Like-Charge Guanidinium Pairing from Molecular Dynamics and *Ab Initio* Calculations

Mario Vazdar[a,b], Jiři Vymětal[a], Jan Heyda[a] , Jiři Vondrášek[a,*] and Pavel Jungwirth[a,*]

[a]*Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Academy of Sciences of the Czech Republic, Center for Biomolecules and Complex Molecular Systems, Flemingovo nám. 2, 16610 Prague 6, Czech Republic*
[b]*Rudjer Bošković Institute, Division of Organic Chemistry and Biochemistry, P.O.B. 180, HR-10002 Zagreb, Croatia*

E-mails: pavel.jungwirth@uochb.cas.cz and jiri.vondrasek@uochb.cas.cz

Abstract

Pairing of guanidinium moieties in water is explored by molecular dynamics simulations of short arginine-rich peptides and *ab initio* calculations of a pair of guanidinium ions in water clusters of increasing size. Molecular dynamics simulations show that in aqueous environment the *di*-arginine guanidinium like-charged ion pairing is sterically hindered, while in the Arg-Ala-Arg tripeptide this pairing is significant. This result is supported by the survey of protein structure databases where it is found that stacked arginine pairs in dipeptide fragments exist solely as being imposed by the protein structure. In contrast, when two arginines are separated by a single amino acid their guanidinium groups can freely approach each other and they frequently form stacked pairs. Molecular dynamics simulations results are also supported by *ab initio* calculations, which show stabilization of stacked guanidinium pairs in sufficiently large water clusters.

Introduction

Ions play a ubiquitous role in all biological systems as essential components of intra- and extra-cellular fluids.[1] Interactions between ions and proteins, influencing proceses such as salting-in and salting-out effects,[2-4] as well as enzymatic activity in different ionic solutions,[5-8] show a high level of ion specificity. An interesting manifestation of ion-specific effects are "electrostatics-defying" structures such as anions bound to nucleic acids[9] or arginine-arginine pairs in proteins.[10-12] Interactions between guanidinium moieties of arginines may also mediate protein-ligand interactions, protein folding, and molecular recognition.[13-18] This seemingly counterintuitive like-charge ion pairing of guanidinum ions has attracted a considerable interest among computational chemists and numerous studies have been performed including *ab initio*,[19,20] Monte Carlo[21] and molecular dynamics simulations,[20,22] which support the fact that a guanidinium homo-ion pair is formed in a water solution. A combined experimental andcomputational study by Mason *et al.* also showed that guanidinium ions in a water solution stack parallel to each other.[23] A recent investigation using dielectric relaxation spectroscopy did not show signs of ion-pairing in guanidinium chloride solutions, however, this may be due to the insensitivity of this technique to short-lived transient ion pairs.[24]

The key to the peculiar behavior of the guanidinium ion is its structure. Guanidinium ion is planar and shows a substantial aromatic character which makes it a member of the so called "Y-aromatics".[25] However, even though it is an ion, its planarity together with charge distribution results in weak hydration of guanidinium faces, with strong hydrogen bonding taking place only in the guanidinium plane.[26] Guanidinium salts are potent protein denaturants since guanidinium ion can readily stack to the hydrophobic protein surfaces, reducing thus the cost of hydration of hydrophobic groups.[27] At the same time, guanidinium ion can also make hydrogen bonds with the protein backbone in the unfolded state and stabilize it.[28]

In this study, we performed molecular dynamics (MD) simulations with different force fields in order to test the robustness of guanidinium-guanidinium pairing in isolated *di*-arginine and arginine-alanine-arginine peptides with respect to the empirical potential employed. We also surveyed different sets of protein structures from Protein Data Bank (PDB) at www.pdb.org, searching for stacked arginine arrangements in proteins. In addition, we performed *ab initio* calculations of microhydrated (with up to 14 water molecules) stacked guanidinium ion pairs. Such a combined approach is needed in order to establish with confidence the existence of guanidinium-guanidinium pairing in aqueous arginine-rich peptides.

Computational Details

We performed 50 ns molecular dynamics simulations (after 1 ns of equilibration) of *di*-arginine, *di*-lysine dipeptides, arginine-alanine-arginine (Arg-Ala-Arg) and lysine-alanine-lysine (Lys-Ala-Lys) tripeptides in water with parameters taken from nonpolarizable parm99,[29] parm99SB,[30] parm03,[31] or parm10[32] force fields. In the case of parm99 force field, its polarizable variant, denoted here as pol-parm99, was used as well. All peptides were terminated with acetyl and N-methyl residues at the N-terminus and C-terminus, respectively. Each of the studied peptides was neutralized with two chloride counterions (without or with polarizability)[33] and solvated in approximately 700 SPC/E[34] nonpolarizable or POL3[35] polarizable water molecules. 3D periodic boundary conditions were used with long range electrostatic interactions beyond the nonbonded cutoff of 10 Å accounted for using the particle-mesh Ewald procedure.[36] The Berendsen thermostat and barostat with temperature of 300 K and pressure of 1 atm was used.[37] The SHAKE algorithm[38] was employed to constrain all bonds containing hydrogen atoms. The time step used in all simulation was set to 1 fs and geometries were collected every 1 ps, which provided 50 000 frames for consequent analysis. The simulation time was long enough to observe several tens of side chain flips, and thus to obtain

statistically converged results concerning the structural distributions of the peptides. Molecular dynamics calculations were performed with the AMBER 11 program.[32]

We analyzed the spatial distribution of side chains in dipeptide and tripeptide protein fragments containing arginines in different protein sets from PDB. The fragments in question were Arg-Arg and Arg-X-Arg where X stands for any non-charged aminoacids (i.e., all aminoacids except Arg, Lys, Asp and Glu). As a reference point for the analysis, the distance between the carbon (CZ) atoms of arginine guanidinium groups was chosen. The secondary structure assignment and evaluation of Ramachandran's dihedral angles φ and ψ populations were performed using the DSSP program.[39] The statistics was collected for three independent protein sets in order to estimate bias in PDB database and the overall reliability of our observations. The first set consists of 691 automatically selected X-ray structures with high resolution (<1.5 Å), low R factor (<1.8 Å) and low sequence homology (<30%) (abbreviated as Set1 in the further text). The second set – the Top500 database[40] (abbreviated as Top500 in the further text) includes 500 high quality X-ray structures provided by Richardson's lab (http://kinemage.biochem.duke.edu/databases/top500.php). The third set was taken from the Dynameomics project[41] (abbreviated as Dyn in the further text) and the extracted structures represent the majority of the known protein folds. It should be mentioned that this set contains also NMR characterized structures.

Ab initio calculations were performed for guanidinium dimers with increasing number of water molecules. We carried out geometry optimizations of clusters with two guanidinium cations (Gdm^+) with increasing number of waters around them, starting from three up to fourteen water molecules using the B-LYP-D/aug-cc-pVDZ method. Vibrational analysis for all clusters was performed in order to verify that a minimum on the potential energy surface was obtained. Single point calculations at the MP2/aug-cc-pVDZ//B-LYP-D/aug-cc-pVDZ levels of theory were performed for comparison. Basis set effects were assessed employing B-LYP-D and MP2 geometry optimizations with aug-cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ basis sets for the cluster containing the guanidinium dimer with three water molecules. Complexation energies of Gdm^+-water clusters were obtained

using the supermolecular approach,[42] i.e., the dimer was decomposed into two fragments with halved number of water molecules around it (or roughly halved for odd number of water molecules). The interaction energy was then calculated as the difference between the dimer energy and the sum of energies of the two individual fragments. The basis set superposition error (BSSE) was corrected by the counterpoise procedure[43] (for the B-LYP-D calculations). Density functional calculations with dispersion correction (B-LYP-D),[44,45] were performed using Turbomole 6.2.[46] MP2 calculations were performed employing Gaussian09.[47]

Results and discussion

Molecular dynamics simulations

We investigated pairing of guanidinium moieties by performing MD simulations of *di*-arginine and Arg-Ala-Arg tripeptide in water with different non-polarizable force fields - parm99, parm99SB, parm03, and parm10 and a polarizable pol-parm99 force field. Pairing was quantified in terms of radial distribution functions (RDFs). The corresponding carbon-carbon RDFs of the Gdm^+ groups in arginine side chains are shown in Fig. 1. Inspection of Fig. 1 clearly shows that a strong first peak in RDFs, which corresponds to the formation of a contact pair of Gdm^+ groups, is present in the case of nonpolarizable parm99 and polarizable pol-parm99 force fields. In the case of the pol-parm99 force field, the height of the peak maximum is somewhat higher with $g(r)$ = 10, in comparison to the nonpolarizable parm99 force field, where $g(r)$ = 7.5 for the Gdm^+ - Gdm^+ distance of ca. 4 Å. The second peak maximum, which describes a formation of a solvent-separated ion pair, occurs at the Gdm^+ - Gdm^+ distance of ca. 7 Å for both force fields. In contrast, there is no contact peak in the Gdm^+ - Gdm^+ RDFs in simulations of aqueous *di*-arginine with newer parm99SB, parm03 or parm10 force fields.

This dependence of results on potential parameters is further exemplified in Fig. 2 which shows spatial distributions of the Gdm^+ group, water molecules, and chloride counterions around the other Gdm^+ moeity in *di*-arginine side chain for parm99 and parm99SB force fields. We can see from Fig. 2 that in parm99 force field simulations, diffuse lobes corresponding to the other Gdm^+ group are present above and below the plane of Gdm^+. In parm99SB simulations there are no such lobes present reflecting the lack of Gdm^+ - Gdm^+ pairing.

Naturally, the question arises why newer force fields like parm99SB, parm03 and parm10 do not predict the Gdm^+ - Gdm^+ pairing? Is the like-charge pairing only an artifact of intermolecular nonbonding parameters which tend to overemphasize guanidinium stacking or is the problem in intramolecular bonding parameters which prevent pairing? In order to resolve this issue, we performed further analysis and additional MD simulations parm99 and parm99SB force fields. Since the results for parm99SB, parm03 and parm10 force fields are quantitatively almost identical, we will not consider any further the results obtained with the parm03 and parm10 force fields. Also, for the sake of simplicity, we will only compare results obtained with parm99SB force fields with results obtained with the nonpolarizable parm99 force field.

We constructed the Ramachandran plots for Arg residues in the dipeptide which show the populations of the two backbone dihedral angles (torsion angles) φ and ψ in each of the two arginine residues during the MD simulation. A definition of backbone angles $\varphi_{1,2}$ and $\psi_{1,2}$ is given in Fig. 3. Results of the Ramachandran analysis for parm99 and parm99SB force fields are shown in Fig. 4. It is clearly visible that completely different basins of the Ramachandran plots for the two arginine residues are populated in simulations with the two force fields. For the parm99 force field, the α-helical region is predominantly occupied, whereas in the case of parm99SB force field, the β-extended region is predominantly occupied. This unambiguously points to the reason for obtaining different results for Gdm^+-Gdm^+ pairing with ff99 and ff99SB force fields, i.e., the different parametrization of the dihedral backbone angle parameters in arginine is responsible for this difference.

Analysis of the Ramachandran plot thus points to the fact that for parm99SB sidechains in *di*-arginine are located on the opposite side of the backbone and they actually do not have the opportunity to interact with each other. As a further check, we restrained the two backbone dihedral angles (φ and ψ) in two arginine residues to values for which Gdm^+ - Gdm^+ pairing can occur, i.e., we placed the two side chains on the same side of the backbone. These restrained values for dihedral angles were ϕ_1 = -50°, ψ_1 = 0°, ϕ_2 = -150° and ψ_2 = 0°. We then performed MD simulations with both parm99 and parm99SB force fields with these restraints. Fig. 5 shows carbon–carbon radial distribution functions of guanidinium groups in *di*-arginine with the above restraints. The first peak in RDFs is strong and clearly shows the like-charge ion pairing for restrained simulations with either parm99 or parm99SB force fields. Therefore, the fact that unconstrained simulations with parm99SB, parm03 and parm10 force field do not reveal like-charge ion pairing is solely due to the backbone dihedral angle parametrization; like-charge guanidinium ion pairing occurs whenever two side chains have a chance to interact with each other.

In order to demonstrate that the restraining procedure does not artificially produce like-charge ion pairs in dipeptides, we also performed MD simulations with parm99 and parm99SB force fields for *di*-lysine (without and with restraining). Results of MD simulations for *di*-lysine are shown in Fig. 6. The nitrogen-nitrogen radial distribution functions for unrestrained *di*-lysine (left panel), do not show any sign of like-charge pairing, in agreement with previous simulations.[20] The first peak maximum for parm99 force field occurs at the NH_3^+ - NH_3^+ distance of ca. 8 Å which corresponds to solvent separated ammonium groups. In the parm99SB force field this peak is not present, which is due to the dihedral parametrization preventing the two side chains to approach each other. Most importantly, there is no pairing in *di*-lysines (i. e. no N-N peak at 4 Å) even after the restraining procedure (right panel of Fig. 6). Only the height of peak maximum which occurs at the NH_3^+ - NH_3^+ distance of 8 Å is now somewhat higher since we have restrained the dihedral backbone angles and thus forced two side chains to be at the same side of the backbone. Therefore, we can conclude that backbone dihedral restraining is not enforcing like-charge ion pairing of NH_3^+ cations in *di*-lysine in any way.

As a final test of our assumptions that differences in guanidinium pairing are due to the different parametrization of the dihedral backbone angles, we performed molecular dynamics simulation of the tripeptides Arg-Ala-Arg and Lys-Ala-Lys with parm99 and parm99SB force fields. In the left panel of Fig. 7 RDFs for the arginine central carbon atoms in the tripeptide are shown. Interestingly, we see the opposite to the situation in *di*-arginine – with parm99 force field there is no pairing, while with parm99SB force field the guanidinium the pairing is present! It is not as strong as in the constrained dipeptide which is due to the fact that it is harder for side chains to mutually interact when alanine residue is inserted between two arginines and keeps the preference for α-helical region to be dominantly occupied (Fig. 8), nevertheless, it is still sizable. Notably, the different parametrization of dihedral backbone angles now prevents the Gdm^+ - Gdm^+ pairing in MD simulations with parm99 force field, while it is feasible for the parm99SB force field. The analysis of Ramachandran plots for the Arg-Ala-Arg tripeptide showed no significant differences in comparison to those for dipeptides, implying hindered Gdm^+-Gdm^+ pairing in the *di*-peptide but not in the tripeptide (for parm99SB) or vice versa (for parm99). This is illustrated in Fig. 8 where representative snapshots from two different simulations are presented. In addition, RDFs for the Lys-Ala-Lys tripeptide shown in the right panel of the Fig. 7 demonstrate that there is no pairing of the ammonium groups for either of the force fields under consideration.

In a recent experimental study, infrared and Raman spectra of capped amino acids have been used to determine relative populations of backbone conformations for a series of dipeptides. It was shown that *di*-arginine occupies mainly the β-extended conformational basin, with the α-helical basin being only weakly populated.[48] This is in accordance with results shown in bottom part of Fig. 4, showing the Ramachadran plot for *di*-arginine with parm99SB force field. Therefore, we can deduce that parm99SB reflects the experimental reality better than parm99 and it is adequate to use it for the conformational sampling of the *di*-arginine phase space. As a result, we can conclude that Gdm^+ - Gdm^+ pairing in *di*-arginine is weaker than previously reported (based on calculations using parm99)[20] due to sterical dihedral constraints. Nevertheless, like-charge guanidinium pairing is

observed even with the newer force fields like parm99SB in systems without such constraints, such as the Arg-Ala-Arg tripeptide.

Protein database survey

We performed an extended search through three PDB derived protein structure sets, analyzing the spatial distribution of side chains in short protein fragments containing arginine residues. Figure 9 shows the guanidinium carbon-guanidinium carbon distance distribution for two Arg residues in dipeptide protein fragments (left panels) and tripeptide protein fragments (right panels) for each protein set. The top panels show all contacts, while the bottom panels depict only situations where the two guanidinium groups are in a parallel (stacked) geometry. We see that close contacts occur both for dipeptide and tripeptide fragments, being more pronounced in the latter case, particularly for stacked geometries.

Most importantly, a detailed analysis of individual structures with two Arg residues in a close contact (below 5 Å) and stacked Gdm^+ groups shows that such arrangements occur in Arg-Arg protein fragments only when at least one residue occupies a α-helical region (see Table 1), which is in accord with the restrained molecular dynamics simulations presented above. In other words, the stacking of Gdm^+ groups in Arg-Arg protein fragments is solely due to the protein environment and it is the structured protein context which drives Gdm^+ groups in a close contact. In contrast, for Arg-X-Arg protein fragments, the same analysis shows that guanidinium pairing is *not* limited to the α-helical region (Table 1), which agrees with the above parm99SB MD results.

Ab initio calculations

In order to further examine Gdm^+ - Gdm^+ pairing as revealed by the MD simulations and the database search, *ab initio* calculations for the guanidinium dimer with an increasing number of water molecules were performed. We chose guanidinium cations as the model for the guanidinium side chain subunit in arginine since *ab initio* calculations for *di*-arginine clusters with a large number of explicit water molecules are rather impractical. In Table 2, complexation energies for guanidinium dimer with three to fourteen solvating water molecules are shown while Figure 10 presents the corresponding optimized geometries. Not surprisingly, guanidinium dimer without water is not stable due to the electrostatic repulsion between two guanidinium cations. Addition of one or two water molecules did not yield stable complexes either. It is, however, striking, that only three water molecules are necessary to obtain a locally stable gas phase dimer. Inspection of data presented in Table 2 reveals that dimer complexation energies with respect to individual fragments obtained at the B-LYP-D/aug-cc-pVDZ level with counterpoise correction are decreasing from a strongly destabilizing value of 58.6 kcal mol^{-1} for the cluster with three water molecules, to a marginally stabilizing value of -1.1 kcal mol^{-1} for the cluster with twelve water molecules. Complexation energies for clusters with thirteen and fourteen water molecules are almost identical to those of the twelve water cluster being -1.0 and -0.3 kcal mol^{-1}, respectively.

The effect of electron correlation on the dimer complexation energy is clearly visible upon comparing HF and MP2 energy values (for geometries at the B-LYP-D level of theory). Since counterpoise correction significantly depends on the geometry of the dimer, it is not included in the single point MP2 calculations and it is calculated only at the B-LYP-D level of theory. The uncorrected complexation energy for the twelve water cluster for the HF method is 22.6 kcal mol^{-1} (in comparison with -1.4 kcal mol^{-1} at the MP2 level and -4.9 kcal mol^{-1} for the B-LYP-D method). This is not surprising, since a significant part of complexation energy is due to the dispersion interaction, which is not accounted for at the Hartree-Fock level.[20,49,50]

In order to study basis set effects on the complexation energy of guanidinium dimers, we performed geometry optimizations of a guanidinium dimer with three water molecules using B-LYP-

D and MP2 methods with different basis sets (Table 3). For the MP2 method, complexation energies are not very sensitive to basis set extensions. The energy difference between cc-pVDZ and aug-cc-pVTZ basis sets is only 2.4 kcal mol^{-1}. In the case of B-LYP-D calculations this difference is slightly larger and equals to 5.6 kcal mol^{-1}. Nevertheless, for augmented basis sets, this difference in complexation energies between aug-cc-pVDZ and aug-cc-pVTZ is only 2.4 kcal mol^{-1}. This indicates that the aug-cc-pVDZ basis set is adequate for B-LYP-D calculations of larger microhydrated clusters which are difficult to calculate with more extended basis sets.

The *ab initio* calculations show that one of the reasons behind the stabilization of guanidinium dimer by water molecules is the fact that number of hydrogen bonds between each guanidinium ion and water molecules is increasing fast with the number of water molecules in the cluster. For the cluster with twelve water molecules, presented in Fig. 11, each guanidinium cation is hydrogen bonded with six water molecules. Six pairs of water molecules are also mutually hydrogen bonded (top view of Fig. 11). Moreover, three of the water pairs at each of the guanidinium cations are additionally hydrogen bonded (side view of Fig. 11). The twelve water cluster thus represents an exceptionally symmetric and well interconnected structure. Adding one or two water molecules to the second solvation shell does not influence the stability of the cluster resulting in little change in complexation energies. The remarkable stability of the guanidinium dimer in water is also due to the cavitation and dispersion effects, as exemplified in previous Polarizable Continuum Model (PCM) calculations.[20]

The equilibrium distance between central carbon atoms in the cluster with twelve water molecules equals to 3.37 Å which is similar to the value of 3.32 Å obtained in PCM optimization of the guanidinium dimer.[20] In comparison, in MD simulations of guanidinium chloride in water solutions the first peak maximum in the radial distribution function corresponds to the distance of ca. 3.75 Å.[20] The difference is mainly due to the fact that B-LYP-D optimized structure represents a minimum on the potential energy surface, while MD dimers at 300 K exhibit somewhat elongated distances due to temperature effects. This qualitative agreement between *ab initio* and MD simulations further

supports the conclusion that like-charge pairing of guanidinium groups is not just a simulation artifact. While it is not possible to directly compare cluster *ab initio* calculations with MD simulations in liquid water, it is encouraging that the strength of Gdm^+ - Gdm^+ pairing is comparable within these two approaches.

Conclusions

In the present computational study we explored pairing between guanidinium moieties in small arginine-rich peptides. We showed that guanidinium like-charged ion pairing is not present in *di*-arginines in a water solution in MD simulations with the newer parm99SB, parm03 and parm10 force field, whereas older parm99 and pol-parm99 force fields exhibit this pairing due to different parametrization of the backbone dihedral angles. However, upon overruling the dihedral constraints this pairing is reestablished in parm99SB force field as well. This supports the conclusion that two guanidinium ions tend to pair in water if they are not sterically prevented from being close to each other. In contrast, no MD simulation showed any of NH_3^+-NH_3^+ pairing in *di*-lysine or in Lys-Ala-Lys.

Results for conformational dynamics of short peptides obtained with parm99SB force field are in qualitatively good agreement with experimental IR and Raman studies[48] contrary to those obtained using parm99. This implies that Gdm^+ - Gdm^+ pairing is likely not to be present in *di*-arginine due to steric backbone constraints. Nevertheless, in the Arg-Ala-Arg tripeptide the backbone steric constraints in newer force fields like parm99SB do not prevent Gdm^+ - Gdm^+ pairing and, consequently, this like-charge pairing is observed in simulations using these potentials. These simulation results are also supported by a survey of three experimental protein structure sets. This search shows that stacking of Gdm^+ groups in Arg-Arg fragments in proteins is a consequence of structured α-helical protein environment which drives them close together. In contrast, stacking of

guanidinium groups in Arg-X-Arg protein fragments is possible also in structures without α-helical restraints.

Further support for the MD results is provided by *ab initio* calculations of clusters containing a guanidinium dimer with increasing number of solvating water molecules. We showed that locally a stable cluster is obtained already for three water molecules. The calculations also demonstrate the gradual increase of the stability of the Gdm^+ - Gdm^+ pair with increasing number of water molecules all the way to globally stable cluster with twelve water molecules. The B-LYP method in conjunction with dispersion correction, gives complexation energy of guanidinium dimer with twelve water molecules with respect to two fragments, containing the cation with six water molecules, of -1.1 kcal mol^{-1}. This value suggests that such a system could be stable at very low temperatures.

Acknowledgments

We thank Phil Mason for valuable discussions. Support from the Czech Ministry of Education (Grants LC512 and LH11020), the Czech Science Foundation (Grant 203/08/0114), and the Academy of Sciences (PraemiumAcademie) is gratefully acknowledged. J. H. thanks the International Max-Planck Research School for support. Part of the work at the Academy of Sciences was supported via Project Z40550506.

References

(1) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*; W H Freeman: New York, 2002.

(2) Collins, K. D.; Washabaugh, M. W. *Quarterly Reviews of Biophysics* **1985**, *18*, 323-422.

(3) Jungwirth, P.; Winter, B. In *Annual Review of Physical Chemistry*; Leone, S. R., Groves, J. T., Ismagilov, R. F., Richmond, G., Eds. 2008; Vol. 59, p 343-366.

(4) Heyda, J.; Vincent, J. C.; Tobias, D. J.; Dzubiella, J.; Jungwirth, P. *Journal of Physical Chemistry B* **2010**, *114*, 1213-1220.

(5) Heyda, J.; Pokorná, J.; Vrbka, L.; Vácha, R.; Jagoda-Cwiklik, B.; Konvalinka, J.; Jungwirth, P.; Vondrášek, J. *Physical Chemistry Chemical Physics* **2009**, *11*, 7599-7604.

(6) Wondrak, E. M.; Louis, J. M.; Oroszlan, S. *FEBS Letters* **1991**, *280*, 344-346.

(7) Zhao, H. *Journal of Molecular Catalysis B: Enzymatic* **2005**, *37*, 16-25.

(8) Bilanicova, D.; Salis, A.; Ninham, B. W.; Monduzzi, M. *Journal of Physical Chemistry B* **2008**, *112*, 12066-12072.

(9) Auffinger, P.; Bielecki, L.; Westhof, E. *Structure* **2004**, *12*, 379-388.

(10) Magalhaes, A.; Maigret, B.; Hoflack, J.; Gomes, J. N. F.; Scheraga, H. A. *Journal of Protein Chemistry* **1994**, *13*, 195-215.

(11) Pednekar, D.; Tendulkar, A.; Durani, S. *Proteins: Structure, Function and Bioinformatics* **2009**, *74*, 155-163.

(12) Vondrasek, J.; Mason, P. E.; Heyda, J.; Collins, K. D.; Jungwirth, P. *Journal of Physical Chemistry B* **2009**, *113*, 9041-9045.

(13) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proceedings of the National Academy of Sciences of the United States of America* **1976**, *73*, 804-808.

(14) Riordan, J. F.; McElvany, K. D.; Borders Jr, C. L. *Science* **1977**, *195*, 884-886.

(15) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *Journal of Molecular Biology* **1994**, *239*, 315-331.

(16) Jairajpuri, M. A.; Azain, N.; Baburaj, K.; Bulliraju, E.; Durani, S. *Biochemistry* **1998**, *37*, 10780-10791.

(17) Onishi, H.; Ohki, T.; Mochizuki, N.; Morales, M. F. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 15339-15344.

(18) Cymes, G. D.; Ni, Y.; Grosman, C. *Nature* **2005**, *438*, 975-980.

(19) No, K. T.; Nam, K. Y.; Scheraga, H. A. *Journal of the American Chemical Society* **1997**, *119*, 12917-12922.

(20) Vondrášek, J.; Mason, P. E.; Heyda, J.; Collins, K. D.; Jungwirth, P. *Journal of Physical Chemistry B* **2009**, *113*, 9041-9045.

(21) Boudon, S.; Wipff, G.; Maigret, B. *Journal of Physical Chemistry* **1990**, *94*, 6056-6061.

(22) Soetens, J. C.; Millot, C.; Chipot, C.; Jansen, G.; Ángyán, J. G.; Maigret, B. *Journal of Physical Chemistry B* **1997**, *101*, 10910-10917.

(23) Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Saboungi, M. L.; Dempsey, C. E.; MacKerell Jr, A. D.; Brady, J. W. *Journal of the American Chemical Society* **2004**, *126*, 11462-11470.

(24) Hunger, J.; Niedermayer, S.; Buchner, R.; Hefter, G. *Journal of Physical Chemistry B* **2010**, *114*, 13617-13627.

(25) Sapse, A. M.; Massa, L. J. *Journal of Organic Chemistry* **1980**, *45*, 719-721.

(26) Mason, P. E.; Neilson, G. W.; Dempsey, C. E.; Barnes, A. C.; Cruickshank, J. M. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 4557-4561.

(27) Greene Jr, R. F.; Pace, C. N. *Journal of Biological Chemistry* **1974**, *249*, 5388-5393.

(28) Mason, P. E.; Brady, J. W.; Neilson, G. W.; Dempsey, C. E. *Biophysical Journal* **2007**, *93*.

(29) Wang, J.; Cieplak, P.; Kollman, P. A. *Journal of Computational Chemistry* **2000**, *21*, 1049-1074.

(30) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function, and Genetics* **2006**, *65*, 712-725.

(31) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *Journal of Computational Chemistry* **2003**, *24*, 1999-2012.

(32) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. In *AMBER 11*.

(33) Perera, L.; Berkowitz, M. L. *Journal of Chemical Physics* **1994**, *100*, 3085-3093.

(34) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *Journal of Physical Chemistry* **1987**, *91*, 6269-6271.

(35) Caldwell, J. W.; Kollman, P. A. *Journal of Physical Chemistry* **1995**, *99*, 6208-6219.

(36) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *Journal of Chemical Physics* **1995**, *103*, 8577-8593.

(37) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684-3690.

(38) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327-341.

(39) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577-2637.

(40) Lovell, S. C.; Davis, I. W.; Arendall III, W. B.; De Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins: Structure, Function, and Genetics* **2003**, *50*, 437-450.

(41) van der Kamp, M. W.; Schaeffer, R. D.; Jonsson, A. L.; Scouras, A. D.; Simms, A. M.; Toofanny, R. D.; Benson, N. C.; Anderson, P. C.; Merkley, E. D.; Rysavy, S.; Bromley, D.; Beck, D. A. C.; Daggett, V. *Structure* **2010**, *18*, 423-435.

(42) Van Duijneveldt, F. B.; Van Duijneveldt-Van De Rijdt, J. G. C. M.; Van Lenthe, J. H. *Chemical Reviews* **1994**, *94*, 1873-1885.

(43) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553-566.

(44) Grimme, S. *Journal of Computational Chemistry* **2004**, *25*, 1463-1473.

(45) Grimme, S. *Journal of Computational Chemistry* **2006**, *27*, 1787-1799.

(46) In *TURBOMOLE V6.2*, 2010, at *http://www.turbomole.com*.

(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.;

Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Kiene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian09, Revision A.1* **2009**.

(48) Grdadolnik, J.; Mohaček-Grošev, V.; Baldwin, R. L.; Avbelj, F. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 1794-1798.

(49) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *Journal of the American Chemical Society* **2002**, *124*, 104-112.

(50) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *Journal of Chemical Physics* **2004**, *120*, 647-659.

Table 1. Numbers of stacked conformations with CZ – CZ distance in Gdm^+ groups below 5 Å vs. total number of conformations for all regions and β-extended region only of Ramachandran plots in Arg-Arg and Arg-X-Arg fragments in proteins extracted from the three protein databases.

	all regions	β-extended region
Arg-Arg – Dyn	1 / 1	0 / 1
Arg-Arg – Set1	6 / 6	0 / 6
Arg-Arg – Top500	2 / 2	0 / 2
Arg-X-Arg – Dyn	4 / 4	0 / 4
Arg-X-Arg – Set1	2 / 10	8 / 10
Arg-X-Arg – Top500	0 / 4	4 / 4

Table 2. Complexation energies ΔE_s for guanidinium dimer clusters with different number of water molecules (given in kcal mol^{-1}) calculated by B-LYP-D/aug-cc-pVDZ (B-LYP-D), HF/aug-cc-pVDZ (HF) and MP2/aug-cc-pVDZ//B-LYP-D/aug-cc-pVDZ (MP2) levels of theory.

	number of water molecules											
ΔE_s	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
B-LYP-D	56.6	50.5	48.9	46.4	38.8	31.2	23.6	16.4	7.0	-4.9	-4.9	-4.4
HF	61.9	-[a]	-[a]	52.4	46.5	40.6	34.7	34.1	30.5	22.6	20.4	18.5
MP2	47.9	51.9	47.9	46.0	38.5	30.8	23.3	17.3	9.3	-1.4	-2.3	-2.7
CP[b]	2.0	2.0	1.9	2.1	1.9	2.0	2.1	3.2	3.8	3.8	3.9	4.1
B-LYP-D_{CP}[c]	58.6	52.6	50.8	48.6	40.7	33.1	25.7	19.6	10.9	-1.1	-1.0	-0.3

a) Not locally stable at the HF/aug-cc-pVDZ level of theory.

b) Counterpoise correction is calculated at the B-LYP-D/aug-cc-pVDZ level of theory.

c) Complexation energies with counterpoise correction.

Table 3. Complexation energies ΔE_c for guanidinium dimer clusters with three water molecules (given in kcal mol^{-1}) calculated by B-LYP-D and MP2 levels of theory with different basis sets.

ΔE_c	MP2	B-LYP-D
cc-pVDZ	55.5	51.5
aug-cc-pVDZ	54.8	56.6
cc-pVTZ	56.7	54.7
aug-cc-pVTZ	57.2	57.1

Figure Captions

Figure 1. Radial distribution functions *g(r)* for the central carbon atoms of the Gdm^+ groups in the arginine residues of *di*-arginine for different force fields.

Figure 2. Spatial distributions of the carbon of the guanidinium group (cyan cloud), water oxygen (red cloud), water hydrogen (gray cloud), and chloride counterions (golden cloud) around the other guanidinium group in diarginine for MD simulations with parm99 vs. parm99SB force fields. This side chain is shown in space-filling representation, while the rest of the dipeptide is depicted in licorice representation. The snapshots represent typical geometries of diarginine within given the two force fields.

Figure 3. A schematic representation of dihedral backbone angles (torsion angles) ϕ_1, ψ_1, ϕ_2, and ψ_2 for *di*-arginine.

Figure 4. Ramachandran plots of populations of dihedral backbone angles ϕ_1, ψ_1, ϕ_2, and ψ_2 for two arginine residues obtained with parm99 and parm99SB force field simulations.

Figure 5. Radial distribution functions *g(r)* for the central carbon atoms of the Gdm^+ groups in the arginine residues of *di*-arginine with restrained (-r) backbone dihedrals.

Figure 6. Radial distribution functions *g(r)* for the central nitrogen atoms of the NH_3^+ groups in the lysine residues of *di*-lysine without (left) and with (right) restrained (-r) backbone dihedrals.

Figure 7. Radial distribution functions *g(r)* for the central carbon atoms of the Gdm^+ groups in the arginine residues in the Arg-Ala-Arg tripeptide (left) and for the central nitrogen atoms of the NH_3^+ groups in the lysine residues of the Lys-Arg-Lys tripeptide (right).

Figure 8. Representative snapshots of tripeptides Arg-Ala-Arg from molecular dynamics simulations with parm99 (left) and parm99SB (right) force fields.

Figure 9. Numbers of counts in three protein structure sets vs. distance of the central carbon atoms of the Gdm^+ groups in the Arg-Arg protein fragments (upper left panel) and Arg-X-Arg protein fragments (upper right panel). Also shown are the numbers of counts for Arg-Arg protein fragments (lower left panel) and Arg-X-Arg protein fragments (lower right panel) of structures which fall outside the helical region in Ramachandran plots.

Figure 10. Optimized structures of the guanidinium dimer with increasing number of water molecules from three to fourteen (3W - 14W).

Figure 11. The symmetric optimized structure of the guanidinium dimer with twelve water molecules in three different perspectives with hydrogen bonding underscored by thin lines.

Figures

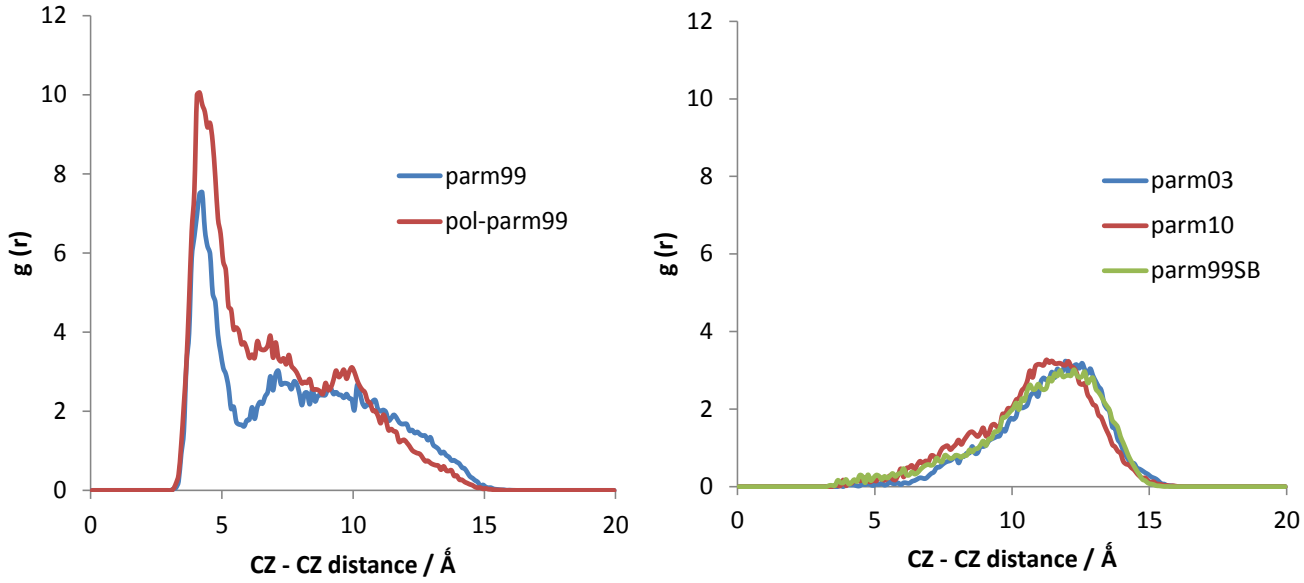

Figure 1.

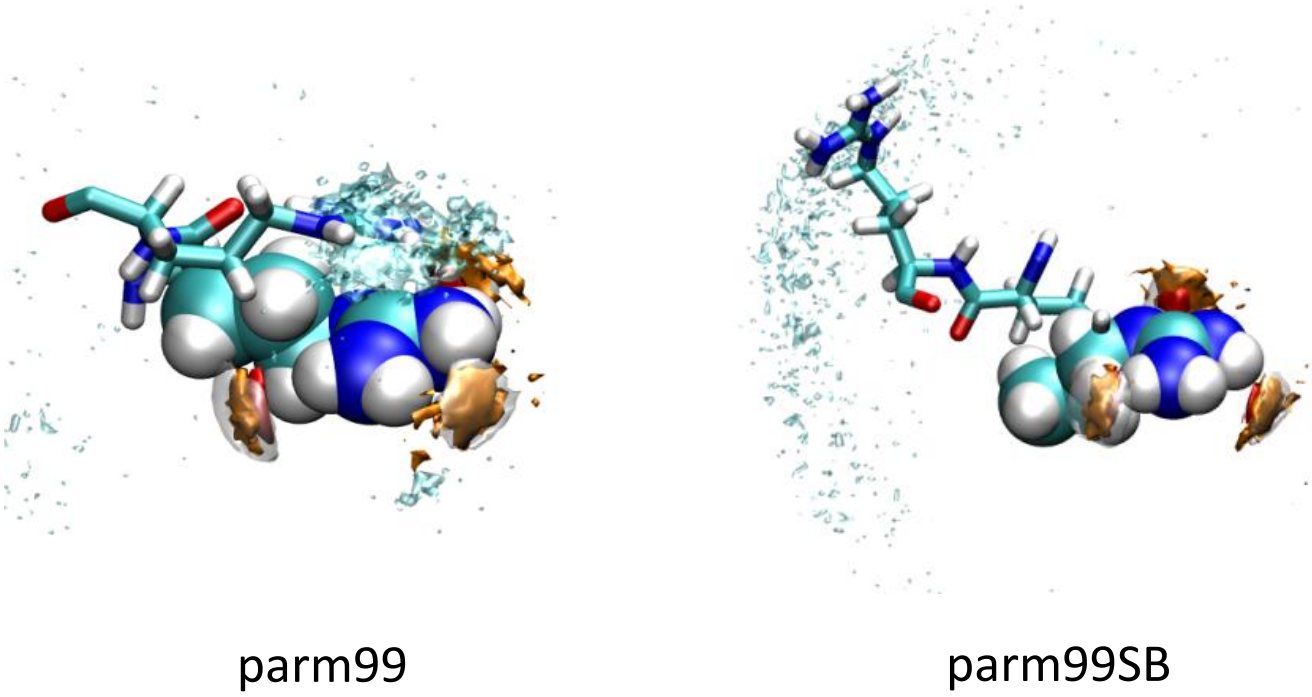

Figure 2.

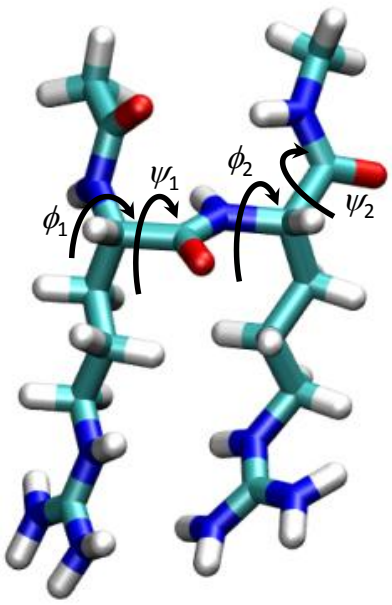

Figure 3.

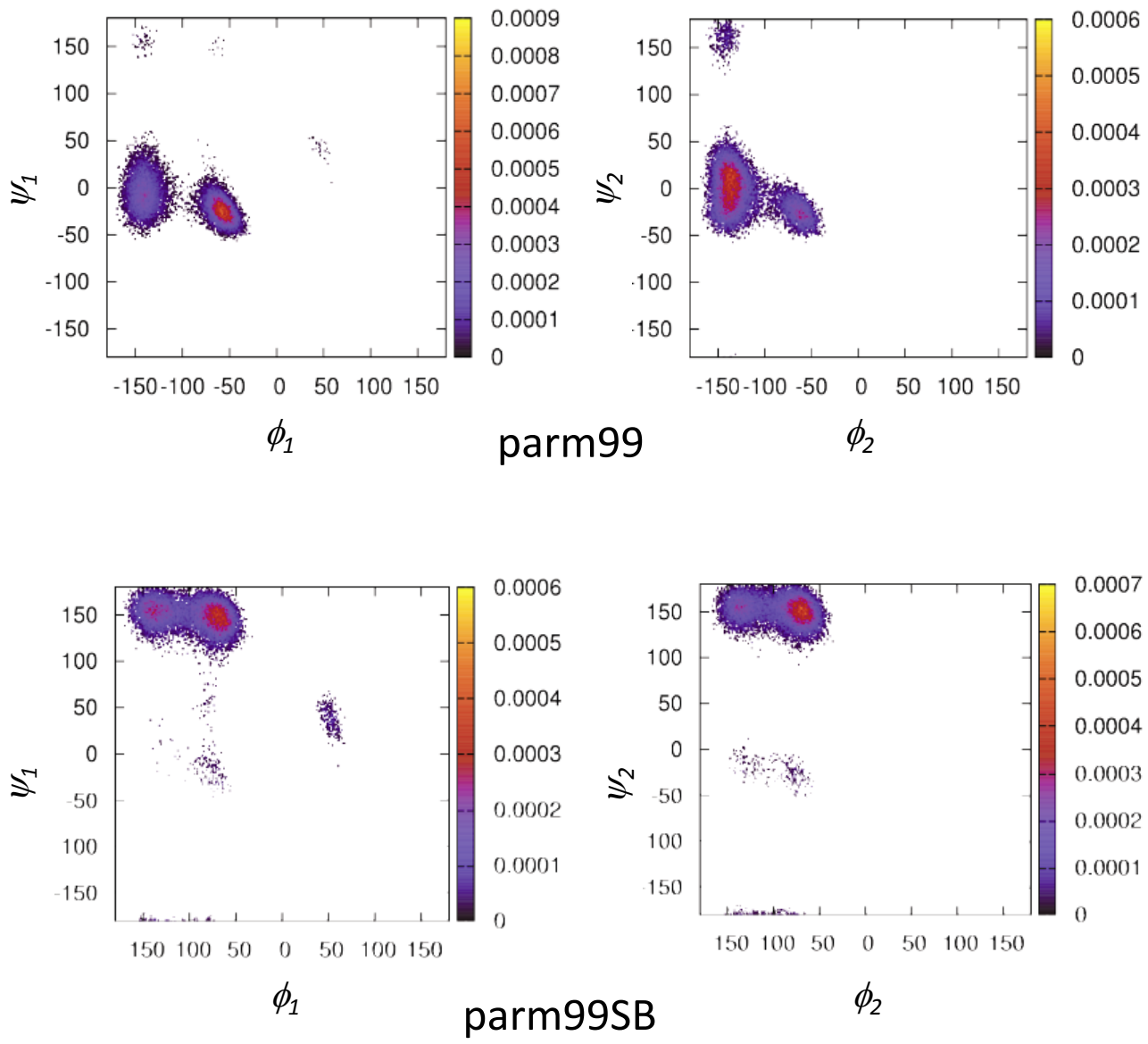
150
100
50
0
-50
-100
-150
ψ_1
-150 -100 -50 0 50 100 150
ϕ_1
0.0009
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0
ψ_2
ϕ_2
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
parm99
parm99SB
0.0007

Figure 4.

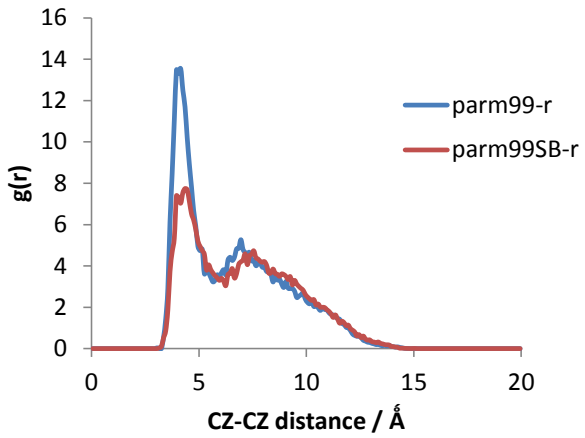
16
14
12
10
8
6
4
2
0
g(r)
0
5
10
15
20
CZ-CZ distance / Å
parm99-r
parm99SB-r

Figure 5.

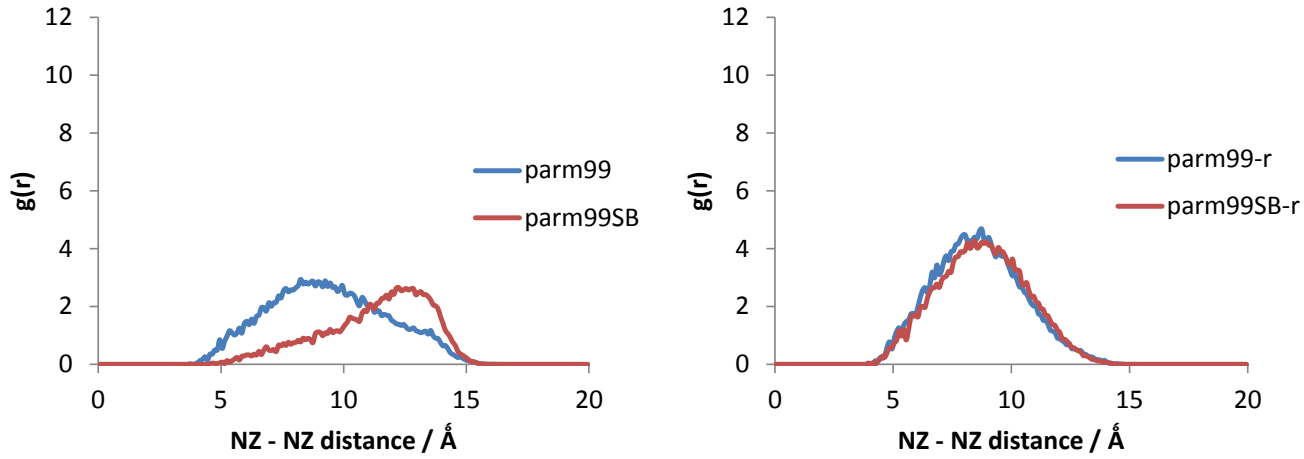
12
10
8
6
4
2
0
g(r)
0
5
10
15
20
NZ - NZ distance / Å
parm99
parm99SB
parm99-r
parm99SB-r

Figure 6.

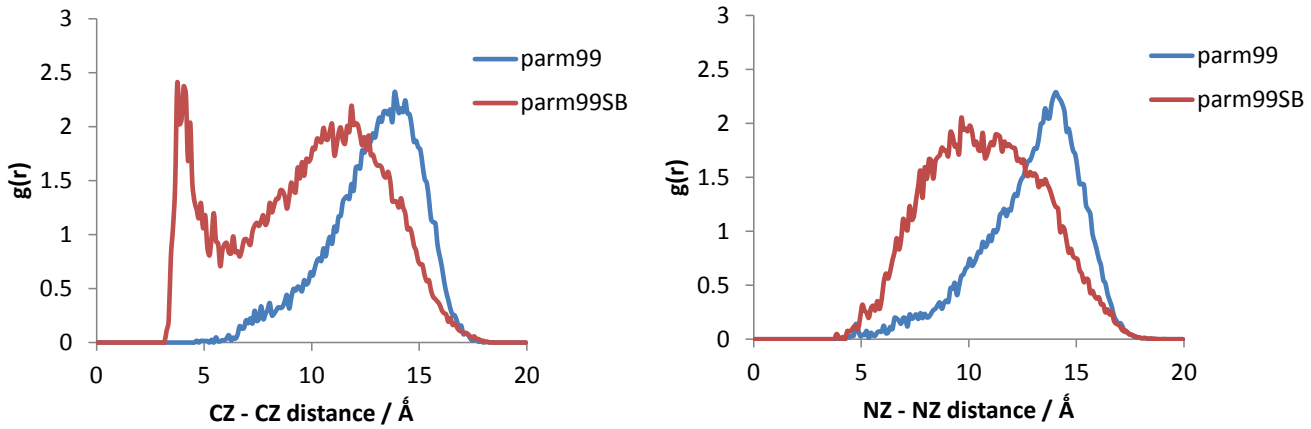
3
2.5
2
1.5
1
0.5
0
g(r)
0
5
10
15
20
CZ - CZ distance / Å
parm99
parm99SB
3
2.5
2
1.5
1
0.5
0
g(r)
0
5
10
15
20
NZ - NZ distance / Å
parm99
parm99SB

Figure 7.

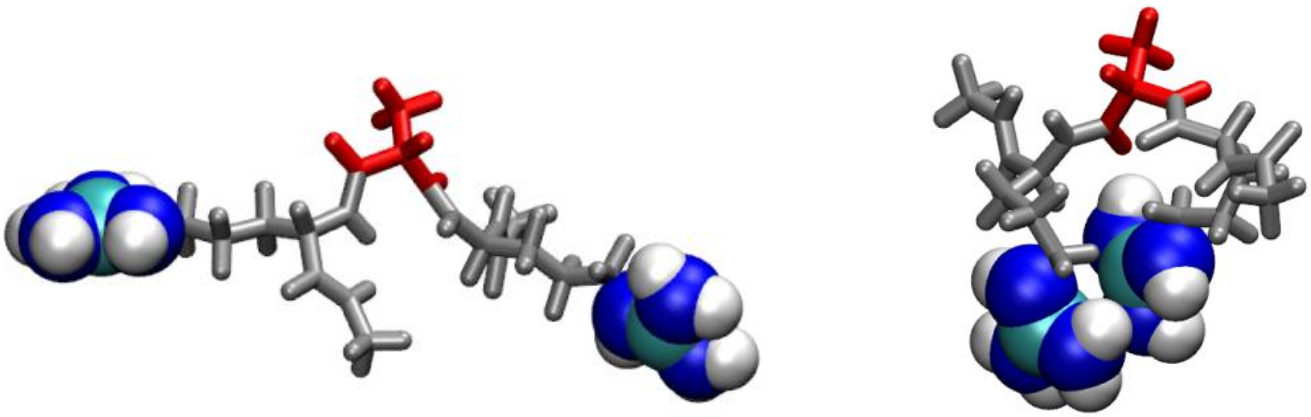

Figure 8.

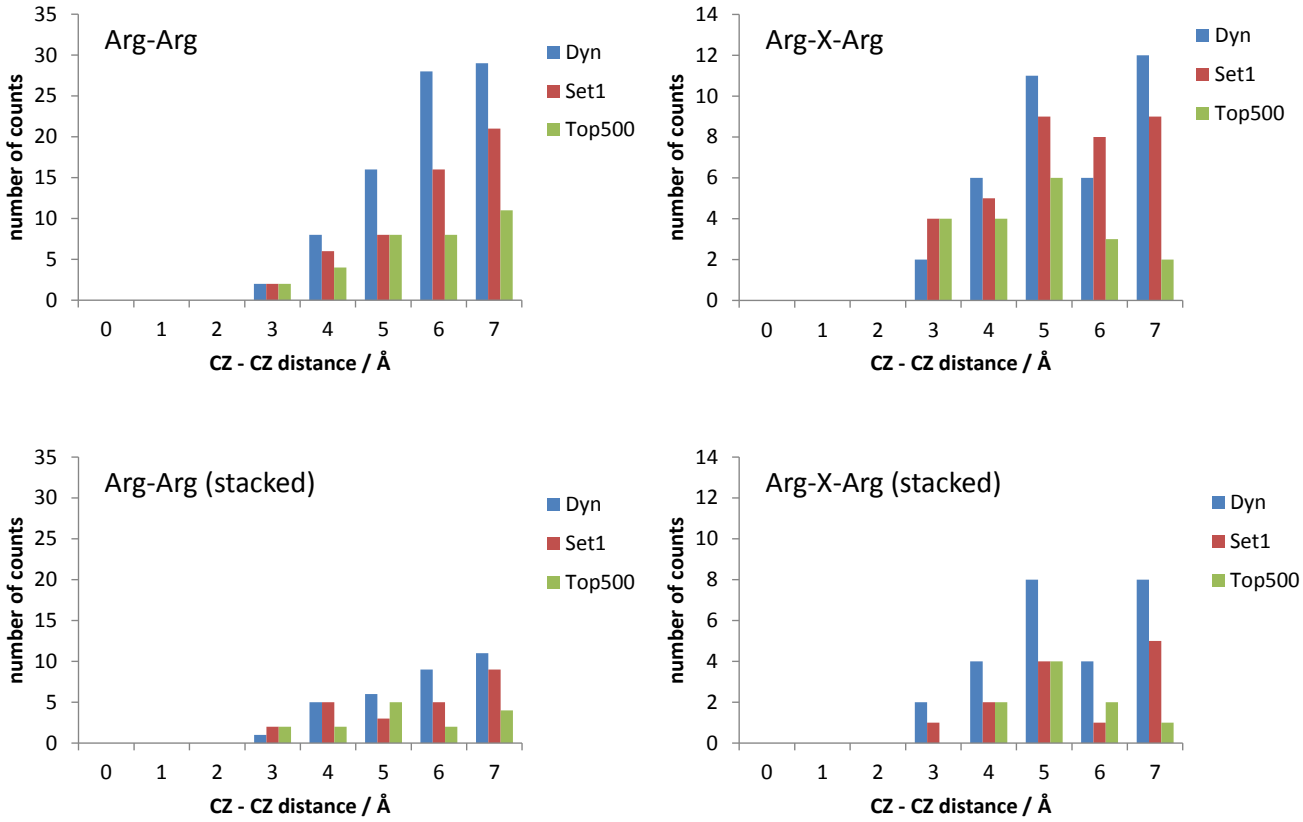
Arg-Arg
number of counts
CZ - CZ distance / Å
Dyn
Set1
Top500
Arg-X-Arg
number of counts
CZ - CZ distance / Å
Dyn
Set1
Top500
Arg-Arg (stacked)
number of counts
CZ - CZ distance / Å
Dyn
Set1
Top500
Arg-X-Arg (stacked)
number of counts
CZ - CZ distance / Å
Dyn
Set1
Top500

Figure 9.

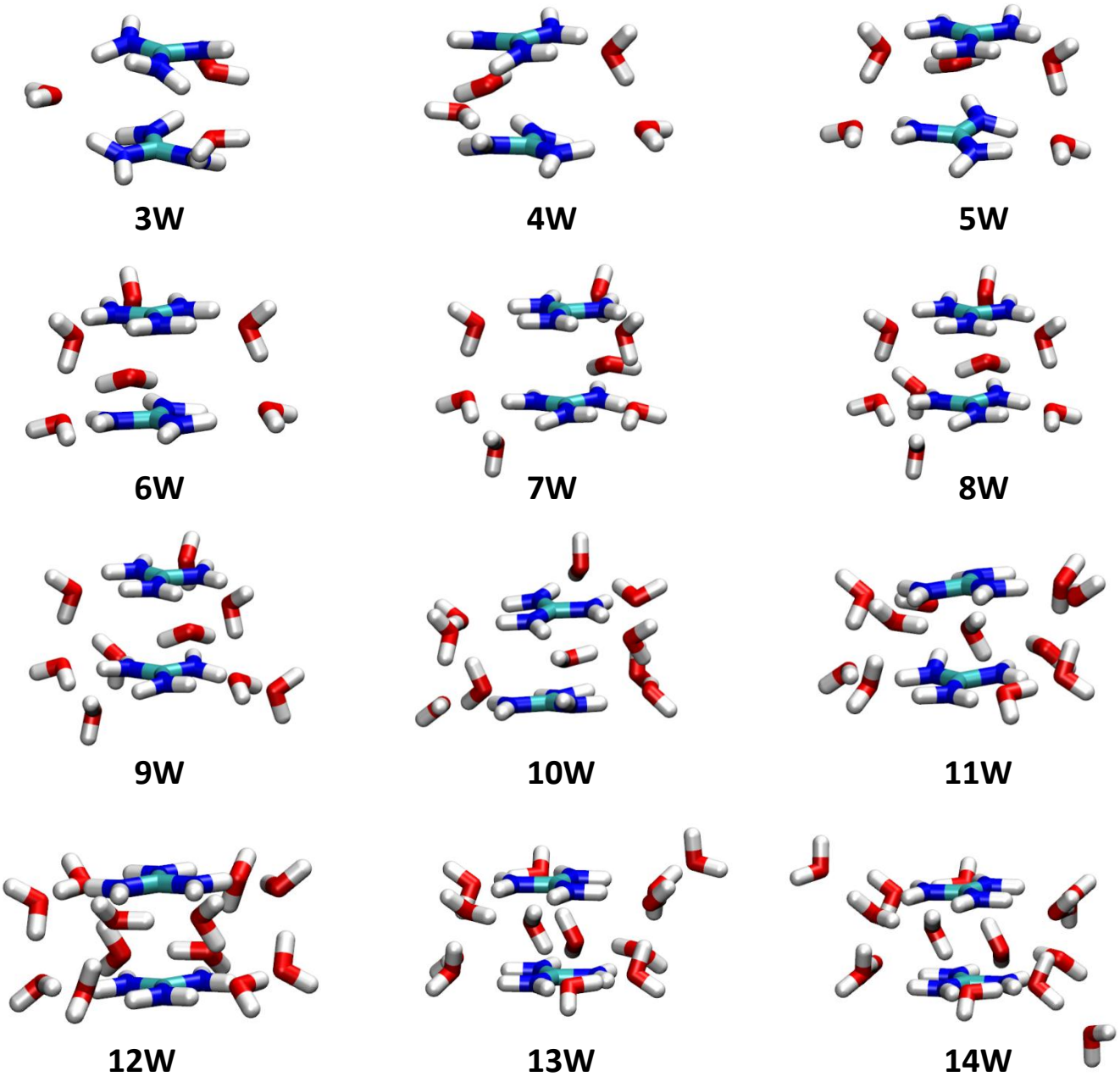

Figure 10.

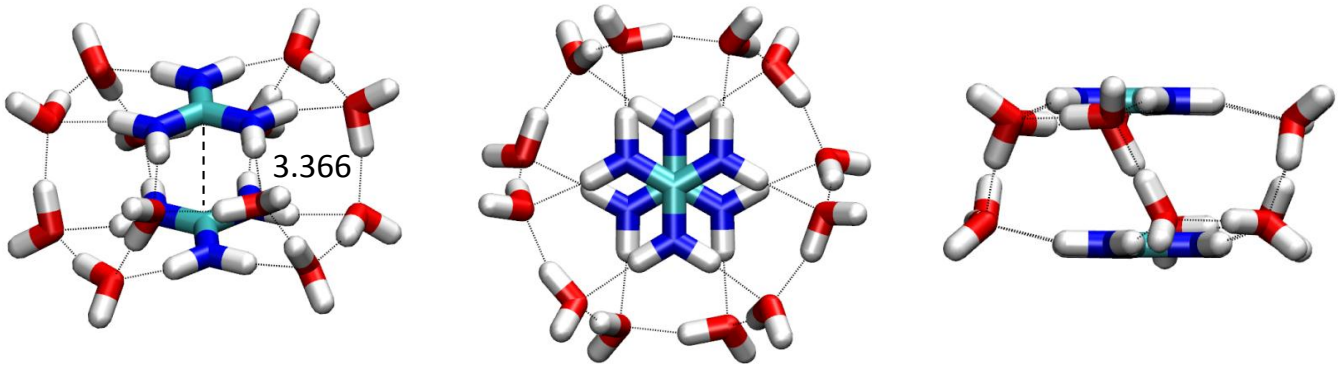

Figure 11.

TOC