational frequencies are computed using ONIOM DFT:AM1 molecular orbital calculations. A minimum set of 27 atoms (over a total number of 47 atoms) should be included in the high layer in order to take into account the most structurally important functional groups, i.e., the peptide backbone and the R and D side-chain atoms involved in hydrogen bonds. In order to obtain an acceptable prediction accuracy for a given set of vibrational frequencies (typically better than 15 cm^{-1}), atoms sharing electronic density with atoms involved into the studied vibrations must also be included. In the present case of the $[RGD + H]^+$ ion, 38 atoms among a total number of 47 are included in the high-layer nevertheless leading to a computational time reduced by a factor of 2 as compared to a pure B3LYP calculation still using the same basis set 6-31G*.

Superimposed structures of a protonated $[RGD + H]^+$ conformer computed at the ONIOM DFT/AM1 and at the pure DFT B3LYP/6-31G* levels of theory are displayed in the Supporting Information. A root-mean-square deviation (rmsd) of only 0.08 Å is observed between the two structures. The

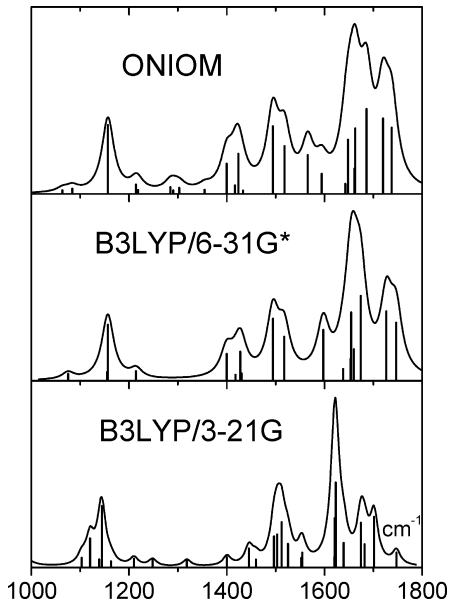

Figure 1. From top to bottom: predicted spectra of the $[RGD + H]^+$ CS1 conformer at the ONIOM B3LYP/6-31G*:AM1, B3LYP/6-31G*, and B3LYP/3-21G levels.

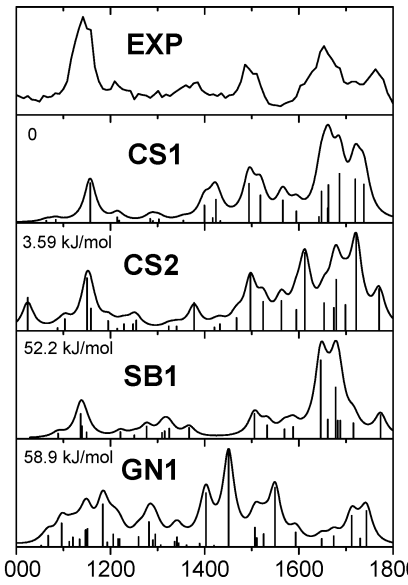

Figure 2. From top to bottom: experimental $[RGD + H]^+$ IRMPD spectrum, predicted ONIOM B3LYP/6-31G*:AM1 infrared absorption spectra of the CS1, CS2, GN1, and SB1 conformers. The relative conformer energies are calculated at the B3LYP/6-31G* level.

simulated spectrum of the RGD peptide CS1 conformer (the CS2 conformer simulated spectrum is provided as Supporting Information) optimized at the ONIOM B3LYP/6-31G*:AM1 level is compared in Figure 1 to the simulated spectra obtained from pure DFT optimizations at the B3LYP/6-31G* and B3LYP/3-21G levels. Those spectra exhibit spectral features in the 1100−1150, 1400−1450, 1500−1550, and 1650−1750 cm^{-1} regions that can be attributed to vibrational modes, respectively, involving O−H, N−H, and $N-H_2$ bending modes and C═O stretch modes. The three methods lead to rather similar simulated spectra in the case of the free O−H bends (1100−1150 cm^{-1}) of the carboxyl groups and that of the amide II (≈1500 cm^{-1}) or amide I (≈1750 cm^{-1}) modes involving atoms engaged into weak hydrogen bonds. The ONIOM B3LYP/6-31G*:AM1 and the pure B3LYP/6-31G* simulated spectra are in fair agreement over the whole spectral range. On the contrary, the B3LYP/3-21G level lacking diffuse functions totally fails to interpret strong hydrogen bonding. This appears in the case of the strongly blue-shifted O−H bends (≈1450 cm^{-1}). The large peak observed around 1620 cm^{-1} at the B3LYP/3-21G level corresponds to the near coincidence of the strongly blue-shifted $N-H_2$ bend of Arg and the strongly red-shifted C═O stretch of the Gly-Asp amide bond both engaged into hydrogen bonding with the C-terminal carboxyl group. At the ONIOM B3LYP/6-31G*:AM1 and pure B3LYP/6-31G* levels, those lines are both predicted at 1655 cm^{-1}.

The IRMPD experimental spectrum of the RGD peptide is compared in Figure 2 to the spectra of the different conformers simulated at the ONIOM B3LYP/6-31G*:AM1 level. The IRMPD spectrum is not strictly equivalent to the calculated linear infrared absorption spectrum. As often observed in IRMPD experiments, intensities are poorly reproduced due to the nonlinear character of the multiphoton absorption process.[53] Moreover, red-shifts due to anharmonic absorption of excited species can also be present.[23] In addition to the low-lying CS1 and CS2 conformers, other conformers can also exist at considerably higher energies. Energy calculations conducted at the ONIOM level contain contributions of the link atoms and thus cannot be used. We here consider relative conformer energies calculated at the pure DFT level. In the neutral guanidine structure (GN), the guanidine side chain of R is neutral and the extra proton is located on the N-terminus that interacts strongly with the guanidine group and with the C-terminal and R-G amide carbonyls. In the salt-bridge (SB) structure, the guanidine group and the N-terminus are both protonated, whereas the D carboxyl group is deprotonated. The interaction between the protonated N-terminus and the D carboxylate group is then very strong. In the simulated spectra of the GN and SB conformers, the calculated amide bands at 1500−1750 cm^{-1} are either too wide or too narrow and do not fit the experimental spectrum, in agreement with their calculated high energies.

3.2. Vancomycin Glycopeptide. The vancomycin glycopeptide is a last resort antibiotic against the *Staphylococcus aureus* bacteria. It binds noncovalently and enantiospecifically to the D-Ala-D-Ala terminus of the bacterial cell-wall peptidoglycan precursor UDP-*N*-acetyl-muramyl-pentapeptide, and this interaction induces cell death. The vancomycin-D-Ala-D-Ala peptide complexes have already been studied in gas phase by means of mass spectrometry.[56,57] We here interpret experimental the IRMPD spectrum of mass-selected $vancomycin^{2+}$ ions (178 atoms) confined in a Paul trap at 300 K[58] that we have recorded in the 1000−2000 cm^{-1} range. These ionic species are the most abundant positive charge states observed with the electrospray ionization (ESI) source. This experimental IRMPD spectrum is displayed in Figure 3 and is still reasonably well resolved although the $vancomycin^{2+}$ ion mass is rather large (1449.4). The full width at half-maximum (fwhm) of the spectral lines have the same order of magnitude (≈40 cm^{-1}) as those observed in the case of small peptides. In order to model the structures of these dications, we use the Protein Data Bank (PDB, entry 1aa5) file as initial structure. We add protons to the two most basic sites in solution, i.e., the $N-H_2$ group in the sugar moiety and the methylated N-terminal of the peptide moiety. In order to explore the conformational space, we have carried out a Systematic Pseudo Monte Carlo (SPMC) torsional sampling of the potential energy surface (PES)[59] implemented in the Macromodel software (MacroModel, version 9.6, Schrödinger, LLC)[60,61] with the MMFF force-field.[62] Ten thousands structures were generated and 200 unique conformations were found within 10 $kcal/mol^{-1}$ of the global minimum. The lowest

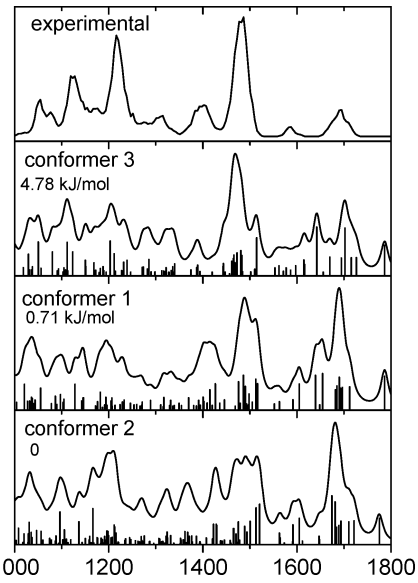

Figure 3. Experimental IRMPD spectrum of [vancomycin + 2H]$^{2+}$ ions and predicted spectra of conformers 3, 1, and 2 simulated at the ONIOM B3LYP/6-31G*:AM1 level. The relative conformer energies are calculated at the B3LYP/6-31G* level.

energy structure was named conformer 1. Clustering of the remaining 100 most stable structures with the Xcluster software[63] indicates the existence of three distinct conformer families. Among those three families, two of them present sizable similarities, and for simplicity, we further consider them as a unique conformer. We are thus left with three conformers, conformer 1 being the lowest-energy one with the MMFF force-field and conformers 2 and 3 representing two other very different families. Conformers 1, 2, and 3 were first optimized at the AM1 level and further reoptimized at the ONIOM B3LYP/6-31G*:AM1 level in order to calculate the vibrational frequencies. The relative energies of these three conformers calculated at the B3LYP/6-31G* level are very close and given in Figure 3. Those energies are only indicative due to the presence of aromatic cycles that are not correctly treated at this level of theory. In between 1000 and 2000 cm^{-1}, a large number of modes exist from strongly coupled modes with concerted atomic movements within the sugar moiety around 1030 cm^{-1} up to C═O stretching modes around 1750 cm^{-1}. As only vibrational modes involving atoms included into the high layer are correctly described, for memory space reasons, we had to carry out two separated ONIOM calculations to get the entire spectrum. The first calculation includes the sugar and the three rings below it in the high layer, whereas the second one includes in the high layer the complementary part, i.e., the peptidic moiety connected by two phenol rings. Those two complementary partitions are provided as Supporting Information. The corresponding rmsd's between the very similar and complementary optimized structures are, respectively, equal to 0.3617, 0.5955, and 0.3711 Å for conformers 1, 2, and 3. Only spectral lines corresponding to vibrational modes involving atoms in the high layers from each complementary calculation are gathered in the simulated spectra displayed in Figure 3. The whole sets of frequencies are convoluted by a Lorentzian function (15 cm^{-1} fwhm). We use the scaling factors obtained in the first part of this study, respectively, 0.957 for C═O stretches, 0.964 for N−H bends, and 0.964 for N−H$_2$ bends. For the remaining modes, mainly O−H bends and C−O stretches of the disaccharide moiety in the 1000−1450 cm^{-1} range, we apply the recommended NIST scaling factor 0.96 for B3LYP/6-31G*. A first important point is that the three spectra corresponding to conformers 1, 2, and 3 are clearly distinguishable. The nonlinearity inherent to the IRMPD process implies that relative intensities are not always well reproduced, but the agreement for band positions and shapes is nevertheless reasonable, especially for conformer 3 in the 1000−1300 cm^{-1} region where mainly C−O stretches in the rather flexible sugar moiety and O−H bends are excited. Interestingly, between 1300 and 1750 cm^{-1}, where bound C═O stretches, N−H bends, and ring modes are resonant with the IR laser, conformer 1 seems to account for the main features. However, none of the three considered conformer spectra entirely matches the experiment. This may be due to the high vibrational temperature of the ions in the trap allowing vancomycin^{2+} ions to widely explore their conformational space. It should also be pointed out that conformers 1, 2, and 3 may not be the lowest-energy structures since they have been obtained through a MMFF force-field exploration and thus may not be the most populated species in our experiment. The free C═O stretching mode at 1790 cm^{-1} is missing in the experimental spectrum. This might be due to the two following reasons. This mode is strongly localized, and its coupling to other lower-frequency modes leading to IRMPD fragmentation is very weak. Besides, the FEL laser intensity sharply decreases at these wavenumbers and the energy threshold induced by the nonlinearity of IRMPD might not have been reached. A superimposition of the PDB 1aa5 crystal structure[64] and of the gas-phase vancomycin^{2+} ion in conformer 1 corresponding to the spectrum displayed in Figure 3 is provided as Supporting Information. The main differences involve the sugar group and the leucine residue. The gas-phase ion structure is more compact than the solution structure due to the absence of the solvent which shields electrostatic interactions.

In a recent paper, Yang et al.[57] published low-energy gas-phase structures of the vancomycin^{+} ion obtained through molecular dynamics (MD) simulations. The lowest-energy conformer of the vancomycin^{+} ion is very different from the three conformers of the vancomycin^{2+} ion. In the vancomycin^{+} ion, the NH_3^+ group is solvated by three carbonyls. In contrast, we here find that in conformer 3 of the vancomycin^{2+} ion, this group is solvated by only two carbonyls sitting outside the binding pocket and that it is free in conformers 1 and 2. The binding pocket is also less opened in the vancomycin^{2+} ion than in the vancomycin^{+} ion. These differences are likely due to the short distance (6 Å) in the vancomycin^{+} ion conformer found by Yang et al.[57] between the NH_3^+ group and the second protonation site of vancomycin as well as to the hydrogen-bonding network stabilizing this site in conformers 1, 2, and 3 of the vancomycin^{2+} ion.

4. Conclusion

Pure DFT calculations turn out to be very efficient for interpretation of infrared spectra. When the size of the studied molecular systems becomes too large, a compromise between calculation requirements (memory and time) versus prediction accuracy must be made. One then faces different choices. Decreasing computer time and memory can be made by lowering the used basis set (e.g., the 3-21 G set instead of the 6-31G* set) or by partitioning the system into a high level containing the most spectroscopically crucial atoms and a low level for the remaining atoms. The ONIOM QM/SE B3LYP/6-31G*:AM1 level requires computer time similar to that required for pure B3LYP/3-21G (Table 3), but its prediction accuracy is nearly twice better. In particular, the B3LYP/3-21G

spectrum fails to reproduce most of the experimental data in the spectral regions involving atoms engaged into strong hydrogen bonding. This turns out crucial in the amide A (N−H stretch) region where even more elaborate QM/SE levels (Table 2) or pure DFT,[30] even taking into account dispersion (DFT-D), can provide somewhat disappointing predictions and the possibility of misassignments of spectral features cannot be excluded. RI-MP2 calculations leading to prediction accuracies better than 7 cm^{-1} must then be preferred.[65]

The use of mode-specific scaling factors makes interpretation of IR spectra somewhat more tedious but can provide a sizable improvement of the prediction accuracy, in particular for amide I and amide II vibrational frequencies that are used in studies of secondary structures of peptides.[34] In the case of small peptides, typical ratios for B3LYP/6-31G* versus ONIOM computer times are around 5. Standard deviations of ca. 8 cm^{-1} are obtained for amide I and amide II modes with specific scaling factors at the ONIOM QM/SE B3LYP/6-31G*:AM1 level as compared to standard deviations obtained with global scaling factors equal to 11 cm^{-1} in the case of a homogeneous family of molecules (linear tetrapyrroles) for the pure B3LYP/6-31G* level.[66] It is interesting to compare those values to the 3% prediction accuracy in the QM/MM theoretical interpretation of IR spectroscopic study of binding of guanosine triphosphate (GTP) in the binding pocket of the large Ras protein.[10]

The choice of the minimum number of atoms that must be included in the high layer is delicate. In the case of the rather small protonated RGD peptide, we observe that this number is quite large (38 atoms over a total number of 47 atoms) and may be considered as somewhat disappointing. However, we notice that the computer time gain for the simulation of infrared spectra using ONIOM as compared to pure DFT calculations scales as the third power of the ratio of the number of atoms in the high layer and the total number of atoms. This gain becomes sizable in the case of vancomycin ions. Taking advantage of the relative stiffness of this molecular system, it has been possible to divide it into two separate moieties and treat them independently. In that case, the interpretation of the IRMPD spectrum can be conducted at the rather reliable B3LYP/6-31G*:AM1 level.

Acknowledgment. The authors thank Dr. Julia Laskin for suggesting the study of the infrared spectrum of vancomycin ions and her generous help.

Supporting Information Available: Bibliography of the experimental data, partitions of the Ac-FGG peptide and vancomycin ions, and structures of vancomycin ion conformers. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
(2) Assfeld, X.; Rivail, J. L. *Chem. Phys. Lett.* **1996**, *263*, 100.
(3) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. *J. Mol. Struct. (THEOCHEM)* **1999**, *461*, 1.
(4) Cerny, J.; Jurecka, P.; Hobza, P.; Valdes, H. *J. Phys. Chem. A* **2007**, *111*, 1146.
(5) Walewski, L.; Bala, P.; Elstner, M.; Frauenheim, T.; Lesying, B. *Chem. Phys. Lett.* **2004**, *397*, 451.
(6) Gerber, R. B.; Chaban, G. M.; Gregurick, S. K.; Brauer, B. *Biopolymers* **2003**, *68*, 370.
(7) Brauer, B.; Chaban, G. M.; Gerber, R. B. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2543.
(8) Rousseau, R.; Kleinscmidt, V.; Schmidt, U. W.; Marx, D. *Angew. Chem., Int. Ed.* **2004**, *43*, 4804.
(9) Nutt, D. R.; Meuwly, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5998.
(10) Klähn, M.; Schlitter, J.; Gerwert, K. *Biophys. J.* **2005**, *88*, 3829.
(11) Kinnaman, C. S.; Cremeens, M. E.; Romesberg, F. E.; Corcelli, S. A. *J. Am. Chem. Soc.* **2006**, *128*, 13334.
(12) Tellez Soto, C. A.; Ramos, J. M.; Rianelli, R. S.; de Souza, M. C. B. V.; Ferreira, V. F. *Spectrochim. Acta, Part A* **2007**, *67*, 1080.
(13) Pires, M. M.; DeTuri, V. F. *J. Chem. Theory Comput.* **2007**, *3*, 1073.
(14) Yan, Y.; Krishnan, G. M.; Kühn, O. *Chem. Phys. Lett.* **2008**, *464*, 230.
(15) Guardia, C. M. A.; Gonzalez Lebrero, M. C.; Bari, S. E.; Estrin, D. A. *Chem. Phys. Lett.* **2008**, *463*, 112.
(16) Dedachi, K.; Ishikawa, Y.; Nakatsu, T.; Natsume, T.; Tsukamoto, T.; Kurita, N. *J. Mol. Struct. (THEOCHEM)* **2008**, *854*, 70.
(17) Linder, R.; Nispel, M.; Haber, T.; Kleinermanns, K. *Chem. Phys. Lett.* **2005**, *409*, 260.
(18) Chin, W.; Piuzzi, F.; Dimicoli, I.; Mons, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1033.
(19) de Vries, M. S.; Hobza, P. *Annu. Rev. Phys. Chem.* **2007**, *58*, 585.
(20) Vaden, T. D.; de Boer, T. S. J. A.; Macleod, N. A.; Marzluff, E. M.; Simons, J. P.; Snoek, L. C. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2549.
(21) Kapota, C.; Lemaire, J.; Maitre, P.; Ohanessian, G. *J. Am. Chem. Soc.* **2004**, *126*, 1836.
(22) Gregoire, G.; Gaigeot, M. P.; Marinica, D. C.; Lemaire, J.; Schermann, J. P.; Desfrançois, C. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3082.
(23) Polfer, N.; Oomens, J.; Suhai, S.; Paizs, B. *J. Am. Chem. Soc.* **2007**, *129*, 5887.
(24) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. *J. Am. Chem. Soc.* **2007**, *129*, 13820.
(25) de Vries, M.; Hobza, P. *Annu. Rev. Phys. Chem.* **2007**, *58*, 585.
(26) Abo-Riziq, A.; Crews, B.; Callahan, M. P.; Grace, L.; de Vries, M. *Angew. Chem., Int. Ed.* **2006**, *118*, 5290.
(27) Oomens, J.; Polfer, N.; Moore, D. T.; van der Meer, L.; Marshall, A. G.; Eyler, J. R.; Meijer, G.; von Helden, G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1345.
(28) Walewski, L.; Bala, P.; Elstner, M.; Frauenheim, T.; Lesying, B. *Chem. Phys. Lett.* **2004**, *397*, 451.
(29) Carauta, A. N. M.; de Carneiro, J. W.; Tellez Soto, C. A. *Int. J. Quantum Chem.* **2005**, *103*, 763.
(30) Bouteiller, Y.; Gillet, J. C.; Gregoire, G.; Schermann, J. P. *J. Phys. Chem.* **2008**, *112*, 11656.
(31) Tschumper, G. S.; Morokuma, K. *J. Mol. Struct. (THEOCHEM)* **2002**, *592*, 137.
(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J. J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, revision B.03; Gaussian, Inc.: Pittsburgh, PA, 2003.
(33) Farnik, M.; Steinbach, C.; Weimann, M.; Buck, U.; Borho, N.; Suhm, M. A. *Phys. Chem. Chem. Phys.* **2004**, *6*, 4614.
(34) Wieczorek, R.; Dannenberg, J. J. *J. Phys. Chem.* **2008**, *112*, 1320.
(35) Chin, W.; Compagnon, I.; Dognon, J. P.; Canuel, C.; Piuzzi, F.; Dimicoli, I.; von Helden, G.; Meijer, G.; Mons, M. *J. Am. Chem. Soc.* **2005**, *127*, 1388.
(36) Chin, W.; Mons, M.; Dognon, J. P.; Mirasol, R.; Chass, G.; Dimicoli, I.; Piuzzi, F.; Butz, P.; Tardivel, B.; Compagnon, I.; von Helden, G.; Meijer, G. *J. Phys. Chem. A* **2005**, *109*, 5281.
(37) Carcabal, P.; Hunig, I.; Gamblin, D. P.; Liu, B.; Jockusch, R. A.; Kroemer, R. T.; Snoek, L. C.; Fairbanks, A. J.; Davis, B. G.; Simons, J. P. *J. Am. Chem. Soc.* **2006**, *128*, 1976.
(38) Jockusch, R. A.; Kroemer, R. T.; Talbot, F. O.; Snoek, L. C.; Carcabal, P.; Simons, J. P.; Havenith, M.; Bakker, J. M.; Compagnon, I.; Meijer, G.; von Helden, G. *J. Am. Chem. Soc.* **2004**, *126*, 5709.
(39) Brutschy, B. *Chem. Rev.* **2000**, *100*, 3891.
(40) Zwier, T. S. *J. Phys. Chem. A* **2001**, *105*, 8827.
(41) LeGreve, T. A.; Baquero, E. E.; Zwier, T. S. *J. Am. Chem. Soc.* **2007**, *129*, 4028.
(42) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(43) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(44) Bruice, T. C.; Kahn, K. *Curr. Opin. Chem. Biol.* **2000**, *4*, 540.
(45) Luque, F. J.; Reuter, N.; Cartier, A.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2000**, *104*, 10923.

(46) Wanno, B.; Ruangpornvisuti, V. *J. Mol. Struct. (THEOCHEM)* **2004**, *685*, 57.
(47) Hall, K. F.; Vreven, T.; Frisch, M.; Bearpark, M. J. *J. Mol. Biol.* **2008**, *383*, 106.
(48) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. *J. Comput. Chem.* **2003**, *24*, 760.
(49) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2005**, *127*, 17216.
(50) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 15645.
(51) Fornili, A.; Moreau, Y.; Sironi, M.; Assfeld, X. *J. Comput. Chem.* **2006**, *27*, 515.
(52) Lemaire, J.; Boissel, P.; Heninger, M.; Mauclaire, G.; Bellec, G.; Mestdagh, H.; Simon, A.; Caer, S. L.; Ortega, J. M.; Glotin, F.; Maitre, P. *Phys. Rev. Lett.* **2002**, *89*, 273002.
(53) Oomens, J.; Meijer, G.; von Helden, G. *Int. J. Mass Spectrom.* **2006**, *249*, 199.
(54) Auernheimer, J.; Hersel, U.; Dahmen, C.; Kantlehner, M.; Jeschke, B.; Enderle, A.; Nies, B.; Goodman, S. L.; Kessler, H. *Biopolymers* **2003**, *71*, 402.
(55) Grégoire, G.; Gaigeot, M. P.; Marinica, D. C.; Lemaire, J.; Schermann, J. P.; Desfrancois, C. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3082.
(56) Heck, A. J. R.; Jorgensen, T. J. D. *Int. J. Mass Spectrom.* **2004**, *236*, 11.
(57) Yang, Z.; Vorpagel, E. R.; Laskin, J. *J. Am. Chem. Soc.* **2008**, *130*, 13013.
(58) Mac Aleese, L.; Simon, A.; McMahon, T. B.; Ortega, J. M.; Scuderi, D.; Lemaire, J.; Maitre, P. *Int. J. Mass Spectrom.* **2006**, *249*, 14.
(59) Goodman, J. M.; Still, W. C. *J. Comput. Chem.* **1991**, *12*, 1110.
(60) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrikson, T.; W. C., S. *J. Comput. Chem.* **1990**, *11*, 440.
(61) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Clark Still, W. *J. Comput. Chem.* **2004**, *11*, 440.
(62) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 550.
(63) Shenkin, P. S.; McDonald, D. Q. *J. Comput. Chem.* **1994**, *8*, 899.
(64) Loll, P. J.; Bevivino, A. E.; Korty, B. D.; Axelsen, P. H. *J. Am. Chem. Soc.* **1997**, *119*, 1516.
(65) Bouteiller, Y.; Poully, J. C.; Desfrançois, C.; Grégoire, G. *J. Phys. Chem. A* **2009**, *113*, 6301.
(66) Magdo, I.; Nemeth, K.; Mark, F.; Hildebrandt, P.; Schaffner, K. *J. Phys. Chem. A* **1999**, *103*, 289.

JP901696D