THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Charge Transfer Effects in the GroEL−GroES Chaperonin Tetramer in Solution

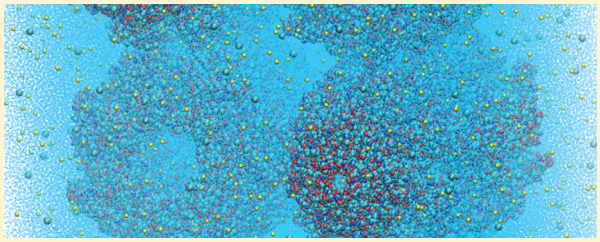
Victor M. Anisimov*[,†] and Andrey A. Bliznyuk[‡]

[†]National Center for Supercomputing Applications, University of Illinois at Urbana—Champaign, 1205 West Clark Street, Urbana, Illinois 61801, United States

[‡]Australian National University, Supercomputer Facility, Leonard Huxley Bld. (#56), Canberra, ACT 0200, Australia

Supporting Information

ABSTRACT: In this work, we present the results of a large-scale, semiempirical LocalSCF quantum mechanical study of GroEL−GroES chaperonin in solution containing 2 481 723 atoms. We find that large biological systems exhibit strong quantum mechanical character, the extent of which was not previously known. Our data show that protein transfers −743 electron units of charge to solvent, which is not described by classical force fields. Contrary to the commonly held belief, which is based on classical mechanics, our computational data suggest that the quantum mechanical effects of charge transfer increase with the size of biological systems. We show that the neglect of charge transfer in classical force fields leads to significant error in the electrostatic potential of the macromolecule. These findings illustrate that a quantum mechanical framework is necessary for a realistic description of electrostatic interactions in large biological systems.



1. INTRODUCTION

1.1. The Nature of Electronic Polarization. Protein−solvent interactions have been the subject of intensive computational studies for more than two decades.[1] Due to the large size of proteins and the accompanying high computational cost of such computations, they are traditionally performed using empirical force fields, which are founded on the principles of classical mechanics.[2] Since the 1970s, the classical force fields represent the main tool in computational biophysics, which speaks highly about their versatility and utility. Classical force fields are very fast and can be remarkably accurate when the necessary parameters are available for the application domain.

Classical force fields employ fixed atomic charges and assign unit or zero charge to amino acids depending on their protonation state. They rely on the assumption that parameters developed for small molecules, representative of functional groups of biomolecules, can be transferred to the larger units, e.g., amino acids, and eventually to proteins.[2] Recently, several attempts have been made to extend the theoretical framework of the traditional fixed-charge force fields to explicitly account for the electronic polarization effect caused by environments of different polarity.[3] This should, in theory, improve the parameter transferability. However, in practice, reaching the goal of better accuracy in biomolecular simulations turned out to be a very difficult problem.[4,5] In part, this is because too little is known about the physics of complex biomolecular systems. Experimental data for large systems are scarce, whereas high-level computational methods, which could provide the reference data to compare against, are typically too demanding to be of practical use. Only recently, with the advent of linear scaling quantum chemistry and due to the progress in computer hardware, realistically sized protein−solvent systems have become accessible for direct quantum mechanical (QM) studies.[6−11]

After three decades of successful application of classical mechanics principles to structural biophysics, the first glimpse of a nonclassical character of protein−solvent interactions pointing to the limitations in the biomolecular simulation theory came from semiempirical QM studies of Cold-Shock protein A by Nagid et al.[12] who showed that this protein transfers about −2 electron units of charge to water. Similar evidence came from studies on other systems.[10,13−17] However, due to the small magnitude of the charge transfer predicted by QM calculations and the uncertainty of its practical implications, the significance of this discovery for biophysics remains unclear. The overwhelming success of classical force fields operating with fixed charges fuels the belief that charge transfer is not a significant phenomenon. Therefore, the force field community aspires a different idea toward the improvement of simulation methodology presuming the significance of charge polarization effects.[3] Among the classical polarizable force field models, the fluctuating charge model is supposedly the most realistic one, since it mimics the real process of charge redistribution happening within functional groups.[18,19]

Before proceeding with further analysis, it is important to clarify the relationship between charge transfer and charge polarization. Despite the seeming differences, they essentially

Received: November 26, 2011
Revised: May 17, 2012
Published: May 17, 2012

   dx.doi.org/10.1021/jp211385e | J. Phys. Chem. B 2012, 116, 6261−6268

represent the same physical phenomenon of electronic density undergoing spatial deformation due to the electric field of the environment. To make the distinction between charge transfer and polarization, one needs to introduce the charge redistribution boundaries. Under the polarization effect, one typically assumes that charge density deformation happens within the confinement of a small chemical fragment. If the charge deformation exceeds the boundary of the fragment, it is called charge transfer, since it results in charge literally traveling in space from one fragment to another. However, making the fragment larger to include the charge donor and acceptor within its boundaries would again reinstate the charge polarization. Since the division of a macromolecule on individual fragments is a subjective process, the distinction between charge transfer and charge polarization is essentially superficial. Nevertheless, from the practical point of view, it is convenient to split the macromolecule polarization into two different components: one is local polarization response of individual fragments composing the macromolecule, and the second one is nonlocal charge redistribution involving two or more fragments. The advantage of such a dual view is in simplicity of treatment of local charge polarization effects by means of classical polarizable force fields, whereas the nonlocal effects are typically considered insignificant.[20] The degree to which it is true will be assessed in the present study.

When a small molecule is subjected to an applied electric field, it responds to the external perturbation by generating an induced dipole. This effect is reasonably well described for small molecules by polarizable force fields.[20] For a large biomolecule, the net polarization effect is believed to be a cooperative response of the individual polarizable fragments composing the macromolecule. The validity of such an assumption cannot be easily tested due to lack of the reference data on macromolecule polarizability. Typically, a projection is made based on the success in description of small molecule polarizability from fitted atomic and fragment polarizabilities,[21−26] thus presuming that the physics of macromolecules is the same as that of small molecules but compounded by a greater number of atoms. However, as shown by Giese and York, even with such simple systems as small molecules and neat liquids, the classical polarizable models cannot simultaneously reproduce gas-phase and condensed-phase polarizability limits.[27] For small molecules, fitting the parameters to the condensed phase is typically sufficient to produce a predictive model, since the latter is representative of the experimental condition the model is designed for. No such flexibility exists for macromolecules in making sure their polarizability is correctly reproduced, since unlike the case with small molecules there are no experimental data on protein polarizability to fit the parameters to. The natural solution would be to generate the reference data from QM calculations.

From the basic QM picture, the electronic density of a macromolecule is completely delocalized in the system. Since proteins are large, the deformation of the electronic density in response to an external electric field may result in charge flow happening between the remote parts of the system, thus raising the issue of nonlocal effects in macromolecule polarization. To this date, the relative significance of local vs nonlocal mechanisms in the polarization response is unknown and can only be determined from QM calculations. Unlike QM methods, modern polarizable force fields are capable of describing only local charge polarization effects at the level of individual amino acids. Since in force fields the amino acid charge is fixed, this precludes the ability of the method to reproduce the nonlocal charge polarization effects. However, there is mounting evidence that charge transfer on the distances larger than the boundary of individual amino acids plays an important role in the protein structure and dynamics. For instance, semiempirical quantum mechanical molecular dynamics (QM MD) studies of ubiquitin in water identified dynamic charge transfer channels in salt bridges.[13] According to these data, individual amino acids in salt bridges exchange up to 0.4 electron units of charge through space in sync with their structural motion. Disabling the charge transfer would lead to MD generating a distorted conformational profile of the protein. In another study performed by Ufimtsev et al.,[10] *ab initio* QM MD simulations were used to show that charge transfer between protein and water serves the purpose of relieving polarization stress on neutral amino acids, which is induced by the electric field of ionized amino acids. According to their work, if protein is treated at the QM level whereas water is treated at the MM level as in popular QM/MM methods, the charge distribution inside the protein becomes unphysical. The correct picture of charge distribution inside the protein is recovered when the protein and water are allowed to exchange electric charge.

Classical force fields effectively avoid this particular problem by using prepolarized charges. However, the mean-field approximation reduces the ability of the method to correctly account for different electrostatic environments experienced by same-type amino acids in complex biological macromolecules. To see the extent of these differences, it may be useful to compare the classical and QM charge models for these amino acids after extracting atomic charges from QM electronic density. Although the atomic charges derived from QM calculations would depend on the choice of the particular method of electronic density decomposition, it would not change the physics of electronic polarization. Even less dependent on the choice of the method would be the magnitude of fluctuation of net amino acid charge in response to the electric field of the environment. Having these data collected, we could identify the areas in the large system where the classical and QM methods agree or disagree.

In planning a meaningful comparison of QM calculations with classical force field methods, it is highly important to adhere to a realistic computational setup representing the experimental conditions. Typical biomolecular simulations deal with the systems encountering tens of thousands of atoms including protein, water, and counterions. So far, to the best of our knowledge, the largest in size QM MD studies of charge transfer in protein−water systems were performed on a 900-atom protein−water system for 8.8 ps at the *ab initio* HF/STO-3G level[10] and on a 12 199-atom system for 20 ps at the semiempirical PM5 level.[13] The largest QM calculations were performed at the *ab initio* level on solvated rubredoxin containing 2825 atoms[28] and at the semiempirical level on the GroEL−GroES monomer containing 119 273 atoms.[9] Due to the high cost of QM methods, performing *ab initio* calculations on real-size systems remains problematic. The computational expense is greatly reduced for semiempirical QM methods. Therefore, in this work, we perform semiempirical QM single-point energy calculations of chaperonin tetramer in water consisting of 2 487 723 atoms and report our findings that shed additional light on the significance of charge transfer in biological systems and the implications of ignoring thereof.

dx.doi.org/10.1021/jp211385e | *J. Phys. Chem. B* 2012, 116, 6261−6268

1.2. Biophysics of GroEL−GroES Chaperonin. Chaperonins are giant molecular machines designed by nature to assist in refolding of improperly folded substrate proteins to their native conformation. Understanding the molecular mechanism of action in chaperonins has important bioengineering implications.[29] Normally after synthesis proteins quickly and efficiently fold on their own to their native conformation in the cell environment. However, due to the presence of hydrophobic molecules and partially folded proteins in the cytosol, the normal process of protein folding may be interrupted and the protein may get trapped in a transient non-native conformation. Chaperonins recognize these misfolded proteins by the high content of hydrophobic residues on their surface and provide an isolated environment of a folding chamber where the substrate protein can refold to its native conformation.

Among all chaperonins, the *E. coli* derived GroEL−GroES chaperonin complex is the best studied by theoretical and experimental methods.[30,31] In the complex, two GroEL rings are stacked one on another creating a folding chamber, and one GroES unit serves as a lid to the chamber. Extensive experimental and computational studies of GroEL−GroES chaperonin revealed remarkable details of its machinery.[32] In its initial state, the folding chamber is lined up by primarily hydrophobic residues that facilitate binding of misfolded proteins. After chaperonin traps the substrate protein in the folding chamber, seven ATP molecules bind to the chaperonin. The binding initiates the twist along the chaperonin main axis and leads to replacement of the previous hydrophobic lining of the folding chamber by polar and ionized residues. The latter introduce a directed electrostatic field inside the chamber, facilitating the substrate protein to adopt its native fold.

This remarkable mechanism of dynamic changing of the lining of the chaperonin chamber from hydrophobic to hydrophilic and back is a very difficult problem for computational studies. The free energy penalty for burying ionized amino acids in the protein interior should be very large. However, the structural change in the giant chaperonin requires only seven ATP molecules.[32] Therefore, there should be factors reducing the free energy penalty for burying ionized residues in the protein interior. For instance, the ionized amino acid might exchange a proton with solvent and assume a neutral form before going into the protein interior. The other possibility might be that charge transfer effectively reduces the net charge on ionized amino acids while they are in the buried state. None of these pathways is easily approachable with classical mechanics models. To study these and other relevant details, one needs to perform QM simulations of chaperonin in solution.

2. COMPUTATIONAL DETAILS

The primitive crystallographic unit of GroEL−GroES chaperonin (PDB id 1AON)[33] contains four tightly packed chaperonin monomers. Its crystal structure is resolved at a temperature of 291 K, and the Cartesian coordinates are deposited to the PDB repository.[33] Since we are interested in the system closely resembling the experimental structure, we took the entire tetramer for the computational study and generated the crystallographic unit by applying coordinate transformations reported in the PDB file. This system carries 28 ADP units, 2408 LYS, 1372 ARG, 2016 ASP, and 2856 GLU amino acids. After adding hydrogen atoms by using the NAMD[34] utility *psfgen* to reflect physiological pH, the tetramer assumes the net share of −1176 electron units. The 28 magnesium atoms, one per ADP unit, were retained in their crystallographic position. The tetramer was solvated in a box of TIP3P water[35] with solvent padding of 15 Å applied on each side of the box using the VMD[36] utility *solvate*, which added 665 081 water molecules to the system. To neutralize the system and to maintain 150 mM physiological salt concentration, 3000 sodium cations and 1880 chloride anions were added by using the VMD utility *ionize*. The resulting box had the edges 315 × 310 × 272 Å^3 and contained 2 481 723 atoms (Figure 1).

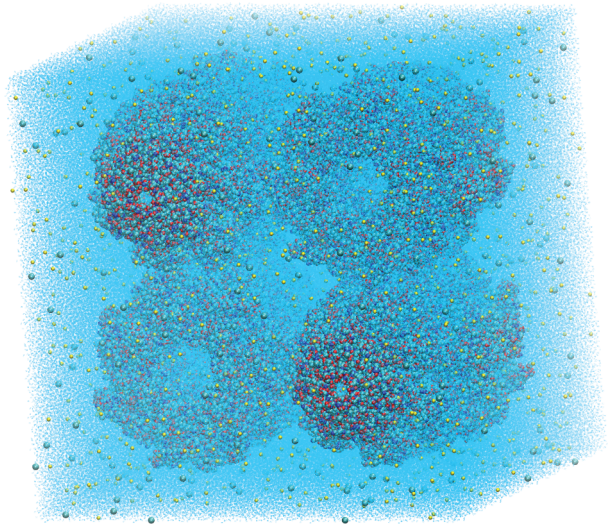


Figure 1. GroEL−GroES tetramer solvated in a water box. Protein and counterions are displayed in VDW mode. Sodium atoms are in cyan, and chlorine atoms are in yellow. The figure is prepared by using the VMD program.[36]

The force field MD simulations were performed using the NAMD 2.6 computer program[34] and CHARMM27 force field[37] under periodic boundary conditions. The nonbond interactions were switched off in the interval from 10 to 12 Å. SHAKE constraints were applied to covalent bonds involving hydrogen atom.[38] The long-range electrostatic interactions were treated using the PME method with a 1 Å grid.[39] The constant temperature of 291 K was maintained by using a Langevin thermostat, and the constant pressure of 1 atm was preserved by applying the Langevin piston method.[40] Before starting the MD simulation, the solvent molecules in the box were relaxed for 10 000 steps by using conjugate gradient minimization while holding protein atoms fixed in their crystallographic position. Additional 10 000 minimization steps were applied while having protein non-hydrogen atoms restrained to their crystallographic position using a force constant of 1 kcal/(mol Å). The system was gradually heated to 291 K under NVT ensemble for 50 ps while keeping the coordinate restraints on protein non-hydrogen atoms. After that, the restraints were removed and the system was simulated under NPT ensemble for the duration of 1 ns. MD simulations were performed using a 1 fs integration time step.

QM single-point energy calculations were performed by using the LocalSCF computer program[41] with AM1,[42] PM3,[43] and PM5[44] semiempirical Hamiltonians on 100 snapshots, each containing 2 481 723 atoms, generated from the classical MD trajectory using a 10 ps interval. The major conclusions in this

dx.doi.org/10.1021/jp211385e | J. Phys. Chem. B 2012, 116, 6261−6268

work were derived from PM5 computations. PM5 is the newest semiempirical Hamiltonian in the sp basis set for which an efficient linear scaling LocalSCF implementation is available. Since PM5 parameters have never been actually published, it warranted performing additional calculations using AM1 and PM3 Hamiltonians. The QM calculations were performed under default program settings using a fast multipole method for the treatment of long-range electrostatics as previously described.[45] The physical memory requirement was 48 GB per job. The average calculation time was 8 h per job for AM1 and PM5 Hamiltonians and 10 h for PM3 Hamiltonian on 8 cores of the 2.93 GHz Intel Nehalem processor.

To study ion solvation, a 1 ns CHARMM27 force field MD simulation was performed on a 50 × 50 × 50 Å^3 water box with 14 dissolved molecules of sodium chloride, which corresponds to 150 mM physiological salt concentration. The classical force field MD simulation details were the same as in the tetramer simulation. Single point semiempirical QM calculations using the PM5 Hamiltonian were performed on 100 snapshots extracted from the 1 ns empirical trajectory using 10 ps intervals. Each such snapshot contained 11 707 atoms.

Throughout this work, QM atomic charges were computed using Mulliken population analysis and averaged over the empirical trajectory. The charges on cations, anions, and tetramer residues were additionally averaged over same type units. The charge on amino acid was computed as the sum over the atoms constituting the residue. The charge on water molecules was computed as the sum of Mulliken charges of atoms constituting water molecule. The net amino acid charges were also computed by using CM2[46] and CM3[47] methods, and the corresponding data are provided in the Supporting Information (Table S1).

3. RESULTS AND DISCUSSION

In the beginning, the system was equilibrated for 500 ps. Such short MD simulation is obviously insufficient to bring the large system to thermodynamic equilibrium, but since we need only to relax the system, the MD simulation provides the necessary means. Running longer MD simulation would have the drawback of pushing the geometry away from the experimental structure due to differences in the computational and experimental setup and because of potential deficiencies in the classical model. On the other hand, it is highly important to remove the high-energy strain coming with the low-resolution experimental structure. Therefore, the purpose of our equilibration is to relax the system from its initial strain while keeping the structure as close as possible to the experimental geometry. The final 500 ps of MD simulation were considered as a production run to minimally account for the flexibility of this large macromolecule.

3.1. Charge Transfer. The progression of tetramer net charge during the 1 ns MD trajectory according to QM calculations is presented in Figure 2. The QM charges on various components of the simulation system averaged over the last 500 ps of the trajectory are presented in Table 1. Here we present the results of AM1, PM3, and PM5 calculations. Before proceeding with the discussion of results, it is useful to briefly compare the ability of different Hamiltonians to accurately describe charge transfer. As it has been shown on small molecules by Merz and co-workers,[48] AM1 and PM3 quantitatively agree with the data obtained from high-level MP2 calculations, with AM1 providing the best agreement with the reference data. However, in our QM MD studies of ubiquitine in explicit water,[49] we found that AM1 strongly overestimates the density of pure water and leads to unphysical spontaneous hydrolysis of water, whereas PM3 and PM5 are essentially free from this error. In our previous work, we were able to show that PM5 very accurately predicts protein−water charge transfer in QM MD calculations.[13] In the present study, as one can see from Table 1, AM1 and PM5 closely agree in their predictions. Therefore, in the absence of PM5, using the AM1 Hamiltonian in such calculations would be fully justified. Our preference to PM5 is to leave open the possibility to compare in our future work the data reported in the current rigid-geometry study with PM5 MD calculations of the same system. Therefore, throughout this work, the reported data are those obtained with the PM5 Hamiltonian unless otherwise mentioned.

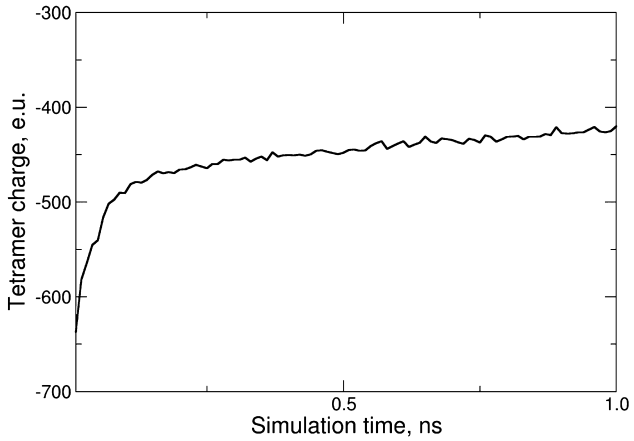


Figure 2. Net charge on tetramer (in electron units, e.u.) along classical force field MD trajectory. The protein charge is obtained from semiempirical single-point QM calculations of the simulation box (2 481 723 atoms), which is performed on 100 snapshots taken with 10 ps intervals. The charge on the tetramer is computed as the sum of atomic Mulliken charges obtained from the QM data.

Table 1. The Net Charge on Tetramer, Water, and Counterions, in Electron Units Assigned in Classical Mechanics, and the One Predicted by Semiempirical QM Methods AM1, PM3, and PM5

model	tetramer	water	chloride	magnesium	sodium
classical	−1176	0	−1880	56	3000
AM1	−574	1392	−1646	19	809
PM3	−280	698	−1635	15	1202
PM5	−433	798	−1583	17	1201

According to the present PM5 calculations, the tetramer sheds −743 electron units of charge to solvent, thus going from the classical net charge of −1176 determined by the protonation state of its ionizable amino acids down to −433 electron units. Water, which is supposed to be electroneutral, according to classical mechanics, acquires 798 units of charge. The charge on 28 magnesium cations, which are tightly bound to 28 ADP molecules, becomes 17 electron units instead of their classically expected value of 56 electron units. The charge on 3000 sodium atoms becomes 1201 electron units, which corresponds to the charge of 0.4 per sodium cation on average. This is in sharp contrast with the value of 1 electron unit assigned to sodium cation in classical force fields. The chloride anions give away their charge at a moderate extent, and instead

dx.doi.org/10.1021/jp211385e | J. Phys. Chem. B 2012, 116, 6261−6268

Table 2. Average Net Charge on Tetramer Residues Based on Semiempirical QM Calculations[a]

residue	number of instances	net charge FF	AM1	PM3	PM5	residue	number of instances	net charge FF	AM1	PM3	PM5
ADP	1400	−3	−2.11 ± 0.05	−1.88 ± 0.06	−2.06 ± 0.05	ILE	103 600	0	0.01 ± 0.04	0.01 ± 0.05	0.02 ± 0.05
ALA	201 600	0	0.01 ± 0.04	0.01 ± 0.05	0.01 ± 0.05	LEU	124 600	0	0.01 ± 0.04	0.01 ± 0.05	0.02 ± 0.05
ARG	68 600	1	0.92 ± 0.04	0.89 ± 0.06	0.91 ± 0.05	LYS	120 400	1	0.92 ± 0.05	0.90 ± 0.06	0.92 ± 0.05
ASN	57 400	0	0.00 ± 0.04	0.01 ± 0.05	−0.01 ± 0.05	MET	49 000	0	0.00 ± 0.04	0.01 ± 0.05	0.01 ± 0.05
ASP	100 800	−1	−0.85 ± 0.07	−0.79 ± 0.07	−0.85 ± 0.07	PHE	21 000	0	0.00 ± 0.04	0.01 ± 0.05	0.01 ± 0.05
CYS	8400	0	−0.02 ± 0.04	−0.02 ± 0.05	−0.02 ± 0.05	PRO	39 200	0	0.00 ± 0.04	0.00 ± 0.05	0.02 ± 0.05
GLN	44 800	0	0.01 ± 0.04	0.02 ± 0.05	0.01 ± 0.05	SER	56 000	0	−0.01 ± 0.04	0.00 ± 0.05	−0.02 ± 0.05
GLU	142 800	−1	−0.84 ± 0.06	−0.78 ± 0.06	−0.83 ± 0.06	THR	96 600	0	−0.01 ± 0.04	0.00 ± 0.05	−0.02 ± 0.05
GLY	154 000	0	0.01 ± 0.04	0.02 ± 0.05	0.02 ± 0.05	TYR	21 000	0	0.00 ± 0.04	0.01 ± 0.05	0.00 ± 0.05
HIS	4200	0	0.02 ± 0.04	0.06 ± 0.05	0.03 ± 0.04	VAL	180 600	0	0.01 ± 0.04	0.01 ± 0.05	0.02 ± 0.04

[a]FF stands for classical force field. The number of residue instances is their quantity in the tetramer multiplied by 50 snapshots. The statistical uncertainties are given by standard deviation. HIS residue is taken in neutral form with the hydrogen atom assigned at the ε-site. Terminal amino acids are excluded from the averaging. The QM values represent the sum of Mulliken atomic charges of the corresponding residues.

of the classical value of −1880, the QM calculation determines their combined charge to be −1583 electron units.

The observed large magnitude of charge transfer in the studied system poses a number of questions. Is it realistic that protein can lose −743 units of charge to the solvent? Can it be a result of a flow in the theory? Unlike the PM3 and DFT methods, which are known for their tendency to overestimate charge transfer,[48,50] the PM5 Hamiltonian is rather conservative in the prediction of magnitude of charge transfer.[13] This fact provides extra credibility to the PM5 data. On the other hand, if the transferred charge of −743 electron units is distributed over 481 572 protein atoms (including ADP), the net effect per atom would become only 0.002 electron units, which is in agreement with the highly accurate *ab initio* data reported for small molecules.[48] Therefore, considering the size of the chaperonin tetramer, the magnitude of −743 electron units for charge transfer is within the expected boundaries. However, there are important differences between large and small molecules to be aware of. In proteins, there are many local areas where charge transfer can be much larger than the average value.[12,13,49] To get the picture of charge distribution in this giant macromolecule, we determined the net charge on each amino acid type by summing up the corresponding atomic Mulliken charges extracted from the QM density matrix of dynamics snapshots, and having the result averaged over the last 500 ps of the empirical trajectory.

The results of AM1, PM3, and PM5 calculations of the net charge of amino acids are presented in Table 2. There is a remarkable similarity between AM1 and PM5 data with PM5 results being presumably more reliable, since the PM5 Hamiltonian is trained on a substantially larger data set. In the table, one can see a perfect agreement between QM and classical pictures on the net charge of neutral amino acids. The QM results show that the average net charge on each neutral amino acid is indeed very close to zero. Despite numerous previous observations that QM calculations of net amino acid charge in proteins typically agree with classical force field values,[13,50,51] this is the first statistically significant evidence of such agreement, which is based on averaging over hundreds of thousands of instances of same type amino acids. As we mentioned before, throughout this work, we deal with Mulliken charges. Other considered alternatives are CM2 and CM3 charge computation schemes. Since we obtained a close agreement between all of these methods, we continue our further discussion with Mulliken charges only. The reader interested in CM2 and CM3 computed charges is referred to the Supporting Information (Table S2).

The obtained large data sets make it possible to compute the probability distribution of the net charge of each amino acid. The distribution function for several neutral amino acids is presented in Figure 3. The plots for remaining neutral amino acids are provided in the Supporting Information (Figure S2). As one can see, the net charge on neutral amino acids fluctuates in the range of ±0.1 electron units. Since the probability distributions are symmetric with respect to zero, classical force fields should be able to describe reasonably well the average electrostatic interaction of neutral amino acids with the electric field of the environment.

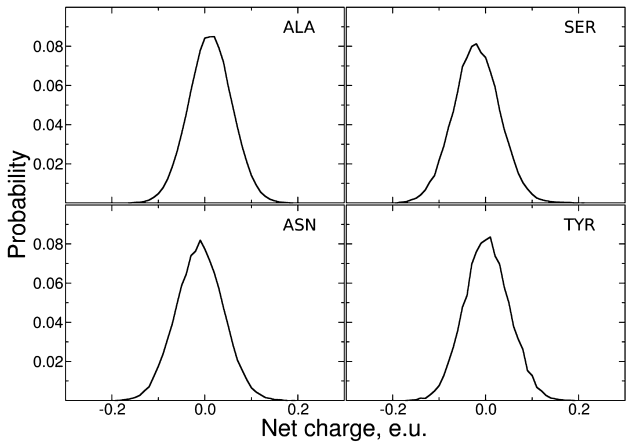


Figure 3. Probability distribution of the net charge on neutral amino acids based on semiempirical QM calculation of 50 snapshots from the last 500 ps of the MD trajectory.

The situation is drastically different for ionized amino acids (Table 2, Figure 4). Here, the average net charge is strongly nonunit. For example, the net charge on GLU is −0.83 ± 0.06, whereas the average net charge on ARG is 0.91 ± 0.05 electron units. Because the probability distribution is centered off the classically expected unit net charge (Figure 4), the existing classical polarizable force fields will not be able to correctly describe the electrostatic interaction of these amino acids with the environment. The shift in the position of the average net charge of ionized amino acids and the significant net charge

dx.doi.org/10.1021/jp211385e | J. Phys. Chem. B 2012, 116, 6261−6268

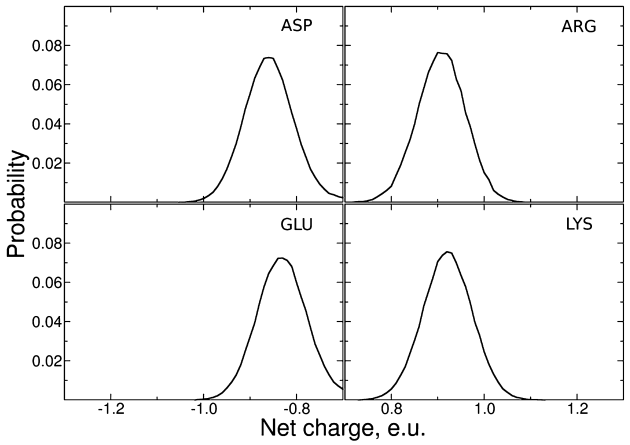


Figure 4. Probability distribution of the net charge on ionized amino acids based on semiempirical QM calculation of 50 snapshots from the last 500 ps of the MD trajectory.

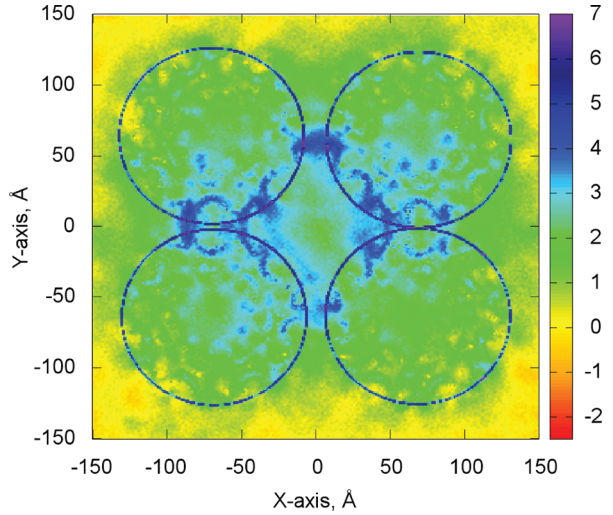


Figure 5. Differential electrostatic potential map ($V_{MM} - V_{QM}$) in volts (V), on cross section of the simulation box by the *xy*-plane at $z = 0$. The origin of the coordinate system is at the center of the box. The circles on the plot represent four GroEL–GroES monomer rings cut by the *xy*-plane. Red color represents the area of the simulation box where MM ESP is more negative than QM ESP. Blue color represents the area of the simulation box where MM predicts more positive ESP than that of QM. The yellow area shows the parts of the system where the force field and QM data quantitatively agree.

fluctuation of all amino acids support our conclusion that charge transfer is the key form of polarization of large biomolecular systems. Therefore, ignoring the charge transfer effects in proteins may be among the critical factors explaining why after so much vested efforts polarizable force fields still cannot improve the accuracy of biomolecular simulations over the level established by nonpolarizable force fields.[4]

3.2. Electrostatic Potential Map. Among macromolecular structures collected in the Protein Data Bank, chaperonin GroEL–GroES has the highest absolute number of ionizable amino acids. For a biomacromolecule of this size, only DNA and RNA could compete with chaperonin on the net charge. This, combined with other experimental evidence, suggests that electrostatic potential (ESP) created by chaperonin should be a very important part of its mechanism of action.[29] The ability of classical force fields to correctly reproduce protein ESP has recently been questioned by Zuegg et al.[52] and Bliznyuk et al.[53] who reported large discrepancies between the empirical and QM ESPs. According to a number of other semiempirical QM calculations, the classical assumption of unit charge assigned to amino acids is correct for about 95% of the amino acids.[13,49,51] Therefore, it is not surprising, that force fields work very well on average. However, for the other 5% of amino acids for which the classical unit charge assumption does not hold, the charge transfer effects, due to exposure to a strong electric field of the environment, must be significant. Such sites could be potentially the source of limited accuracy of the current nonpolarizable and polarizable force fields. At such sites, the errors of force fields are locally higher than on average. To get the notion of differences between the classical and QM electrostatic models of chaperonin tetramer in solution, we compared the ESP maps generated by the respective methods.

The differential ESP map representing the difference between the empirical ESP and QM ESP is displayed in Figure 5. It is computed on a 1 Å grid on the plane cutting the simulation box in half. Perspective view on this plane is provided in the Supporting Information (Figure S3). The contribution of atoms close to the grid points by less than 1 Å is ignored. The QM ESP is computed from PM5 Mulliken atomic charges. The part of the plot where the empirical and QM ESP values agree is shown in yellow. The largest patch on the plot where the empirical value is more positive than QM data by 1.5 V is shown in green, which is the place occupied by protein units. The part where the difference between classical and QM ESPs reaches 4 V is shown in blue. This is the area of protein–protein interaction accompanied by the strongest electronic polarization effects involving charge transfer. The part of the simulation box where empirical ESP is more negative than that of QM by −1 V is shown in orange.

The presence of blue and orange patches in Figure 5 in the interior and exterior of the tetramer is the consequence of charge transfer effects being ignored by classical force fields. The orange patch signifies that QM ESP dissipates faster with distance than the ESP of the classical model. Since we are dealing with the distances at the order of 100 Å, only the monopole component of the amino acid potential can be visible at such a scale. From this, it follows that all nonpolarizable and polarizable force fields assigning fixed unit charge to ionized amino acids will have the same characteristic deficiency in their ESP map. The presence of a blue patch on the ESP plot suggests that the empirical model strongly overestimates the strength of electrostatic interactions inside the highly ionized tetramer, which could lead to unphysically strong electrostatic repulsion between the monomers in the tetramer. Charge transfer from protein to solvent would be a natural mechanism reducing the electrostatic repulsion between the highly charged chaperonin monomers, making feasible their tetramerization and eventual crystallization.

3.3. Nonclassical Charge on Counterions. There are a considerable number of studies in the literature that provide important insights on the significance of charge transfer in biological systems.[12−14,49,51] However, in all of these reports, protein was simulated in a pure water solution; i.e., no counterions were present in the simulation system. However, counterions are important components of any biophysical system stabilizing the ionized state of titratable amino acids. They are an important modulator of the electrostatic properties of the solvated biomolecule and thus of its structure and

dx.doi.org/10.1021/jp211385e | J. Phys. Chem. B 2012, 116, 6261−6268

function. In the absence of counterions, the protein would have only water to transfer the charge to, whereas the presence of counterions provides an additional degree of freedom for charge to flow, i.e., from protein to cations. Since in our case the protein is negatively charged, its electron density can go either to water or sodium cations. Resolving the relative significance of these two mechanisms is important for understanding the physics of protein−solvent interaction.

According to our QM calculations, water molecules carry the overall positive net charge of 798 electron units, and the charge on sodium atoms is reduced from 1.0 to 0.4 electron units (Table 1). On the basis of these data, one may conclude that the protein electron density goes entirely to sodium cations. However, there are reasons to believe that the picture is a bit more complicated. According to *ab initio* HF/6-31G* calculations, the charge on sodium cation in pure water is 0.41 electron units.[17] From these data, it appears that sodium cation carries the same charge in pure water as the value we obtained in the presence of protein. This surprising observation warrants additional studies.

To address the limited amount of available QM data on solvation of counterions in pure water, we performed a series of semiempirical PM5 calculations on 100 snapshots representing 150 mM physiological salt solution of sodium chloride in a large water box containing 11 707 atoms prepared by classical MD simulation. According to our semiempirical QM calculations, the average charge on sodium and chloride ions in this system is +0.44 and −0.85 electron units, respectively, which is in close agreement with the *ab initio* data for ions in pure water[16,17] and close to the charge on sodium cation we obtained in the presence of chaperonin tetramer. Thus, we come to the conclusion that on average the charge on sodium cations is roughly insensitive to the presence of protein in solution. Presumably, sodium cation is such a strong acceptor, and there is plenty of water around, so the sodium cation always takes what it needs from the surrounding water molecules. In turn, water seems to be a very flexible electron donor and acceptor simultaneously, so it readily gives up its electron density to sodium cations and easily accepts the excess charge from the protein, thus serving as a conductor of electric charge.

Despite QM calculations predicting a significant deviation of the charge on sodium cation from the classically expected unit value, the success of molecular mechanics simulations of sodium chloride in water[54] suggests that it is certainly possible to develop a classical model of ion solvation by employing the fixed-charge electrostatic approximation.[55] The feasibility of such a model may be explained by reference to the mean-field theory. The charge transferred from sodium cation to water and averaged over the molecular dynamics trajectory cannot go far from the cation and would stay nearby roughly in the form of a spherical charge cloud centered on the cation atom. Such a charged sphere will be sensed at a large distance as a monopole in much the same way as the potential of a classical point-charge. The discrepancy in cation−water interaction energy due to the classical mechanics approximation is compensated in force fields by fitting the Lennard-Jones terms to experimental condensed phase data, thus making the model suitable for practical molecular modeling applications. However, one must be aware that, due to the simplicity of the model, such compensation is strictly valid for isotropic systems only, where mean-field approximation has a solid physical ground.

Unlike the case of single-atom ions, the charge lost by an ionized amino acid in protein will be transferred to adjacent amino acids, which have considerably lesser mobility than water molecules. Under such conditions, trajectory averaging cannot make the transferred charge assuming a spherical form around the donor. Thus, charge transfer in proteins produces a nonuniform charge distribution around the interacting amino acids, leading to correction terms containing higher multipoles. Such a change in the charge distribution of amino acids will be unique for each protein, and hence, the interaction energy cannot be corrected by a single set of Lennard-Jones parameters.[56] This is a serious limitation of the classical force field models, since it is practically impossible to adjust LJ parameters for each protein.

The magnitude, and hence significance, of charge transfer in biomacromolecular systems suggests that future improvement in the accuracy of biomolecular simulations is unlikely without the use of QM methods where the charge transfer effects are naturally included.

4. CONCLUSIONS

In this work, we performed a QM study of a 2 481 723-atom system of chaperonin tetramer in solution. The computations predict a very large charge transfer of −743 electron units happening between protein and solvent, and show a significant difference between the classical fixed-charge ESP and QM ESP profiles of this macromolecular system. Substantial differences from the classical mechanical model are observed at the level of individual amino acids. Although the average zero net charge is well reproduced for neutral amino acids, our QM calculations show substantial deviation from the unit value for ionizable amino acids. The large degree of amino acid net charge fluctuation, which is observed for all amino acids in the protein, suggests that nonlocal charge transfer effects serve as the primary means of electrostatic polarization of this large system. This invalidates the classical mechanics assumption that macromolecule polarizability can be described on the basis of fragment polarizability. It also serves as evidence that the fixed atomic charges adopted by classical force fields in the context of small molecules, where charge transfer is negligible, become progressively less suitable for large systems. This may explain why classical polarizable force fields experience difficulty in raising the accuracy bar of biomolecular simulations. The present study also raises a concern that the representation of counterions by unit charge may be inconsistent with the real physics of ion solvation where strong charge transfer happens between ion and water molecules. The unusual charge-conducting role of water serving as a charge donor with respect to sodium counterions and as a charge acceptor with respect to protein is reported. This is an entirely quantum mechanical effect for which classical representation does not exist. Unlike classical force fields, QM methods explicitly take into account both local and nonlocal polarization effects; therefore, our results suggest that QM methods are irreplaceable in computational studies of large biomolecular systems.

■ ASSOCIATED CONTENT

Supporting Information

Backbone rmsd, charge density distribution, and cross section plane position plots. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: anisimov@illinois.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are grateful to the Australian National Computational Infrastructure (NCI) facilities for providing supercomputer resources for this project.

REFERENCES

(1) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557−7561.
(2) MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584−1604.
(3) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.
(4) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515−1531.
(5) Baker, C. M.; Anisimov, V. M.; MacKerell, A. D. *J. Phys. Chem. B* **2011**, *115*, 580−596.
(6) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133−146.
(7) Dixon, S. L.; Merz, K. M., Jr. *J. Chem. Phys.* **1996**, *104*, 6643−6649.
(8) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744−2750.
(9) Anikin, N. A.; Anisimov, V. M.; Bugaenko, V. L.; Bobrikov, V. V.; Andreyev, A. M. *J. Chem. Phys.* **2004**, *121*, 1266−1270.
(10) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Phys. Chem. Lett.* **2011**, *2*, 1789−1793.
(11) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904−6914.
(12) Nadig, G.; Van Zant, L. C.; Dixon, S. L.; Merz, K. M. *J. Am. Chem. Soc.* **1998**, *120*, 5593−5594.
(13) Anisimov, V. M.; Bugaenko, V. L.; Cavasotto, C. N. *ChemPhysChem* **2009**, *10*, 3194−3196.
(14) Komeiji, Y.; Ishida, T.; Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.* **2007**, *28*, 1750−1762.
(15) Peraro, M. D.; Raugei, S.; Carloni, P.; Klein, M. L. *ChemPhysChem* **2005**, *6*, 1715−1718.
(16) Thompson, W. H.; Hynes, J. T. *J. Am. Chem. Soc.* **2000**, *122*, 6278−6286.
(17) Tanaka, M.; Aida, M. *J. Solution Chem.* **2004**, *33*, 887−901.
(18) Patel, S.; Mackerell, A. D., Jr; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1504−1514.
(19) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141−6156.
(20) Luque, F. J.; Dehez, F.; Chipot, C.; Orozco, M. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 844−854.
(21) Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8533−8542.
(22) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341−350.
(23) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399−2407.
(24) Ewig, C. S.; Waldman, M.; Maple, J. R. *J. Phys. Chem. A* **2001**, *106*, 326−334.
(25) Stout, J. M.; Dykstra, C. E. *J. Phys. Chem. A* **1998**, *102*, 1576−1582.
(26) Zhou, T.; Dykstra, C. E. *J. Phys. Chem. A* **1999**, *104*, 2204−2210.
(27) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2004**, *120*, 9903−9906.
(28) Guidon, M.; Hutter, J.; VandeVondele, J. *J. Chem. Theory Comput.* **2009**, *5*, 3010−3021.
(29) Ellis, R. J. *Nature* **2006**, *442*, 360−362.
(30) Thirumalai, D.; Lorimer, G. H. *Annu. Rev. Biophys.* **2001**, *30*, 245−269.
(31) Sigler, P. B.; Xu, Z.; Rye, H. S.; Burston, S. G.; Fenton, W. A.; Horwich, A. L. *Annu. Rev. Biochem.* **1998**, *67*, 581−608.
(32) Hyeon, C.; Lorimer, G. H.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18939−18944.
(33) Xu, Z.; Horwich, A. L.; Sigler, P. B. *Nature* **1997**, *388*, 741−750.
(34) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(35) Jorgensen, W. L. *J. Phys. Chem.* **1983**, *87*, 5304−5314.
(36) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(37) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(39) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(40) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(41) Bugaenko, V. L.; Bobrikov, V. V.; Andreyev, A. M.; Anikin, N. A.; Anisimov, V. M. *LocalSCF*, 2.1 ed.; Fujitsu Ltd.: Tokyo, 2005.
(42) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(43) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.
(44) Stewart, J. J. P. *MOPAC 2002*; Fujitsu Ltd.: Tokyo, 2002.
(45) Panczakiewicz, A.; Anisimov, V. M. The Linear Scaling Semiempirical LocalSCF Method and the Variational Finite LMO Approximation. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, The Netherlands, 2011; Vol. *13*, pp 409−437.
(46) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.
(47) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291−1304.
(48) van der Vaart, A.; Merz, K. M. J. *J. Chem. Phys.* **2002**, *116*, 7380−7388.
(49) Anisimov, V. M.; Cavasotto, C. N. Quantum-Mechanical Molecular Dynamics of Charge Transfer. In *Kinetics and Dynamics: From Nano- to Bio-Scale*; Paneth, P., Dybala-Defratyka, A., Eds.; Springer: Dordrecht, The Netherlands, 2010; Vol. *12*; pp 247−266.
(50) van der Vaart, A.; Merz, K. M. *J. Am. Chem. Soc.* **1999**, *121*, 9182−9190.
(51) Liu, H.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. *Proteins* **2001**, *44*, 484−489.
(52) Zuegg, J.; Bliznyuk, A. A.; Gready, J. E. *Mol. Phys.* **2003**, *101*, 2437−2450.
(53) Bliznyuk, A. A.; Rendell, A. P. *J. Phys. Chem. B* **2004**, *108*, 13866−13873.
(54) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D.; Schulten, K.; Roux, B. *J. Phys. Chem. Lett.* **2011**, *2*, 87−92.
(55) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6*, 774−786.
(56) Baker, C. M.; Lopes, P. E. M.; Zhu, X.; Roux, B.; MacKerell, A. D. *J. Chem. Theory Comput.* **2010**, *6*, 1181−1198.