

Oxoanion Binding by Guanidiniocarbonylpyrrole Cations in Water: A Combined DFT and MD Investigation

Daive Moiani,^[a] Carlo Cavallotti,*^[a] Antonino Famulari,^[a] and Carsten Schmuck*^[b]

Abstract: Structures and properties of nonbonding interactions involving guanidinium-functionalized hosts and carboxylate substrates were investigated by a combination of ab initio and molecular dynamics approaches. The systems under study are on one hand intended to be a model of the arginine–anion bond, so often observed in proteins and nucleic acids, and on the other to provide an opportunity to investigate the influence of molecular structure on the formation of supramolecular complexes in detail. Use of DFT calculations, including extended basis sets and implicit water treatment, allowed us to determine minimum-

energy structures and binding enthalpies that compared well with experimental data. Intermolecular forces were found to be mostly due to electrostatic interactions through three hydrogen bonds, one of which is bifurcate, and are sufficiently strong to induce a conformational change in the ligand consisting of a rotation of about 180° around the guanidiniocarbonylpyrrole axis. Free binding energies of the com-

plexes were evaluated through MD simulations performed in the presence of explicit water molecules by use of the molecular mechanics Poisson–Boltzmann solvent accessible surface area (MM-PBSA) and linear interaction energy (LIE) approaches. LIE energies were in quantitative agreement with experimental data. A detailed analysis of the MD simulations revealed that the complexes cannot be described in terms of a single binding structure, but that they are characterized by a significant internal mobility responsible for several low-energy metastable structures.

Keywords: density functional calculations • molecular dynamics • molecular recognition • receptors • supramolecular chemistry

Introduction


The design and development of artificial hosts that will selectively bind a given target under physiological conditions—that is, in water—still remains challenging despite the progress that has been achieved in recent years. In general, multiple weak binding interactions such as hydrogen bonds (H bonds), electrostatic, and hydrophobic interactions have to be combined within a single recognition motif to allow

substrate complexation even in a protic environment (“Gulliver effect”).^[1] However, it is not possible to determine the individual contributions of the various interactions for substrate binding experimentally, as only the overall affinity can be determined. One possible means to learn more about the contributions of the individual binding interactions, at least in a semiquantitative way, is to compare systematically varying series of closely related binding motifs.^[2,3] We have used this approach to study knock-out analogues of a self-complementary zwitterion, showing strong dimerization interactions even in water, both experimentally^[4] and theoretically.^[5] Such studies help us to acquire better understanding of the underlying molecular recognition process, also with respect to being able to improve the binding motifs further.

The synthesis of complexes involving guanidinium-functionalized hosts and the study of their properties have been the subject of several investigations in recent years.^[6] In the past, however, most artificial model studies were restricted to organic solvents. In fact, simple ion pairs between guanidinium cations and oxoanions are normally stable only in solvents of low polarity. In aqueous solutions, for example, the competing solvation of both donor and acceptor sites by

[a] D. Moiani, Prof. Dr. C. Cavallotti, Dr. A. Famulari
Politecnico di Milano
Dipartimento di Chimica, Materiali
e Ingegneria Chimica “G. Natta”/CIIRCO
Via Mancinelli 7, 20131 Milano (Italy)
E-mail: carlo.cavallotti@polimi.it

[b] Prof. Dr. C. Schmuck
Institut für Organische Chemie, Universität Würzburg
Am Hubland, 97074 Würzburg (Germany)
Fax: (+49)931-8884-626
E-mail: schmuck@chemie.uni-wuerzburg.de

 Supporting information for this article is available on the WWW under <http://www.chemeurj.org/> or from the author.

individual solvent molecules significantly decreases ion pair stability. For nature this normally does not pose any problems, because of the rather hydrophobic interiors of the proteins where such dedicated pairing usually occurs or because of clustering of several such small interactions. For chemical receptors, however, the weakness of simple ion pairs in polar solvents represents a severe limitation both in terms of the interpretation of the results obtained from such in the study of model systems and, even more importantly, with regard to any dealing with potential application, which necessarily has to take place under physiological conditions.^[1]

In 2000, one of us reported that guanidiniocarbonylpyrrole cations can be used to bind *N*-acetyl amino acid carboxylates even in aqueous solvents.^[7] Since then, guanidiniocarbonylpyrroles have emerged as one of the most efficient hosts for oxoanions.^[8] A comparative experimental thermodynamic study of a series of six related hosts suggested that the affinity is mainly due to formation of a directed ion pair between the carboxylate and the guanidiniocarbonylpyrrole cation. However, complex affinity is then significantly increased by further additional H-bond donors such as the pyrrole NH or an amide NH. The combination of all these interactions increases the affinity by more than a factor of 30 relative to that of a simple guanidinium cation. It has not yet been possible to determine the structures of these complexes unambiguously by experiment. The structural analysis was therefore based on NMR results obtained with a more simple substrate—acetate—and on molecular mechanics (MM) calculations. We now report here a state-of-the-art theoretical study using a combined density functional theory (DFT) and molecular dynamics (MD) approach with explicit solvation treatment to analyze complex formation between these hosts and *N*-acetyl alanine carboxylate as the substrate (Figure 1). We evaluated different theoretical approaches, based on *ab initio* and molecular dynamics computational methodologies, for the determination of free binding energies of the supramolecular complexes in aqueous solvents. The results obtained from the calculations are not only in excellent agreement with the experimentally determined stabilities of the complexes, but also provide detailed information on the structures of the complexes, which

as it turned out are different from those initially anticipated on the basis of the earlier molecular modeling studies. Furthermore, MD simulations allowed us to explain the observed differences in stability between two closely related hosts: the glycine and the valine derivative **13** and **14**, respectively. Even though the structures of the complexes and hence the number of binding interactions are similar, the valine derivative is the better host because of its capability to form multiple metastable structures with the substrate, due to the difference of its side chain mobility from that of the glycine derivative.

Method and theoretical background

The theoretical investigation of the interactions of guanidiniocarbonylpyrrole hosts with carboxylate substrates is particularly challenging because of the high conformational mobility of the molecules involved, which is often a problem in van der Waals complexes, and because of difficulties in the accurate determination of binding energies and entropies. To address the problem suitably, different computational approaches were adopted, so that binding energies, minimum energy structures, and atomic charges were determined by density functional calculations, while free binding energies and the time-dependent conformational evolution of the complexes were studied by molecular dynamics simulations.

Starting structures of isolated molecules and complexes were obtained with the aid of two commercial codes distributed by Schrödinger, LigPrep, and Glide,^[9] which exploit minimum-energy search algorithms and docking protocols to determine molecular structures including solvent effect treatments. The most stable geometries were used in further calculations. Though the structures predicted by Glide and LigPrep have proved in many cases to match experimental data, their reliability rests on the accuracy of the MM force field employed. As the torsion and angular parameters pertaining to carbonyl and amino groups' interactions with pyrrole moieties are not well known, guess structures were tested by high-level calculations by density functional theory. DFT calculations were accomplished by use of the Becke 3 parameters^[10] and the Lee—Yang—Parr^[11] functionals for exchange and correlation energies as implemented in the Gaussian03 suite of programs.^[12] In order to achieve the best compromise between accuracy and computational efficiency, two basis sets were adopted.

A first refinement of the structures determined with the Schrödinger MM force fields was performed in the gas-phase approximation at the B3LYP/6-31G(d,p) level.^[13] Minima were checked by frequency calculations, and zero-point energy corrections were determined in the harmonic approximation. Gas-phase optimized structures were successively employed as starting points for geometry optimization in solution in water, which was the solvent used for most of the experiments. Water was modeled with the integral equation formalism polarizable continuum model (IEF-PCM) at a temperature of 300 K.^[14] Energies were refined by single-point energy calculations with adoption of the augmented

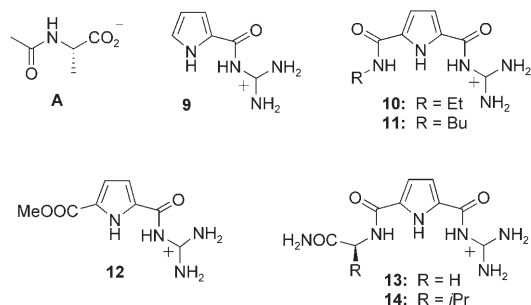


Figure 1. Structures of the substrate, *N*-acetyl alanine carboxylate **A**, and the cationic guanidiniocarbonylpyrrole hosts **9–14** under study (the numbering of the hosts was taken from the original experimental work [ref. [7]]).

correlation-consistent aug-cc-pVDZ basis set,^[15] with water effects being taken into account by use of the IEF-PCM implicit solvation model. It is worth pointing out here that, although DFT calculations are not able to determine all the nonbonding energy terms, such as hydrophobic contributions, correctly, these effects are indirectly reintroduced through the implicit solvation method adopted (vide infra).^[16]

Despite the high accuracies of the structures determined by DFT calculations, several minimum structures with similar stabilization energies can coexist on the intermolecular potential energy hypersurface of the complex when the interaction between molecules is not particularly strong. If energy barriers connecting minima are not too high, several conformations can be visited periodically as the temperature increases, so that the system oscillates between different structures. In order to test whether this might be the case for the system under investigation, we performed molecular dynamics simulations explicitly accounting for water effects. All MD simulations were accomplished by use of the ff03 force field^[17] as implemented in the Amber simulation package.^[18] For each compound, atomic-type assignments, connectivities, and interatomic distances and angles were collected in a library and assigned on the basis of similarity with atomic types defined in the ff03 force field. Much care was devoted to the assignment of the atomic charges, which were determined for each molecule by the same approach as adopted in the ff03 development. Electrostatic potentials (ESPs) were calculated at the B3LYP/6-31g(d) level, and charges were fitted by use of the RESP formalism.^[19] ESP values were determined on a grid of 1 point per Å² at 1.4, 1.6, 1.8, and 2.0 times the van der Waals radii and were then fitted to atomic charges through a two-step procedure. In the first step, a charge of +1 was assigned to the guanidinium cation, while the charge of the carboxylate was held fixed at -1, while in the second step, the charge equivalence for chemically equivalent atoms and the same charge as in step 1 were imposed.

Complexes and isolated compounds were solvated by use of explicit TIP3P water molecules^[20] with addition of a cubic solvent box with a lateral size of 20 Å. A dielectric constant of 1 was used for all simulations, and the non-bonded cutoff was set to 15 Å. All simulations were performed with use of periodic boundary conditions, according to which the system is partitioned into unit cells of equal size. Long-range electrostatic interactions were evaluated by the particle mesh Ewald method, so that a particle within a unit cell interacts with molecules in the same cell as well as with periodic images in neighboring cells. The advantages and disadvantages of Ewald boundary conditions have recently been discussed by Hunenberger and Cammon.^[21] The computational protocol adopted in MD simulations was as follows. Firstly, a 2000-cycle minimization, in which the compounds were restrained with a harmonic potential $k(\Delta x)^2$ (where Δx is the displacement and k is the force constant, held fixed at 500 kcal mol⁻¹ Å⁻²) was performed to remove the initial unfavorable close contacts arising from randomly placed sol-

vent molecules. This was followed by a second 3500-cycle minimization step without restraints. The temperature was then raised from 0 to 300 K by a simulated annealing of 20 ps at constant volume. To avoid wild fluctuations a weak restraint was imposed on the solute at this stage ($k=10$ kcal mol⁻¹ Å⁻²). In order to permit the water density to relax, after the heating of the system a 100 ps run was performed at constant pressure. Finally, molecular dynamics simulations were performed for a standard evolution time of 1 ns. The SHAKE algorithm was used for all covalent bonds involving hydrogen atoms, which allowed use of a time step of 2.0 fs. All simulations were performed at 300 K and constant atmospheric pressure. Electrostatic and van der Waals solute-solvent interaction energies were calculated by use of the Anal program of the Amber 8.0 computational suite,^[18] after re-imagining to the original cell water molecules moved to neighbor cells. Energies were averaged on sets of 125 snapshots for MM/GBSA and MM/PBSA simulations (see below), corresponding to a time span of 1 ns. A smaller time span of 0.5 ns was considered for LIE evaluations, to avoid loss of periodicity caused by diffusion of the solute outside of the simulation box, which can be relevant for the smallest molecules after re-imagining water molecules to the original simulation box. The convergence of the results with the simulation time was confirmed by performing 2 ns simulations for some complexes, which verified that the sums of the calculated electrostatic and van der Waals solute-solvent interaction energies had converged within ± 1 kcal mol⁻¹. Standard deviations of electrostatic and van der Waals energies were about 10% of the absolute value.

Binding free energies were determined by three different approaches often adopted in the literature: the linear interaction energy (LIE), the MM/GBSA (Generalized Born/Surface Area),^[22] and the MM/PBSA (Poisson-Boltzmann/Surface Area)^[23] methods. The advantages and disadvantages of these computational techniques have recently been reviewed by Brandsdal et al.^[24] Views of the optimized structures reported in the paper were produced with Molden 4.4,^[25] VMD 1.8.2,^[26] and the Maestro visualization tool of the Schrödinger suite of programs.

Results and Discussion

The results of the calculations are presented and discussed as follows. We first describe the computed minimum-energy structures of receptor, ligand, and complexes under investigation. We then report the corresponding DFT and MM energies. Finally, we discuss in detail the conformational evolution of two key complexes (**A-13** and **A-14**) to provide an explanation of the differences between measured and calculated free binding energies.

Structures

Receptors and ligand: In principle, each host can assume several conformations, arising from the rotation about the

pyrrole carbonyl amide bonds (Figure 2). The dipole of the amide NH groups can either point in the same direction as the bond dipole of pyrrole NH (“in conformation”) or in

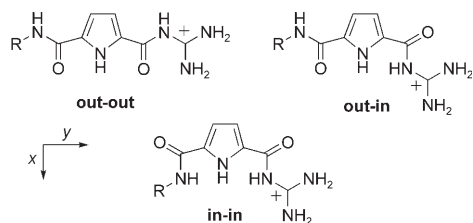


Figure 2. Conformations of guanidiniocarbonylpyrrole hosts.

the opposite direction (“out conformation”). The B3LYP/6-31g(d,p) geometries of anion **A** and of guanidinium compounds **9–14** (optimized structures including implicit water molecules treatment) are outlined in Figure 3, while those

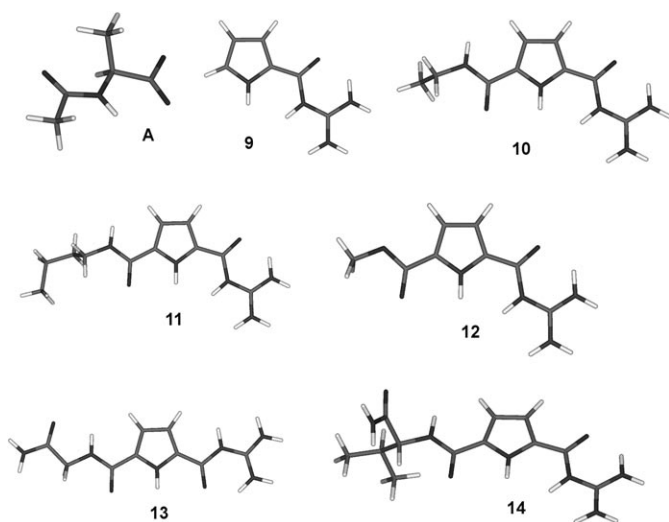


Figure 3. B3LYP/6-31g(d,p) structures of the receptors under investigation (implicit water environment is included).

of their conformers are reported as Supporting Information. In general, it can be observed that, because of dipole–dipole intramolecular interactions, **in** conformations are energetically disfavored with respect to the corresponding **out** conformations. As can be observed in Figure 2, to reduce the molecular dipole, **out-out** conformations possess lateral NH and CO bond dipoles aligned and coupled with the central pyrrole NH group, minimizing the total molecular dipole magnitude perpendicularly to its main chain (axis *X* in Figure 2). **Out-in** and **in-in** con-

formations lose this alignment so that the molecular dipole increases in the *X* direction, whereas it decreases, however, along the *Y* direction. The overall molecular dipole is a combination of these two effects. Dipoles and energetic analyses of the different isomers are summarized in Table 1.

As can be observed, more pronounced differences are observed in the gas phase. All the hosts present **out-in** conformations (guanidinium cation rotated inwards) that are less stable (relative energies of about 3.5–4 kcal mol⁻¹) than the corresponding **out-out** conformations. A water environment decreases the energy gap to 2–2.5 kcal mol⁻¹. **In-in** conformations—that is, the amide group in position 5 rotated inwards (see Figure 2)—are destabilized with respect to **out-out** conformations by about 18 kcal mol⁻¹ in the gas phase (host **10**) and by about 5 kcal mol⁻¹ in water. This conformation is energetically too unstable to be of any importance for complex formation. Therefore, the high-energy **in-in** structures are also not considered any further in successive calculations for the other hosts.

Complexes: Figure 4 shows B3LYP/6-31g(d,p) structures of the complexes formed by the interaction of the substrate **A** with guanidiniocarbonylpyrrole cation hosts **9–14** (see Figure 3) including implicit water environment. Key geometric parameters as well as RESP charges of atoms involved in H-bond formation (see Scheme 1) are reported in Table 2 and Table S1. Complete geometrical details are furnished as Supporting Information. Table 7 (below) gives Natural Charges calculated at the B3LYP/6-31g(d,p) level, while the corresponding ESP charges and complex charge transfer are presented in Table S2. Table S3 summarizes Mulliken B3LYP/6-31g(d,p) charges.

The geometrical parameters given in Table 2 show that H-bond 1 is the strongest interaction, because of the close contact distance between O and N atoms, well within the sums of their van der Waals radii (i.e., 3.07 Å). In contrast, interaction distances pertaining to H-bond 4 are longer than the sums of the van der Waals radii (see Table 2). Hydrogen bonds 2 and 3 are bifurcated interactions. As can be observed in Table 2, the H-bonds display longer interaction distances in water than in the gas phase, with the only exception being H-bond 4, which seems to be disfavored in the gas phase.

Table 1. Relative energies of **out-out** and **out-in** conformers of the ligands calculated at the indicated level of the theory. Energies are reported in kcal mol⁻¹ and are relative to **out-out** conformations (absolute minima). B3LYP/aug-cc-pVDZ dipole moments are also reported (gas-phase values in parentheses).

| | Enthalpy change out-out → out-in | | | Dipole moments [Debye] | |
|-----------|--|------------|-------------|------------------------|---------------|
| | Gas | Water | aug-cc-pVDZ | out-out | out-in |
| | 6-31g(d,p) | 6-31g(d,p) | | | |
| 9 | unstable | 2.45 | 2.50 | 10.0 (7.5) | 13.2 (9.4) |
| 10 | 3.27 | 2.00 | 2.26 | 20.9 (16.4) | 17.0 (18.4) |
| 11 | 3.16 | 1.98 | 2.07 | 25.0 (20.1) | 18.3 (15.5) |
| 12 | 4.44 | 2.17 | 2.43 | 19.6 (15.1) | 16.2 (12.5) |
| 13 | 3.69 | 2.09 | 2.30 | 20.4 (16.6) | 18.7 (15.1) |
| 14 | 3.71 | 2.30 | 2.35 | 26.1 (21.7) | 23.6 (19.6) |

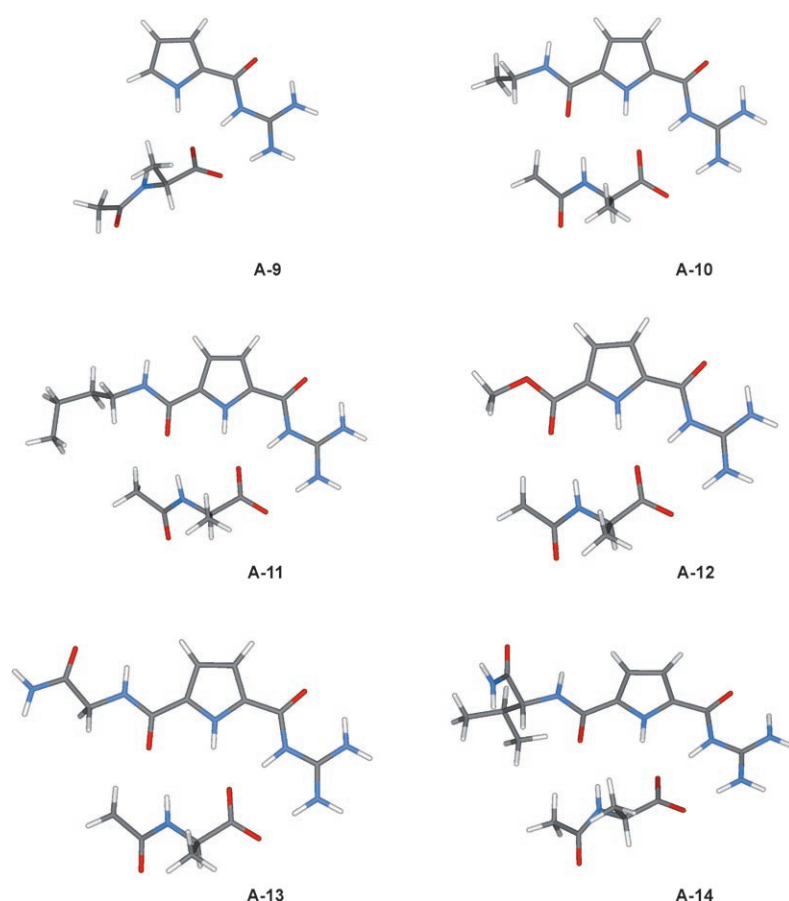
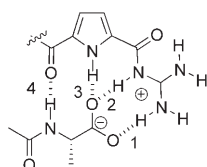


Figure 4. B3LYP/6-31g(d,p) structures of complexes under investigation; implicit water treatment is included.



Scheme 1. Numbering of atoms involved in hydrogen bond interactions.

Table 2. Hydrogen bond lengths and O–N distances^[a] in the complexes under investigation. The structures were optimized in the gas phase and in water solution at the B3LYP/6-31g(d,p) level.

| Distance [Å] | | 1 | | 2 | | 3 | | 4 | |
|--------------|---------|------|-------|------|-------|------|-------|------|-------|
| | | Gas | Water | Gas | Water | Gas | Water | Gas | Water |
| A-9 | NH...OC | 1.55 | 1.69 | 1.65 | 1.66 | 1.80 | 1.78 | – | – |
| | N–OC | 2.62 | 2.74 | 2.71 | 2.71 | 2.77 | 2.75 | – | – |
| A-10 | NH...OC | 1.49 | 1.69 | 1.57 | 1.66 | 1.77 | 1.78 | 2.20 | 2.08 |
| | N–OC | 2.59 | 2.74 | 2.64 | 2.71 | 2.73 | 2.75 | 3.17 | 3.06 |
| A-11 | NH...OC | 1.50 | 1.72 | 1.57 | 1.63 | 1.77 | 1.77 | 2.20 | 1.97 |
| | N–OC | 2.59 | 2.76 | 2.64 | 2.68 | 2.73 | 2.74 | 3.17 | 2.99 |
| A-12 | NH...OC | 1.50 | 1.70 | 1.58 | 1.67 | 1.75 | 1.75 | 2.29 | 2.14 |
| | N–OC | 2.59 | 2.74 | 2.65 | 2.72 | 2.72 | 2.73 | 3.25 | 3.12 |
| A-13 | NH...OC | 1.50 | 1.70 | 1.58 | 1.66 | 1.77 | 1.78 | 2.20 | 2.07 |
| | N–OC | 2.59 | 2.75 | 2.64 | 2.71 | 2.73 | 2.75 | 3.16 | 3.05 |
| A-14 | NH...OC | 1.50 | 1.71 | 1.58 | 1.65 | 1.77 | 1.78 | 2.21 | 2.09 |
| | N–OC | 2.59 | 2.76 | 2.64 | 2.70 | 2.73 | 2.75 | 3.18 | 3.08 |

[a] The sum of the van der Waals radii of N and O atoms is 3.07 Å.

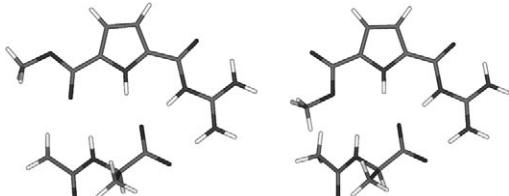
The optimized structures of the complexes display **out-in** conformations of the hosts, which are not the global minima of isolated molecules (energies higher than the **out-out** arrangements by about 2.5 kcal mol⁻¹), but they allow better binding interactions with the anions. In this conformation the guanidinium cation forms a bidentate hydrogen-bonded ion pair with the carboxylate group while the pyrrole NH group forms an additional H-bond to the inner oxygen of the carboxylate. From host **10** onwards, the carbonyl CO of the additional amide moiety accepts an H-bond from the amide NH of the substrate. In this respect all complex structures are very similar. Furthermore, hosts **11**, **13**, and **14** display additional nonbonding interactions with the alkyl side chains of the substrate. The high-level DFT calculations presented here show that previous model structures of the complexes have to be refined. Previously

calculated MM structures contained the hosts in the even less stable **in-in** host conformations, which was predicted to allow for a further H-bond between the NH of the additional amide group and the bound carboxylate.^[7] Our DFT calculations now show that the complexes are not made up of energetically disfavored **in-in** host conformations, but rather contain **out-in** arrangements. As already underlined, this is not the most stable conformation in the isolated molecules, but it allows additional H-bonds with the substrate, besides the ion pair formation weakening the energy of the complexes by about 2.5 kcal mol⁻¹. Also, instead of the amide NH forming a H-bond to the anion as predicted by the MM calculations, the DFT calculations show that the amide CO accepts a H-bond through the N-acetyl amide group of the substrate.

For host **12** an alternative complex structure was also calculated. If the ester group in the complex **A-12** in the **in-**

out arrangement is rotated by 180° (see Table 3), the alcohol oxygen could form an H-bond with the NH group of the substrate. As expected, the energy of the system is increased about 2.5–3 kcal mol⁻¹ because the carbonyl oxygen is a better donor (see Table 3).

Table 3. Minimized structures of complex **A-12** together with corresponding relative energies (data reported in kcal mol⁻¹ and referred to the absolute minimum). Method a: 6-31g(d,p). Method b: aug-cc-pVDZ

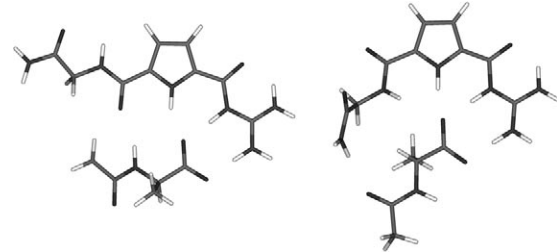


| | Gas | Water | Water | Gas | Water | Water |
|----------|-----|-------|-------|-------|-------|-------|
| method | a | a | b | a | a | b |
| <i>E</i> | 0.0 | 0.0 | 0.0 | +2.92 | +2.45 | +2.34 |

The complexes involving hosts **13** and **14** were studied in detail by B3LYP/aug-cc-pVDZ calculations including implicit water treatment. The calculated minimum-energy structures are shown in Tables 4 and 5, together with those of some of the most interesting stable isomers found during minimization steps. The experiments show that the complex containing host **14** is more stable than the corresponding complex with host **13** (**A-13**) system. A possible explanation is that the terminal NH of the amide group can form an additional H-bond with the carboxylate moiety. Because of the higher flexibility of the unsubstituted side arm in the case of host **13**, it was assumed that this interaction would be more important in the valine derivative **14** rather than **13**. However, the energy minimum found by DFT calculations does not show any interaction of the terminal amide group of the host with the carboxylate group of the substrate, so this alternative structure was found in the case of host **13** to be higher in energy by about 4–5 kcal mol⁻¹. For the valine derivative **14**, however, a similar structure possesses an even higher energy (+8–20 kcal mol⁻¹). The high energy differences between the conformers calculated with the 6-31g(d,p) and aug-cc-pVDZ basis sets can be ascribed mostly to repulsive interactions between the side arms, which are better described by the larger basis set. Hence, the DFT calculations show that the observed difference in complex stability between **13** and **14** cannot be ascribed to the formation of an additional H-bond by the terminal amide group as initially proposed.^[7]

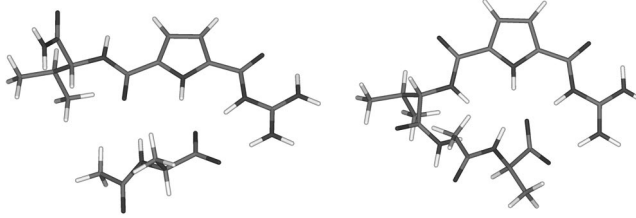
The analysis of partial atomic charges reported in Tables S1–S4 shows that the formation of the bond between the alanine anion and the guanidiniocarbonylpyrrole cation involves a significant intermolecular charge transfer, which is maximum for host **9**. As complex **A-9** was experimentally determined to be the least stable of those examined, as is discussed more in detail in the next section, this indicates that the amount of charge transfer cannot be regarded as a

Table 4. Minimized structures of complex **A-13** (together with corresponding relative energies (data reported in kcal mol⁻¹ and referred to the absolute minimum).



| | Gas | Water | Water | Gas | Water | Water |
|----------|-----|-------|-------|-------|-------|-------|
| method | a | a | b | a | a | b |
| <i>E</i> | 0.0 | 0.0 | 0.0 | +4.25 | +3.86 | +4.63 |

Table 5. Minimized structures of complex **A-14**, together with corresponding relative energies (data reported in kcal mol⁻¹ and referred to the absolute minimum)



| | Gas | Water | Water | Gas | Water | Water |
|----------|-----|-------|-------|------|-------|-------|
| Method | a | a | b | a | a | b |
| <i>E</i> | 0.0 | 0.0 | 0.0 | +8.3 | +11.5 | +20.4 |

measure of the binding energy. Moreover, the decrease in charge transfer between **A-9** and the other complexes confirms, as expected, that the increase in binding energy is determined by electrostatic and van der Waals interactions between the host side arms and the substrate and not by an increase of the binding energy between the guanidinium and the carboxylate groups. As already reported,^[27] while ESP and natural charges are consistent with the expected qualitative behavior, Mulliken population analysis is not suitable to describe intermolecular charge transfer adequately, not even from a qualitative standpoint as it predicts negative charge transfers for almost all complexes.

Energetic analysis

The structural analysis based on DFT calculations discussed in the previous section revealed some important new insights concerning anion binding by guanidiniocarbonylpyrrole cations. However, the analysis is so far based more or less solely on enthalpic contributions to the complex stability. We therefore performed energetic and entropic analyses of complex formation by the various hosts **9–14** to see how well the relative trend of the experimentally determined complex stabilities could be reproduced.

Method assessment: In order to test the accuracy of the adopted computational protocol, energies of compound **A9** were calculated both at the DFT level with use of three different basis sets and at the MP2 level with extension to complete basis set^[28] on the B3LYP/6-31g(d,p) structures optimized in water. Absolute energies and energy changes in solution, not including electrostatic terms (cavitation, dispersion, and repulsion energies), are reported in Table 6.

Table 6. Absolute (Hartree) and interaction energies (kcal mol⁻¹) for compound **A-9** calculated at two levels of theory (B3LYP and MP2) with the assumption of B3LYP/6-31g(d,p) geometries optimized in water with selected basis sets (a: 6-31g(d,p); b: cc-pVDZ; c: cc-pVTZ; d: aug-cc-pVDZ, e: aug-cc-pVTZ) and, in the case of MP2, extended to the CBS limit (CBSt: with use of the b and c basis sets; CBSat: with the augmented d and e basis sets).

| Theory | A | 9 | A-9 | ΔE [kcal mol ⁻¹] |
|---------------|------------|------------|-------------|--------------------------------------|
| B3LYP/a | -475.96721 | -528.22700 | -1004.22362 | -18.45 |
| B3LYP/d | -476.04921 | -528.28543 | -1004.34729 | -7.93 |
| B3LYP/e | -476.16926 | -528.41489 | -1004.59559 | -7.17 |
| MP2/b SCF | -473.24292 | -525.09918 | -998.35862 | -10.37 |
| Corr | -1.39107 | -1.61037 | -3.02055 | -11.99 |
| MP2/c SCF | -473.38624 | -525.24080 | -998.63272 | -3.56 |
| Corr | -1.74008 | -1.98206 | -3.74251 | -12.78 |
| MP2/d SCF | -473.28841 | -525.12815 | -998.41968 | -1.96 |
| Corr | -1.47846 | -1.68399 | -3.18302 | -12.90 |
| MP2/e SCF | -473.39736 | -525.24670 | -998.64578 | -1.08 |
| Corr | -1.77901 | -2.01642 | -3.81593 | -12.87 |
| MP2/CBSt SCF | -473.43454 | -525.28852 | -998.72509 | -1.27 |
| Corr | -1.98229 | -2.24001 | -4.24355 | -13.33 |
| MP2/CBSat SCF | -473.43407 | -525.28664 | -998.72197 | -0.79 |
| Corr | -1.98758 | -2.24712 | -4.25517 | -12.84 |

While B3LYP/6-31g(d,p) calculations seem to overestimate the complex stability considerably, the B3LYP/aug-cc-pVDZ and B3LYP/aug-cc-pVTZ interaction energies are very similar, differing only by 0.6 kcal mol⁻¹. This indicates that the B3LYP/aug-cc-pVDZ energies are already approaching the DFT basis set limit, and so they were used below to estimate binding energies of larger compounds. MP2 interaction energies were determined with basis sets of increasing size and extended to the basis set limit (CBS) by the procedure suggested by Truhlar.^[28] Two CBS estimates are given, the first based on a cc-pVDZ \rightarrow cc-pVTZ extrapolation, and the second performed with the augmented basis sets. The HF and MP2 correlation contributions to the interaction energy are reported separately in Table 6, giving clear evidence that most of the binding energy is due to electron correlation. This is counterintuitive because, as is shown later (section on LIE free energy evaluation), these complexes are dominated by electrostatic interactions, while correlation energy is usually associated with dispersion forces. However, as correlation takes account of both intermolecular dispersion and intramolecular correlation interaction energies, it is possible that the intramolecular term changes significantly upon binding, thus mixing with the dispersion force contribution. The best DFT and MP2 estimates of interaction energies differ substantially, by almost 5 kcal mol⁻¹. Though this is at least partly a result of

B3LYP's known underestimation of noncovalent hydrogen-bonded complexes,^[29] the MP2 interaction energies are about 2–3 kcal mol⁻¹ higher than those determined through MD simulations (about 10 kcal mol⁻¹ in view of the **in-out** \rightarrow **out-out** isomerization energy) that were used to predict free binding energies in good agreement with experimental data. However, the comparison between MP2 and MD interaction energies should be made with care, as MD data are averaged over multiple snapshots at 300 K while the MP2 energy is calculated on minimum-energy structures at 0 K. Thus it is reasonable that MP2 might be slightly overbinding with respect to the calculated MD interaction energy.

A possible source of error in the QM calculations is the use of an implicit solvation approach. It is in fact known that the introduction of explicit water molecules is sometimes required in order to describe hydrogen bonds correctly,^[14d] and this likely to be the case in the system considered here, at least as far as the carboxylate and guanidinium groups are concerned. While on one hand it is reasonable that part of the error due to the inaccurate description of the interaction between charged groups and water will be self-compensated when interaction energies are calculated as the difference between the complex and the separated host and guest molecules, on the other hand the presence of some charge transfer, as discussed above (see section on Complexes), is likely to decrease the interaction energies in water with respect to those of the separated host and guest, so that the error compensation can only be partial. The interaction of water with complexes and individual compounds has, however, been explicitly evaluated in MD simulations, which are in good quantitative agreement with experimental data, as reported below (LIE free energy evaluation section). MP2 interaction energies can thus be viewed as reasonable approximations of the noncovalent bond energies, with possible overestimation of not more than 2–3 kcal mol⁻¹, which can be considered the maximum error of the implicit solvation approach, while B3LYP interaction energies are underestimated by about 4 kcal mol⁻¹.

DFT free energy evaluation: Energies and entropies both of individual compounds and of complexes were calculated by use of B3LYP/6-31g(d,p) structures and different basis sets. Energies of solvated molecules include non-electrostatic terms (cavitation, dispersion, and repulsion energies), while gas-phase energies are corrected with zero-point energies. Table 7 gives the gas-phase entropy contributions estimated from frequency calculation performed at the B3LYP/6-31g(d,p) level. The same table also shows solvation free energies calculated with 6-31g(d,p) and aug-cc-pVDZ basis sets. The free solvation energies are only slightly different, with the smaller basis set underestimating the augmented data by about 1–2 kcal mol⁻¹. The reaction energy changes calculated in the gas phase and in solution at the B3LYP/6-31g(d,p) level and corrected for zero-point energies are given in Table 8. The experimental data for complexation determined by Schmuck et al. from NMR titration experiments^[7] are reported in the last column.

Table 7. Total gas phase entropy changes (S°) and vibrational contributions (S_{vib}) of all compounds and complexes. Free energy changes were calculated on B3LYP/6-31g(d,p) structures with use of two different basis sets: 6-31g(d,p) and aug-cc-pVDZ.

| Compound | S° (gas) | S_{vib} (gas) | ΔG_s [kcal mol ⁻¹] | ΔG_s [kcal mol ⁻¹] |
|-------------|-----------------|------------------------|--|--|
| | Basis set | 6-31g(d,p) | 6-31g(d,p) | aug-cc-pVDZ |
| A | 97.9 | 28.1 | -55.5 | -55.3 |
| 9 | 101.0 | 29.7 | -62.3 | -63.0 |
| 10 | 135.0 | 59.9 | -62.3 | -63.6 |
| 11 | 150.4 | 73.9 | -63.1 | -62.2 |
| 12 | 125.4 | 51.0 | -60.3 | -61.4 |
| 13 | 144.6 | 68.3 | -70.5 | -74.0 |
| 14 | 162.3 | 84.5 | -65.6 | -67.3 |
| A-9 | 157.9 | 80.6 | -7.5 | -8.6 |
| A-10 | 187.9 | 108.6 | -8.5 | -10.5 |
| A-11 | 202.0 | 121.9 | -6.0 | -8.4 |
| A-12 | 178.6 | 99.7 | -4.1 | -6.0 |
| A-13 | 198.8 | 118.6 | -14.8 | -16.4 |
| A-14 | 216.2 | 135.2 | -10.5 | -13.5 |

The calculated interaction energies (ΔE) in solution are substantially equivalent to enthalpy changes and so cannot be directly compared with experimental data, as the entropic contribution is not included. However, as free solvation energies and gas-phase entropy changes are similar for all complexes, it is likely that the binding entropies are approximately constant for all complexes, and so the measured relative differences in binding constants might be interpreted in terms of differences between binding energies. The calculations in fact correctly predict that host **9** should form the least stable complex, as was also found experimentally. The calculated interaction energies for hosts **10–14** are significantly larger but do not differ much ($\Delta E = -7$ kcal mol⁻¹ for **9** versus ca. $\Delta E = -10$ kcal mol⁻¹ for **10–14**). This is reasonable and in good agreement with the calculated complex structures. Hosts **10–14** each form a further H-bond from the additional amide CO to the substrate, which increases complex stability relative to the case of host **9**. At least in the calculated structures, however, there are no significant differences found between **10–14**, as is reflected in similar interaction energies. This was also observed experimentally, as hosts **10–13** all have similar stabilities. However, host **14** was found in the experiments to form an even stronger complex, which is only partially reflected in the calculated data. The interaction energy for host **14** is larger than those of hosts **13** and **11** but not than that of **10**. A possible reason

Table 8. Reaction energy changes calculated for all complexes in the gas phase and in solution, together with reaction solvation free energy changes. Gas-phase energy changes are corrected for zero-point energies. Energy changes in solution are calculated at 300 K. All data were calculated at the B3LYP level on B3LYP/6-31g(d,p) structures with use of 6-31g(d,p) and aug-cc-pVDZ (AUG) basis sets.

| | ΔE_{gas} | $T\Delta S_{\text{gas}}$ | $\Delta\Delta G_s$ | $\Delta\Delta G_s$ | ΔE_{wat} | ΔE_{wat} | ΔG |
|-------------|-------------------------|--------------------------|--------------------|--------------------|-------------------------|-------------------------|--------------------|
| | 6-31g(d,p) | 6-31g(d,p) | 6-31g(d,p) | AUG | 6-31g(d,p) | AUG | exp ^[4] |
| A-9 | -123.45 | 12.26 | 110.37 | 109.75 | -17.41 | -7.01 | -2.90 |
| A-10 | -127.30 | 12.81 | 109.29 | 108.35 | -21.71 | -10.23 | -3.94 |
| A-11 | -126.22 | 13.32 | 110.37 | 109.13 | -20.56 | -8.98 | -3.89 |
| A-12 | -128.46 | 13.03 | 111.68 | 110.73 | -20.87 | -9.71 | -4.06 |
| A-13 | -126.09 | 12.57 | 112.69 | 112.94 | -21.54 | -9.10 | -3.86 |
| A-14 | -126.01 | 12.72 | 110.63 | 109.19 | -20.49 | -9.53 | -4.37 |

for this difference might be the neglected entropic contributions to the complex stability. Indeed, as discussed below, MD analysis of the complex conformational evolution in water showed that complex **A-14** and **A-13** have higher mobilities than all the other complexes, which should lead to higher entropies.

Since reaction enthalpy changes can be calculated at high levels of accuracy, it would also be of great interest to determine reaction entropy changes in solution directly, as this would allow solution free energy changes to be evaluated and hence calculated and experimental data to be compared directly. However, though progress in this direction has been reported in the literature, it is still a complicated matter to evaluate solute vibrational and solvent entropies theoretically.^[30] Alternatively, reaction entropy changes can be determined indirectly through a thermodynamic cycle as $T\Delta S_{\text{wat}} = \Delta E_{\text{gas}} + \Delta\Delta G_{\text{sol}} - T\Delta S_{\text{gas}} - \Delta E_{\text{sol}}$. The entropy changes calculated in this way at the B3LYP/6-31g(d,p) level from the data reported in Table 7 are, however, very high at about -10 kcal mol⁻¹, which is probably due in part to the large error associated with the evaluation of the binding energies by use of the 6-31g(d,p) basis set, and also in part to the uncertainty associated with gas-phase entropies determined in the harmonic approximation.

An alternative approach is to determine free binding energies through molecular dynamics simulations. There are several possible methodologies to estimate ΔG_{sol} directly by exploiting MD, differing in their levels of approximation, and the most accurate among them are the thermodynamic integration and free energy perturbation approaches. However, these approaches require a reference state for free energy change calculation, while in this work we are interested in the estimation of an absolute value. Thus, in the following paragraphs we report the results concerning evaluation of free binding energies by the two approaches most commonly used in the literature: the linear interaction energy (LIE) and the MM-GB/PBSA methods.

LIE free energy evaluation: The aim of the molecular dynamics simulations presented here is to investigate the conformational stabilities of the complexes under study at room temperature when immersed in water. Simulations were performed in a cubic volume with periodic boundary conditions considering explicit water molecules by the protocol described above. The lateral dimension of the simulated cubic domain was 20 Å for all complexes and molecules. In order to prevent the lighter (and thus translationally faster) molecules from diffusing out of the periodic box during the simulation time, an increased lateral size of 30 Å was employed for molecules **9** and **A** and complex **A-9**. At the end of the simulations, water molecules

that had diffused outside the simulation box were re-imaged to the reference unit cell. Intermolecular interaction energies with the environment were then determined by use of the Anal code of the Amber suite.^[18] The time span considered for the energetic analysis was between 40 and 280 ps, corresponding to 70 coordinates. Electrostatic (E^{el}) and Van der Waals (E^{vdw}) interaction energies were then averaged over the whole time span of 240 ps.

The computed interaction energies of the complexes, the hosts, and the alanine-carboxylate substrate with the environment are summarized in Table 9. Free energies were computed with use of two different sets of scaling factors for

Table 9. Average interaction energies of receptors with the environment (water+alanine and water only) computed for MD simulations of 240 ps at 300 K and averaged over 70 trajectories. The two sets of data were obtained from two different simulations. LIE free binding energies were calculated as $\Delta G_{\text{LIE}} = \alpha \Delta E^{\text{vdw}} + \beta \Delta E^{\text{el}}$, with model A: $\alpha = 0.16$ and $\beta = 0.5$, and model B: $\alpha = 0.63$ and $\beta = 0.43$. In model C the energy differences between the **in-out** and the **out-out** configuration given in Table 1 have been added to the data of column A. All data are reported in kcal mol^{-1} .

| | Receptor/water +alanine | | Receptor/water | | ΔE [kcal mol^{-1}] | | ΔG_{LIE} | | | ΔG_{exp} |
|-------------|----------------------------|-----------------|------------------|-----------------|--|------------------------|-------------------------|-------|-------|-------------------------|
| | E^{vdw} | E^{el} | E^{vdw} | E^{el} | ΔE^{vdw} | ΔE^{el} | A | B | C | |
| A-9 | -9.78 | -85.73 | -9.25 | -73.90 | -0.53 | - | -6.00 | -5.42 | -3.50 | -2.90 |
| A-10 | -17.07 | -96.02 | -15.76 | -83.66 | -1.31 | - | -6.39 | -6.14 | -4.13 | -3.94 |
| A-11 | -20.63 | -94.96 | -19.34 | -84.56 | -1.00 | - | -5.36 | -5.10 | -3.29 | -3.89 |
| A-12 | -15.29 | -91.58 | -14.64 | -78.55 | -0.65 | - | -6.62 | -6.01 | -4.19 | -4.06 |
| A-13 | -16.05 | -114.40 | -14.71 | -103.55 | -1.34 | - | -5.64 | -5.51 | -3.34 | -3.86 |
| A-14 | -20.02 | -118.35 | -18.84 | -104.90 | -1.18 | - | -6.91 | -6.33 | -4.56 | -4.37 |

the electrostatic and VdW interaction energies. The first are the standard LIE factors, while the others are the results of a recent fit of experimental data performed with the Amber95 force field.^[31] Ligand energies were computed for their **out-in** configurations, because **out-out** \rightarrow **out-in** isomerization energies are available at a higher level of theory (see Table 1), and were used to correct free energies calculated with the standard LIE scaling parameters in column C of Table 9, which therefore represents our best estimate of the complexes free binding energies.

The calculated ΔG values correlate well with the experimental data for all complexes from both quantitative and qualitative points of view. Of course, one cannot expect the calculations to give exactly the same values, for several reasons. For one thing, the calculations report the complex formation between two single molecules in a indefinitely dilute solution, whereas the experiments were performed at millimolar concentrations. However, it is well known—especially for ionic interactions, such as we are dealing with here—that the concentrations of salts (=ionic strength) has a significant impact on absolute complex stability: increasing of solution concentration decreases the complex stability, as has also previously been demonstrated experimentally.^[32] To test

the impact of the salt on the results of our calculations we performed MD simulations of both ligands and complexes in explicit water, adding Cl^- anions (about 1 M). The effect was negligible, as we observed in all cases that Cl^- tended to diffuse away both from the complex and from the ligand during the MD simulations, though they had been positioned in proximity to the guanidinium group at the beginning of the simulations. A detailed study of the first instants of the simulations showed, however, that if Cl^- is positioned in the vicinity of the solvated molecule, it can stabilize the cation significantly, increasing its interaction energy with the environment by 1–2 kcal mol^{-1} . This would likely be the case

should the salt concentration be significantly higher than that considered in this study.

Though the agreement with experimental data is within $\pm 0.6 \text{ kcal mol}^{-1}$, there are some discrepancies that require discussion. The first is the lower stability of complex **A-9** with respect to complexes **A-11** and **A-13**. This might be determined by an overestimation of the electrostatic interaction energy ΔE^{el} . In fact, complex **A-9** has the highest charge transfer among the complexes considered, as evidenced by the ESP data outlined in Table S3, which are not accounted for in the force field adopted for the simulations. The inclusion of charge

transfer—by, for example, adoption of a polarizable force field—would probably lead to a decrease in the **A-9** electrostatic interaction energy and thus in the stability of this complex. The second discrepancy concerns complexes **A-13** and **A-14**, which are predicted to differ in binding free energy by more than 1 kcal mol^{-1} , while experimentally this difference is much smaller. The reason for such a disagreement is probably the higher conformational mobility of complex **A-13**, which is likely to determine an entropic stabilization effect, as is discussed in more detail below (MD analysis of conformational evolution of key complexes).

MM-GBSA/PBSA free energy evaluation: An alternative approach to the LIE method is what is usually referred to as the MM-GB/PBSA scheme.^[23] This is an indirect approach allowing the determination of binding free energies by taking advantage of the evaluation of free solvation energies by means of an implicit approach through a thermodynamic cycle as expressed in:

$$\langle \Delta G \rangle = \langle \Delta E_{\text{MM}} \rangle + \langle \Delta \Delta G_{\text{sol}} \rangle - T \Delta S_{\text{gas}} \quad (1)$$

where ΔE_{MM} is molecular energy change in the gas phase for the reaction of dissociation of the complex, and is determined by the sum of three contributions:

$$\langle \Delta E_{MM} \rangle = \langle \Delta E_{int} \rangle + \langle \Delta E_{El} \rangle + \langle \Delta E_{vdw} \rangle \quad (2)$$

which correspond to changes in internal energy (consisting of bond stretching, angular deformation, and torsion energy), electrostatic energy (expressed as Coulomb interaction between point charges), and Van der Waals energy (expressed through Lennard Jones potentials). Molecular mechanics energies were calculated by use of the ff03 Amber force field.^[18] $\Delta \Delta G_{sol}$ is the difference between the solvation free energy of the complex and that of the ligand and receptor. Solvation free energies were implicitly determined by the Poisson–Boltzmann (PB) and generalized Born (GB) approaches. Finally, $T\Delta S_{gas}$ is the entropy difference between complex, ligand, and protein and was calculated by considering rotational, translational, and vibrational contributions, which were determined by computing normal mode frequencies with the Nmode code of the Amber suite.^[33] The computed free binding energies, reported in Table 10, do not include gas-phase entropy change contributions, which are outlined separately in Table 11. The free binding energies calculated as defined in Eq. (1) are reported on the right-hand side of Table 11. Standard deviations are about 2 kcal mol⁻¹ both for MM-GBSA and for MM-PBSA free binding energies, respectively, except for complex **A-13**, for which the standard deviation is 3.5 kcal mol⁻¹.

The interaction energy data reported in Table 10, computed from a single trajectory simulation as defined in Eq. (1) without consideration of the gas-phase entropic contribution, match the experimental data qualitatively well, with

Table 10. Sum of gas-phase binding energies and free solvation energy differences calculated by the PBSA and GBSA approaches and averaged over 125 snapshots of a single-trajectory MD simulation performed with a cut-off for long-range interactions of 15 Å for 1 ns. Energies (kcal mol⁻¹) are defined as in Eq. (1), but do not include gas-phase contributions.

| | PBSA | | | | COM | GBSA | | |
|-------------|---------|--------|---------|-------------|---------|--------|---------|-------------|
| | COM | REC | LIG | ΔPB | | REC | LIG | ΔGB |
| A-9 | -191.62 | -76.33 | -102.18 | -13.11 | -191.13 | -78.52 | -100.19 | -12.42 |
| A-10 | -198.99 | -76.82 | -108.19 | -13.98 | -197.95 | -79.00 | -105.73 | -13.22 |
| A-11 | -190.40 | -78.12 | -98.44 | -13.84 | -189.21 | -80.29 | -95.83 | -13.09 |
| A-12 | -155.18 | -76.01 | -65.52 | -13.65 | -154.65 | -79.91 | -60.59 | -14.15 |
| A-13 | -215.31 | -77.63 | -125.84 | -11.84 | -215.57 | -79.83 | -124.41 | -11.33 |
| A-14 | -220.08 | -77.36 | -127.91 | -14.81 | -219.40 | -79.46 | -125.77 | -14.17 |

Table 11. Gas-phase entropies were calculated by averaging vibrational, rotational, and translational contributions evaluated through a normal mode analysis of 50 snapshots obtained from MD simulations performed with a cut-off for long-range interactions of 15 Å and for 1 ns. Free binding energies were calculated by subtracting the entropic contributions to the energies reported in Table 10. Data are reported in kcal mol⁻¹.

| | S_{COM} | S_{REC} | S_{LIG} | $T\Delta S$ | ΔG PBSA | ΔG GBSA | $\Delta G_{exp}^{[7]}$ |
|-------------|-----------|-----------|-----------|-------------|-----------------|-----------------|------------------------|
| A-9 | 48.04 | 29.64 | 29.32 | -10.92 | -2.19 | -1.50 | -2.9 |
| A-10 | 56.92 | 29.64 | 38.55 | -11.27 | -2.71 | -1.95 | -3.94 |
| A-11 | 59.86 | 29.64 | 43.07 | -12.85 | -0.99 | -0.24 | -3.89 |
| A-12 | 54.00 | 29.64 | 35.86 | -11.50 | -2.15 | -2.65 | -4.06 |
| A-13 | 59.96 | 29.64 | 41.53 | -11.21 | -0.63 | -0.12 | -3.86 |
| A-14 | 63.76 | 29.64 | 46.80 | -12.68 | -2.13 | -1.49 | -4.37 |

complex **A-14** having the highest interaction energy, followed by complexes containing hosts **10** and **12**. Also, the significant binding energy difference between systems **A-14** and **A-13** is well predicted by these data. The only qualitative inconsistency concerns complex **A-9**, the binding energy of which is out of scale with respect to the experimental trend. This overestimation is similar to that observed for the LIE analysis, and so the considerations above are also valid in this case.

In absolute terms, after the entropy contributions are included, the calculated complex stabilities are now significantly smaller than those obtained from the experimental data. There are two possible factors that might justify the observed quantitative disagreement. The first is an erroneous evaluation of gas-phase vibrational frequencies. In fact, 15–20 vibrational frequencies of the complexes calculated by DFT at the B3LYP/6-31g(d,p) level are smaller than 150 cm⁻¹, and so the harmonic approximation is probably not adequate to describe most of these intramolecular motions. Also, from the data outlined in Table 11 it is easy to calculate that the contribution of vibrational entropies to the overall gas-phase entropy change is about 50%, so an error in the evaluation of the vibrational entropy would be reflected significantly in the overall free energy change. A second possibility is that the error associated with the evaluation of the solvation free energies of charged molecules might significantly influence the free energy change evaluation. It has in fact been shown that the Amber force field tends to overestimate the solvation energies of positively charged amino acids and to underestimate those of negatively charged amino acids.^[34] Though the two errors should balance out in the present case, it is likely that a significant uncertainty would still remain. Similar systematic errors have

been found in the literature when the MM-PBSA approach was used to determine absolute free binding energies of complexes involving DNA and proteins,^[35] and were attributed either to errors in the radii used for the evaluation of solvation free energies, so that re-fitting of the PARSE radii has been suggested, or to uncertainties in the estimation of entropy contributions.

MD analysis of conformational evolution of key complexes

The comparison between DFT and MD calculations seems to indicate that most of the experimentally measured free binding energies can be explained in terms of a relatively strong binding enthalpy of

about 8 kcal mol^{-1} produced by the formation of three hydrogen bonds between the guanidiniocarbonylpyrrole cation and the negatively charged carboxylate of the substrate. Another 2 kcal mol^{-1} in complex stability is then gained through the addition to the pyrrole of the second amide group, which can form a further H-bond with the substrate. Despite the copious information gained by the conjunct DFT and MD analyses, however, it is difficult to understand precisely how the addition of an isopropyl group to host **13** enhances the affinity of host **14** towards alanine. Even more so, as MD free binding energy simulations predict the difference in stability between the two complexes to be even higher than that experimentally measured. In order to investigate the origin of the different binding affinities of hosts **13** and **14** further, we analyzed the conformational evolution of the respective complexes over time in detail.

The first significant difference observed between complexes **A-14** and **A-13** is that **A-14** periodically changes its conformation between the minimum-energy structure during the MD simulations (structure **X**; see Figure 5) determined by DFT calculations and used as starting point of the simulations and a second structure (**Y**; see Figure 5) in which the *N*-acetyl group of alanine interacts with the isopropyl group of host **14** and with the pyrrole N atom. The fourth H-bond from the pyrrole amide CO to the NH of the alanine amide is not present in this structure. A third structure, in which only the carboxylate group is still interacting with the receptor (structure **Z**; see Figure 5), could also be observed. Similar structures were also observed for complex **A-13** (see Figure 5). It is interesting to note that structures **X** and **Y** were observed more frequently for complex **A-14** than for **A-13**.

To determine the relative stabilities of structures **X** and **Y** in complexes **A-14** and **A-13**, DFT calculations were performed with the MD structures shown above as starting geometries. Because of the weak natures of the interactions present in complex **Y**, which involves relatively distant atoms, geometry optimizations had to be performed with

the computationally expensive aug-cc-pVDZ basis set. We found that for complex **A-13** the optimization of conformation **Y** leads to structure **X**, which means that **Y** is not a local minimum for complex **A-13**. In contrast, in complex **A-14** we were able to determine a minimum-energy structure similar to structure **Y**, which is less stable than structure **X** by about 4 kcal mol^{-1} and thus represents a second energy well on the potential energy surface of **A-14**.

To determine the times spent by each complex in configurations **X** and **Y** during the MD simulations quantitatively, we monitored the distance between the carbon atom of the *N*-acetyl group of alanine and the pyrrole nitrogen atom. The results of these analyses are outlined in Figures 6 and 7 for complexes **A-13** and **A-14**, respectively. The total number of trajectories screened was 250 (one every four ps).

The relative populations of the three structures **X**, **Y**, and **Z** for the complexes **A-13** and **A-14** are summarized in Table 12, together with the absolute MM-PBSA energies

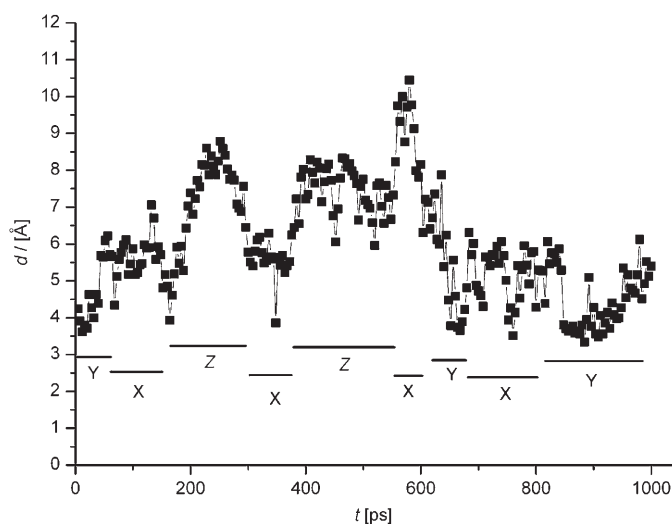


Figure 6. Graphical analysis of the distance between the carbon atom of the terminal methyl group of alanine and the nitrogen atom of pyrrole measured for complex **A-13** during 1 ns of MD simulation.

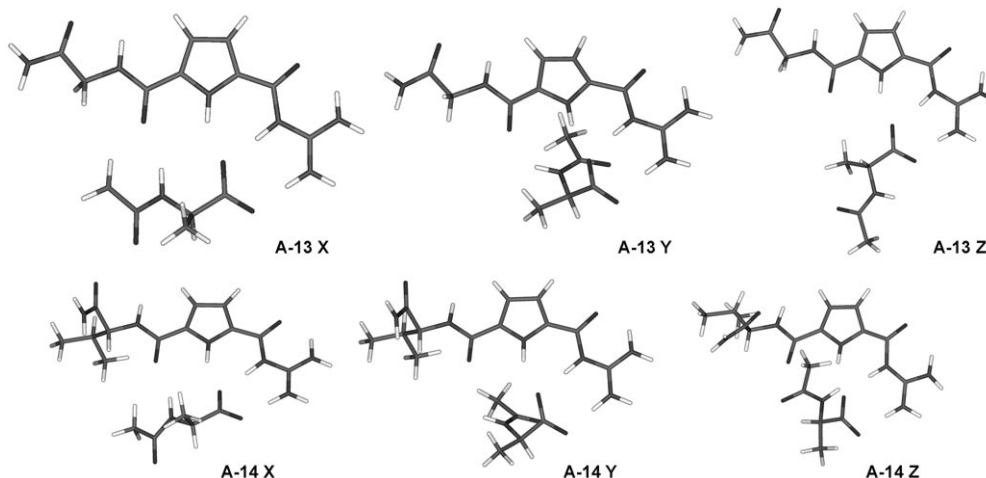


Figure 5. Key intermediate conformations observed during the 1 ns molecular dynamics of complexes **A-13** and **A-14**.

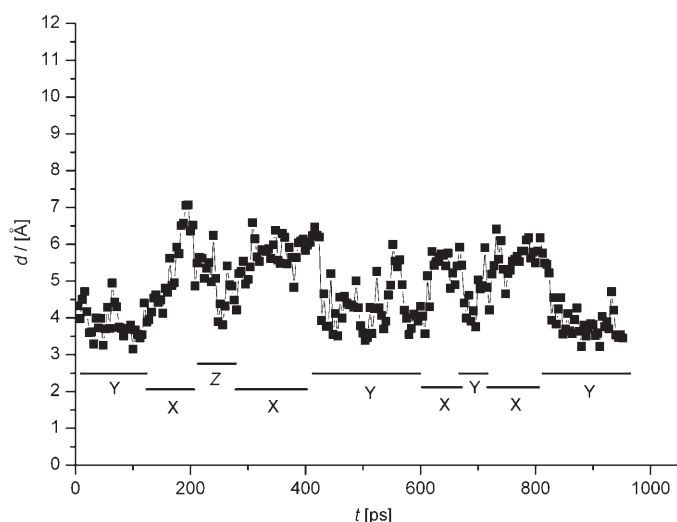


Figure 7. Graphical analysis of the distance between the carbon atom of the terminal methyl group of alanine and the nitrogen atom of pyrrole measured for complex **A-14** during 1 ns of MD simulation.

Table 12. Average MM-PBSA energies (kcal mol^{-1} , the sums of gas-phase and solvation free energies calculated with the PBSA solvation model) and free energy changes (calculated as defined in Eq. (1), without inclusion of the entropic contributions) of complexes **A-13** and **A-14** calculated during a 1 ns MD simulation.

| | t [ns] | Conf | PB | Δ PB |
|-------------|-----------|------------|--------|-------------|
| A-13 | 0.08–0.20 | X | –219.3 | –14.0 |
| | 0.28–0.40 | X-Z | –217.2 | –13.2 |
| | 0.40–0.60 | Z | –209.1 | –7.0 |
| | 0.60–0.80 | X/Z | –211.8 | –10.4 |
| | 0.80–1.0 | Y | –218.9 | –15.1 |
| A-14 | 0.14–0.22 | X | –221.4 | –14.0 |
| | 0.30–0.44 | X | –221.9 | –14.7 |
| | 0.44–0.60 | Y | –221.6 | –15.8 |
| | 0.60–0.80 | X/Y | –220.8 | –15.3 |
| | 0.80–0.94 | Y | –221.3 | –16.4 |

and free energy changes computed with the MM-PBSA, uncorrected for gas-phase entropy changes. The data in Figure 6 and Figure 7 clearly show that the monitored distance has smaller fluctuations in complex **A-14** than in complex **A-13**. The maximum changes are about 4 Å for **A-14** and 7 Å for **A-13**, which might indicate that complex **A-14** is more stable. Also, the energetic analysis in Table 12 indicates that the predicted decrease in binding energy for complex **A-13** through population of the higher-energy conformations with respect to complex **A-14** can be ascribed mostly to the fact that complex **A-13** spends part of its time in configuration **Z**, which has a significantly reduced binding energy. The change of conformation between **X**, **Y**, and **Z** is determined by two internal rotations of the alanine carboxylate constituents around its stereogenic α -carbon, and in particular around the C–N and the C–COO bonds. For complex **A-13** the rotation about the C–COO leads to structure **Z** without any intermediate minimum for structure **Y**. The structure **Z** has one H-bond fewer than the corresponding global minimum **X**, so reduced stability is obviously to

be expected. For complex **A-14** the same rotation brings the N-acetyl group into close proximity to the bulky isopropyl side chain, leading to an energy minimum for structure **Y**, which is therefore periodically reached during the simulations. Further rotation to the less stable structure **Z** is obviously then prohibited by the steric interaction between the two groups, and the complex returns to structure **X**. According to MD simulations, complex **A-14** hardly passes through conformation **Z**, in sharp contrast to complex **A-13**. It is interesting to observe that while the MM-PBSA energy of **A-14** is essentially constant during the simulation time, independently of the oscillations between the two conformations, that of **A-13** varies significantly, with the interaction minima and absolute MM-PBSA energies reached in conformation **Z**. The high mobility of complex **A-13** increases its conformational entropy, even more so as the reduced moment of inertia of the two mutually rotating moieties is particularly high given their size and weight, thus leading to the previously invoked entropic stabilization of the complex.

Conclusion

The combined DFT and MD investigation of structures and energies of guanidiniopyrrole–carboxylate complexes has allowed us to improve our understanding of the origin of the stability of these compounds in water. Minimum-energy structures of complexes, substrates, and ligands were determined computationally with the aid of state-of-the-art DFT calculations, revealing complexes characterized by three strong H bonds involving the carboxylate, the guanidinium, and the pyrrole NH groups. The use of the DFT approach has been justified with selected MP2 calculations. The complexes are further stabilized by additional interactions between the ligand side groups and the substrate. The host–guest interaction is sufficiently strong to induce a conformational change in the ligand, consisting of a 180° rotation around the guanidiniocarbonyl–pyrrole bond, leading to a structure that is about 2.5 kcal mol^{-1} higher in energy. Binding energies computed with implicit water molecule effects taken into account were significantly affected by basis set size, confirming the need for diffuse functions with high angular momentum in order to describe supramolecular complexes. The predicted binding enthalpy of about 9 kcal mol^{-1} is mostly determined by electrostatic interactions, comparing well with the experimentally measured trend of the complexes' stabilities. Complementary DFT calculations with MD simulations led to a deeper understanding of this system. At 300 K in water several torsion vibrations degenerate into hindered internal rotors, so that different binding structures are observed during the simulations. The stabilization of one of these structures, determined by VDW interactions between the guanidinium side arm and the carboxylate substrate, played a key role in increasing the strength of the most stable complex considered in this study. The reason is that it allowed two internal rotations of the alanine carboxylate constituents around its stereogenic α -carbon to be lim-

ited, leading to the formation of unstable structures. Finally, it was interesting to find out that the LIE method allowed binding energies of the complexes in almost quantitative agreement with experimental data to be calculated. This indicates that the fundamental hypothesis on which LIE is based, the linear response approximation, is reasonable for the systems under investigation. Some minor deviations from experimental data should probably be correctable by improving the quality of the force field by adopting, for example, polarizable force fields.

Acknowledgement

Support by European COST-D31 action is gratefully acknowledged. A.F. acknowledges the support of the MURST (Ministero Università e Ricerca Scientifica e Tecnologica) PRIN2006 and of the Cariplo Foundation. C.S. thanks the Deutsche Forschungsgemeinschaft (DFG) for ongoing financial support of his work.

- [1] a) L. J. Prins, D. N. Reinhoudt, P. Timmerman, *Angew. Chem.* **2001**, *113*, 2446; *Angew. Chem. Int. Ed.* **2001**, *40*, 2382; b) G. Cooke, V. M. Rotello *Chem. Soc. Rev.* **2002**, *31*, 275; c) G. V. Oshovsky, D. N. Reinhoudt, W. Verboom, *Angew. Chem.* **2007**, *119*, 2418; *Angew. Chem. Int. Ed.* **2007**, *46*, 2366.
- [2] One concept used in this context are chemical-mutant cycles: S. L. Cockroft, C. A. Hunter, *Chem. Soc. Rev.* **2007**, *36*, 172.
- [3] For a fascinating experimental case study involving peptide binding by vancomycin antibiotics see: a) C. C. McComas, B. M. Crowley, D. L. Boger *J. Am. Chem. Soc.* **2003**, *125*, 9314; b) B. M. Crowley, D. L. Boger, *J. Am. Chem. Soc.* **2006**, *128*, 2885.
- [4] C. Schmuck, W. Wienand, *J. Am. Chem. Soc.* **2003**, *125*, 452.
- [5] S. Schlund, C. Schmuck, B. Engels, *J. Am. Chem. Soc.* **2005**, *127*, 11115.
- [6] a) M. D. Best, S. L. Tobey, E. V. Anslyn, *Coord. Chem. Rev.* **2003**, *240*, 3; b) P. A. Gale, *Coord. Chem. Rev.* **2003**, *240*, 191; c) R. J. Fitzmaurice, G. M. Kyne, D. Douheret, J. D. Kilburn, *J. Chem. Soc. Perkin Trans. 1* **2002**, 841; d) F. P. Schmidchen, M. Berger, *Chem. Rev.* **1997**, *97*, 1609; e) P. Blondeau, M. Segura, R. Perez-Fernandez, J. de Mendoza, *Chem. Soc. Rev.* **2007**, *36*, 198; f) K. A. Schug, W. Lindner, *Chem. Rev.* **2005**, *105*, 67.
- [7] C. Schmuck, *Chem. Eur. J.* **2000**, *6*, 709.
- [8] a) C. Schmuck, *Coord. Chem. Rev.* **2006**, *205*, 3053; b) C. Schmuck, D. Rupprecht, W. Wienand, *Chem. Eur. J.* **2006**, *12*, 9186; c) C. Schmuck, P. Wich, *Angew. Chem.* **2006**, *118*, 4383; *Angew. Chem. Int. Ed.* **2006**, *45*, 4277; d) C. Schmuck, T. Rehm, L. Geiger, M. Schäfer, *J. Org. Chem.* **2007**, *72*, 6162; e) C. Schmuck, V. Bickert, *J. Org. Chem.* **2007**, *72*, 6832.
- [9] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739.
- [10] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
- [11] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
- [12] Gaussian 03, Revision C01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, 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. Ochtersk, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, A. L. G. Liu, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc, Pittsburgh, **2003**.
- [13] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724.
- [14] a) E. Cancès, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032; b) B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *106*, 5151; c) B. Mennucci, E. Cancès, J. Tomasi, *J. Phys. Chem. B* **1997**, *101*, 10506; d) J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.
- [15] D. E. Woon, T. H. Dunning, Jr., *J. Chem. Phys.* **1993**, *98*.
- [16] a) A. Castleman, Jr., P. Hobza, *Chem. Rev.* **1994**, *94*, 1721; b) B. Brutschy, P. Hobza, *Chem. Rev.* **2000**, *100*, 3861.
- [17] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang, P. Kollman, *J. Comput. Chem.* **2003**, *24*, 1999.
- [18] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, H. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross, P. A. Kollman, Amber, 8.0 ed., University of California San Francisco, **2004**.
- [19] W. D. Cornell, P. Cieplak, C. I. Bayly, P. A. Kollman, *J. Am. Chem. Soc.* **1993**, *115*, 9620.
- [20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- [21] P. H. Hunenberger, J. A. McCammon, *Biophys. Chem.* **1999**, *78*, 69.
- [22] V. Tsui, D. A. Case, *Biopolymers* **2000**, *56*, 275.
- [23] J. Wang, P. Morin, W. Wang, P. A. Kollman, *J. Am. Chem. Soc.* **2001**, *123*, 5221.
- [24] B. O. Brandsdal, F. Osterberg, M. Almlöf, I. Fejerberg, V. B. Luzhkov, J. Aqvist, *Protein Simul.* **2003**, *66*, 123.
- [25] G. Schaftenaar, J. H. Noordik, *J. Cold Reg. Eng. J. Comput.-Aided. Mol. Des.* **2000**, *14*, 123.
- [26] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.
- [27] J. W. Zou, Y. J. Jiang, M. Guo, G. X. Hu, B. Zhang, H. C. Liu, Q. S. Yu, *Chem. Eur. J.* **2005**, *11*, 740.
- [28] D. G. Truhlar, *Chem. Phys. Lett.* **1998**, *294*, 45.
- [29] a) J. Sponer, P. Jurecka, P. Hobza, *J. Am. Chem. Soc.* **2004**, *126*, 10142; b) Y. Zhao, D. G. Truhlar, *J. Chem. Theor. Comput.* **2007**, *3*, 289.
- [30] J. Carlsson, J. Aqvist, *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385.
- [31] M. Almlöf, B. O. Brandsdal, J. Aqvist, *J. Comput. Chem.* **2004**, *25*, 1242.
- [32] C. Schmuck, S. Graupner, *Tetrahedron Lett.* **2005**, *46*, 1295.
- [33] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668.
- [34] S. B. Dixit, R. Bhasin, E. Rajasekaran, B. Jayaram, *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 1105.
- [35] a) H. Gohlke, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 238; b) N. Spackova, T. E. Cheatham, III, F. Ryjacek, F. Lankas, L. van Meerwelt, P. Hobza, J. Sponer, *J. Am. Chem. Soc.* **2003**, *125*, 1759; c) C. S. Page, P. A. Bates, *J. Comput. Chem.* **2006**, *27*, 1990.

Received: November 6, 2007

Revised: February 22, 2008

Published online: April 22, 2008