The molecular origin of like charge arginine-arginine pairing in water

Jiří Vondrášek,[1] Philip E. Mason,[2] Jan Heyda,[1] Kim D. Collins,[3] and Pavel Jungwirth[1*]

[1]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and Center for Biomolecules and Complex Molecular Systems. Flemingovo nám. 2,16610 Prague 6, Czech Republic.[2]Department of Food Sciences, Stocking Hall, Cornell University, Ithaca, NY, 14853, USA. [3]Center of Marine Biotechnology and Medical Biotechnology Center, University of Maryland Biotechnology Institute, 725 W. Lombard, Baltimore, MD 21201, USA.

RECEIVED DATE (automatically inserted by publisher); pavel.jungwirth@uochb.cas.cz

Abstract

Molecular dynamics simulations show significant like-charge pairing of guanidinium side chains in aqueous poly-arginine, while this effect is absent in aqueous poly-lysine. This behavior of the guanidinium group is revealed also by protein database searches, having important biochemical implications. Combination of molecular dynamics simulations with explicit solvent and ab initio calculations employing a polarizable continuum model of water allows to rationalize the formation of contact ion pairs between guanidinium cations in terms of individual interactions at the molecular level.

Introduction

Ions play critical roles in biological systems as components of metabolites and of macromolecules such as proteins, membranes and nucleic acids, as well as in the form of small mobile species such as Na^+, K^+, and Ca^{2+}, which provide electroneutrality and create concentration gradients across membranes.[1] In an aqueous environment, biological macromolecules are subject to a mixture of forces arising from water affinity, cavitation, dispersion, and other effects that can outweigh direct electrostatic effects. This leads to a series of "electrostatics-defying" biological structures such as anions bound to anionic protein surfaces, and the bases of DNA, occurrence of almost completely anionic protein surfaces, and arginine-arginine pairing via positively charged guanidinium (Gdm^+) groups within and between protein subunits.[2-5] While it is prohibitively expensive in energetic terms to bury a charged arginine residue in the low dielectric region at the middle of a cellular membrane[6,7] or in the tightly packed nonpolar matrix connecting the cross beta-strands of protein amyloid,[8] many arginines within and between protein subunits appear to experience the high dielectric constant that allows non-electrostatic effects to dominate. In water, Gdm^+ has been shown by neutron diffraction to be weakly hydrated[9] and calculations have also indicated that in aqueous solutions Gdm^+ ions tend to form like-charged contact ion pairs.[10-12] The flat surface of the guanidinium group is large enough that its dehydration releases about two interfacial water molecules from each side to become strongly interacting bulk water.

In this study, we present results from molecular dynamics (MD) simulations of poly-arginine showing pairing of like-charged side chains. We contrast this behavior to that of poly-lysine which exhibits a lack of such pairing. Combining MD simulations employing explicit solvent with ab initio calculations of ions in a polarizable continuum model (PCM) of water we rationalize this effect at a molecular level. Additionally, using analysis of structural databases, we relate the present findings to arginine-arginine interactions in proteins. Despite the fact that these interactions are abundant,[5] their role for protein function is not fully understood yet.

Methods

We have performed 50 ns molecular dynamics trajectories of *di*-arginine and *di*-lysine and 10 ns trajectories of *deca*-arginine and *deca*-lysine (all with a 1 fs time step, with Berensen coupling[13] constraining the system at 300 K and 1 atm) in water, after 1 ns equilibration. The periodic box contained a single peptide described using the parm99 force field,[14] 866 (*di*-peptides) or 4513 (*deca*-peptides) SPC/E water molecules,[15] and two or ten chloride counterions.[16] All water O-H bonds were constrained to their optimal values using the SHAKE algorithm.[17] Long range electrostatic interactions beyond the 10 Å cutoff were accounted for via the particle-mesh Ewald procedure[18] with the relative error due truncation of the reciprocal vector space set below $5x10^{-5}$ and the relative error in the direct part of the Ewald sum set below 10^{-5}. For each trajectory, 10000 snapshots provided input for consequent analysis.

Ab initio calculations of homo-ion pairing were performed emplying the Polarizable Continuum Model (PCM) of water[19] and, for comparison and robustness check, also using the COSMO model.[20] For cationic pairs we employed the cc-pvtz basis, while for the nitrate pair the aug-cc-pvdz basis set. By comparison to test CCSD(T) calculations for the guanidinium-guanidinium pair we found the results to be converged within 0.5 kcal/mol at the MP2 level, which was then employed for all ion pairs. The ab initio calculations allow a dissection of the factors that determine the energetic viability of a homo-ion pair. MD and ab initio calculations were performed using Amber 8[21] and Gaussian 03,[22] respectively.

Results and Discussion

MD simulations of both $(Arg)_2$ and $(Arg)_{10}$ reveal that the guanidinium groups of the side chains tend to associate. This is apparent both from a visual inspection of the trajectory and from a strong first peak (around 4 Å) of the radial distribution function of the central carbon atoms of the guanidinium groups (Figure 1). In contrast, neither $(Lys)_2$ or $(Lys)_{10}$ exhibit any direct pairing of ammonium-containing side chains, despite the fact that they are one hydrophobic CH_2 group longer (Figure 1).

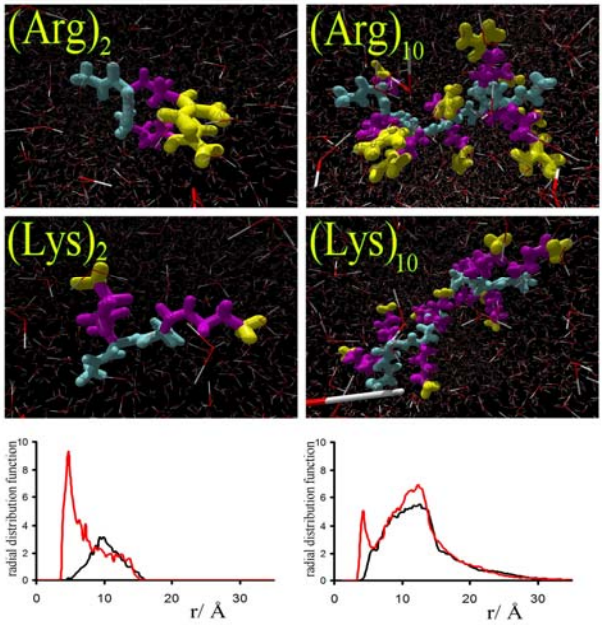


Figure 1: Snapshots from the MD simulation of *di-* and *deca-*arginine and lysine (cationic group in yellow, side chain in purple and backbone in cyan). Lower panels show the radial distribution functions g(r) for the central atom of the cationic group (left for the dimer, right for the decamer: in both cases the arginine species is shown in red and the lysine species in black).

Figure 2 further demonstrates the like-charge pairing in $(Arg)_2$ by plotting the distribution of the Gdm^+ group, as well as water molecules and chloride counterions around the other Gdm^+ group. The two diffuse lobes occurring above and below the Gdm^+ group are a clear signature of like-charge pairing. In contrast, water molecules and counter-ions tend to bind in the plane of the Gdm^+ group.

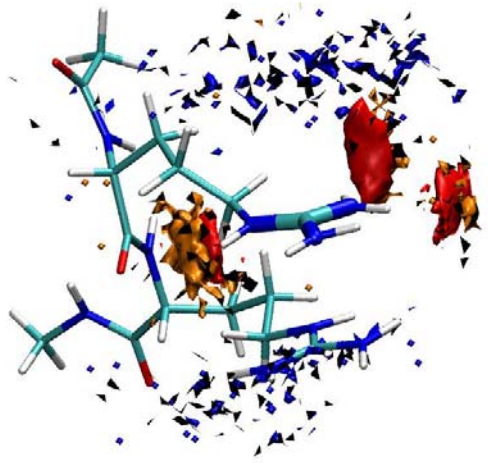


Figure 2: Distribution of the Gdm^+ group (blue diffuse regions demonstrating like charge pairing), water oxygens (red lobes), and chloride counterions (gold lobes) around the other Gdm^+ group of *di-*arginine (thick licorice). The molecular structure depicts a representative geometry of *di*-arginine.

While the observed repulsion between the ammonium groups of $(Lys)_2$ (Figure 1) can be relatively easily rationalized on electrostatic grounds, the attraction observed between the positively charged guanidinium based side chains of $(Arg)_2$ may seem counter-intuitive. Nevertheless this result is born out not only from classical MD simulations but also from ab initio calculations employing a PCM solvent presented below.

Table 1 shows the free energy of association of like-charged ions at center-of-mass distance of 3.32 Å corresponding to the Gdm^+-Gdm^+ minimum in PCM water. The free energy minimum of a Gdm^+-Gdm^+ pair in water amounts to -2.1 kcal/mol at the MP2/cc-pvtz level. This number is stable within 0.5 kcal/mol with respect to further basis set extension and further inclusion of correlation effects at the CCSD(T) level. However, test calculations at the Hartree-Fock level which lacks correlation effects such as dispersion bring the free energy of association close to zero.

Table 1 The ab initio energy (in kcal/mol) of like-charged ion pairs optimized at a separation of 3.32 Å in water and in the gas phase. Last line which represents a direct application of the Coulomb law.

	in water	in the gas phase
$Gdm^+\cdots Gdm^+$	-2.1	+65.6
$NH_4^+\cdots NH_4^+$	+7.4	+91.8
$Na^+\cdots Na^+$	+5.6	+99.9
$NO_3^-\cdots NO_3^-$	+2.3	+79.3
$+\cdots+$	+1.2	+99.8

The attractive potential of the aqueous Gdm^+-Gdm^+ homo-ion pair turns out to be rather exceptional (at least among ions which do not contain hydrophobic groups). In contrast, the NH_4^+- NH_4^+ pair is strongly repulsive by 7.4 kcal/mol, slightly more than a Na^+-Na^+ pair. A NO_3^--NO_3^- pair, which is geometrically similar to the Gdm^+-Gdm^+ pair, is much less repulsive (+2.3 kcal/mol) than pairs of spherical or almost spherical ions such as sodium or ammonium. Note that a pair of point charges at a separation of 3.32 Å is repelled by 1.2 kcal/mol according to the Coulomb law at $\varepsilon = 80$, however, this repulsion would increase assuming a lower effective dielectric constant in the vicinity of the ion pair.

In addressing the origin of the Gdm^+-Gdm^+ attraction in water it is useful to examine also the potential of these ion pairs in the gas phase (Table 1). If two single positive charges are brought in vacuum 3.32 Å apart they have a repulsive Coulomb potential of 99.8 kcal/mol. When these point charges are replaced with sodium atoms there is effectively no change in the repulsive potential of this ion pair, while for an ammonium pair this repulsion is slightly reduced by about 8 kcal/mol. For the Gdm^+ and NO_3^- ion pairs, however, there is a substantial reduction in the gas phase repulsive potential by about 35 and 20 kcal/mol, respectively. While NH_4^+ is an almost spherical ion, Gdm^+ and NO_3^- are flat dendritic structures which can decrease the Coulomb repulsion in a like charge pair by assuming a staggered geometry, which optimizes the quadrupole-quadrupole interactions (as does, albiet to a lesser extent, a T-shaped configuration). The optimal staggered structure of the Gdm^+-Gdm^+ pair in PCM

water, together with the electron density and charge distribution, is depicted in Figure 3.

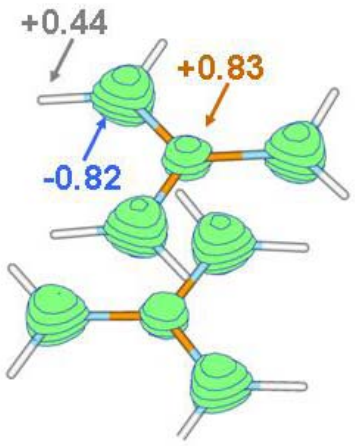


Figure 3: Geometry, charge distribution (from natural population analysis), and electron density of the optimal staggered Gdm^+-Gdm^+ pair in PCM water.

The interaction of the homo-ion pairs in PCM water can be broken down into individual contributions, i.e., mean field ion-ion interaction (at the Hartree-Fock level) plus change in solvent polarization, change in cavitation (i.e., the energy to form a cavity for the solute), ion-ion dispersion and other correlation effects (i.e., the difference between MP2 and Hartree-Fock energies), and change in ion-solvent dispersion, as presented in Table 2. As a robustness check we perfomed analogous calculations employing the COSMO model with slightly different ionic radii, which semi-quantitatively reproduces both homo-ion pairing and decomposition into individual interaction terms.

Table 2 Individual contributions to the like charge interaction (in kcal/mol) from ab initio calculations with a PCM water.

	ion-ion mean field & change in solvent polarization	change in cavi-tation	ion-ion dispersion & other correlation effects	change in ion-solvent dispersion
$Gdm^+\cdots Gdm^+$	+2.7	-3.1	-5.0	+3.6
$NH_4^+\cdots NH_4^+$	+8.6	-1.8	-0.9	+1.5
$Na^+\cdots Na^+$	+6.0	0.0	0.0	-0.4
$NO_3^-\cdots NO_3^-$	+4.3	-0.8	-3.6	+2.7

It follows from Table 2 that the main difference between interactions of disc-shaped (Gdm^+ and NO_3^-) and sphere-like (NH_4^+ and Na^+) ions comes from the ion-ion electrostatics and solvent polarization. Note that the more favorable electrostatic interactions of the former species due to ion geometry considerations were present already in the gas phase. In water these are screened due to the high dielectric constant of the solvent, nevertheless qualitatively the effect pertains (see Figure 3). The sum of the last two columns in Table 2 essentially gives the excess dispersion of the ions in water. We see that net dispersion effects are stabilizing by about -1 kcal/mol for the quasi-aromatic ions Gdm^+ and NO_3^-, while they are negligible for Na^+ and NH_4^+. Finally, cavitation is stabilizing for all investigated ions (being, however, negligible for Na^+ where the cavities practically do not overlap at a separation of 3.32 Å). This solvent exclusion effect, which is the most stabilizing for the Gdm^+-Gdm^+ pair is, in the PCM framework, due to the reduction of the total cavity volume upon ion complexation.

We have also performed PCM calculations of guanidinium-guanidinium pairing in solvents less polar than water, possibly mimicking a more hydrophobic environment of the proteins. As a matter of fact, it is hard to assign a single dielectric constant to the protein environment (and it might not even be fully meaningful due to the electrostatic non-homogeneity of proteins). Nevertheless, we did calculations for two additional values of solvent relative permitivity, 4 and 20. While in the former case, the Coulomb repulsion between two guanidinium cations was reduced from the gas phase value of 66 kcal/mol to only 10 kcal/mol, in tha latter case attraction of -0.5 kcal/mol between the two guanidinum ions was already established.

In summary, several effects acting in concert are responsible for the occurrence of the free energy minimum of the Gdm^+-Gdm^+ pair in water. While reduction in electrostatic repulsion dominates, without further interactions such as cavitation and dispersion the balance does not tip to negative free energy values. Note that these additional stabilizing interactions are somewhat weaker for the NO_3^--NO_3^- pair and, as a result there is still a small residual repulsion between two nitrates. For the NH_4^+-$NH_4^+$and Na^+-Na^+ pairs there is a strong electrostatic repulsion which, with the lack of appreciable additional stabilization, leads to a very unfavorable energetics of like charge pairing.

The analysis of Gdm^+-Gdm^+ pairing based on ab initio calculations with PCM water is consistent with results from MD with explicit solvent.[12,23] While one cannot expect a fully quantitative agreement between these two very different approaches, qualitatively, MD simulations also reveal the importance of ion geometry and charge distribution, as well as solvent exclusion for homo-ion pairing. The latter can be described in more detail with explicit solvent than is cavitation within PCM, nevertheless both approaches provide the same general picture. Namely, test MD simulations with turned-off London interactions for Gdm^+ result in a sizable decrease in Gdm^+-Gdm^+ pairing which again points to the non-negligible role of dispersion for this effect (see radial distribution functions in Figure 4).

Additionally, we have also performed simulations in a system where chloride counterions were not present (as in ab initio PCM calculations). This simulation demonstrates the relatively weak effect the Cl^- counterions have on the strength of the guanidinium-guanidinium pairing (Figure 4). For comparison, the same figure also depicts the lack of homo-ion paring in MD simulatons of aqueous ammonium chloride.

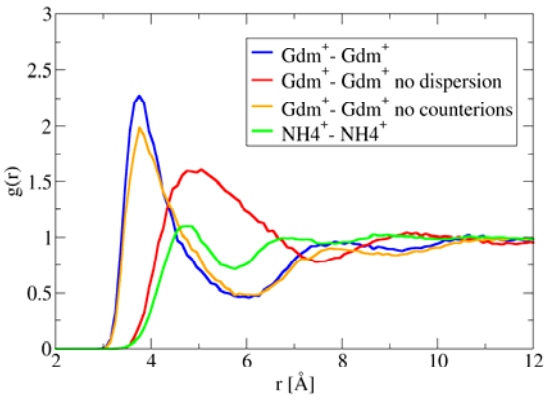


Figure 4: Guanidinium-guanidinium and ammonium-ammonium radial distribution functions demonstrating formation of contact homo-ion pairs in the former but not in the latter case. For guanidinium simulations were also performed with the attractive Lennard-Jones term turned off on the cations and in a system where chloride counterions were removed.

The polar side chains of arginine and lysine are more complex than the bare Gdm^+ and NH_4^+ ions as the former contain additional short aliphatic chains. While these can contribute to the side chain association it appears that this is not a determining factor. Namely, lysine has the longer aliphatic chain, yet the charged groups are still found to repel each other. In the arginine case the hydrophobic side chain is one CH_2 group shorter, yet the Gdm^+ based head-groups still tend to associate (Figure 1).

The present computational results can be directly compared with data extracted from protein structural database. To obtain a representative set of Arg-Arg side chain pairs, we utilized recent version of the Atlas of Protein Side-Chain Interactions (SCA).[24] The atlas is extracted from the structures in the Protein Data Bank (PDB).[25] having mutual sequence identity less than 20% and being determined by X-ray with resolution of 2.0Å or better. Interacting Arg side chains are considered to be those having a center-to-center distance between their closest two atoms of less than the sum of their van der Waals radii (plus 1Å to allow for coordinate error) and the two amino acids must be at least four residues apart.

The cluster representatives for a given distribution are determined as follows: The root-mean-square distance (RMSD) to all other side chains in the distribution is computed using atoms that define the side chain's frame of reference. Any side chain with an RMSD of less than 1.5 Å from the selected side chain is considered a "neighbor". For Arg-Arg we detected six distinct distributions represented by corresponding cluster representatives.

Results of this analysis are displayed in Figure 5. The side chain in the center is the reference side chain. It is apparent that most of the guanidinium groups in the cluster representatives have tendency to associate in a stacking-like arrangement above and below the reference side chain. This result demonstrates that the association of positively charged Gdm^+ based side chain of arginine is widely observed in the Protein Data Bank, and is clearly of significance for the structure and association of proteins.[5] In contrast, the other positively charged amino acid – lysine, behaves quite differently. Normalized occurrences of Lys-Lys contacts (which in this case correspond to solvent-separated pairs[26]) in protein structures have only half the abundance of those of Arg-Arg and there is no orientational preference for the corresponding cluster representatives.

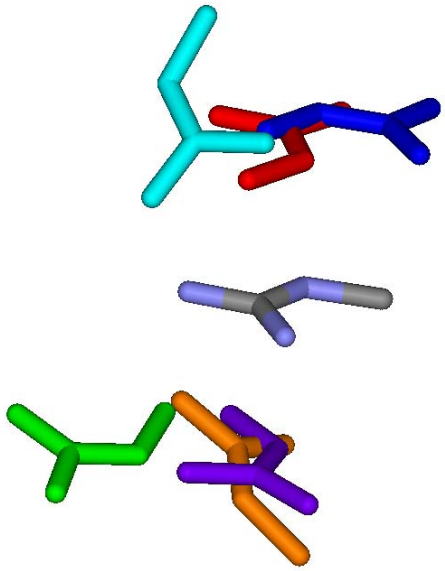


Figure 5: Side Chain Representatives of six Arg-Arg structural clusters obtained from SCA .

Conclusions

In summary, we have demonstrated the cationic side chain association in MD simulations of aqueous oligo-arginines, but not oligo-lysines. This effect can be traced to the different behavior of aqueous homo-ion pairs of Gdm^+ and NH_4^+, which are the charge carriers of the side chains of Arg and Lys. While the association of Gdm^+ ions has been documented in previous calculations[10-12,23] the molecular origin of this effect has not been addressed in detail. The ab initio PCM water calculations presented here both support the previous MD results concerning the energetically favorable formation of the Gdm^+-Gdm^+ homo-ion pair, and also allow this attraction to be dissected into its individual components. It is found that a combination of factors results in a favorable Gdm^+ homo-ion pair, but an unfavorable NH_4^+ ion pairing.

In particular, the most important factor for stabilization of the Gdm^+-Gdm^+ homo-ion pair is the reduction in electrostatic repulsion upon moving from quasi-spherical ions to ions possessing a flat geometry with a very non-homogeneous internal distribution of charge. Additional two factors which bring two Gdm^+ ions together are i) appreciable gains in cavitation (solvent exclusion) energy and ii) in dispersion interactions between the two ions upon association. The present results thus provide a molecular rationalization of the

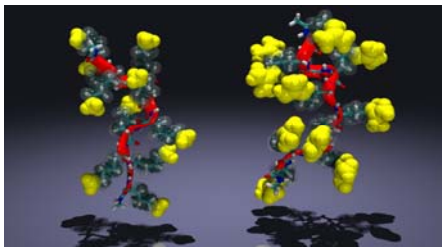
electrostatically counter-intuitive Arg-Arg pairing which plays an important role both within and between proteins.

Acknowledgement: We thank Mikael Lund for valuable comments. We are grateful to the Czech Ministry of Education (grant LC512) and the Czech Science Foundation (grant 203/08/0114) for support.

References:

[1]Collins, K. D.; Washabaugh, M. W. Quart. Rev. Biophys. 1985, 18(4), 323-422.
[2]Auffinger, P.; Lukasz, B.; Westhof, E. *Structure* **2004**, *12*, 379-388.
[3]Ledvina, P.S.; Yao, N.; Choudhary, A.; Quiocho, F. A. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6786-6791
[4]Vyas, N. K.; Vyas, M. N.; Quiocho, F.A. *Structure* **2003**, *11*, 765-774.
[5]Pednekar, D.; Tendulkar, A.; Durani, S. *Proteins* **2009**, *74*, 155-163.
[6]Li, L.; Vorobyov, I.; Allen, T. W. *J. Phys. Chem. B* **2008**, *112*, 9574-9587.
[7]Vorobyov, I.; Li. L.; Allen, T. W. *J. Phys. Chem. B* **2008**, *112*, 9588-9602.
[8]Wang. L.; Maji, S. K.; Sawaya, M. R.; Eisenberg, D.; Riek, R. *PLoS. Biol.* **2008** *6(8)*, 1791-1801.
[9]Mason, P. E.; Neilson, G. W.; Dempsey, C. E.; Barnes, A. C.; Cruickshank, J. M. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4557-4561.
[10]Boudon, S.; Wipff, G.; Maigret, B. *J. Phys. Chem.* **1990**, *94*, 6056-6061.
[11]No, K. T.; Nam, K-Y.; Scheraga, H. A. *J. Am. Chem. Soc*., **1997**, *119*, 12917-12922.
[12]Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Saboungi, M. L.; Dempsey, C. E.; MacKerell, A. D.; Brady, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 11462 – 11470.
[13]Berendsen, H.J.C.; Postma, J. P. M.; VanGunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
[14]Wang, J.M.; Cieplak, P.; Kollman, P. A. *J. Comp. Chem.*, **2000**, *21*, 1049-1074.
[15]Berendsen, H.J.C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.*, **1987**, *91*, 6269-6271.
[16]Dang, L.X.; Chang, T. M. *J. Phys. Chem. B*, **2002**, *106*, 235-238.
[17]Ryckaert, J.P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.*, **1977**, *23*, 327-341.
[18]Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.*, **1995**, *103*, 8577-8593.
[19]Tomasi, J.; Cammi, R.; Mennucci, B. *Int. J. Quantum Chem.*, **1999**, *75*, 783-803.
[20]Barone, V.; Cossi, M. *J. Phys. Chem. A*, **1998**, *102*, 1995-2001.
[21]Case, D.A.D.; Cheatham, III, T. E; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A., *Amber 8*. 2004, Amber 8, University of California: San Francisco.
[22]Frisch, M.J.T.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr. J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, H.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J.J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, K. J.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., *Gaussian03*. 2003: Gaussian03, Pittsburgh PA.
[23]Mason, P. E.; Brady, J. W.; Neilson, G. W.; Dempsey, C. E. *Biophys. J.* **2007**, *93*, L04–L07.
[24]Singh J.; Thornton J. M. *Atlas of Protein Side-Chain Interactions, Vols. I & II*, IRL press, Oxford, 1992.
[25]see http://www.ebi.ac.uk/thornton-srv/databases/sidechains
[26]Villareal, M.; Montich, G. *Protein Sci.* **2002**, *11*, 2001-2009.

TOC Graphics



Snapshots from molecular dynamics simulations of deca-lysine and deca-arginine showing like-charge pairing in the latter but not in the former case.