J Mol Model (2011) 17:3195–3207
DOI 10.1007/s00894-011-0983-x

ORIGINAL PAPER

Aggregation of Kanamycin A: dimer formation with physiological cations

Johannes M. Dieterich · Ulrich Gerstel · Jens-Michael Schröder · Bernd Hartke

Received: 27 July 2010 / Accepted: 21 January 2011 / Published online: 2 March 2011

Abstract Global cluster geometry optimization has focused so far on clusters of atoms or of compact molecules. We are demonstrating here that present-day techniques also allow to globally optimize clusters of extended, flexible molecules, and that such studies have immediate relevance to experiment. For example, recent experimental findings point to production of larger clusters of an aminoglycoside closely related to Kanamycin A (KA), together with certain preferred physiological cations, by *Pseudomonas aeruginosa*. The present study provides first theoretical support for KA clustering, with a close examination of the monomer, the bare dimer, and dimers with sodium and potassium cations, employing global cluster structure optimization, in conjunction with force fields, semiempirical methods, DFT and ab-initio approaches. Interestingly, already at this stage the theoretical findings support the experimental observation that sodium cations are preferred over potassium cations in KA clusters, due to fundamentally different cationic embedding. Theoretically predicted NMR and IR spectra for these species indicate that it should be possible to experimentally detect the aggregation state and even the cationic embedding mode in such clusters.

Keywords Aminoglycoside clustering · Evolutionary computation · Genetic algorithms · Global cluster structure optimization · Pathogen-associated molecules

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-0983-x) contains supplementary material, which is available to authorized users.

J. M. Dieterich · B. Hartke (✉)
Institut für Physikalische Chemie,
Christian-Albrechts-Universität,
Olshausenstraße 40,
24098 Kiel, Germany
e-mail: hartke@phc.uni-kiel.de

U. Gerstel · J.-M. Schröder
Department of Dermatology,
University-Hospital Schleswig-Holstein,
Campus Kiel, Arnold-Heller-Str. 3, House 19,
24105 Kiel, Germany

Introduction

Clusters are recognized as important objects of scientific study in chemistry and physics [1–3]. Theoretical determination of their most probable structures, however, is a difficult task [4], requiring stochastic-heuristic global optimization algorithms for clusters of interesting sizes. Despite considerable progress in this area [5], most studies still deal with clusters of atoms, or with clusters of small, rigid molecules that do not change their outer shape strongly when turned around, as evidenced, e.g., by database entries [6]. Notably, the situation is different in ab-initio crystal structure prediction, where it has become common to treat more difficult flexible molecules [7–9]. In this contribution, we demonstrate that global cluster structure optimization can also deal with larger, flexible molecules.

In addition, we show that such cluster studies do have considerable direct practical relevance. To this end, we have selected clusters of Kanamycin A (KA) as application example. To describe their practical relevance, we briefly digress into medicinal biochemistry in the following paragraph.

The adaptive immune system of higher multicellular organisms, developing antibodies against antigens presented by pathogens, is supported by the evolutionary older innate immune system. Human skin appears to have a "chemical barrier" function, recognizing bacterial coloniza-

tion via so-called pathogen-associated molecules (PAMs) which trigger various defense mechanisms [10]. Certain PAMs induce production of cytokines in epithelial cells, leading to inflammatory reactions. Other PAMs induce antimicrobial peptides (AMPs) that kill bacteria without inflammatory side reactions, e.g., by forming pores in bacterial cell membranes [10]. Recently, Schröder and coworkers discovered in a mucoid clinical isolated of *Pseudomonas aeruginosa* potent AMP human beta-defensin-2 (hBD-2)-inducing activity of yet unknown origin [11]. Preliminary purification experiments revealed the existence of an hBD-2-inducing factor by *Pseudomonas aeruginosa*, which did not induce proinflammatory cytokines in epithelial cells, thus excluding the origin of the known Toll-like-receptor(TLR)-5-binding bacterial flagellin. In initial attempts to purify the hBD-2-inducer they found in several hBD-2-inducing activity-containing high performance liquid chromatography (HPLC) fractions the aminoglycoside Kanamycin A (KA), or an isomer of it. Nanospray ESI-MS and ^{1}H-NMR have lead to the current working hypothesis that in these active fractions KA may occur in stable dimers and higher aggregates (up to 26 units) together with certain physiological cations, preferentially sodium.

The absolute configuration of the KA monomer was determined via Xray crystallography of the monosulfate-monohydrate species [12]. However, literature on aggregation of KA or related aminoglycosides is extremely sparse, both on the experimental and on the theoretical side. A 1:1 aggregate of KA with Cu^{2+} could be identified via NMR and EPR [13]. A 2:1 complex could be detected in an electrochemical study [14]. Earlier, similar aggregates of the related compound Gentamicin were found with various methods [15]. One study [16] reported formation of long KA fibers on negatively charged surfaces. Other than that, no experimental evidence of KA aggregation has come to our attention. Theoretical studies on KA are rarer still: One group has performed molecular dynamcis and docking studies involving aminoglycosid monomers and RNA [17, 18], using standard force fields and scoring functions. Apparently only in two cases electronic structure theory methods were applied to KA: One of them again in the context of KA binding to RNA [19], the other one for the isolated monomer in the gas phase [20]. Both studies, however, only employed HF calculations with small basis sets. Theoretical studies on the aggregation of KA or similar aminoglycosides appear to be non-existent.

In a previous paper [21], two of the present authors described the recent development of the OGOLEM program suite for global structure optimization of arbitrarily mixed clusters of flexible molecules. As exemplary application, that work already contained first results of applying this program suite to KA aggregation with Na^{+}, using the GAFF force field, identifying a 2:1 aggregate as particularly stable. In the present article, we would like to put this application on a more solid foundation, by (1) comparing different levels of theory for the KA monomer, (2) an in-depth analysis and comparison of $(KA)_2M^{+}$, M = Na,K, and (3) providing contact points to experiment via calculation of NMR and IR spectra, pointing out experimentally accessible signatures of aggregation.

For systems of biochemical importance such as the present one, it certainly is mandatory to also include an implicit or even explicit solvent description. To ease the burden of the present initial steps, however, we defer this task to the next stage of our studies. A methodical reason for this is that it is easier to understand aggregation of a ternary system (KA, cations, water) by first examining the pure main ingredient (KA) and its interaction with the second ingredient (cations), as we do here. Furthermore, experiments indicate that the cation-aminoglycoside clusters are perfectly stable without any solvent molecules under the rather stringent mass-spectrometry conditions.

Methods

Our OGOLEM program suite has been presented in full detail in a recent publication [21]. Therefore we just highlight the features and techniques used in this work.

OGOLEM is loosely based on genetic algorithms as described in Ref. [22], using a pool variant as described in Ref. [23], and has been constructed to be universal by design. This includes both a universality in the allowed type and composition of building blocks as well as in the methods used to evaluate energies of structures. For this work, the interfaces toward AMBER[24] and MOPAC[25] have been used, allowing for energy evaluations on classical mechanical and semiempirical levels.

All global optimization calculations mentioned in this paper have been carried out on the classical mechanical AMBER/GAFF level of theory with implicitly relaxed building blocks. As starting structures, KA monomers optimized with B3LYP/TZVP have been used. OGOLEM then created randomized starting dimers including an ion, if applicable. These starting structures where locally optimized using GAFF and added to the genetic pool. The genetic pool has then been used for mating of two structures, where one was picked randomly and the other fitness weighted. These structures where then crossed using the *genotype*, real-number-based *Germany* operator and mutated with a 5% probability. To reduce the number of local optimizations, OGOLEM then checks the resulting children structures for physical sensibility using a collision (CD) and dissociation detection (DD) based on distance criteria and graphs. These are designed to “localize“ the search space around possible minimum funnels, drastically

reducing search space size. Crossing and mutation are attempted again if either CD or DD flag the structure as unphysical. Only children structures that are neither clashing nor dissociated are then again locally optimized.

Only the fitter of the two children structures is then returned to the genetic pool and only added if it does not violate a fitness diversity criterion, designed to reduce the risk of premature convergence of the genetic pool.

These primitive global optimization steps are then repeated, yielding a pool of minima.

To assess the quality of GAFF for this class of applications, a representative pool for each system containing 1000 structures was then re-optimized using MOPAC/PM3. The resulting structures were again ranked by energy, structures which converged into the same minimum were eliminated and for the best ten structures of each system, the energy was again evaluated using MOLPRO[26] and the DF-LMP2/cc-pVDZ level of theory.

The lowest-energy structure of each system can then be considered to be a good candidate for the global minimum structure.

For these resulting, lowest energy structures, electronic properties were calculated using Orca [27] and the RI-BP86/TZVP level of theory, namely NMR and IR spectra. We are fully aware of the shortcomings of GGA-based DFT for long-ranged interactions, as appearing in clustered systems. Considering the performance penalty of post-HF *ab initio* methods, we need to trade accuracy for speed. Additionally, it can be argued that the dominant intra- and intermolecular interactions in pure KA clusters can be expected to be hydrogen bonds, for which GGA-DFT has a reasonable chance of providing qualitatively correct results. KA clusters with cations like Na^+ and K^+ can be expected to simplify the situation further, due to additional, strong interactions between the cationic charge and the partial charges on the KA molecules, which should be described fairly well by GGA-DFT.

Results and discussion

In order to highlight the effects of aggregation, we first present results for the KA monomer, followed by results for the KA dimer systems, without and with sodium or potassium ions.

Kanamycin A monomer

The starting point for any kind of attempt toward the global optimization of a cluster is a proper characterization of the building block in its monomer form. While this is relatively trivial for atomic and well-known, small molecular building blocks, it is getting difficult with bigger molecules such as KA. We therefore present both results on the local optimization of a monomer unit as well as monomer properties such as NMR and IR spectra, as a reference for the dimer results to be obtained later.

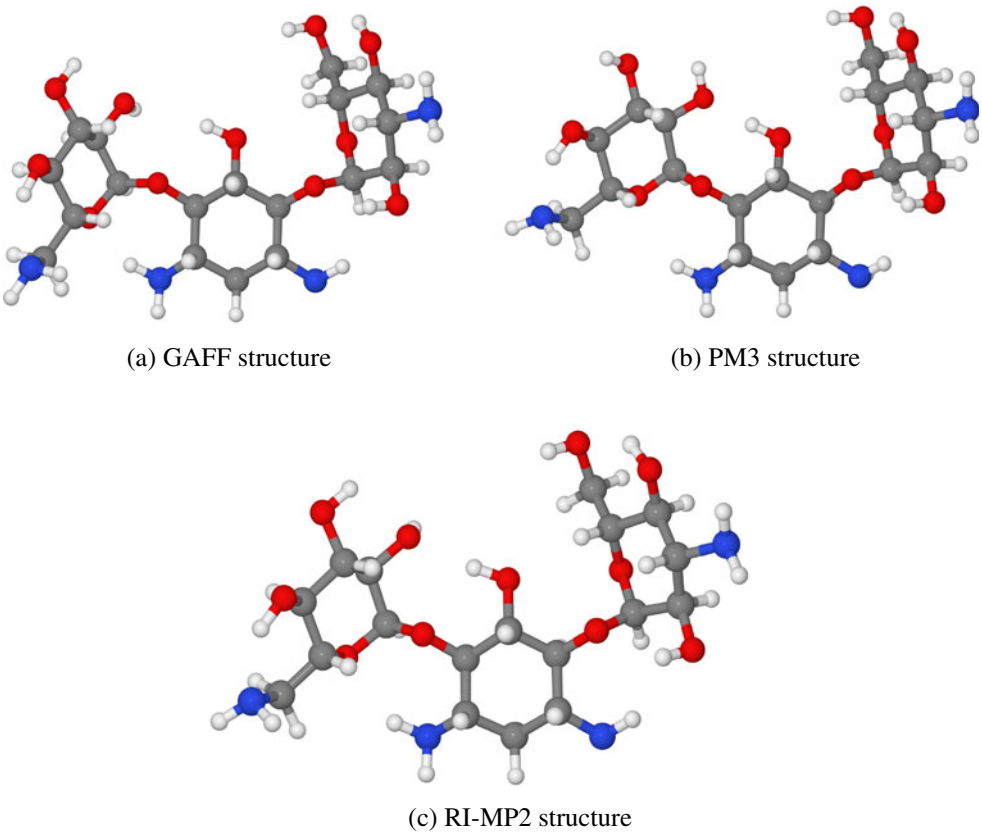
Starting with the task of local optimization, we compare results for the local optimization of the monomer structure using the semiempirical PM3 method, RI-BP86 and B3LYP with the TZVP basis set as well as RI-MP/def2-TZVPP and the GAFF force field. Since weak long-range forces should not have too big of an impact on the monomer structure, we consider density functional theory to be an appropriate level of theory for comparative purposes.

As can be seen in Table 1, there is no geometrical difference between the choice of a GGA functional like BP86 and the hybrid B3LYP, the RMSD value being below 0.1 Å. PM3 does increase the RMSD value to 0.238 and 0.291 Å in comparison to RI-BP96 and B3LYP respectively. The RMSD to the MP2 reference are below 0.2 Å for both DFT methods and PM3, clearly proving that all three hold as valid descriptions of the monomer. Bigger differences arise to the AMBER/GAFF level of theory, bringing the RMSD values up to 5.7 Å. This is neither surprising nor a reason to worry if GAFF is a valid description of KA. Of course, a classical-mechanical description of any molecular or atomic system neglects important quantum mechanical contributions resulting in a noticable error. Furthermore, a seemingly large value of a single numerical measure like the RMSD value is not necessarily a reliable indicator for model breakdown: Obviously, the RMSD value cannot discern between chemically relevant distortions and minor deviations in uninteresting bond length or angles.

Therefore, for better illustration, the monomer structures of GAFF, PM3 and MP2 are depicted in Fig. 1. As can be seen, the qualitative differences between these structures is still negligible, despite the RMSD values of up to 5.7 Å. None of the presented deviations allows the conclusion that any of the methods would be not suitable for the local optimization of the monomer molecule. This is of special

Table 1 RMSD values of the resulting local minima of the monomer. All values in Å

	PM3	RI-BP86/TZVP	B3LYP/TZVP	RI-MP2/def2-TZVPP
GAFF	5.697	5.724	5.702	4.777
PM3		0.238	0.291	0.186
RI-BP86			0.062	0.145
B3LYP				0.119



(a) GAFF structure

(b) PM3 structure

(c) RI-MP2 structure

Fig. 1 Minimum structures using different levels of theory. Graphical representations of these structures and all following were produced with Jmol [28] and POV-Ray[29]

importance since our ultimate target is the optimization of larger aggregates, using cheap descriptions like force fields. Any method already failing in the monomer process of course would have to be discarded instantly.

As indicated in the introduction, there already is NMR data on (possibly large) aggregates of KA in experiment. For comparison with our future theoretical studies on larger KA aggregates, we have also calculated NMR chemical shifts for the KA monomer and dimers. For this purpose, we have both used the hybrid B3LYP and the GGA BP86 functional in the ORCA program package, predicting NMR peaks in the hydrogen and carbon spectra. Here and in the following, a complete listing of all data can be found in the electronic supplementary information.

For comparative purposes, we carried out several test runs to assess the impact of certain variations in the theoretical approach, in particular, we tested different functionals (GAA vs. hybrid), different basis sets (IGLOIII vs. TZVP) and different geometries (reoptimized vs. non-reoptimized). The full set of chemical shifts can be found in the supplementary information for both hydrogen and carbon. A representative collection can be found in Table 2 for hydrogen shifts and Table 3 for carbon shifts. For the hydrogen shifts, the replacement of the hybrid B3LYP functional with the GGA BP86 functional causes a shift of the peaks to high field. On the other hand, the substitution

Table 2 Representative hydrogen shifts of the KA monomer. All results in ppm. Reference: hydrogen peak of chloroform 23.9 ppm (B3LYP/IGLOIII), 23.2 ppm (RI-BP86/IGLOIII), 24.4 ppm (RI-BP86/TZVP)

Core	B3LYP IGLOIII reopt	RI-BP86 IGLOIII reopt	RI-BP86 TZVP reopt	RI-BP86 TZVP PM3 geometry
2	30.2	29.8	30.2	30.5
15	27.6	27.1	27.7	27.1
32	28.4	28.0	28.3	27.8
34	29.2	28.7	29.1	30.0
67	27.9	27.5	27.8	27.4

Table 3 Representative carbon shifts of the KA monomer. All results in ppm. Reference: carbon peak of chloroform 73.2 ppm (B3LYP/IGLOIII), 63.5 ppm (RI-BP86/IGLOIII), 98.8 ppm (RI-BP86/TZVP)

Core	B3LYP IGLOIII reopt	RI-BP86 IGLOIII reopt	RI-BP86 TZVP reopt	RI-BP86 TZVP PM3 geometry
4	128.1	124.0	162.3	162.3
22	70.0	62.7	106.8	103.8
26	100.4	95.6	135.4	129.3
36	139.2	135.6	173.3	170.2

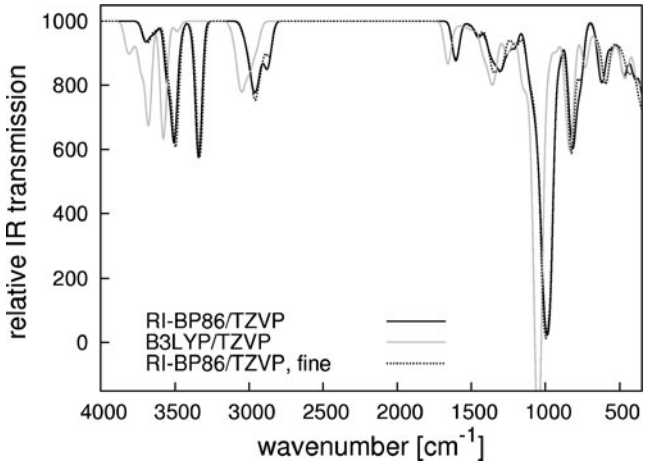


Fig. 2 Predicted infrared spectra of the KA monomer

of the specialized IGLOIII basis set with the general-purpose TZVP basis set causes a shift to low field. In the case of hydrogen shifts, this yields a good error compensation, causing the resulting RI-BP86/TZVP peaks to be in excellent agreement with the B3LYP/IGLOIII peaks (no deviation bigger than 0.3 ppm). Carrying the chemical shift calculations out on the PM3 geometry causes deviations up to 1 ppm, or approx. 4%, indicating that shielding is less sensitive toward non-minimum structures than other properties like vibrational spectra. This indicates that for larger KA aggregates, qualitative ^{1}H-NMR spectra could be obtained from force field optimized structures without reoptimization at the DFT level. According to further test calculations, we consider these results to be robust under all tested circumstances.

In the carbon shifts, similar trends are observed. The substitution B3LYP to BP86 causes a shift to high field, the substitution IGLOIII to TZVP causes a shift to low field. Unfortunately, the combined RI-BP86/TZVP method does not profit from the error compensation encountered in the case of hydrogen shifts. The shift to low field is signifiantly bigger than the shift to high field, in total causing low-field shifted peaks. Since the chloroform reference shows the same behavior (low-field shift of 15.6 ppm against B3LYP/IGLOIII), the overall robustness of the RI-BP86/TZVP level of theory is reasonable. The chemical shifts calculated on the PM3 structure now introduce a deviation of up to 20.2 ppm or approx. 14%. Although this is a bigger deviation than in the case of hydrogen shifts, the general agreement and robustness is acceptable.

Another standard analytical tool are infrared spectra. Since their computational generation is relatively sensitive toward both the shape of the hypersurface in close vicinity to the minimum structure as well as the structural shape, this is also a rather good benchmark for the applied level of theory. In Fig. 2, a comparison is shown between different predictions using the standard BP86 GGA functional and the hybrid B3LYP functional. Trying to estimate the accuracy of the ORCA default settings for the spectral prediction, we thightened the SCF convergence threshold from $5\,x\,10^{-6}$ E_h to $1\,x\,10^{-6}$ E_h, enabled central differences to be calculated, reduced the numeric increment to $0.001\,a_0$ and refined the DFT grid. Introducing a performance penalty of a factor of approx 6.5, these settings should allow for rather converged spectra within the applied level of theory. Comparing the so-obtained spetra with the one with default settings, one can see that the default settings are well sufficient for the KA system, changing only slightly the shape of individual peaks.

The next step then of course needs to be the comparision with another functional. In this case, the B3LYP functional can probably be considered to resemble a somewhat superior choice. Comparing the B3LYP spectra obtained with default settings on a B3LYP optimized geometry to the one obtained with RI-BP, one clearly sees a shift of peaks. Albeit this shift is clearly visible, it is not changing peak orders or shapes in a qualitative way, therefore allowing the conclusion that the BP86 functional allows for a qualitatively correct desription in the context of density functional theory whilst cutting the computational cost by a factor of approx. 14.5 when additionally applying the density fitting approximation to BP86.

Drawing the conclusion that the RI-BP86/TZVP description is sufficient for our purposes, the prediction of Raman spectra is computationally feasible. The monomer spectrum obtained with ORCAs default settings can be found in Fig. 3.

The results presented here are meant as a reference for the effects of aggregation. They definitely have shortcomings stemming from the level of theory (DFT) and numerical approximations (numerical spectra) but comparisons shows that they seem to be sufficiently stable.

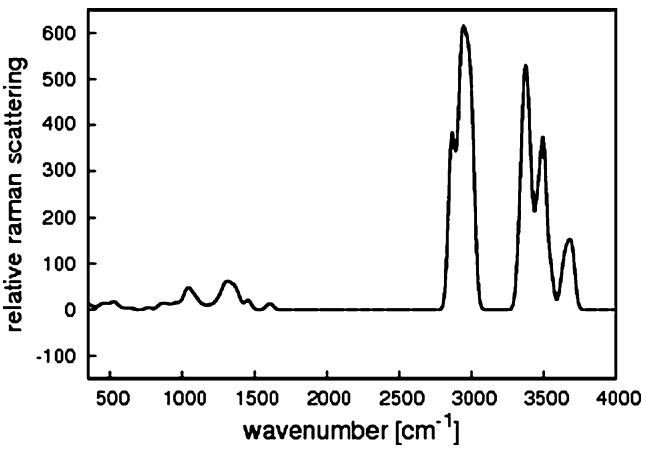


Fig. 3 Predicted Raman spectrum of the KA monomer, at the RI-BP86/TZVP level of theory

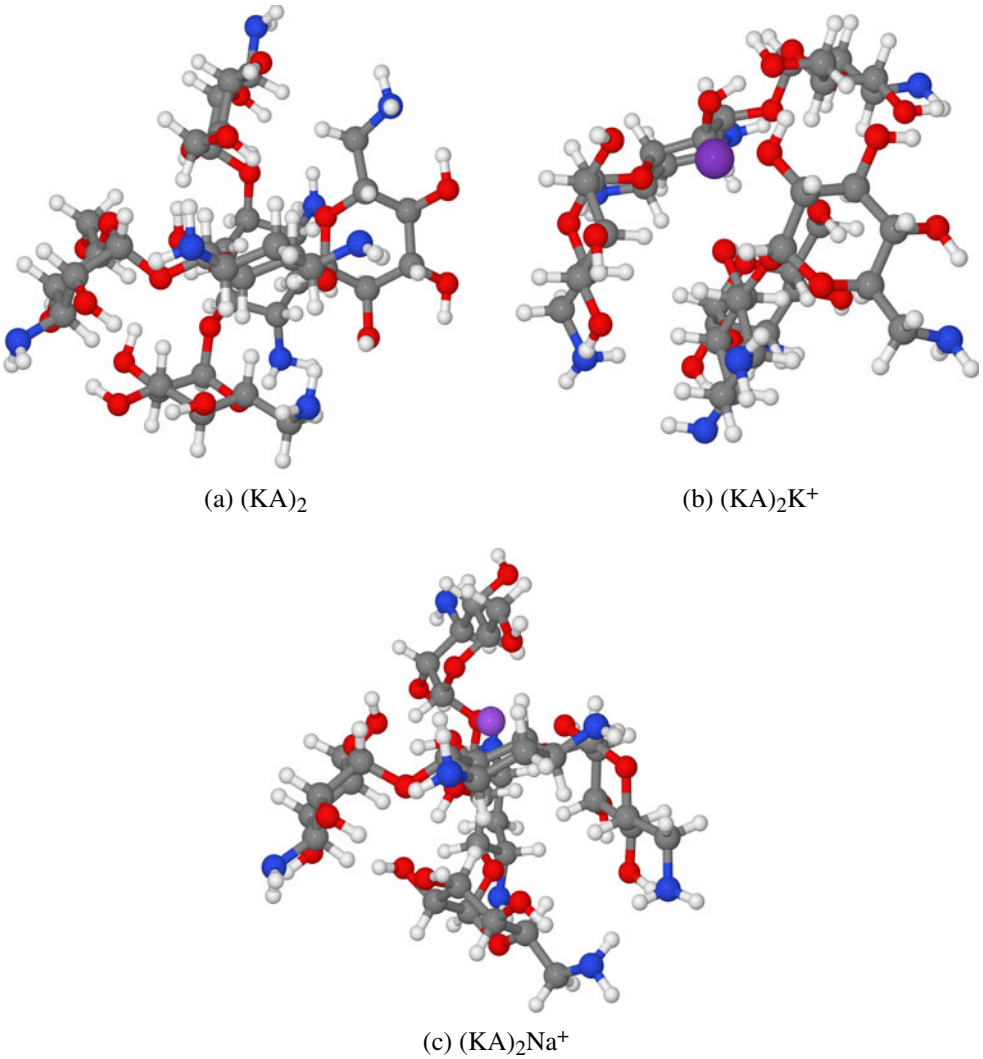
Kanamycin A dimer

The global optimization of the KA dimer with different physiological cations was carried out using OGOLEM and AMBER with the GAFF force field. Multiple successive runs were carried out, employing a number of global optimization iterations large enough to yield the same result in most of them. Based on this the assumption can be made that a converged result and a good candidate structure for the global minimum was obtained. These global minimum candidate structures are displayed in Fig. 4, for the bare KA dimer and for the KA dimer with a sodium and a potassium cation, respectively.

A noticable difference is encountered. While the dimer without ions and with the sodium ion forms a crossed structure with the ion (if applicable) centered in between the two monomer units, the dimer with the potassium ion forms a more open structure, exposing the ion to the surface. On the one hand, this is an intriguing result since experimental data point to a preference for sodium over potassium in KA aggregates with physiological ions. On the other hand, this result obviously cannot be taken as is, since there exists no knowledge on whether or not the chosen level of theory provides even qualitative agreement with higher levels of theory for these systems.

Therefore, a representative final pool from the GAFF runs is taken completely, containing a thousand different minimum structures for each system, and reoptimized using MOPACs PM3 implementation. Correlations between the obtained PM3 and GAFF energies can be found in Fig. 5.

While some remnants of qualitative agreement are visible, the overall correlation is in all cases far from optimal. Additionally, as can be seen from the horizontal accumulations of points (most pronounced in the KA dimer without ion), a lot of GAFF minima seem to disappear on the PM3 hypersurface, leaving a smaller total number of minima. Contrary to first impression, these findings do not rule out the use of this methodical ladder for this system.



(a) $(KA)_2$ (b) $(KA)_2K^+$

(c) $(KA)_2Na^+$

Fig. 4 Minimum structures of the KA dimer on the GAFF level of theory

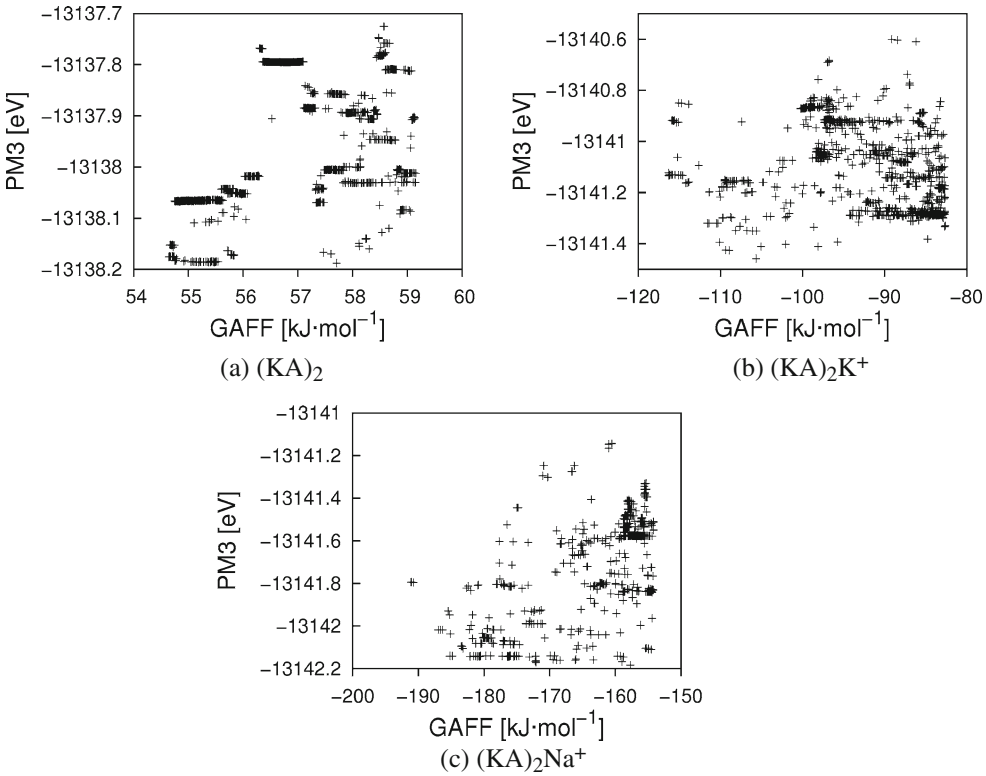


Fig. 5 Correlations between the GAFF and PM3 energies of optimized structures

Correlations of similarly poor quality have been successfully used earlier in studies of other systems; a rare example where bad correlations were admitted and shown was a study on silicon clusters [30]. In such a situation, the key to success is to use a very large number of low-level results as input for the higher-level method, as we do here. Of course, a more correct way to improve matters would be a recalibration of the lower-level method. We reserve this step for future work.

The lowest PM3 structure of each system can be found in Fig. 6. Obviously, the dimer system without ion changes significantly between the GAFF and the PM3 level of theory. The other two systems do not change qualitatively, allowing for the conclusion that despite the unfavorable correlation of energy values found above the GAFF level of theory provides some qualitative agreement with the PM3 level of theory for the best structures. Nevertheless, apparently some caution is in order.

The resulting minimum structures were ranked by PM3 energy and from each system, the lowest ten minimum structures with substantial structural difference were picked manually. The energies of these structures were then recalculated with the linear scaling DF-LMP2/cc-pVDZ level of theory. We are fully aware of the errors introduced by local approximations and the limited basis set size. However, these are the biggest possible calculations for us to-date. As an additional bonus, this local method eliminates basis set superposition error (BSSE) by design, which enhances the reliability of the results for clustered molecules.

As can be seen from the data depicted in Fig. 7, and in comparison to the GAFF-PM3 comparisons presented above, the correlation between PM3 and LMP2 is rather good for $(KA)_2Na^+$, moderate for $(KA)_2$ and non-optimal in the $(KA)_2K^+$ case. Of course, with only 10 data points the statistics is much worse than for the GAFF–PM3 comparison above. Nevertheless, the conclusion seems to be that PM3 may be usable as predictor for ab-initio results, provided some caution is exercised, as explained in detail above.

This situation is illustrated even better when checking the best LMP2 structures in Fig. 8 in a qualitative manner. Just as in the PM3 and GAFF case, the sodium ion is tightly encapsulated by the two KA monomers interacting to form a hydrogen-bonded network of OH groups around the cation. In contrast, such a tight and extended OH group network is not possible around the larger potassium ion. Instead, in this case the two KA molecules group themselves more loosely around the cation. Actually, in comparison to the sodium cation, the potassium cation is not in the center of the cluster but exposed to its surface. With this observation being robust across strongly different levels of theory, we consider this to be a major clustering tendency of the KA system, worth further investigations targeted at the relative stability of KA aggregates of various sizes with ions of different types and numbers. Similar differences in coordinative propensities between these two cations have been noted frequently [31–33]. One simple reason (among others) for this different behavior is the different energy cost for re-orientation [33] in the interac-

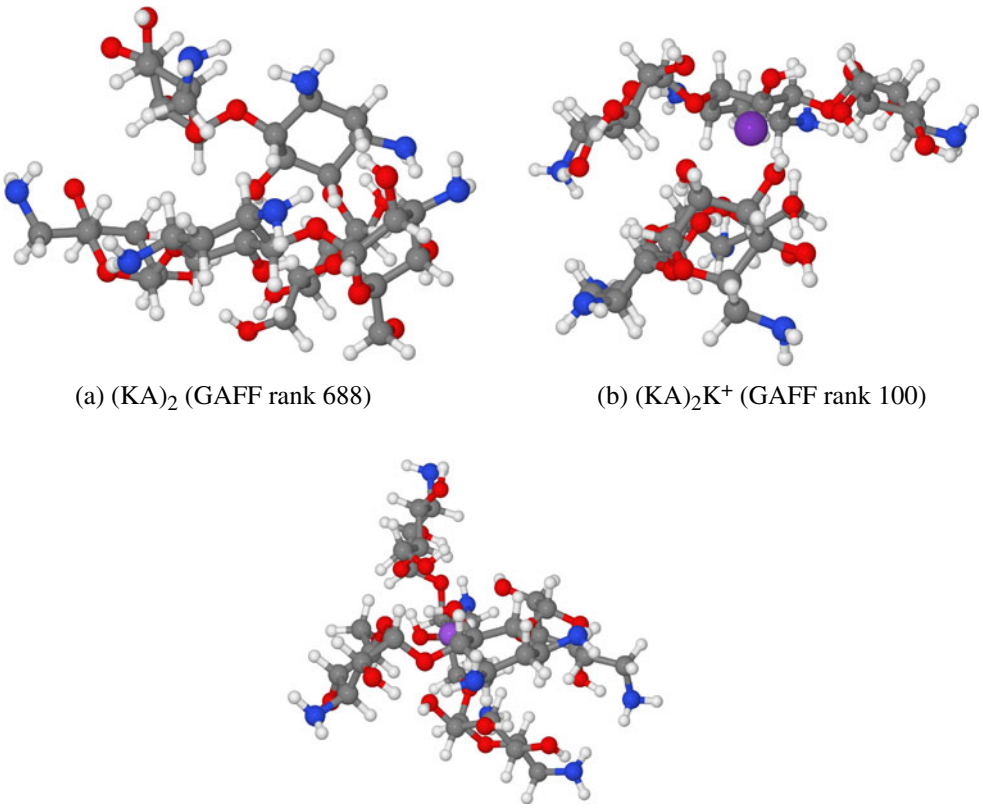


Fig. 6 Minimum structures of the KA dimer on the PM3 level of theory

(a) $(KA)_2$ (GAFF rank 688)

(b) $(KA)_2K^+$ (GAFF rank 100)

(c) $(KA)_2Na^+$ (GAFF rank 583)

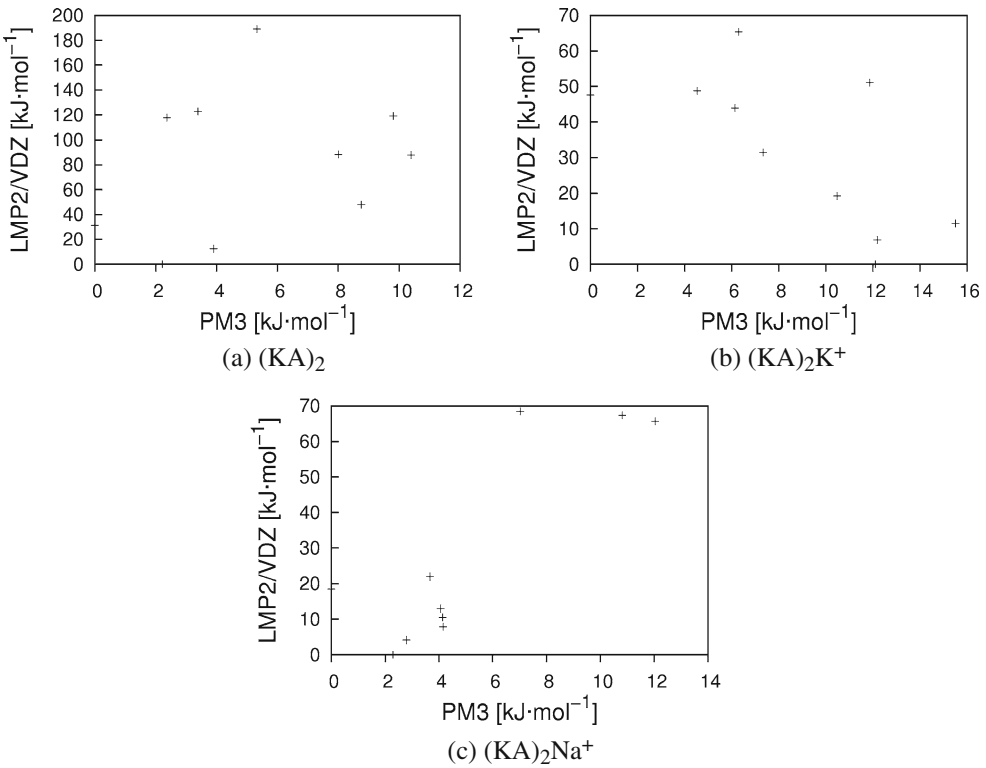


Fig. 7 Correlations between PM3 and DF-LMP2/cc-pVDZ energies of optimized PM3 structures. Energies referenced to the lowest energy structure

(a) $(KA)_2$

(b) $(KA)_2K^+$

(c) $(KA)_2Na^+$

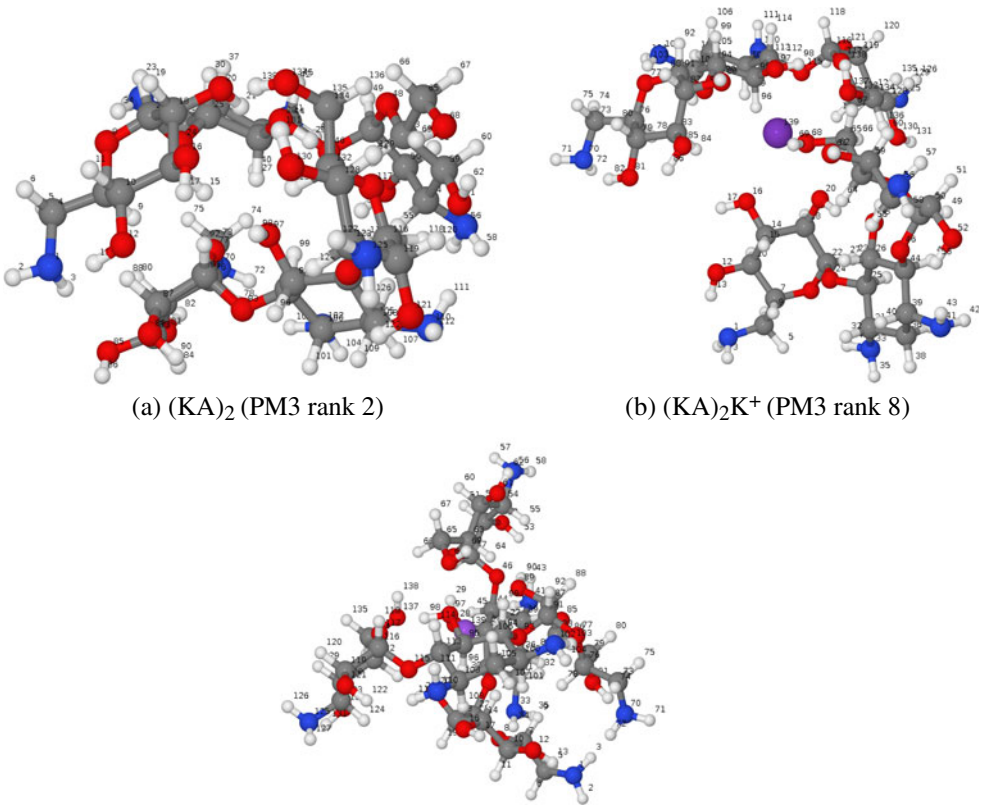


Fig. 8 Minimum structures of the KA dimer on the DF-LMP2/cc-pVDZ level of theory

(a) $(KA)_2$ (PM3 rank 2)

(b) $(KA)_2K^+$ (PM3 rank 8)

(c) $(KA)_2Na^+$ (PM3 rank 2)

tion between the cation and its OH-group coordination environment. Interestingly, one central experimental finding on KA aggregates is that they indeed clearly prefer sodium cations over potassium cations.

Since we want to provide hints for the experiments on how to detect clustering of these molecules, spectral data is of utmost importance. Through easy dilution experiments, an NMR study of aggregation of these systems should be possible. Therefore, the robust chemical shifts are a good and computational feasible choice for prediction of spectral data.

In Tables 4 and 5, representative results of the calculation of chemical shifts are compiled for all studied dimers and, as reference, for the monomer. The full set of chemical shifts is available from the supplementary information. Starting with the $(KA)_2$ dimer, the biggest

Table 4 Representative hydrogen chemical shifts of the dimers. All values in ppm, calculated with RI-BP86/TZVP on reoptimized structures

Core	KA	$(KA)_2$	$(KA)_2K^+$	$(KA)_2Na^+$
2	30.2	30.1	29.7	29.9
21	29.7	25.3	23.8	28.5
69	30.8	26.5	28.5	26.9
80	27.6	27.4	27.7	27.5
82	26.1	25.0	23.9	22.9
90	29.7	29.2	26.2	25.8
136	27.8	27.6	26.9	26.8
138	30.8	25.1	29.6	25.5

Table 5 Representative carbon chemical shifts of the dimers. All values in ppm, calculated with RI-BP86/TZVP on reoptimized structures

Core	KA	$(KA)_2$	$(KA)_2K^+$	$(KA)_2Na^+$
4	162.3	161.7	161.9	162.5
14	138.4	130.8	133.4	133.3
25	113.0	119.4	116.0	116.2
39	156.6	152.4	156.9	156.5
44	123.5	123.3	110.2	111.0
63	143.5	141.3	136.7	129.8
65	142.1	142.3	142.5	143.9
94	113.0	128.7	114.1	108.4
95	135.4	127.8	135.4	136.7
113	123.5	123.0	109.7	109.4
116	109.1	108.3	102.2	101.6
132	143.5	136.9	132.8	127.8
134	142.1	143.6	143.6	145.9

changes are observed for 14C, 25C, 94C, 95C and 132C in the carbon shifts. These are all in close range to the other building block and therefore subject to interaction with it. In the hydrogen shifts, the biggest changes are observed for 21H which forms a hydrogen bond to 28O, 69H forming a hydrogen bond to 117O and 138H forming a hydrogen bond to 20O. This is in good agreement with the observed networks of hydrogen bonds, making the NMR chemical shifts a good indicator of aggregation.

For the $(KA)_2K^+$ system, the biggest changes in carbon shifts are observed for 44C, 113C and 132C which are in close vicinity to the potassium ion. In the hydrogen shifts, the biggest differences occur for the 21H again forming a hydrogen bond to 28O, 69H being subject to attraction of both 61O and 89O as well as 90H forming a hydrogen bond to 61O.

The $(KA)_2Na^+$ system behaves similar to the $(KA)_2K^+$ system, showing the biggest differences for 44C, 63C, 113C and 132C, which are all subject to interaction with the sodium ion. The biggest differences in hydrogen shifts occur for 69H, forming a hydrogen bond to 89O, 82H interacting with 70N, 90H interacting with both 20O and 28O, and 138H being both in the immediate vicinity to the sodium ion and to 68O.

Similarly clear signatures of aggregation are also visible in the IR spectra. In order to promote the experimental use of vibrational spectroscopy on aminoglycoside clusters, we discuss our findings in this area in greater detail. Figure 9 displays an overview of our simulated IR spectra for the bare KA dimer, and for $(KA)_2Na^+$ and $(KA)_2K^+$, respectively, in comparison to the KA monomer spectrum which was already shown in Fig. 2. At first sight, all four IR spectra appear to be very similar. However, a closeup of the OH fingerprint region (cf. Fig. 10) reveals striking differences. The three strong peaks at 3338 cm^{-1}, 3495 cm^{-1} and 3551 cm^{-1} in the monomer spectrum correspond to stretching vibrations of the three OH groups that can form hydrogen bonds to neighboring OH or NH_2 groups in the optimal KA monomer structure. The strongly differing peak positions correspond to the different environments: One OH group at one terminal ring is hydrogen-bonded to a neighboring OH group; it corresponds to the peak at 3551 cm^{-1}. The second OH group sits on the middle ring and forms a hydrogen bond to an OH group at a terminal ring; it is responsible for the peak at 3495 cm^{-1}. The third OH group is at the other terminal ring and hydrogen-bonds to an NH_2 group there, leading to the peak at 3338 cm^{-1}.

In the KA dimer, the signals from the two terminal OH groups survive but are shifted to 3190 cm^{-1} and 3311 cm^{-1}, respectively, due to changed monomer-internal hydrogen-bond surroundings, which in turn are induced by the relatively close packing of the two monomers in the dimer. Interestingly, the remaining four strongest peaks actually

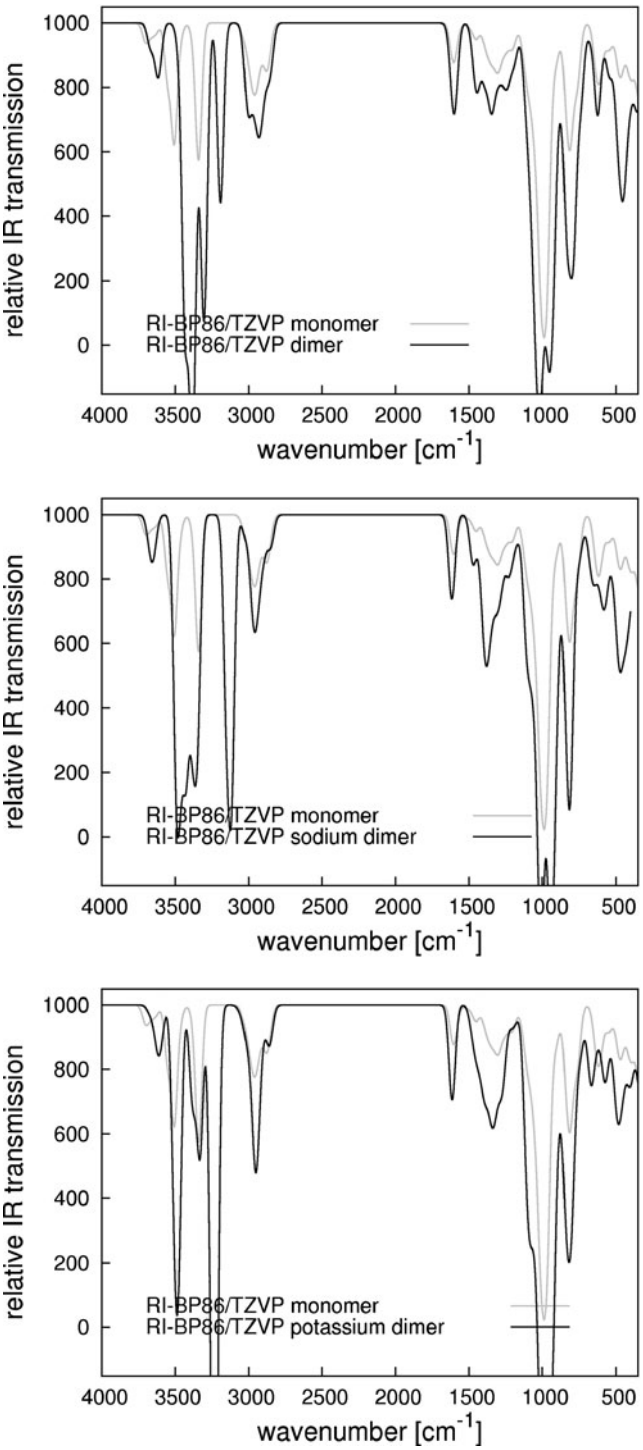


Fig. 9 Predicted infrared spectra of the bare KA dimer (top), and the KA dimers with one sodium cation (middle) and one potassium cation (bottom), respectively

are two peak pairs (in symmetric and antisymmetric versions) involving strong contributions from two or three OH groups hydrogen-bonded to each other and sitting on different monomers. In other words, they provide a specific signature of the KA–KA aggregation.

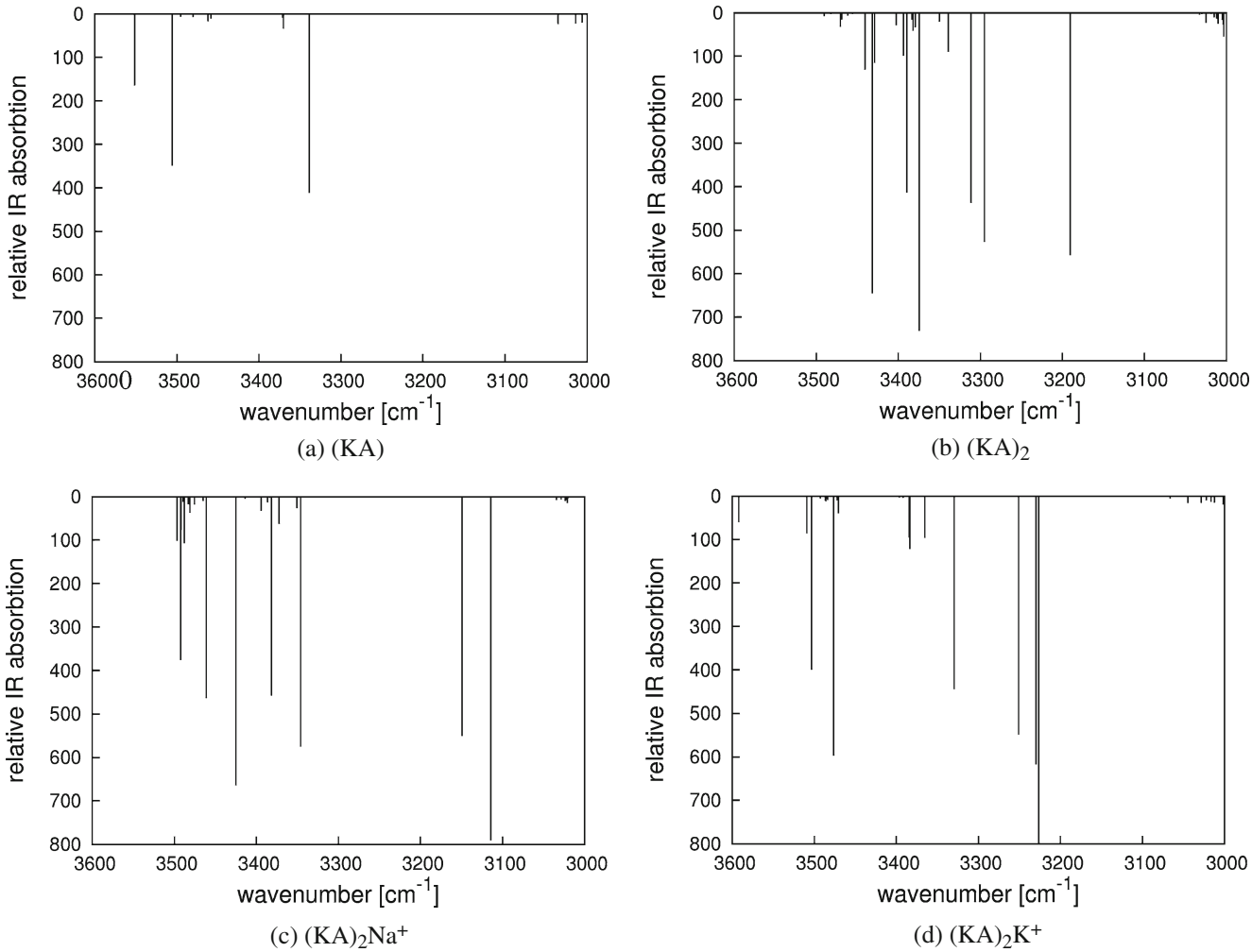


Fig. 10 OH-stretch fingerprint region of predicted IR spectra. Panel (**a**): KA monomer, (**b**): $(KA)_2$, (**c**): $(KA)_2Na^+$, (**d**): $(KA)_2K^+$

Even more interestingly, aggregation signatures in the dimer complexes with cations exhibit both similarities and differences, with respect to the bare dimer. Thus, they provide a clear signature of how the two KA monomers encapsulate the sodium cation. They do this exclusively via the OH groups mentioned above, forming a hydrogen-bonded network around the sodium cation. The three strong peaks in the interval between 3400 and 3500 cm^{-1} in the $(KA)_2Na^+$ spectrum (at 3425 cm^{-1}, 3461 cm^{-1} and 3492 cm^{-1}) involve strong contributions from two or even three such OH groups, and remarkably these always come from both KA molecules. The vibrations of the remaining OH groups in the network around the sodium cation are strongly redshifted and split up into peak pairs, one for each KA in the complex (the KA monomer peak at 3338 cm^{-1} reappears as the peak pair at 3114 cm^{-1} and 3149 cm^{-1}; the mid-ring OH group moves from 3495 cm^{-1} to 3346 cm^{-1} and 3381 cm^{-1}).

Several remaining spectral features are common to both complexes, $(KA)_2Na^+$ and $(KA)_2K^+$: Dangling OH groups pointing to the cluster outside are far away from the cation and exhibit correspondingly small IR intensities, they are responsible for small signals above 3600 cm^{-1}. The NH2 groups also point to the outside of the cluster. Their symmetric and antisymmetric strech vibrations produce the tiny background signals in the OH-group cation coordination fingerprint spectral region (3300–3500 cm^{-1}) discussed above. The other end of the depicted range, at 3050 cm^{-1} and below, exclusively belongs to CH-stretch vibrations.

The most prominent peaks of the $(KA)_2K^+$ complex, however, show a strikingly different pattern, compared to the KA monomer, the KA dimer and the $(KA)_2Na^+$ cases. The two differentiating structural features of the $(KA)_2K^+$ complex are that the potassium cation is larger than the sodium cation and much less tightly packed in a less extensively connected hydrogen-bonded OH-group network within the KA dimer complex. These structural features produce two effects in the cation-neighboring OH-stretch normal modes: (1) Modes that largely resemble

KA monomer ones show a smaller redshift. For example, the peaks at 3229 cm^{-1} and 3250 cm^{-1} essentially correspond to the monomer peak at 3338 cm^{-1}. Thus, they are redshifted by about 100 cm^{-1}, only half of the amount in the Na^+ case. (2) The remaining modes are of very different character, compared to the sodium cation case. Many of them have significant contributions only from one OH bond, and if a second OH bond contributes it does not belong to the closer neighborhood of the cation anymore and thus has lower intensity. Therefore, while it may be difficult to differentiate between the bare dimer and $(KA)_2Na^+$ without close support from theory, it may be possible to detect the significantly less tight packing in $(KA)_2K^+$ without detailed analysis, merely from the absence of strong peaks in the interval between 3300 cm^{-1} and 3500 cm^{-1} that corresponds to OH-combination bands in the hydrogen-bond network spanning the two monomers and the cation.

In summary, as discussed above, both the NMR and the IR results provide good hints on, and reflect the structural changes occuring in, the different aggregation structures of KA with different ions. Therefore, these simulation data provide valuable support for future experimental studies of aminoglycoside clusters.

Summary and outlook

Larger clusters of an aminoglycoside closely related to KA, together with certain physiological cations, have been found in recent experiments on innate immune system response of the human skin. It was unknown, however, if KA forms resonably stable clusters at all, let alone what their properties could be. In this article, we have provided the necessary ground work to launch a large-scale study to provide the badly needed theoretical support for this line of work.

Specifically, with benchmark calculations for the KA monomer, and global optimization studies of the KA dimer and of the systems $(KA)_2Na^+$ and $(KA)_2K^+$, we have tested the performance of the standard GAFF force field and the standard PM3 semiempirical method against DFT and MP2. The obtained method correlations are only partly satisfactory, indicating the need to depart from standard low-level methods and to attempt system-specific recalibrations for future studies of these systems. Nevertheless, a robust finding across all three methods is that Na^+ induces the KA dimer to take on a well-ordered, chelate-like shape around the cation, forming a stable aggregate and shielding the cation from the environment. This does not happen for the potassium cation, where only a loose, unshielded aggregate is formed. This correlates nicely with the experimental finding that Na^+ is preferred over K^+ in these systems.

To provide further convenient contact points with experiment, we have also calculated NMR and IR data for the same systems. Detailed analysis of the results reveals in both cases clear and experimentally accessible indicators for KA-KA and KA-cation aggregation, including the possibility to differentiate between tightly encapsulated and loosely bound structures.

In order to improve the correlations between different calculational levels, future work will address a force field recalibration against the ab-initio data provided here. Our findings indicate that improving the description of the potassium ion should be the most important and perhaps already sufficient ingredient. Besides implicit and/or explicit solvent modeling already mentioned as obligatory extension in the introduction, further important goals to intensify contact with experiment are larger KA aggregates, also including other physiological cations (both experimentally preferred ones like Cu^{2+} and other obvious possibilities like Mg^{2+}, Ca^{2+} or NH_4^+ that do not seem to fit into the KA clusters), and comparisons to minor constitutional isomers of KA (again following experimental indications).

Acknowledgements B.H. would like to thank the German Research Foundation (DFG) for strong financial support via grant Ha-2498/10. J.M.D. and B.H. thank the North-German Supercomputing Alliance (HLRN) for a generous grant of computer time, and would like to acknowledge the friendly and competent efforts of the HLRN support staff. J.M.D. would like to acknowledge Mats Eriksson for hints on AMBER's (fixed) input formatting.

References

1. Campbell EEB, Larsson M (eds) (2000) The physics and chemistry of clusters, no. 117 in Proc of Nobel Symposium. World Scientific
2. Johnston RL (2002) Atomic and molecular clusters. Taylor and Francis, New York
3. Catlow CRA, Bromley ST, Hamad S, Mora-Fonz M, Sokol AA, Woodley SM (2010) Phys Chem Chem Phys 12:786
4. DJ Wales (2003) Energy landscapes. Cambridge University Press, Cambridge
5. Hartke B (2002) Angew Chem Int Ed 41:1468
6. The cambridge cluster database. http://www-wales.ch.cam.ac.uk/CCD.html
7. Görbitz CH, Dalhus B, Day GM (2010) Phys Chem Chem Phys 12:8466
8. Kim S, Orendt AM, Ferraro MB, Facelli JC (2009) J Comput Chem 30:1973
9. Karamertzanis PG, Kazantsev AV, Issa N, Welch GWA, Adjiman CS, Panetlides CC, Price SL (2009) J Chem Theor Comput 5:1432
10. Schröder JM, Harder J (2006) Cell Mol Life Sci 63:469
11. Harder J, Meyer-Hoffert U, Teran LM, Schwichtenberg L, Bartels J, Maune S, Schröder JM (2000) Am J Respir Cell Mol Biol 22:714
12. Puius YA, Stievater TH, Srikrishnan T (2006) Carbohydr Res 341:2871
13. D'Amelia N, Gaggelli E, Gaggelli N, Molteni E, Baratto MC, Valensin G, Jezowska-Bojczuk M, Szczepanik W (2004) Dalton Trans. p. 363

14. Lu XQ, Zhang M, Kang JW, Wang XQ, Zhuo L, Liu HD (2004) J Inorg Biochem 98:582
15. Lesniak W, Harris WR, Kravitz JY, Schacht J, Pecoraro VL (2003) Inorg Chem 42:1420
16. Kopaczynska M, Lauer M, Schulz A, Wang T, Schaefer A, Fuhrhop JH (2004) Langmuir 20:9270
17. Hermann T, Westhof E (1999) J Med Chem 42:1250
18. Moitessier N, Westhof E, Henssian S (2006) J Med Chem 49:1023
19. Huang L, Massa L, Karle J (2007) Proc Nat Amer Soc 104:4261
20. Monajjemi M, Heshmata M, Haeria HH (2006) Biochem (Moscow) 71:S113
21. Dieterich JM, Hartke B (2010) Mol Phys 108:279
22. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Kluwer, Dordrecht
23. Bandow B, Hartke B (2006) J Phys Chem A 110:5809
24. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossv´ary I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008) Amber 10, Univ California, San Francisco
25. Stewart, JJP (2009) Mopac, version 10.055l
26. Werner HJ, Knowles PJ, Lindh R, Manby FR, Schütz M et al (2010) Molpro, development version, a package of ab initio programs
27. Orca: an ab initio, DFT and semiempirical electronic structure package
28. Jmol: an open-source java viewer for chemical structures in 3d. http://www.jmol.org
29. Pov-ray - the persistence of vision raytracer. www.povray.org
30. Jackson KA, Horoi M, Chaudhuri I, Frauenheim T, Jackson KA (2004) Phys Rev Lett 93:013401
31. Dudev T, Lim C (2010) J Am Chem Soc 132:2321
32. Eriksson M, Lindhorst TK, Hartke B (2008) J Chem Phys 128:105105
33. Schulz F, Hartke B (2002) Chem Phys Chem 3:98