Gas-phase salt bridge interactions between glutamic acid and arginine

Jaeqx, S.; Oomens, J.; Rijs, A.M.

DOI
10.1039/c3cp52508b

Publication date
2013

Document Version
Final published version

Published in
Physical Chemistry Chemical Physics

PCCP

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2013, **15**, 16341

Gas-phase salt bridge interactions between glutamic acid and arginine†

Sander Jaeqx,[a] Jos Oomens[ab] and Anouk M. Rijs*[a]

Received 17th June 2013,
Accepted 6th August 2013

DOI: 10.1039/c3cp52508b

www.rsc.org/pccp

The gas-phase side chain–side chain (SC–SC) interaction and possible proton transfer between glutamic acid (Glu) and arginine (Arg) residues are studied under low-temperature conditions in an overall neutral peptide. Conformation-specific IR spectra, obtained with the free electron laser FELIX, in combination with density functional theory (DFT) calculations, provide insight into the occurrence of intramolecular proton transfer and detailed information on the conformational preferences of the peptides Z-Glu-Ala$_n$-Arg-NHMe (n = 0,1,3). Low-energy structures are obtained using molecular dynamics simulations *via* the simulated annealing approach, resulting in three types of SC–SC interactions, in particular two types of pair-wise interactions and one bifurcated interaction. These low-energy structures are optimized and frequency calculations are performed using the B3LYP functional, for structural analysis, and the M05-2x functional, for relative energies, employing the 6-311+G(d,p) basis set. Comparison of experimental and computed spectra suggests that only a single conformation was present for each of the three peptides. Despite the increasing spacing between the Glu and Arg residues, the peptides have several types of interactions in common, in particular specific SC–SC and dispersion interactions between the Arg side chain and the phenyl ring of the Z-cap. Comparison with previous experiments on Ac-Glu-Ala-Phe-Ala-Arg-NHMe as well as molecular dynamics simulations further suggest that the pairwise interaction observed here is indeed energetically most favorable for short peptide sequences.

Introduction

Intramolecular interactions in small isolated peptides have been under extensive study over the past decade. Interactions such as hydrogen bonds and dispersion interactions dictate the 3-dimensional shape of proteins and peptides to a large extent.[1] By studying these interactions under isolated conditions, the intrinsic folding properties of the peptide can be disentangled from the effects of the environment,[2] *e.g.* solvent molecules, on the conformational structure of biomolecules.[3]

Another type of interaction influencing the folding structure of a peptide is the salt bridge interaction, a non-covalent bond between a positively and a negatively charged residue.[4] This zwitterionic interaction is extensively investigated in complexes of peptides with metal cations, mainly using infrared multiphoton dissociation,[5–9] but also with collision induced dissociation,[10] electron capture dissociation[11] and in theoretical studies.[12] The metal ion can either have a charge solvated (CS) interaction with the peptide, or it can form a salt bridge (SB) interaction. In the SB interaction, usually the carboxylic acid C-terminus of the peptide is deprotonated, forming the CO_2^- group. The studies address the relation between size/charge of the cation and the relative stability of SB and CS structures. Additionally, a gas-phase SB is observed in ArgArgH$^+$, where a SB is formed between the protonated side chain (SC) of arginine and the deprotonated C-terminus of the peptide.[13] However, salt bridges can also be formed between two oppositely charged residues within an overall neutral peptide.[14]

Here, the gas-phase salt bridge formation between two amino acid SC's, namely the basic arginine (Arg) and acidic glutamic acid (Glu) residue in an overall neutral system will be investigated focusing on possible SC–SC interactions. The interaction between the charged SC's of these amino acids is stronger than for any other pair of amino acids.[15] However, for this type of interaction to occur, intramolecular proton transfer is required, leading to the formation of a zwitterion, which is not trivial under isolated conditions. Under physiological conditions, polar solvent molecules and metal ions stabilize the positive and negative charges on the biomolecule[16,17] or facilitate intramolecular proton transfer by acting as a solvent bridge.[18,19] Obviously, this charge stabilizing effect is absent under isolated

[a] *Radboud University Nijmegen, Institute for Molecules and Materials, FELIX Facility, Toernooiveld 7, 6525 ED Nijmegen, The Netherlands. E-mail: a.rijs@science.ru.nl*

[b] *Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3cp52508b

conditions, resembling to some extent the situation in hydrophobic protein pockets (in the absence of water molecules). In such environments, ionized residues can be stabilized by the formation of salt bridges as well as by ion–dipole interactions with the protein.[20,21]

These salt bridge interactions are not only important for the overall structure of the peptide; salt bridge interactions are also present in active sites of proteins. The amino acid residues studied in this paper, Glu and Arg, can for instance be found in the active sites of F_0F_1-ATPase,[22] Ricin[23] and the human chloride intracellular channel (CLIC).[24] In these active sites, the amino acid residues need to be in a specific conformation to perform their task, to recognize and bind co-factors. In these three systems, the SC's of the Glu and Arg residues interact in a different manner, *i.e.* electrostatic interactions are formed between different atoms. Therefore, it is of interest to obtain a detailed understanding of the formation of these SC–SC interactions.

Formation of gas-phase zwitterions in neutral peptides has been observed upon binding of solvent molecules to amino acids and peptides.[16,25] For example, Blom *et al.* observed a transition from the canonical to the zwitterionic form of tryptophan upon addition of five water molecules to the isolated molecule.[26] Formation of a zwitterionic structure in a neutral peptide without solvent molecules attached was first observed in Ac-Glu-Ala-Phe-Ala-Arg-NHMe.[14] In this system, proton transfer occurs from the acidic carboxylic acid SC of Glu to the basic guanidine SC of Arg. However, we showed recently that the formation of a zwitterion is not trivial, for example it does not occur in Z-Arg-OH.[27] In this system a similar proton transfer would be possible from the C-terminal carboxylic acid of Arg to its guanidine SC. Nevertheless, this proton transfer is not observed due to competition with dispersion interactions and strain on the SC.

In the present paper, the formation of zwitterionic structures is investigated for Z-Glu-Ala$_n$-Arg-NHMe (n = 0,1,3, see Fig. 1). By increasing the number of alanine (Ala) residues between Glu and Arg, the backbone of the peptide becomes more flexible and the zwitterionic structure may be expected to be formed more easily. With Glu, one of the most acidic amino acids, and Arg, the most basic amino acid, proton transfer is expected to occur readily.[14] The Z-cap (Z = benzyloxycarbonyl) is incorporated as a UV-chromophore, enabling us to perform conformation-specific IR-UV ion-dip spectroscopy.

In addition to the understanding of zwitterion formation, gas-phase IR spectroscopy provides the resolution that allows us to study the SC–SC interactions in detail. By comparing experimental IR-UV ion dip spectra with computed spectra of low energy conformations, obtained *via* high-level quantum-chemical calculations, it is possible to distinguish between the different modes in which the SC's can interact. Additionally, the effect of the spacing between the Glu and Arg residues on their SC–SC interactions will be investigated here.

IR-UV ion dip spectra are obtained in the region between 1000 and 1800 cm^{-1}, which is found to be structurally informative. Of particular interest are the C=O stretch (Amide I) and NH bend (Amide II) vibrations, which are strongly dependent on the hydrogen bond environment. Additionally, the Amide I region provides insights into the occurrence of proton transfer. In combination with the fingerprint region (1000 – 1400 cm^{-1}), these vibrations give detailed information on the 3-dimensional structure of the peptide.

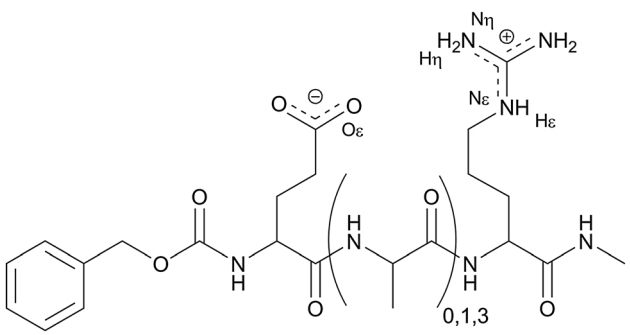

Fig. 1 Chemical structure of zwitterionic Z-Glu-Ala$_n$-Arg-NHMe. Atom labels used to describe the intramolecular interactions are: the C=O group of the Z-cap is O0, the NH and the C=O group of the ith residue is called N$_i$ and O$_i$, respectively. The NH group of the methylamide is N$_j$, where j equals the total number of residues +1.

Experimental section

Experimental methods

The samples Z-Glu-Arg-NHMe, Z-Glu-Ala-Arg-NHMe and Z-Glu-Ala$_3$-Arg-NHMe (>95% purity) were purchased from GL Biochem Ltd. (Shanghai, China) and were used without further purification. The samples were mixed with graphite powder and applied on a solid graphite bar. A pulsed near-IR laser (Polaris Pulsed Nd:YAG Laser System, New Wave research) with a pulse energy of about 1.5 mJ and a wavelength of 1064 nm was used to desorb the sample molecules from the graphite substrate as intact neutral molecules. The gas-phase molecules are entrained in a supersonic molecular beam of argon, produced by a pulsed valve (R.M. Jordan Co.) and a backing pressure of 3 bar.

The molecular beam travels through a skimmer about 10 cm downstream to enter a differentially pumped chamber housing a reflectron time-of-flight mass spectrometer (R.M. Jordan Co.). Here, the molecules interact with a UV beam produced by a pulsed Nd:YAG laser (either Innolas GmbH Spitlight 1200 or Spectra-Physics Quanta-Ray Lab Series) coupled to a frequency doubled dye laser (Radiant Dyes Narrowscan, laser dye: coumarin 153). The UV laser was operated at 10 Hz with typical pulse energies of 1–2 mJ. The molecules are 2-photon ionized *via* a (1 + 1) REMPI scheme.[28]

IR absorption spectra are obtained by employing IR-UV ion dip spectroscopy. In this technique, a constant ion signal is produced with a UV laser *via* the (1 + 1) REMPI scheme. The IR beam is spatially overlapped with the UV beam and precedes the UV pulse by ~200 ns. Whenever the IR laser excites a vibrational transition of the molecule of interest, the ground state is depleted and a dip in the ion signal is observed. By measuring the ion yield as a function of the IR wavelength, an IR ion-dip spectrum is obtained. To correct for long term UV power drifts and changing source conditions, alternating IR-on

and IR-off signals are measured by operating the IR laser at 5 Hz and the UV laser at 10 Hz. The IR radiation is produced by the Free Electron Laser for Infrared eXperiments (FELIX).[29]

Theoretical methods

Conformational searches were performed by applying the simulated annealing (SA) approach using the GROMACS4 package[30] and the amber99sb force field.[31] The maximum temperature used in the simulations was 1300 K; the simulation lengths were 20 ns with time steps of 2 fs. The temperature was lowered exponentially to 5 K in 20 ps. The structures at 5 K were stored. About 30 low-energy conformations found in the SA approach are optimized with the B3LYP functional[32] and the 6-31G(d,p) basis set using the Gaussian 09 program package.[33] From these structures, the 15 lowest energy structures are selected to be optimized using the B3LYP and M05-2x[34] functionals with the 6-311+G(d,p) basis set. For the largest system studied (Z-Glu-Ala_3-Arg-NHMe), a smaller basis set (6-311G(d,p)) has been employed to save computation time. Fig. S1 in the ESI† shows that omission of the diffuse functions has little impact on the calculated IR absorption spectra.

Additional low temperature SA simulations were performed on promising structures that showed good agreement with the experimental IR spectrum in the Amide I and the Amide II region, but poor agreement in the fingerprint region ($\sim$1000–1450 cm^{-1}). By running low temperature SA simulations, small conformational changes are induced, while keeping the hydrogen bond network unchanged.

The obtained conformations are optimized using DFT (B3LYP) and DFT-D (M05-2x) functionals. Frequency calculations are performed for the optimized structures using the B3LYP/6-311+G(d,p) level of theory. To compare predicted and experimental spectra, calculated frequencies were scaled by 0.9845 for the 6-311+G(d,p) basis set (0.9793 for the 6-311G(d,p) basis set) and convoluted with a Gaussian line shape with a FWHM of 15 cm^{-1}. The frequency calculations of M05-2x are used to determine the zero point energies (ZPE) and the Gibbs free energies at 300 K (ΔG). It was shown that the room temperature ΔG can play a crucial role in determining the gas-phase conformational population at low temperatures.[35] Although B3LYP has been proven to be a suitable functional for frequency calculations on peptide systems, its performance in predicting relative energies of the conformational structures is not as accurate. Our previous paper[27] addressing the conformational preferences of Z-Glu-OH and Z-Arg-OH shows that the DFT-D functionals do find the correct energetic ordering of the various structures, comparable with the ordering predicted by the higher correlation method MP2.

Neutral Glu, neutral Arg and the Z-cap are not present in the amber99sb force field, and therefore had to be implemented manually. The new amino acid residues have been implemented in a manner consistent with how the rest of the force field was originally derived.[36] Briefly, the molecular structure was first constructed in Chemcraft[37] with an alanine residue on both the N- and the C-terminus to simulate a peptide environment. The atomic charges were determined by the AM1-BCC charge method implemented in AmberTools.[38] Parameters for new bond lengths, bond angles, dihedrals and impropers were copied from existing analogues in the Ambersb99 force field.

The potential energy surface (PES) of peptides becomes very complex with increasing peptide size.[39] Even for the smallest system studied in this paper, Z-Glu-Arg-NHMe, the PES is too complex for all local minima to be explored. Not only does it possess three peptide bonds, the SC's of Glu and Arg are very flexible and contain many degrees of freedom. As a consequence, the conformational search performed here is necessarily incomplete. However, we are confident that the structures assigned to the experimental spectrum are accurate, based on the agreement between theoretical and experimental spectra, as well as on the computed relative free energies of the assigned structures, which are found to be the lowest in our conformational search.

REMPI spectra

REMPI spectra of Z-Glu-Arg-NHMe, Z-Glu-Ala-Arg-NHMe and Z-Glu-Ala_3-Arg-NHMe are presented in Fig. 2. The REMPI spectrum of Z-Glu-Arg-NHMe shows a partially resolved doublet, with peaks at 37 578 cm^{-1} and 37 591 cm^{-1}. However, the REMPI spectrum of Z-Glu-Ala-Arg-NHMe is broad and only shows three unresolved absorption bands (at 37 567 cm^{-1}, 37 592 cm^{-1} and 37 613 cm^{-1}), while the REMPI spectrum of Z-Glu-Ala_3-Arg-NHMe features only one very broad and unresolved band at around 37 550 cm^{-1}.

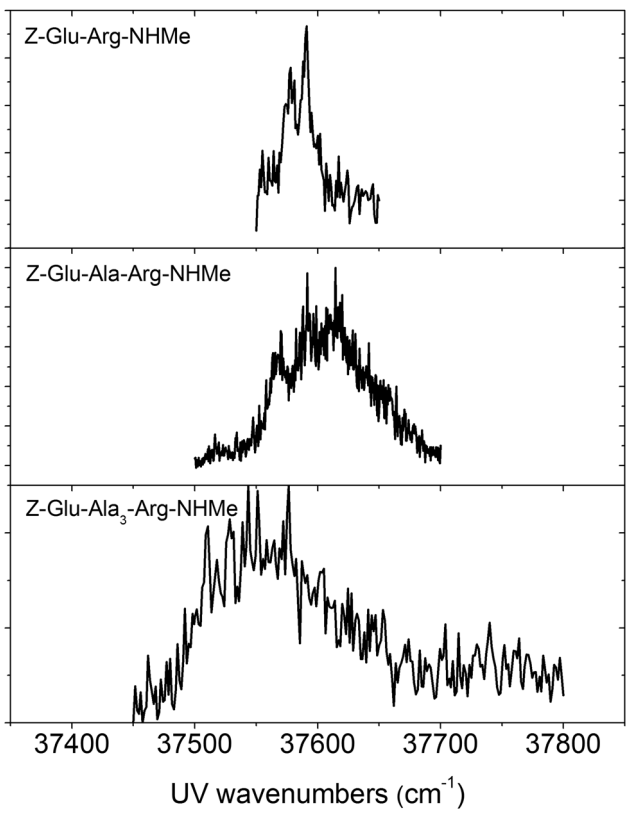

Fig. 2 REMPI spectra of Z-Glu-Arg-NHMe (top trace), Z-Glu-Ala-Arg-NHMe (middle trace) and Z-Glu-Ala_3-Arg-NHMe (bottom trace).

With increasing peptide size, the REMPI spectra become broader, which is probably due to incomplete jet-cooling of the increasing degrees of freedom of the molecules. As observed previously, REMPI spectra of larger and complex systems are often unresolved, as for instance observed for rotaxanes[40] and for various peptides.[41–43] Additionally, the floppiness of the arginine side chain might reduce the lifetime of the excited state due to increased IVR rates, and consequently result in broadened REMPI spectra of Arg containing peptides. However, despite the unresolved REMPI spectra, the IR spectra recorded by IR-UV double resonance show well-resolved absorption bands.

Z-Glu-Arg-NHMe

Conformational search

For the conformational search on Z-Glu-Arg-NHMe, three different classes of input structures were generated: zwitterionic structures, with the Glu and Arg SC deprotonated and protonated, respectively, and two types of non-zwitterionic structures. The non-zwitterionic structures differ in the tautomeric form of the Arg SC, as shown in Fig. 3.

For zwitterionic Z-Glu-Arg-NHMe, three different types of SC–SC interactions are found; two different types of pairwise interactions (type A and B) and one bifurcated interaction (type C) as shown in Fig. 4. In the type A interaction, one of the Oε atoms of glutamate is hydrogen bonded to Hε of Arg and the other Oε atom is hydrogen bonded to an Arg Hη atom (see Fig. 1 for atom labels). The type B interaction has a similar pairwise interaction; however, here the two Oε atoms are hydrogen bonded to Hη1 and Hη2 of the Arg SC. In the type C interaction, one of the Oε atoms has a bifurcated interaction with Hη1 and Hη2, while the second Oε is hydrogen bonded to amide NH groups of the peptide backbone.

Z-Glu-Arg-NHMe contains three peptide bonds, so that various backbone–backbone (BB–BB) interactions are possible. A commonly found backbone interaction is a C5 interaction involving the C=O and NH groups of Glu. Another C5 interaction is possible, between the C=O and the NH group of the Arg residue. This interaction is encountered only once, in conformer ER_C8, forming an extended β-sheet structure with two C5 interactions. Two C7 interactions are possible, one connecting the C=O group of the Z-cap with the amide NH group of Arg and one connecting the amide C=O of Glu with the NH group of the C-terminal cap. In the zwitterionic structures, only the latter C7 interaction is observed. The only zwitterionic structure where both types of C7 interactions are present is ER_Z16. Double C7 interactions are frequently observed in the non-zwitterionic structures. Generally, non-zwitterionic structures exhibit more BB–BB interactions, probably due to lower interference of the non-charged SC's. Finally a C10 BB–BB interaction is possible between the C=O group of the Z-cap and the NH group of the C-terminal NHMe cap. This BB–BB interaction can be classified as a "β-turn (type I)" according to the Φ and Ψ dihedral angles, although slightly distorted; it is observed in the four lowest energy structures. An overview of the observed BB–BB interactions is shown in Fig. 5.

In addition to the pure BB–BB and SC–SC interactions, also BB–SC interactions have been found in our conformational search. These BB–SC interactions appear more frequently in zwitterionic than in canonical and tautomeric structures. The locally charged SC's of glutamate and Arg can interact with the backbone C=O and NH groups. The non-zwitterionic structures, with neutral SC's, also interact with these backbone groups, although not as strongly. The lone pair on the Nε atom readily forms hydrogen bonds with the backbone in the tautomeric structures, however, interactions between the lone pair and the backbone are not observed for canonical structures. Additionally, the phenyl ring of the Z-cap gives rise to the occurrence of dispersion interactions. As shown in our previous paper, these interactions are important in small biomolecules.[27]

A total of 38 structures have been optimized and their vibrational frequencies have been computed using the B3LYP

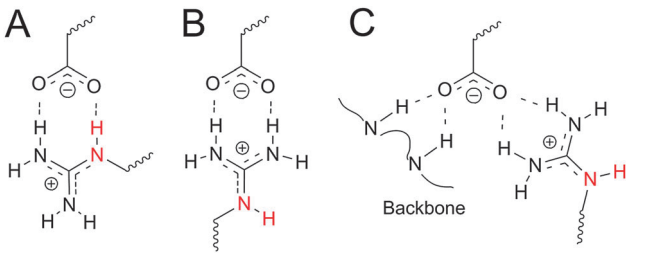

Fig. 4 Three different types of SC–SC interactions found in the conformational search for zwitterionic Z-Glu-Ala$_n$-Arg-NHMe.

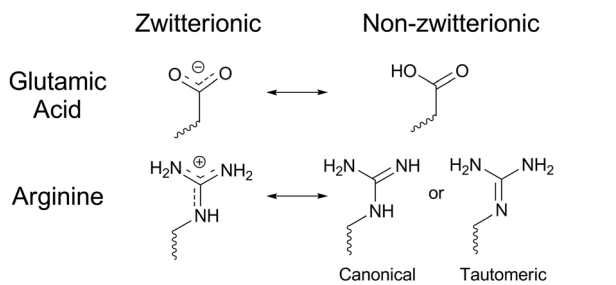

Fig. 3 Molecular structures for the Glu and Arg SC's in their zwitterionic and non-zwitterionic states.

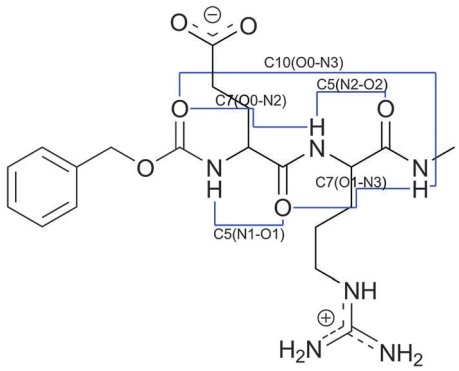

Fig. 5 Backbone–backbone interactions observed in the conformational search of Z-Glu-Arg-NHMe.

Table 1 ZPE-corrected energies (ZPE), Gibbs free energies at 300 K (ΔG) and intramolecular interactions for the optimized structures of Z-Glu-Arg-NHMe. The employed basis set for the M05-2x and B3LYP functionals is 6-311+G(d,p)

	M05-2x		B3LYP		Interactions			
	ZPE	Gibbs	ZPE	Gibbs	SC – SC	Disp. Int.	BB–BB	BB–SC
ER_Z1	0.00	0.00	4.06	5.79	A	NH_2	C10(O0–N3)	N1–O
ER_Z2	3.76	3.24	5.89	5.01	C	N2	C10(O0–N3)	N1–O + O1–H_2N
ER_C1	3.81	2.00	5.41	5.03	OH–NH	NH_2	C10(O0–N3)	N1–O
ER_Z3	4.08	4.06	5.82	8.22	A	2 × NH_2	C10(O0–N3)	N1–O^1 + N2–O^2 + O2–H_2N
ER_Z4	4.28	0.54	0.00	0.00	A		C5(N1–O1) + C7(O1–N3)	N2–O + O2–H_2N
ER_C2	6.59	4.41	5.88	3.74	OH – NH	NH_2	C7(O1–N3)	N1–O
ER_Z10	7.84	5.23	4.64	3.82	A		C5(N1–O1)	N2–O^1 + N3–O^2
ER_Z12	9.08	8.07	10.02	9.55	A		C10(O0–N3)	N2–O
ER_C3	9.75	6.06	8.37	7.56	OH – NH	N2	C10(O0–N3)	N1–O
ER_T1	9.98	7.92	10.83	10.76	NH_2–OH	NH_2	C5(N1–O1)	N1–O + O3–HO + N2–Nε
ER_T2	11.12	8.16	7.66	7.29	O–$(H_2N)_2$		C7(O1–N3)	N2–Nε + N3–Nε + O0–HO
ER_C4	11.41	9.52	9.76	9.41			C5(N1–O1) + C7(O1–N3)	O0–HO
ER_T3	12.80	10.21	11.96	10.69			C10(O0–N3)	N1–O + N3–Nε + O2–HO
ER_C5	12.90	9.87	10.48	9.84	OH–NH		C5(N1–O1) + C7(O1–N3)	N2–O
ER_Z16	20.33	18.60	16.20	16.03	A		C7(O0–N2) + C7(O1–N3)	

Energies are given in kcal mol^{-1}.

and M05-2x functionals employed with the 6-311+G(d,p) basis set. Zero-point corrected energies, 300 K Gibbs free energies as well as intramolecular interactions for selected structures are shown in Table 1. Energetics and intramolecular interactions for all structures are shown in the ESI,† Table S1. The corresponding structures and IR spectra are presented in Fig. S2 and S3 (ESI†). 16 out of the 38 computed structures show proton transfer. Of the remaining 22 non-zwitterionic structures, 12 feature the Arg SC in its canonical form and 10 in its tautomeric form. The structures exhibiting proton transfer are typically lower in energy, the first non-zwitterionic structure is found at a relative energy of 3.8 kcal mol^{-1} with respect to the lowest-energy proton transfer structure using the M05-2x functional. However, close examination of the lowest energy canonical structures reveals that there is a strong interaction between the carboxylic acid OH group and the nitrogen lone pair of the NηH group of the Arg SC. This interaction is found for conformers ER_C1, ER_C2, ER_C3, ER_C5, ER_C6 and ER_C7. The stability of these structures is questionable; due to the high acidity and basicity of the Glu and Arg SC, respectively, it is expected that proton transfer readily occurs when both groups are in close proximity. This is confirmed by the observation that the three lowest energy canonical structures are almost identical to zwitterionic ones, only differing in the position of the proton. ER_C1 is almost identical to ER_Z1, ER_C2 to ER_Z6 and ER_C3 to ER_Z2. In all three cases, the zwitterionic structures are lower in energy by an amount ranging from 2.2 to 6.0 kcal mol^{-1}. ER_T1 is the lowest-energy non-zwitterionic structure not exhibiting this interaction, and lies about 10 kcal mol^{-1} higher in energy using the M05-2x functional. The lowest canonical structure without the SC–SC interaction is 11.4 kcal mol^{-1} higher in energy than the lowest energy zwitterionic structure. Based on the computed energies of the non-proton transferred structures, observing these structures in the experiment is not expected.

The lowest energy proton transfer structure, ER_Z1, exhibits an A-type interaction (see Fig. 4a), a backbone C10 interaction and dispersion interaction between the $N\eta H_2$ group of Arg and the phenyl ring of the Z-cap (see Fig. 11). In addition, a backbone NH group interacts with the glutamate SC.

IR spectrum

IR-UV double resonance spectra of Z-Glu-Arg-NHMe are recorded with the UV laser frequency fixed on the most intense peak in the REMPI spectrum at 37 591 cm^{-1}. The experimental IR absorption spectrum in the 1000–1850 cm^{-1} region is shown as the black trace in the top panel of Fig. 6. Four dominant features are observed. According to the calculations, the feature at around 1700 cm^{-1} in the Amide I region consists of six vibrational modes. Three of these vibrations originate from the three backbone C═O stretch modes. In the case of a zwitterionic structure, the remaining modes originate from two $N\eta H_2$ scissors and an amine NεH bend vibration, while for the canonical structures the remaining three modes are the result of a carboxylic acid C═O stretch of the Glu SC, Cζ═Nη stretch and $N\eta H_2$ scissor modes of the Arg SC. For the non-zwitterionic tautomeric structure, the feature in the Amide I region consists of seven vibrations; three backbone C═O stretch modes, one carboxylic acid C═O stretch vibration of the Glu SC, two $N\eta H_2$ scissor modes and one Cζ═Nε stretching mode of the Arg SC.

The Amide I region provides a first glimpse at the occurrence of proton transfer. A free carboxylic acid C═O stretch vibration is usually observed at around 1780 cm^{-1} and can be red shifted by ~30 cm^{-1} due to hydrogen bonding. The absence of a strong absorption band in this region suggests that proton transfer has occurred in Z-Glu-Arg-NHMe. However, a full conformational assignment is needed to confirm that this proton transfer indeed has taken place. The Amide II band, ranging from 1450 cm^{-1} to 1580 cm^{-1}, consists mainly of backbone NH bend vibrations and CH_2 scissor vibrations. The Amide III

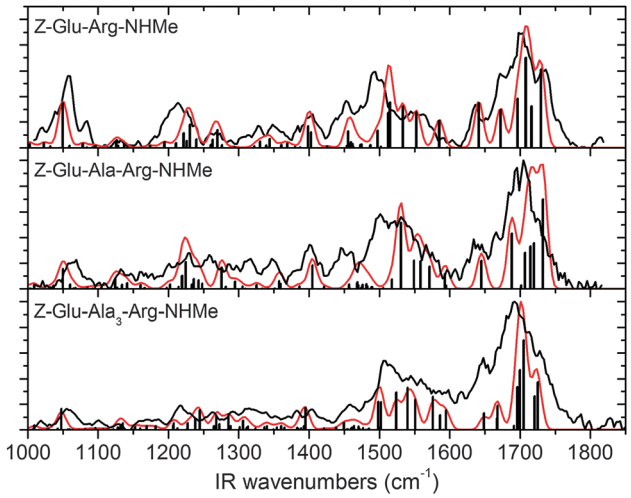

Fig. 6 IR-UV ion dip spectra of Z-Glu-Arg-NHMe (top panel), Z-Glu-Ala-Arg-NHMe (middle panel) and Z-Glu-Ala_3-Arg-NHMe (bottom panel) in black. The red traces are the theoretical spectra of the assigned structures. Z-Glu-Arg-NHMe is assigned to ER_Z1, Z-Glu-Ala-Arg-NHMe is assigned to EAR_Z1 and Z-Glu-Ala_3-Arg-NHMe is assigned to EA3R_Z1. UV wavelength used for Z-Glu-Arg-NHMe is 37 591 cm^{-1}, for Z-Glu-Ala-Arg-NHMe at various wavelengths (see the text) and for Z-Glu-Ala_3-Arg-NHMe 37 530 cm^{-1}.

band observed at around 1200 cm^{-1} can be mainly attributed to backbone NH bend vibrations and CH_2 twisting modes. The intense band at 1050 cm^{-1} is due to the C–O(C) stretch vibration in the ester group of the Z-cap. The IR absorption at around 1400 cm^{-1} is also diagnostically important, as here the carboxylate symmetric O–C–O stretch vibration of the zwitterionic structures is expected.[14]

Structural assignment

Fig. 7 shows the experimental IR band of Z-Glu-Arg-NHMe in the Amide I region, which clearly consists of multiple vibrational modes. This feature is deconvoluted using six Gaussian functions, each corresponding to one of the vibrational modes in the Amide I region, to assist in the structural assignment. The FWHM of the Gaussian functions is set at 20 cm^{-1}, to match the observed bandwidth, which is the consequence of the IR laser bandwidth and insufficient cooling, which is typical for Arg containing peptides.[14,27] The floppy SC of Arg is not easily frozen into a single conformation, making cooling less efficient. The conformational assignment of the experimental spectrum is supported by comparing the fitted frequencies and intensities of the deconvoluted Gaussians with the computed frequencies of the low energy structures (see Table S2 of the ESI†). The sum of the six Gaussian functions is normalized, to compare the experimental values with the computed values. The non-zwitterionic tautomeric structures are excluded in this analysis, since careful examination of the computed spectra of the tautomeric structures in Fig. S3 of the ESI† (ER_T structures) clearly shows that these structures cannot be responsible for the observed experimental spectrum. They all show a blue shifted carboxylic acid C═O stretch vibration, which is not observed in the experiment. Moreover, the calculated energies

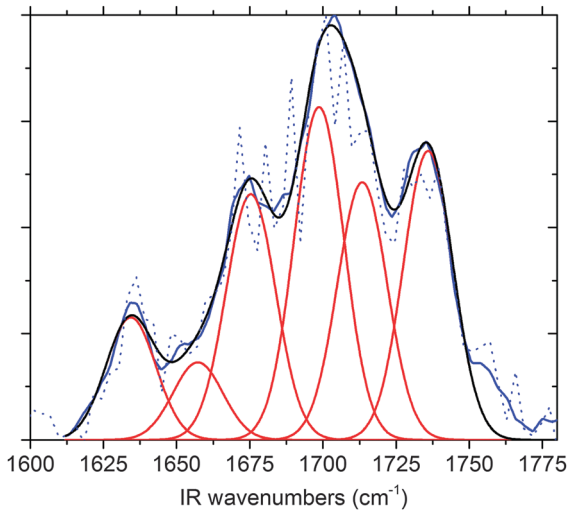

Fig. 7 Deconvolution of the experimental Amide I feature of Z-Glu-Arg-NHMe (blue) into six Gaussian functions (red). The sum of the six Gaussian functions is shown in black. The raw, unsmoothed experimental spectrum is shown in the dotted, blue trace.

for these structures are at least 9.98 kcal mol^{-1} above those of the zwitterionic structures.

The computed spectra of five selected structures are shown in Fig. 8. The selection is based on deviations between the computed frequency and intensity of the selected structures with the frequencies and intensities of the Gaussian functions obtained by the deconvolution of the Amide I absorption band (see ESI†, Table S2). Based on these values, as well as the energetics, ER_Z1 is suggested to provide the best match. Also in the lower frequency region of the spectrum, the agreement between theory and experiment is satisfactory. The main bands are nicely reproduced, although the Amide II (~1500 cm^{-1}) and Amide III (~1200 cm^{-1}) bands are slightly shifted. The peak at around 1400 cm^{-1} corresponds to the symmetric carboxylate O–C–O stretch vibration and is in good agreement

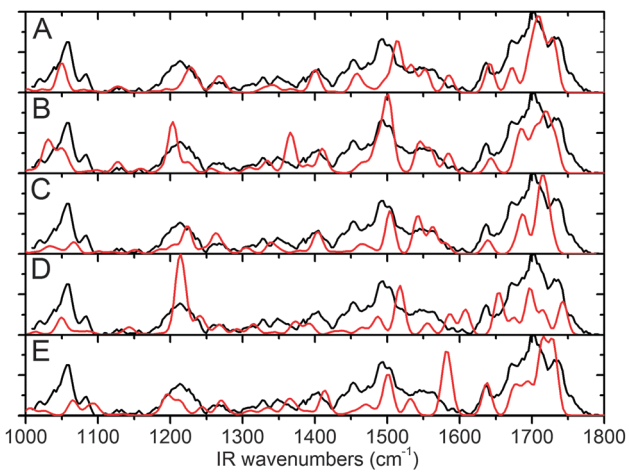

Fig. 8 IR-UV ion dip spectrum of Z-Glu-Arg-NHMe (black trace), and theoretical spectra of (A) ER_Z1, (B) ER_Z10, (C) ER_Z12, (D) ER_C5 and (E) ER_Z16. Theoretical traces are shown in red.

with experiment. Examining the remaining structures in Fig. 8 reveals that the computed spectra of ER_Z10, ER_Z12 and ER_C5 all deviate further from the experimental spectrum, so that we feel it is secure to exclude them. For example, ER_Z10 and ER_C5 exhibit a strong absorption at around 1200 cm^{-1}, while for ER_Z12 the regions at around 1050 cm^{-1} and 1550 cm^{-1} are in disagreement with experiment. We therefore assign the experimental spectrum of Z-Glu-Arg-NHMe to the lowest energy structure ER_Z1.

Z-Glu-Ala-Arg-NHMe

IR spectrum

The IR-UV ion dip spectra of Z-Glu-Ala-Arg-NHMe are recorded with the UV laser fixed at 37 526.5 cm^{-1}, 37 554 cm^{-1} and 37 618.5 cm^{-1}. All IR-UV spectra were identical, indicating that a single conformation is present in our molecular beam. The middle panel of Fig. 6 (black trace) shows the experimental spectrum of Z-Glu-Ala-Arg-NHMe in the 1000–1850 cm^{-1} range. The Amide I and Amide II bands are broad and hardly show any resolved structure. Between 1100–1400 cm^{-1} weaker absorptions are observed with partly resolved structure.

Conformational search

An elaborate conformational search was performed for Z-Glu-Ala-Arg-NHMe. The resulting structures are similar to those obtained for Z-Glu-Arg-NHMe. As proton transfer already occurred in Z-Glu-Arg-NHMe, the emphasis in the conformational search was on the zwitterionic structures, although also canonical and tautomeric structures were included. A total of 43 structures were optimized using DFT, of which 32 are zwitterionic. From the remaining 11 structures, the Arg SC was in its tautomeric form in 7 of them. The lowest energy non-zwitterionic structure is 12 kcal mol^{-1} higher than the lowest energy zwitterionic structure.

The zwitterionic structures can be classified according to their SC–SC interactions, as is shown in Fig. 4. In the conformational search of Z-Glu-Ala-Arg-NHMe an additional type of a bifurcated interaction is found, which will be referred to as SC–SC interaction C*. As compared to type C, in this type of C* interaction the Glu Oε-atom interacts with one $N\eta H_2$ group and with the NεH group of Arg, instead of with both $N\eta H_2$ groups. The energies and intramolecular interactions of selected structures are shown in Table 2. A complete list of the energies and intramolecular interactions of the computed structures is given in the ESI,† Table S3. In addition, all structures and calculated infrared spectra are shown in Fig. S5 and S6 of the ESI.†

Structural assignment

Based on the computed thermochemistry, observing non-zwitterionic structures experimentally is not expected. Inspection of the computed spectra of these structures (Fig. S6, ESI†) reveals that none of these structures are responsible for the experimental spectrum. All structures either exhibit a carboxylic acid C=O stretch vibration between 1760–1800 cm^{-1} or an intense absorption at 1200 cm^{-1}, which are not observed in the experiment. We therefore conclude that proton transfer has occurred in Z-Glu-Ala-Arg-NHMe, as in Z-Glu-Arg-NHMe.

The calculated spectrum of the lowest energy structure, EAR_Z1, is shown in the top panel of Fig. 9. There is a fairly good agreement between the calculated and the experimental spectrum, *i.e.* the Amide I peak is slightly blue shifted and the peak at 1500 cm^{-1} corresponding to one of the Amide II vibrations is absent. The fingerprint region, however, is reproduced fairly well. For example, the deviation of the carboxylate symmetric O–C–O stretch at 1400 cm^{-1} is only 3 cm^{-1}. The experimental spectrum of Z-Glu-Ala-Arg-NHMe is therefore assigned to EAR_Z1 based on both the computed energetics and on the peak positions of most bands in the IR spectrum.

As shown in the middle trace of Fig. 2, the REMPI spectrum of Z-Glu-Ala-Arg-NHMe is largely unresolved. The conformational selectivity arises when two different conformations can be selectively ionized, *i.e.* when the $S_1 \leftarrow S_0$ transitions of the conformations are sufficiently separated in frequency. Conformational selectivity for

Table 2 ZPE-corrected energies (ZPE), Gibbs free energies at 300 K (ΔG) and intramolecular interactions for the optimized structures of Z-Glu-Ala-Arg-NHMe. The employed basis set for the M05-2x and B3LYP functionals is 6-311+G(d,p)

	M05-2x		B3LYP		Interactions			
	ZPE	Gibbs	ZPE	Gibbs	SC–SC	Disp. Int.	BB–BB	BB–SC
EAR_Z1	0.00	0.00	2.56	5.86	A	2 × NH_2	C10(O0–N3)	N1–O^1 + N2–O^2
EAR_Z2	0.73	0.47	3.47	6.28	B	NH_2 + Nε	C10(O0–N3)	N1–O^1 + N2–O^2
EAR_Z3	1.11	2.54	4.07	7.73	B	NH_2	C10(O0–N3) + C10(O1–N4)	O3–(H_2N + Hε)
EAR_Z4	1.85	1.31	2.67	5.61	A	2 × NH_2	C7(O2–N4)	N1–O^1 + N2–O^2
EAR_Z5	1.90	2.09	5.16	8.20	A	2 × NH_2	C10(O1–N4)	N1–O^1 + N2–O^1 + N3–O^1 + O4–H_2N
EAR_Z6	2.13	1.96	4.65	7.88	A	2 × NH_2	C10(O1–N4)	N3–O + O3–H_2N
EAR_Z7	2.85	2.82	3.46	7.12	A	2 × NH_2	C7(O2–N4)	N1–O^1 + N3–O^2 + O3–H_2N
EAR_Z18	6.79	5.86	7.45	8.46	A	NH_2	C5(N1–O1)	N2–O^1 + N3–O^1 + N4–O^1 + O0–H_2N
EAR_T1	11.96	11.01	12.17	14.04	O–H_2N		C7(O2–N4)	O0–$(NH_2)_2$ + N2–H_2N + N3–Nε + O3–HO
EAR_T2	18.03	15.54	14.53	14.59			C7(O0–N2) + C7(O2–N4)	N1–O + O1–$(H_2N)^1$ + O3–$(H_2N)^2$ + N3–Nε
EAR_C1	18.30	17.31	15.73	17.00			C5(N1–O1)	N3–NηH + N4–NηH + O1–H_2N + O2–HO
EAR_C2	19.72	16.17	14.36	13.66			C7(O1–N3) + C7(O2–N4)	N1–O + O1–Hε + O3–HO

Energies are given in kcal mol^{-1}.

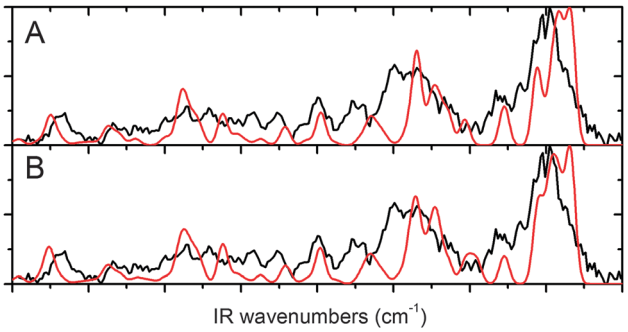

Fig. 9 IR-UV ion dip spectrum of Z-Glu-Ala-Arg-NHMe (black trace), and theoretical spectra of (a) EAR_Z1 and (b) 60–40 mixture of EAR_Z1 and EAR_Z2.

Table 3 Frequencies and intensities of the seven Gaussians for the fit in the Amide I region for the experimental spectrum, and for the match structure of Z-Glu-Ala-Arg-NHMe

Fit experimental Amide I region		EAR_Z1 (assigned structure)		
ν (cm^{-1})	I	ν (cm^{-1})	I	Assignment
1636.7	0.33	1645.2	0.29	NH_2 scissor (phen)
1664.1	0.34	1688.3	0.57	C═O stretch Arg
1688.3	0.71	1701.5	0.03	NH imine + NH_2 (COO^-)
1688.3	0.01	1706.9	0.37	C═O stretch Glu
1707.7	0.85	1714.8	0.43	C═O stretch Ala
1729.7	0.41	1719.9	0.47	NH imine + NH_2 (COO^-)
1751.3	0.13	1732.5	0.92	C═O stretch Z-cap

Z-Glu-Ala-Arg-NHMe is therefore not warranted. In addition, the temperature of the molecule can be relatively high due to insufficient cooling, as already mentioned in the REMPI section above. At higher temperatures conformational barriers can be crossed making it possible for two or more conformations to coexist in the gas phase, so that the observed IR spectrum may be a combination of more conformations.

The bottom panel of Fig. 9 shows the computed spectrum of EAR_Z1 and EAR_Z2 mixed in a 60/40 ratio together with the experimental spectrum. These structures are almost identical, differing only in the orientation of the guanidinium SC of Arg. This results in different SC interactions, being type A for EAR_Z1 and type B for EAR_Z2. In the supersonic cooling process both interactions may be formed and then cool into the same overall structure with a C10(O0-N3) interaction, a dispersion interaction and two SC–BB interaction between the oxygen atoms of the Glu SC and the NH groups of the backbone.

For Z-Glu-Arg-NHMe, deconvoluting the Amide I region assuming six Gaussian-shaped bands supported the structural assignment. The same approach was used for Z-Glu-Ala-Arg-NHMe, however, fitting the Amide I region to seven vibrations turned out to be very difficult. The results of the fit are shown in Fig. S7 (ESI†) and listed in Table 3. At first glance, the agreement between the experimental fit and the computed vibrations of the assigned structure is poor. The deconvoluted band at 1664.1 cm^{-1} is absent in the EAR_Z1 computed spectrum. Moreover, the deconvolution yields an absorption band at 1751.3 cm^{-1}, which is the result of the tail on the blue side of the experimental spectrum. The other vibrations, however, show a good agreement. The deconvoluted band at 1688.1 cm^{-1} corresponds to the C═O stretch vibration of the Arg backbone. The computed vibrations at 1706.9 cm^{-1}, 1714.8 cm^{-1} and 1719.9 cm^{-1} merge and are responsible for the experimental band at 1707.7 cm^{-1}. Lastly, the C═O stretch of the Z-cap at 1732.5 cm^{-1} is reproduced nicely at 1729.7 cm^{-1}, although the computed intensity for this vibration is much higher.

Z-Glu-Ala-Ala-Ala-Arg-NHMe

The IR absorption spectrum of Z-Glu-Ala$_3$-Arg-NHMe is shown in the bottom panel of Fig. 6 (black trace) recorded with the UV excitation wavelength at 37 530 cm^{-1}. The spectrum shows two intense bands, the Amide I and Amide II bands. Both bands show little structure and consist of a large number of overlapping transitions. The remainder of the spectrum shows a large number of weak broad absorptions.

Because of the size of Z-Glu-Ala$_3$-Arg-NHMe, the smaller basis set 6-311G(d,p) is employed here instead of 6-311+G(d,p). Although diffuse functions are considered to be important for DFT calculations[44] and for describing hydrogen bonding,[45] the effect of omitting diffuse functions from the basis set on the computed IR spectra 1000–1850 cm^{-1} regime is minimal for Z-Glu-OH, Z-Arg-OH, Z-Glu-Arg-NHMe and Z-Glu-Ala-Arg-NHMe, as shown in Fig. S1 (ESI†).

A total of 18 structures are optimized for Z-Glu-Ala$_3$-Arg-NHMe, of which 11 structures exhibit proton transfer, covering all three forms of SC–SC interactions, three structures are canonical non-zwitterionic and four are tautomeric non-zwitterionic. The type C* interaction observed in Z-Glu-Ala-Arg-NHMe is not found for Z-Glu-Ala$_3$-Arg-NHMe. The relative free energies and intramolecular interactions of selected structures are tabulated in Table 4, a complete list is shown in the ESI,† Table S4. In Fig. S8 (ESI†) all optimized structures are shown. The lowest non-zwitterionic structure found is 14.8 kcal mol^{-1} higher in energy than the lowest energy structure found with proton transfer. It is thus again unlikely that structures without proton transfer are observed in the gas-phase experiments.

Examination of the theoretical spectra of the optimized non-zwitterionic structures (Fig. S9, ESI†) confirms this expectation. Nearly all non-zwitterionic structures exhibit strong absorptions at around 1200 cm^{-1}, which are not observed in the experiment. Moreover, the Amide II band is predicted to be stronger than experimentally observed. The only non-zwitterionic structure lacking a strong absorption at around 1200 cm^{-1} is EA3R_T2. Overall, this structure has a decent match with the experimental spectrum, except for the band at 1050 cm^{-1}. However, the calculated energy of this structure is 17.7 kcal mol^{-1} higher than the lowest energy zwitterionic structure, which makes it unlikely that this structure is present in the molecular beam.

Of the 11 computed zwitterionic structures, four have a paired "A-type" interaction, four a paired B-type and two a bifurcated C-type interaction (see Fig. 4). The lowest energy structure, at both the B3LYP and M05-2x levels, has a paired

Table 4 ZPE-corrected energies (ZPE), Gibbs free energies at 300 K (ΔG) and intramolecular interactions for the optimized structures of Z-Glu-Ala-Ala-Ala-Arg-NHMe. The employed basis set for the M05-2x and B3LYP functionals is 6-311G(d,p)

	M05-2x		B3LYP		Interactions			
	ZPE	Gibbs	ZPE	Gibbs	SC – SC	Disp. Int.	BB – BB	BB – SC
EA3R_Z1	0.00	0.00	0.00	0.00	A	2 × NH_2	C10(O0–N3) + C7(O2–N4) + C5(N5–O5)	O4–H_2N + N1–O^1 + N6–O^2
EA3R_Z2	0.07	0.89	3.71	2.32	B	NH_2 + Nε	C7(O1–N3) + C10(O2–N5) + C10(O3–N6)	N1–O^1 + N4–O^2 + O4– H_2N
EA3R_Z3	0.36	0.49	3.93	1.15	A	NH_2	C5(N1–O1) + C7(O1–N3) + C10(O1–N4) + C10(O2–N5) + C10(O3–N6)	N2–O + O0–(H_2N)^1 + O4–(H_2N)^2
EA3R_Z4	3.18	2.03	3.69	−0.75	A		C5(N1–O1) + C7(O1–N3) + C10(O2–N5) + C10(O3–N6)	N2–O + O0–(H_2N)^1 + O4–(H_2N)^2
EA3R_Z5	3.35	3.52	5.89	3.93	B	NH_2 + Nε	C7(O1–N3) + C10(O2–N5) + C13(O2–N6)	N1–O^1 + N4–O^2 + O4– H_2N
EA3R_T1	14.84	13.95	18.09	15.00			C10(O0–N3) + C11(N2–O4) + C7(O2–N4)	N5–Nε + N6–Nε + N1–O + O5–HO
EA3R_C1	16.69	15.59	16.41	11.82			C7(O0–N2) + C10(O1–N4)	N3–O + N6–NηH + O4–(Hε + H_2N) + O5–OH
EA3R_T2	17.73	14.62	18.28	13.54			C10(O0–N3) + C14(N2–O5) + C7(O2–N4) + C5(N5–O5)	N1–O + N6–Nε
EA3R_C2	33.79	28.46	24.48	16.24			C7(O0–N2) + C7(N3–O5) + C7(N4–O6)	O1– H_2N + N3–NηH

Energies are given in kcal mol^{-1}.

A-type interaction and a dispersion interaction between both NηH$_2$ groups and the phenyl ring of the Z-cap (EA3R_Z1). According to the M05-2x functional results, at least two additional structures exist within less than 1 kcal mol^{-1}. Structure EA3R_Z2 has a paired B-type interaction and a dispersion interaction, while conformer EA3R_Z3 has a paired A-type interaction, but lacks a dispersion interaction.

Of the zwitterionic structures, EA3R_Z7, EA3R_Z8, EA3R_Z9, EA3R_Z10 and EAR3_Z11 can be excluded because they predict strong absorptions at around 1200 cm^{-1} (see Fig. S8, ESI†), which are not observed experimentally. The six remaining structures are presented in Fig. 10. The lowest energy structure, EA3R_Z1 (Fig. 10a), shows a reasonably good match in the Amide I and II region. The bands in the experimental spectrum appear broader than those for the smaller systems, which can be an indication of insufficient cooling. The computed spectrum is convoluted with a Gaussian line shape function with a FWHM of 15 cm^{-1} to better reflect the actual bandwidth of the IR source FELIX. The temperature of small peptides is often estimated to be around 15 K,[46] and it is expected that the temperature of Z-Glu-Ala$_3$-Arg-NHMe in the molecular beam is substantially higher. At higher temperatures, conformational dynamics can become important, so that it is again not warranted that the experimental absorption spectrum is conformation specific. Of the two structures within 1 kcal mol^{-1} of the global minimum, it appears unlikely that EA3R_Z2 contributes to the experimental spectrum since its computed spectrum contains a moderately intense absorption at 1250 cm^{-1} and a C=O stretch vibration at 1758 cm^{-1}. However, a minor contribution of EA3R_Z3 cannot be excluded.

EA3R_Z1 and EA3R_Z6 (Fig. 10f) differ only in the orientation of the guanidinium SC. In EA3R_Z1 the gaunidinium group is orientated towards the phenyl ring, forming a strong dispersion interaction, which is not the case for EA3R_Z6. This results in very similar calculated spectra. However, the energetics clearly predict EA3R_Z1 to be more favorable (3.8 kcal mol^{-1}), which is due to stabilization by the dispersion interaction.

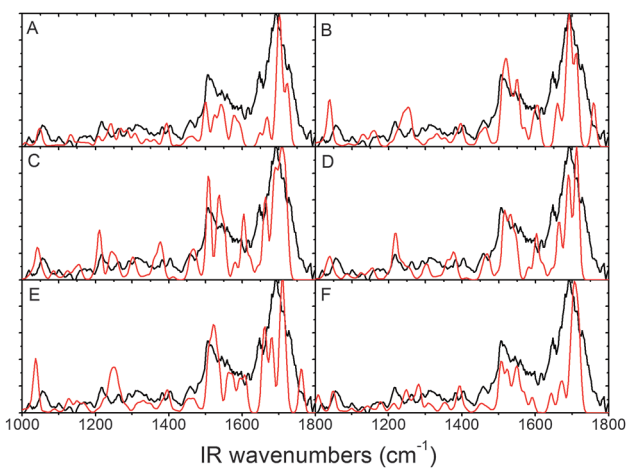

Fig. 10 IR-UV ion dip spectrum of Z-Glu-Ala-Ala-Ala-Arg-NHMe (black trace), and theoretical spectra of (A) EA3R_Z1, (B) EA3R_Z2, (C) EA3R_Z3, (D) EA3R_Z4, (E) EA3R_Z5 and (F) EA3R_Z6. Theoretical traces are shown in red.

Discussion

The assigned structures for Z-Glu-Arg-NHMe, Z-Glu-Ala-Arg-NHMe and Z-Glu-Ala$_3$-Arg-NHMe are shown in Fig. 11a. There are some striking similarities between the structures. They all show a paired A-type SC–SC interaction, where one Glu Oε atom is hydrogen-bonded to an Arg Hη atom and the other Oε atom is hydrogen-bonded to the Arg Hε atom. Moreover, they all exhibit a dispersion interaction between the guanidinium SC and the phenyl ring of the Z-cap, although the exact form of the interaction differs between the systems. While in Z-Glu-Arg-NHMe only one NH_2 group participates in the dispersion interaction, in Z-Glu-Ala$_3$-Arg-NHMe both NH_2 groups participate. Z-Glu-Ala-Arg-NHMe forms an intermediate case with one

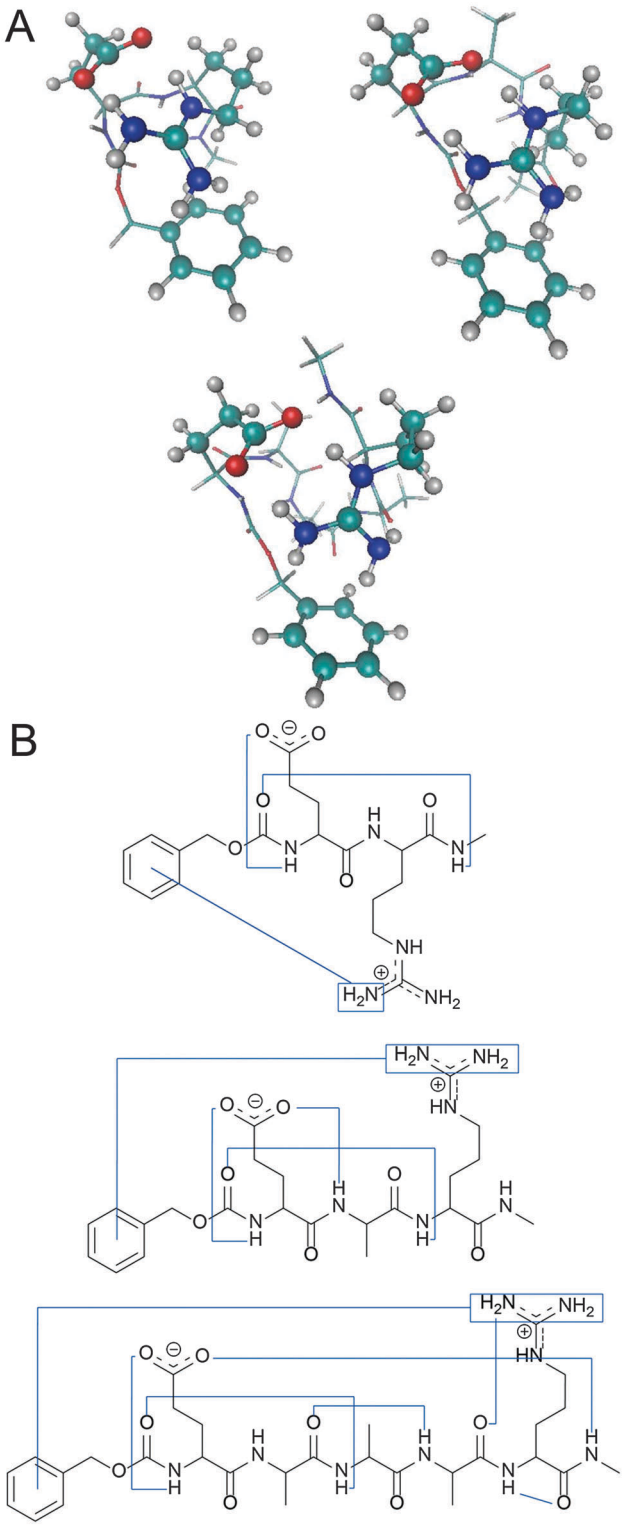

Fig. 11 (a) Assigned structures for Z-Glu-Ala$_n$-Arg-NHMe (n = 0,1,3), highlighting the SC–SC interaction and dispersion interaction. (b) Molecular structure and intramolecular interactions for Z-Glu-Ala$_n$-Arg-NHMe (n = 0, 1, 3).

of the NH_2 groups having a much stronger interaction than the other. This suggests that with increasing backbone length, the increased flexibility of the backbone allows the molecule to find the optimal geometry for maximizing this interaction energy.

Besides the identical SC–SC interaction, there are other interactions that are observed in all three molecules (see Fig. 11b). All systems show a C10 interaction between the C═O group of the Z-cap and the NH group of the third amino acid residue (C10(O0-N3)). Hence, the backbone folding appears not to depend on the identity of the participating residues. The Ψ- and Φ-dihedral angles of the first two residues in all three molecules lie very close to the typical values of a β-turn (type I), except for the Ψ-dihedral angle of the second residue (typical β-turn (type I) values are 0° *versus* 17° for Z-Glu-Arg-NHMe, 31° for Z-Glu-Ala-Arg-NHMe and 46° for Z-Glu-Ala$_3$-Arg-NHMe). This deviation is probably the result of the interacting SC's. Moreover, in Z-Glu-Ala-Arg-NHMe there is a strong interaction between the backbone NH of Ala and one of the Glu side-chain oxygen atoms (N2-O). In Z-Glu-Ala$_3$-Arg-NHMe, a deviation from the typical Ψ-dihedral angle value is probably caused by the additional C7(O2-N4) interaction. Furthermore, in all three structures, the Glu backbone NH group interacts with one of the oxygen atoms of its SC (see Fig. 11b).

The similarities between the molecules are also reflected in their IR spectra. The Amide I peak is located close to 1700 cm^{-1} for all three molecules, with a shoulder at 1670 cm^{-1}. The peak at ~1645 cm^{-1} also appears in all three spectra. In addition, the maximum of the Amide II band lies very close to 1500 cm^{-1} in all three spectra, with an unresolved tail to the blue and a peak at 1455 cm^{-1}. Lastly, all three molecules show an absorption at 1050 cm^{-1}, which corresponds to the asymmetric COC stretch vibration of the ester group.

The symmetric COO^- stretch vibration is very sensitive to its surroundings[47,48] and hence received special emphasis in the structural assignment procedure. For the assigned structures, this mode is found at 1399 cm^{-1} in Z-Glu-Arg-NHMe, at 1403 cm^{-1} in Z-Glu-Ala-Arg-NHMe and at 1393 cm^{-1} in Z-Glu-Ala$_3$-Arg-NHMe. For Z-Glu-Arg-NHMe and Z-Glu-Ala-Arg-NHMe, this band matches excellently with the experiment, suggesting that the assigned SC–SC interactions are indeed correct.

The occurrence of proton transfer in a similar system, capped Glu-Lys, has been investigated theoretically using Born–Oppenheimer Molecular Dynamics (BOMD) by Pluhařová *et al.*[25] They did not observe proton transfer in this peptide under isolated conditions. However, the addition of a single water molecule triggered zwitterionization in this peptide. This is in contrast with the results in this paper, where proton transfer was already observed in isolated Z-Glu-Arg-NHMe. This deviation is possibly the result of different pK_a values of the Arg and Lys SC, being 12.5 and 10.8, respectively.

Okur *et al.* and Sugita *et al.* performed molecular dynamics simulations on peptides containing glutamate and Arg. Replica Exchange Molecular Dynamics (REMD) on Ac-Arg-Ala-Ala-Glu-NH_2[49] identified the A-type interaction in agreement with our gas-phase experiments. In contrast, REMD simulations in ref. 50 found a "B-type" interaction,[50] perhaps as a result of the much larger size of the system studied, *i.e.* a 7-residue spacing between glutamate and Arg. Although the B-type interaction

provides more stabilization than the A-type interaction, it induces more strain in the backbone so that the overall energy of these structures is higher for the short peptides investigated here. The energy difference between the lowest energy A-type interaction structure and lowest energy B-type interaction structure decreases with increasing spacing between Glu and Arg, going from 4.37 kcal mol^{-1} in Z-Glu-Arg-NHMe to 0.73 kcal mol^{-1} in Z-Glu-Ala-Arg-NHMe to almost isoenergetic in Z-Glu-Ala_3-Arg-NHMe.

For Ac-Glu-Ala-Phe-Ala-Arg-NHMe, an A-type interaction was also found.[16] This also suggests that the SC–SC interactions of our assigned structures are indeed the one present in the experiment. However, placing the phenyl group in the center of the peptide, in contrast to the terminal position in the present study, yields a 3_{10} helical structure, whereas this study shows a mixture of C10, C7 and C5 backbone interactions. The central position of the phenyl group may be at the origin of this structural difference. In the former, a dispersion interaction between the Arg SC and the phenyl ring is difficult to form due to steric restrictions, while for the latter this interaction can form without much strain.

Dean *et al.* have explored the conformational preferences for Z-Gly_n-OH (n = 1,3,5) and Z-Gly_5-NHMe.[51] This allows us to investigate the effect of the charged residues on the backbone conformation. It appears that the charged SC's give rise to a more compact BB conformation. In addition, the dispersion interaction between the guanidinium SC and the phenyl ring disrupts the ''ring-like'' structure observed for Z-Gly_5-NHMe.

In principle, the same number of BB interactions is possible for Z-Gly_3-OH and Z-Glu-Arg-NHMe. However, the uncapped C-terminal OH group of Z-Gly_3-OH has a strong interaction with the phenyl ring, which cannot be formed in Z-Glu-Arg-NHMe. Moreover, the dispersion interaction with the Arg SC in Z-Glu-Arg-NHMe cannot be formed in Z-Gly_3-OH. These interactions have a large effect on the backbone arrangement, so that comparing these structures is not useful.

Finally, the gas phase SC–SC interactions between Glu and Arg found here are compared with SC–SC interactions observed in the biological systems Ricin,[23] ATPase[22] and the human chloride intracellular channel (CLIC).[24] The interaction between Glu and Arg is different in each of these proteins. In ATPase, a bifurcated interaction is found between one of the Glu Oε atoms and the Arg Hε and Hη atoms. This is the type C* interaction, which is not experimentally observed for the peptides studied here. In Ricin, the same type of interaction is found as in EA3R_Z11, with one Glu Oε atom interacting with one Arg NH_2 group. The SC–SC interaction of Glu and Arg in CLIC is of A-type, hence similar to what is observed for the peptides investigated here.

Thus, in biological systems a variety of SC–SC interactions are found. Many other factors, apart from the stabilization energy of the SC–SC interaction, play a role in these systems, including the overall backbone conformation, other amino acid residues in the proximity, water molecules and possibly metal ions. Therefore, to mimic a specific SC–SC interaction in the computations, the influence of the biological environment on these local interactions should be included, which is also experimentally possible using the methods employed here.[26,52] Such studies are planned in the near future.

Conclusions

The intramolecular charge transfer between glutamic acid and arginine in isolated, overall neutral peptides has been studied to shed light on the factors contributing to the various SC–SC interactions present in active sites of proteins. Therefore, we have performed a conformational analysis of Z-Glu-Ala_n-Arg-NHMe (n = 0,1,3) using IR-UV ion dip spectroscopy. For all three molecules, only a single conformation was identified in the gas phase. The assigned structures were the lowest energy structures found in the conformational search and optimized with the M05-2x functional. The three assigned structures all exhibit proton transfer and a paired A-type interaction, a dispersion interaction between the Arg SC and the Z-cap and a C10 interaction involving the C=O group of the Z-cap and the NH group of the third amino acid residue.

The interactions between the SC's are identical to those observed previously for the capped pentapeptide EAFAR. Previously reported molecular dynamics simulations have also shown this type of interaction for comparable amino acid spacings. However, such simulations suggest a paired B-type interaction for peptides with a larger spacer (7 residues). The increased strain on the backbone for shorter spacers is suggested to prevent the formation of the paired B-type interaction. Further gas-phase experiments on systems with a more flexible backbone are planned to validate this hypothesis.

Acknowledgements

We gratefully acknowledge the expert assistance of the FELIX staff, in particular Drs A.F.G. van der Meer and B. Redlich. This work is part of the research program of FOM, which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). We acknowledge in particular the NWO Physical Sciences Division (EW) and SurfSara for the use of the LISA supercomputer facilities. J.O. thanks the Stichting Physica for support.

References

1 L. Stryer, J. L. Tymoczko and J. M. Berg, *Biochemistry*, Freeman, New York, 5th edn, International edition, 2002.
2 M. S. de Vries and P. Hobza, *Annu. Rev. Phys. Chem.*, 2007, **58**, 585–612.
3 M. F. Jarrold, *Annu. Rev. Phys. Chem.*, 2000, **51**, 179–207.
4 H. R. Bosshard, D. N. Marti and I. Jelesarov, *J. Mol. Recognit.*, 2004, **17**, 1–16.
5 M. Citir, C. S. Hinton, J. Oomens, J. D. Steill and P. B. Armentrout, *J. Phys. Chem. A*, 2012, **116**, 1532–1541.
6 R. C. Dunbar, A. C. Hopkinson, J. Oomens, C.-K. Siu, K. W. M. Siu, J. D. Steill, U. H. Verkerk and J. Zhao, *J. Phys. Chem. B*, 2009, **113**, 10403–10408.
7 R. C. Dunbar, N. C. Polfer and J. Oomens, *J. Am. Chem. Soc.*, 2007, **129**, 14562–14563.
8 R. C. Dunbar, J. D. Steill and J. Oomens, *Phys. Chem. Chem. Phys.*, 2010, **12**, 13383–13393.

9 J. S. Prell, T. G. Flick, J. Oomens, G. Berden and E. R. Williams, *J. Phys. Chem. A*, 2009, **114**, 854–860.
10 P. B. Armentrout, Y. Chen and M. T. Rodgers, *J. Phys. Chem. A*, 2012, **116**, 3989–3999.
11 T. Flick, W. Donald and E. Williams, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 193–201.
12 M. H. Khodabandeh, H. Reisi, M. D. Davari, K. Zare, M. Zahedi and G. Ohanessian, *ChemPhysChem*, 2013, **14**, 1733–1745.
13 J. S. Prell, J. T. O'Brien, J. D. Steill, J. Oomens and E. R. Williams, *J. Am. Chem. Soc.*, 2009, **131**, 11442–11449.
14 A. M. Rijs, G. Ohanessian, J. Oomens, G. Meijer, G. von Helden and I. Compagnon, *Angew. Chem., Int. Ed.*, 2010, **49**, 2332–2335.
15 K. Berka, R. A. Laskowski, P. Hobza and J. Vondrasek, *J. Chem. Theory Comput.*, 2010, **6**, 2191–2203.
16 T.-K. Hwang, G.-Y. Eom, M.-S. Choi, S.-W. Jang, J.-Y. Kim, S. Lee, Y. Lee and B. Kim, *J. Phys. Chem. B*, 2011, **115**, 10147–10153.
17 S. Im, S. W. Jang, S. Lee, Y. Lee and B. Kim, *J. Phys. Chem. A*, 2008, **112**, 9767–9770.
18 S. Lee, *Bull. Korean Chem. Soc.*, 2011, **32**, 1117–1124.
19 E. Vöhringer-Martinez and A. Toro-Labbé, *J. Comput. Chem.*, 2010, **31**, 2642–2649.
20 J. Bush and G. I. Makhatadze, *Proteins: Struct., Funct., Bioinf.*, 2011, **79**, 2027–2032.
21 S. Kumar and R. Nussinov, *J. Mol. Biol.*, 1999, **293**, 1241–1255.
22 M. Dittrich, S. Hayashi and K. Schulten, *Biophys. J.*, 2004, **87**, 2954–2967.
23 S. A. Weston, A. D. Tucker, D. R. Thatcher, D. J. Derbyshire and R. A. Pauptit, *J. Mol. Biol.*, 1994, **244**, 410–422.
24 S. J. Harrop, M. Z. DeMaere, W. D. Fairlie, T. Reztsova, S. M. Valenzuela, M. Mazzanti, R. Tonini, M. R. Qiu, L. Jankova, K. Warton, A. R. Bauskin, W. M. Wu, S. Pankhurst, T. J. Campbell, S. N. Breit and P. M. G. Curmi, *J. Biol. Chem.*, 2001, **276**, 44993–45000.
25 E. Pluharova, O. Marsalek, B. Schmidt and P. Jungwirth, *J. Chem. Phys.*, 2012, **137**, 185101–185108.
26 M. N. Blom, I. Compagnon, N. C. Polfer, G. von Helden, G. Meijer, S. Suhai, B. Paizs and J. Oomens, *J. Phys. Chem. A*, 2007, **111**, 7309–7316.
27 S. Jaeqx, W. Du, E. J. Meijer, J. Oomens and A. M. Rijs, *J. Phys. Chem. A*, 2012, **117**, 1216–1227.
28 P. M. Johnson and C. E. Otis, *Annu. Rev. Phys. Chem.*, 1981, **32**, 139–157.
29 D. Oepts, A. F. G. van der Meer and P. W. van Amersfoort, *Infrared Phys. Technol.*, 1995, **36**, 297–308.
30 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
31 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.
32 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
33 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. R. J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, H. P. H. A. F. I. X. Li, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. D. A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, in Gaussian, Inc., Wallingford CT2009.
34 Y. Zhao, N. E. Schultz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2006, **2**, 364–382.
35 E. Gloaguen, B. de Courcy, J. P. Piquemal, J. Pilme, O. Parisel, R. Pollet, H. S. Biswal, F. Piuzzi, B. Tardivel, M. Broquier and M. Mons, *J. Am. Chem. Soc.*, 2010, **132**, 11860–11863.
36 A. Howard, B. Ross, AMBER User Manual, Appendix C100: Parameter Development.
37 Chemcraft, http://www.chemcraftprog.com.
38 D. A. Case, T. A. Darden, I. T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, San Francisco, 2010.
39 T. S. Zwier, *J. Phys. Chem. A*, 2006, **110**, 4133–4150.
40 A. M. Rijs, B. O. Crews, M. S. de Vries, J. S. Hannam, D. A. Leigh, M. Fanti, F. Zerbetto and W. J. Buma, *Angew. Chem., Int. Ed.*, 2008, **47**, 3174–3179.
41 I. Hunig and K. Kleinermanns, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2650–2658.
42 T. D. Vaden, S. A. N. Gowers and L. C. Snoek, *Phys. Chem. Chem. Phys.*, 2009, **11**, 5843–5850.
43 A. Abo-Riziq, J. E. Bushnell, B. Crews, M. Callahan, L. Grace and M. S. de Vries, *Chem. Phys. Lett.*, 2006, **431**, 227–230.
44 B. J. Lynch, Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2003, **107**, 1384–1388.
45 E. Papajak, J. Zheng, X. Xu, H. R. Leverentz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2011, **7**, 3027–3034.
46 G. Meijer, M. S. Devries, H. E. Hunziker and H. R. Wendt, *Appl. Phys. B*, 1990, **51**, 395–403.
47 W. Lewandowski, M. Kalinowska and H. Lewandowska, *J. Inorg. Biochem.*, 2005, **99**, 1407–1423.
48 J. Oomens and J. D. Steill, *J. Phys. Chem. A*, 2008, **112**, 3281–3283.
49 A. Okur, L. Wickstrom and C. Simmerling, *J. Chem. Theory Comput.*, 2008, **4**, 488–498.
50 Y. Sugita and Y. Okamoto, *Biophys*, 2005, **88**, 3180–3190.
51 J. C. Dean, E. G. Buchanan and T. S. Zwier, *J. Am. Chem. Soc.*, 2012, **134**, 17186–17201.
52 M. Cirtog, A. M. Rijs, Y. Loquais, V. Brenner, B. Tardivel, E. Gloaguen and M. Mons, *J. Phys. Chem. Lett.*, 2012, **3**, 3307–3311.

