

The Stabilization Effect of Dielectric Constant and Acidic Amino Acids on Arginine–Arginine (Arg–Arg) Pairings: Database Survey and Computational Studies

Zhengyan Zhang,^{†,‡,||} Zhijian Xu,^{‡,||} Zhuo Yang,[‡] Yingtao Liu,[‡] Jin'an Wang,[‡] Qiang Shao,[‡] Shujin Li,^{*,†} Yunxiang Lu,^{*,§} and Weiliang Zhu^{*,‡}

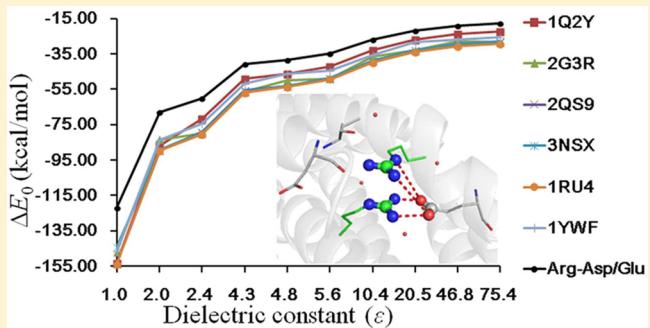
[†]College of Chemistry, Chemical Engineering and Materials Science of Soochow University, Soochow University, Suzhou, Jiangsu, 215123, China

[‡]Drug Discovery and Design Center, CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China

[§]Department of Chemistry, East China University of Science and Technology, Shanghai, 200237, China

S Supporting Information

ABSTRACT: Database survey in this study revealed that about one-third of the protein structures deposited in the Protein Data Bank (PDB) contain arginine–arginine (Arg–Arg) pairing with a carbon···carbon (CZ···CZ) interaction distance less than 5 Å. All the Arg–Arg pairings were found to bury in a polar environment composed of acidic residues, water molecules, and strong polarizable or negatively charged moieties from binding site or bound ligand. Most of the Arg–Arg pairings are solvent exposed and 68.3% Arg–Arg pairings are stabilized by acidic residues, forming Arg–Arg–Asp/Glu clusters. Density functional theory (DFT) was then employed to study the effect of environment on the pairing structures. It was revealed that Arg–Arg pairings become thermodynamically stable (about -1 kcal/mol) as the dielectric constant increases to 46.8 (DMSO), in good agreement with the results of the PDB survey. DFT calculations also demonstrated that perpendicular Arg–Arg pairing structures are favorable in low dielectric constant environment, while in high dielectric constant environment parallel structures are favorable. Additionally, the acidic residues can stabilize the Arg–Arg pairing structures to a large degree. Energy decomposition analysis of Arg–Arg pairings and Arg–Arg–Asp/Glu clusters showed that both solvation and electrostatic energies contribute significantly to their stability. The results reported herein should be very helpful for understanding Arg–Arg pairing and its application in drug design.



INTRODUCTION

The guanidinium ion $\text{C}(\text{NH}_2)_3^+$, called “Y-aromaticity”,¹ is of great biochemical interest.² Arginine (Arg) is a derivative of guanidinium cation³ and is suggested as an important factor in the thermostability of proteins.⁴ As shown in Figure 1a, two positively charged arginines could form a pairing structure,⁵ namely, Arg–Arg pairing, and this seemingly counterintuitive like-charge ion pairing has attracted considerable interest among computational chemists. Numerous studies have been performed, including *ab initio*, Monte Carlo, molecular dynamics simulations and *ab initio* molecular dynamics, which support the fact that the pairing is stable in water solution.^{6–10} However, the stabilization of the pairing in water is attributed to different factors in different literature, including cavitation effects, dispersion effects,¹¹ van der Waals interactions,¹⁰ quadrupole–quadrupole interactions, hydrogen bond, and $\pi\cdots\pi$ interactions.^{6,9,12} Therefore, the main contribution to the stability of Arg–Arg pairing residing in complex biological systems needs to be further clarified.

Proteins are exceedingly complex, and their dielectric properties are poorly understood and difficult to calculate quantitatively.¹³ The local dielectric of the protein is highly inhomogeneous and strongly dependent on the local chemical environment.¹⁴ In general, the dielectric constant (ϵ) is low in the protein interior, and then increases in the vicinity of the surface of the proteins.¹⁵ The protein dielectric constant can vary from as little as 2 in the protein center to a value of about 80^{15–17} due to the local environment.^{18,19} In calculations, the dielectric properties of the protein and solvent phase are represented in terms of dielectric constant.²⁰ Thus, systematic study of the effect of different dielectric constants on the pairing stability is of importance.

Our Protein Data Bank (PDB) survey in this study showed a high frequency of occurrence of acidic residues, i.e., aspartic

Received: January 6, 2013

Revised: March 25, 2013

Published: April 1, 2013

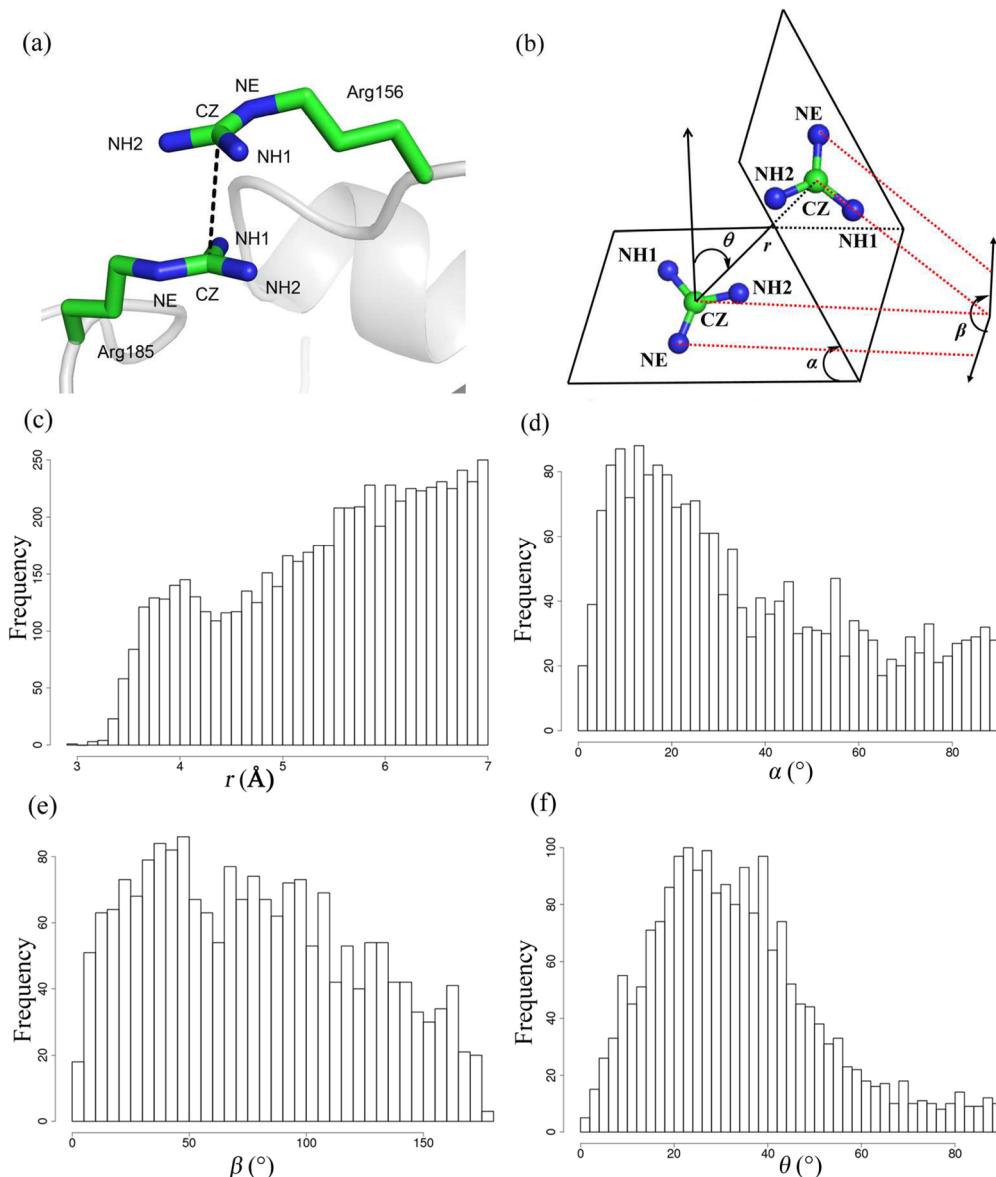


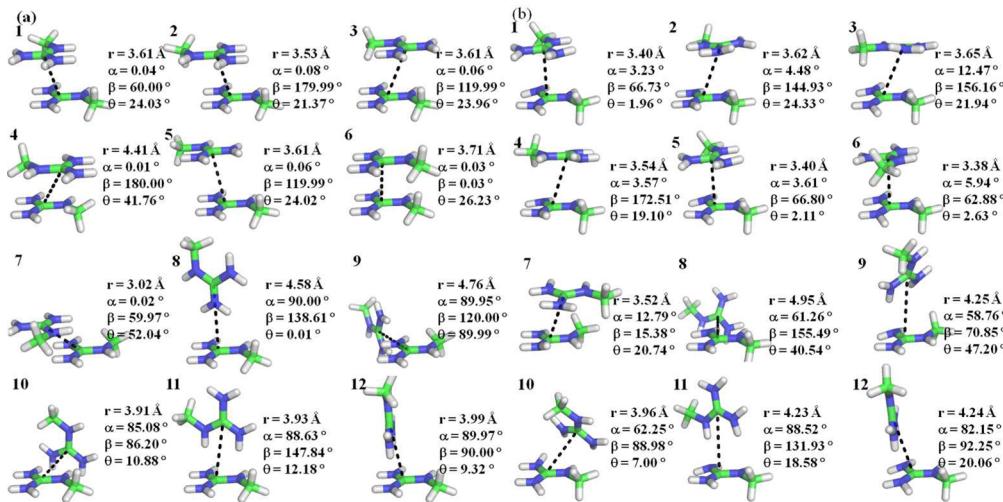
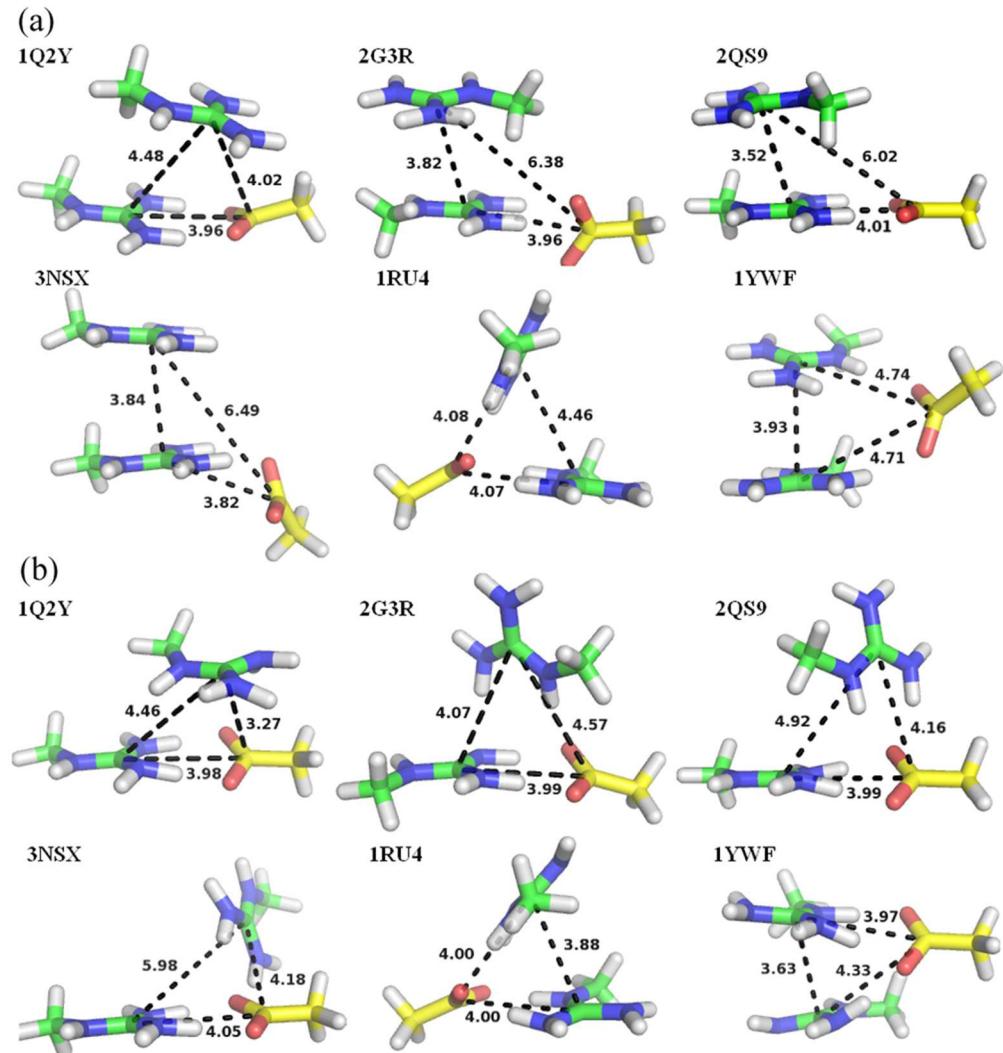
Figure 1. Arg–Arg interactions in the Protein Data Bank. (a) A selected example of Arg–Arg pairing. Arg-156 and Arg-185 form a stable pair with a CZ···CZ distance of 3.48 Å in pyridoxamine kinase (PDB code: 3PZS). (b) Geometrical parameters of Arg–Arg interaction. The definition for distance r , interplanar angle α , side chain orientation angle β , and angular displacement θ . (c) r distribution for 6160 pairs of Arg–Arg interactions from Dunbrack Lab’s 6344 nonredundancy structures with $r < 7$ Å. Frequency distributions of 1975 Arg–Arg pairings ($r < 5$ Å) for α (d), β (e), and θ (f).

acid (Asp) and glutamic acid (Glu), around the Arg–Arg pairings. Therefore, the possible stabilization effect of the acidic residues on the Arg–Arg pairing should not be ignored.²¹ However, few studies have been reported to date on the interaction between the Asp/Glu and the Arg–Arg pairing. To explore the possible stabilization effect of different dielectric constants and acidic residues on Arg–Arg pairing, some typical configurations of Arg–Arg pairings and Arg–Arg–Asp/Glu clusters were designed or extracted from PDB and studied by using density functional theory (DFT) approaches. The results showed that Arg–Arg binding, which might be attractive in a strong polar environment ($\epsilon \geq 46.8$), could be attributed mainly to a solvation effect, while the carboxyl group could significantly stabilize the Arg–Arg pairing with electrostatic energy as a main contribution. Therefore, it is understandable that the Arg–Arg pairings are often located on a protein

surface. When the Arg–Arg pairing is buried inside the protein, a carboxyl group or strong polar moiety usually forces the pairing to form an Arg–Arg–Asp/Glu cluster.

METHODS

Database Survey. The PDB (<http://www.pdb.org>) houses all kinds of experimental structures produced by X-ray crystallography, nuclear magnetic resonance (NMR), neutron diffraction, electron microscopy, powder diffraction, fiber diffraction, solution scattering, and so on. To achieve reliable results, only the X-ray structures with a resolution of 2.0 Å or better were used in this study. Up to July 2011, there were 30 254 structures that satisfy this criterion. Two arginine residues were considered to be an Arg–Arg pairing structure if the carbon···carbon (CZ···CZ) distance is less than 5 Å (Figure 1a and b). However, the redundancy in PDB may bias the analysis.

Figure 2. Arg–Arg pairings: (a) 12 initial structures; (b) optimized structures in DMSO ($\epsilon = 46.8$).Figure 3. Arg–Arg–Asp/Glu cluster: (a) initial structures extracted from PDB; (b) optimized structures in DMSO ($\epsilon = 46.8$). Distance between CZs of arginine and distance between CZ (arginine) and CG (aspartic acid)/CD (glutamic acid) are shown in angstroms.

Therefore, a small data set of 6344 chains was chosen for further analyses, which was downloaded from Dunbrack Lab²² with a percentage identity cutoff of 30%, a resolution cutoff of

2.0 Å, and an R-factor cutoff of 0.25 (http://dunbrack.fccc.edu/Guoli/pisces_download.php). After removing the data missing guanidinium NH1 or NH2 nitrogen atom due to the high

flexibility, 1975 Arg–Arg pairings were used for the study. All the atoms surrounding the center of Arg–Arg pairings in a sphere of 5 Å radius were extracted for analyzing possible forces that stabilize the Arg–Arg pairing.

Interaction Geometries. The definition for distance r , interplanar angle α , side chain orientation angle β , and angular displacement θ for Arg–Arg interactions was shown in Figure 1b.³ The atoms to define the planes are CZ, NH1, and NH2 in arginine. On the basis of α , the Arg–Arg interactions were termed as parallel ($0^\circ \leq \alpha \leq 30^\circ$), oblique ($30^\circ < \alpha \leq 60^\circ$), and perpendicular ($60^\circ < \alpha \leq 90^\circ$).³ Side chain orientation angle β is formed by the two CZ–NE vectors indicating the relative orientation of the two guanidinium groups.²³ On the basis of θ , the Arg–Arg interactions were termed as centered ($0^\circ \leq \theta \leq 30^\circ$), offset ($30^\circ < \theta \leq 60^\circ$), and edge ($60^\circ < \theta \leq 90^\circ$).³

Residues on Protein Surface. A Pymol script was prepared to identify surface residues. After removing non-polymer atoms and adding hydrogen atoms to both polar and nonpolar heavy atoms, the solvent accessible surface (SAS) area of the side chain was calculated by using a probe size of 1.4 Å. The residues with a SAS area of side chain greater than 8.0 Å² would be considered as surface residues.

Geometry Optimization and Interaction Energy Calculation. Consistent with prior work,³ our PDB survey also shows that the distance between two CZ atoms is around 4 Å (Figure 1c), indicating a possible van der Waals contribution in the pairing. Therefore, a DFT-D method,^{24–26} namely, B97-D^{27,28} with a reparameterization of Becke's B97 functional for empirical van der Waals correction, was applied in this study to optimize the Arg–Arg pairing geometries and interactions in nine different solvents with solvent model CPCM, which systematically mimic various atomic environments in protein, viz., cyclohexane ($\epsilon = 2.0$), toluene ($\epsilon = 2.4$), diethyl ether ($\epsilon = 4.3$), chloroform ($\epsilon = 4.8$), chlorobenzene ($\epsilon = 5.6$), dichloroethane ($\epsilon = 10.4$), acetone ($\epsilon = 20.5$), DMSO ($\epsilon = 46.8$), and water ($\epsilon = 78.4$) (Table S1, Supporting Information). Then, solvated harmonic frequency calculations were performed to verify energy minima and to calculate the interaction energy ΔE_0 ($\Delta E + ZPE$). Four different basis sets, namely, 6-31G(d,p), 6-311++G(d,p), cc-pVTZ, and aug-cc-pVTZ, were used (Table S2, Supporting Information). These calculations were performed with the help of the Gaussian 09 software.²⁹

Taking into account the conformations of Arg–Arg dimer from the PDB survey, 12 different initial structures of Arg–Arg pairings were designed, which can be classified into two groups on the basis of α , viz., parallel (1, 2, 3, 4, 5, 6, and 7) and perpendicular (8, 9, 10, 11, and 12) (Figure 2a). In order to save computational resources, methylguanidinium–methylguanidinium structures were used to mimic the Arg–Arg pairing, while the carboxymethyl group was retained for the acidic residues. Therefore, the methylguanidinium–methylguanidinium–carboxymethyl structure was used to mimic the Arg–Arg–Asp/Glu systems (Figure 3a).

Energy Decomposition Analysis (EDA). The Morokuma-type energy decomposition method and Slater type orbital (STO) analysis were used for EDA at the DFT level of Perdew–Burke–Ernzerhof (PBE) and double- ζ polarized (DZP) basis sets.³⁰ The bond dissociation energy is partitioned into two major components ΔE_{prep} and ΔE_{int} :

$$\Delta E_b = \Delta E_{\text{prep}} + \Delta E_{\text{int}} \quad (1)$$

ΔE_{prep} in eq 1 is the amount of energy that the fragments distorted from their equilibrium geometries to the structures when they formed a stable pair.^{31,32} In this paper, ΔE_{prep} is ignored because the energy disparity between the free monomer and the monomer in a pair is inapparent. In a vacuum, the interaction energy ΔE_{int} is further decomposed into three physically meaningful terms:^{33–35}

$$\Delta E_{\text{int}} = \Delta E_{\text{elstat}} + \Delta E_{\text{Pauli}} + \Delta E_{\text{orb}} \quad (2)$$

where ΔE_{elstat} is the electrostatic interaction energy and ΔE_{Pauli} refers to the Pauli repulsive interactions, while ΔE_{orb} represents the orbital interactions between the fragments. In this study, the EDA is calculated in solvent and the solvation energy ΔE_{sol} should be added into the interaction energy ΔE_{int} .³⁶

$$\Delta E_{\text{int}} = \Delta E_{\text{elstat}} + \Delta E_{\text{pauli}} + \Delta E_{\text{orb}} + \Delta E_{\text{sol}} \quad (3)$$

The calculations were performed using the Amsterdam Density Functional (ADF) program.^{33,37}

RESULTS AND DISCUSSION

Arg–Arg Pairings. There are 25 446 Arg–Arg pairings found in 10 319 crystal structures, manifesting that 34.1% (10 319/30 254) of the protein structures contain at least one Arg–Arg pairing in PDB. Meanwhile, it was found that 2231 Arg–Arg pairings occur in 1599 chains from Dunbrack Lab's 6344 nonredundant chains, suggesting that 25.2% (1599/6344) of the nonredundant structures have at least one Arg–Arg pairing. A bimodal distribution of r values with a peak near 4 Å and a minimum at ~4.3 Å was observed (Figure 1c), which is similar to usual π – π stacking interactions.³⁸ The α distribution has a peak near 10° (Figure 1d) and the θ distribution has a peak near 30° (Figure 1f), corresponding to an off-centered parallel Arg–Arg configuration. On the other side, the β distribution is more flat, ranging from 5 to 135° compared to α and θ , indicating that the two CZ–NE bonds may have very different orientations (Figure 1e).

Arg–Arg Pairings Are Mostly Solvent Exposed. The statistic result further shows that 82.2% Arg–Arg pairings are located on a protein surface, strongly suggesting that the Arg–Arg pairing prefers a polar environment. The exposed nature of the Arg–Arg pairing also implies a role on molecular recognition, e.g., protein–protein interaction.³

The Residues to Stabilize Arg–Arg Pairings. Scheraga et al. analyzed all the atoms surrounding Arg–Arg pairing in a sphere of 8 Å radius centered at the mass center of the pair of guanidinium moieties and found that oxygen atoms play a central role in the stabilization of the pairs of groups of similar charge.² Durani et al. examined the residues within a 10 Å cutoff of the mean position between CZ atoms and found that Arg–Arg pairing tends to occur in an environment enriched in Trp, Tyr, Phe, Met, Leu, Glu, and Asp.³ To further detail the possible force stabilizing the Arg–Arg pairing, the micro-environment within 5 Å radius around the center of the pair of guanidinium moieties was examined. We found that all the Arg–Arg pairings are always accompanied by water molecules, acidic residues, or other polar and polarizable moieties in agreement with the previous report,³⁹ indicating that the pairing is instable in nonpolar environment. As shown in Figure 4, the water molecule has the highest occurrence around the Arg–Arg pairing. This observation is in agreement with the supposed obligatory role for solvent to stabilize the pairing.^{2,7} Figure 4 also shows high occurrences of Asp and Glu around

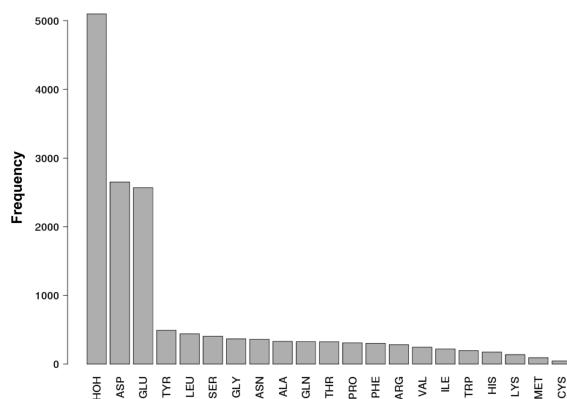


Figure 4. Residues to occur within a 5 Å cutoff of Arg–Arg pairings.

Arg–Arg pairings, viz., 68.3% of the Arg–Arg pairings form salt bridges with at least one anionic residue. Especially for the Arg–Arg pairings residing in protein interior, 86.7% of them form salt bridges with at least one anionic residue, indicating that anionic residues play an essential role in stabilizing the pairing structure.

Structures and Energies of the Optimized Arg–Arg Pairings. Table 1 shows six representative results of Arg–Arg pairings (**1**, **2**, **3**, **5**, **11**, and **12**) that were optimized in different solvents with 6-311++G(d,p) basis sets. We also found that the different basis sets produced different interaction energies (Table S2, Supporting Information). To balance accuracy and computational costs, the 6-311++G(d,p) basis set which yielded similar results to cc-pVTZ was selected for further study.

In the solvent with very low dielectric constant, e.g., cyclohexane ($\epsilon = 2.0$), only 1 of 12 structures was found to have a stationary point, which is a perpendicular-centered structure (**10**, $r = 4.55 \text{ \AA}$, $\alpha = 73.63^\circ$, $\theta = 26.03^\circ$) with a positive value of the binding energy, indicating the optimized geometry is only a metastable structure (Figure 5 and Table S1, Supporting Information). Therefore, a low dielectric constant environment could only embrace Arg–Arg pairings in perpendicular form as metastable structures. When the dielectric constant increased to 2.4 (toluene), another perpendicular-centered metastable structure (**12**, $r = 4.47 \text{ \AA}$, $\alpha = 73.44^\circ$, $\theta = 23.90^\circ$) was obtained besides **10** ($r = 4.46 \text{ \AA}$, $\alpha = 80.38^\circ$, $\theta = 22.10^\circ$) (Figure 5 and Table S1, Supporting Information). One more metastable perpendicular structure was obtained from an initial parallel structure **7** as the dielectric constant further increased (diethyl ether, $\epsilon = 4.3$; chloroform, $\epsilon = 4.8$) (Figure 5 and Table S1 and Figure S1, Supporting Information).

With the increase of dielectric constant, the binding energies decreased for all of the Arg–Arg pairings (Figure 5a and Table S1, Supporting Information). Impressively, when the dielectric constant increased to 46.8 (DMSO), the interaction in all structures is attractive (Figures 2b and 5a and Table S1, Supporting Information). Therefore, it is logical to deduce that the solvent effect contributes essentially to the attraction. To verify this hypothesis, the interaction energies for the 12 Arg–Arg pairings were decomposed as classical electrostatic (ΔE_{elstat}), Pauli repulsion (ΔE_{Pauli}), attractive orbital (ΔE_{orb}), and solvation (ΔE_{sol}) components. The decomposition result of the six representative Arg–Arg pairings (**1**, **2**, **3**, **5**, **11**, and **12**) in acetone ($\epsilon = 20.5$), DMSO ($\epsilon = 46.8$) (Figure 2b), and

Table 1. Geometrical Parameters and Binding Energies (kcal/mol) of Six Selected Arg–Arg Structures (**1**, **2**, **3**, **5**, **11**, and **12**) Calculated with the B97-D/6-311++G(d,p) Approach in a Series of Solvents (Diethyl Ether, $\epsilon = 4.3$; Acetone, $\epsilon = 20.5$; DMSO, $\epsilon = 46.8$; water, $\epsilon = 78.4$)

solvent	parameter	1	2	3	5	11	12
initial	r^a	3.61	3.53	3.61	3.61	3.93	3.99
	α^a	0.04	0.08	0.06	0.06	88.63	89.97
	β^a	60.00	179.99	119.99	119.99	147.84	90.00
	θ^a	24.03	21.37	23.96	24.02	12.18	9.32
$\epsilon = 4.3$	r^b	— ^c	4.36	4.91	—	4.87	4.34
	α^b	—	0.02	23.38	—	77.63	77.94
	β^b	—	179.90	150.66	—	98.78	94.95
	θ^b	—	18.40	26.81	—	30.82	20.53
$\epsilon = 20.5$	ΔE_0	—	12.4	11.6	—	12.1	12.5
	r^b	3.41	3.63	3.69	3.42	4.25	4.25
	α^b	4.51	5.10	11.14	4.60	87.64	81.87
	β^b	66.76	144.80	154.37	80.82	132.17	92.51
$\epsilon = 46.8$	θ^b	2.38	23.84	22.25	8.88	19.08	20.20
	ΔE_0	1.2	0.6	-0.1	1.1	1.0	1.3
	r^b	3.40	3.62	3.65	3.40	4.23	4.24
	α^b	3.23	4.48	12.47	3.61	88.52	82.15
$\epsilon = 78.4$	β^b	66.73	144.93	156.16	66.80	131.93	92.25
	θ^b	1.96	24.33	21.94	2.11	18.58	20.06
	ΔE_0	-0.5	-1.0	-1.9	-0.5	-0.7	-0.4
	r^b	3.39	3.48	3.64	3.53	4.09	4.24
	α^b	2.60	0.24	12.25	0.05	72.58	82.23
	β^b	66.65	180.00	156.35	107.66	127.19	92.24
	θ^b	1.70	14.38	21.98	21.26	13.92	20.09
	ΔE_0	-0.7	-1.4	-2.1	-0.5	-0.6	-0.6

^aParameter in the designed initial structures. ^bParameter in the optimized initial structures. ^cStationary point not found.

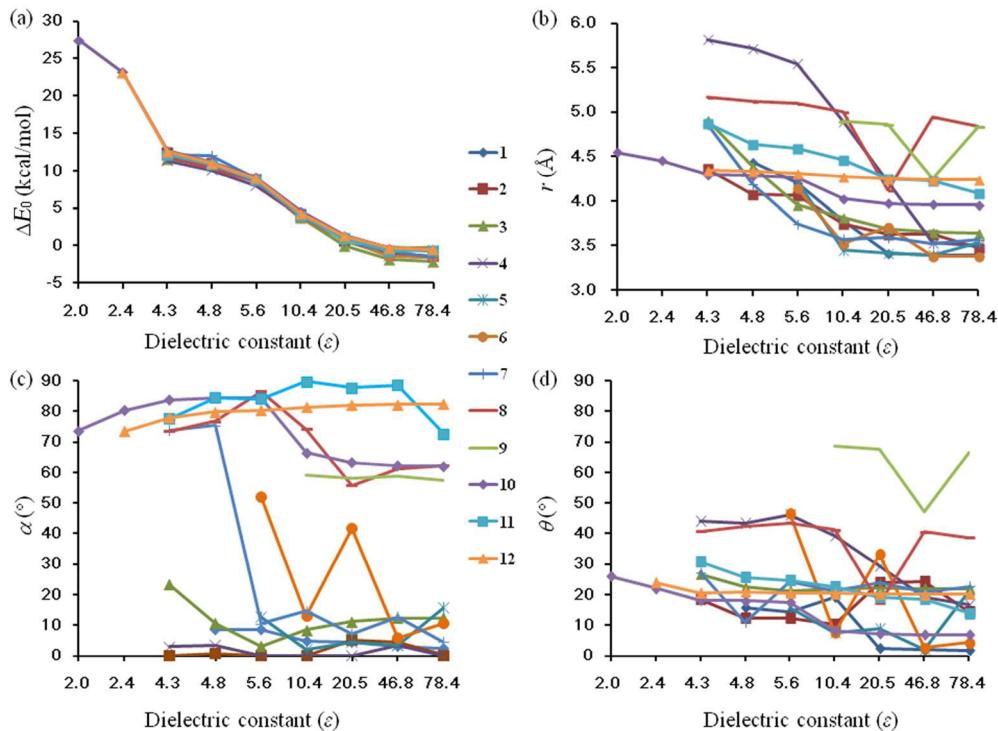


Figure 5. Binding energies (a) and geometrical parameters ((b) r , (c) α , (d) θ) of Arg–Arg dimers, optimized with B97-D/6-311++G(d,p) in a series of solvents (cyclohexane, $\epsilon = 2.0$; toluene, $\epsilon = 2.4$; diethyl ether, $\epsilon = 4.3$; chloroform, $\epsilon = 4.8$; chlorobenzene, $\epsilon = 5.6$; dichloroethane, $\epsilon = 10.4$; acetone, $\epsilon = 20.5$; DMSO, $\epsilon = 46.8$; water, $\epsilon = 78.4$).

Table 2. Energy Components in kcal/mol Calculated Using the PBE Method at the DZP Basis Set of the B97-D/6-311++G(d,p) Optimized Arg–Arg Structures in Acetone ($\epsilon = 20.5$), DMSO ($\epsilon = 46.8$), and Water ($\epsilon = 78.4$)

config	dielectric constant	ΔE_{int}	ΔE_{Pauli}	ΔE_{elstat}	ΔE_{orb}	ΔE_{sol}
1	$\epsilon = 20.5$	3.1	4.3	60.3	-3.8 (6.2)	-57.7 (93.8) ^a
	$\epsilon = 46.8$	1.6	4.5	61.0	-3.9 (6.1)	-60.0 (93.9)
	$\epsilon = 78.4$	0.6	4.5	61.2	-3.8 (5.8)	-61.4 (94.2)
2	$\epsilon = 20.5$	3.1	3.9	60.5	-4.1 (6.6)	-57.2 (93.4)
	$\epsilon = 46.8$	1.7	4.8	61.1	-4.2 (6.5)	-59.4 (93.5)
	$\epsilon = 78.4$	0.9	5.1	62.0	-4.3 (6.5)	-62.0 (93.5)
3	$\epsilon = 20.5$	2.8	5.0	59.1	-4.3 (7.0)	-57.0 (93.0)
	$\epsilon = 46.8$	1.5	5.4	59.9	-4.4 (6.9)	-59.4 (93.1)
	$\epsilon = 78.4$	0.6	5.5	60.1	-4.3 (6.6)	-60.7 (93.4)
5	$\epsilon = 20.5$	3.0	3.9	59.9	-3.5 (5.7)	-57.4 (94.3)
	$\epsilon = 46.8$	1.6	4.5	60.9	-3.9 (6.1)	-59.9 (93.9)
	$\epsilon = 78.4$	1.3	4.4	60.9	-3.7 (5.8)	-60.3 (94.2)
11	$\epsilon = 20.5$	3.5	5.2	58.1	-6.7 (11.1)	-53.2 (88.9)
	$\epsilon = 46.8$	1.9	5.5	58.2	-6.8 (11.0)	-55.0 (89.0)
	$\epsilon = 78.4$	0.4	5.2	58.5	-6.0 (9.5)	-57.2 (90.5)
12	$\epsilon = 20.5$	2.9	4.3	57.5	-6.1 (10.3)	-52.9 (89.7)
	$\epsilon = 46.8$	1.4	4.5	57.6	-6.1 (10.1)	-54.6 (89.9)
	$\epsilon = 78.4$	0.0	4.6	57.5	-5.9 (9.4)	-56.2 (90.6)

^aThe values in parentheses are the percentage contribution to the total attractive interactions $\Delta E_{\text{orb}} + \Delta E_{\text{sol}}$.

water ($\epsilon = 78.4$) (Figure S2, Supporting Information) was shown in Table 2 (Table S3, Supporting Information). Indeed, the solvation energy is the dominant contribution to the attractions, with the contribution more than 90% while the other 10% come from the orbital term (ΔE_{orb}). It is not surprising that there were some discrepancies of the interaction energies between ΔE_{int} (ADF program) and ΔE_0 (Gaussian program), considering the difference of the program, method, and basis sets.

With the increase of dielectric constant, the optimized interaction distance, r , becomes shorter and shorter while the predicted binding strength, ΔE_0 , becomes stronger and stronger (Figure 5a and b, Table 1, and Table S1, Supporting Information). The parallel structures (1, 2, 3, 4, 5, 6, and 7) have shorter interaction distances, ranging from 3.38 to 3.65 Å in DMSO ($\epsilon = 46.8$), than the perpendicular structures (8, 10, 11, and 12), ranging from 3.96 to 4.95 Å, in the same solvent. An oblique structure (9) with an interaction distance of 4.25 Å was obtained via the optimization of initial perpendicular

structure **9** (Figure 2b). As shown in Figure 5d, with the increase of dielectric constant, an overall decreased trend for θ is observed, indicating that centered conformations are more preferred considering the exposed nature of Arg–Arg pairings (Figure 1f).

In high dielectric constant solvents, parallel Arg–Arg pairings (**1**, **2**, **3**, **4**, **5**, **6**, and **7**) yielded stronger attractive interactions than the perpendicular structures. For instance, the binding energies in DMSO ($\epsilon = 46.8$, Figure 2b) are -0.5 to -1.9 kcal/mol for the different parallel structures, while they are -0.3 to -0.4 kcal/mol for the various perpendicular structures (**8**, **10**, **11**, and **12**) and -0.2 kcal/mol for the oblique structure (**9**) (Figure 5a and Table S1, Supporting Information). In water ($\epsilon = 78.4$), the predicted binding energies for the parallel Arg–Arg pairings are -0.7 to -2.1 kcal/mol, while they are -0.3 to 0.6 kcal/mol for the perpendicular counterparts (Figure 5a and Table S1 and Figure S2, Supporting Information).

In order to estimate the difference caused by different solvent models, a test was performed with structure **3** as an example using the PCM model. The binding energy for **3** in water was calculated to be -2.5 kcal/mol, which is in good agreement with the CPCM model that is -2.1 kcal/mol (Table 1), suggesting no significant difference between the PCM and CPCM models for the Arg–Arg pairing study. BSSE is another factor affecting the calculation results, although the CPCM model does not support BSSE calculation. We select **3** again to estimate its BSSE in a vacuum that is 0.7 kcal/mol at the B97-D/6-311++G(d,p) level, suggesting that BSSE should not affect the conclusion achieved in this study. For comparison between different methods, MP2/6-311++G(d,p) was applied to fully optimize **3**. The calculated binding energy based on the optimized **3** is -2.8 kcal/mol, very close to the result by the B97-D/6-311++G(d,p) method, -2.1 kcal/mol (Table 1).

Interaction of the Arg–Arg–Asp/Glu Clusters. To explore the effect of anionic residues on the stabilization of Arg–Arg pairings, six Arg–Arg–Asp/Glu clusters (Figure 3a) were extracted from the PDB (PDB code: **1RU4**, **2G3R**, **1Q2Y**, **2QS9**, **3NSX**, and **1YWF**) and optimized using the B97-D method with the 6-311++G(d,p) basis set in a vacuum ($\epsilon = 1.0$) and nine different solvents. The optimized structures revealed that the Arg–Arg–Asp/Glu clusters are rather stable with RSMD values less than 2 \AA to the crystal structures in most cases (Figure 3b and Table S4, Supporting Information). The binding energy was estimated using a cluster of a methylguanidinium pair and an acetate anion as a model. As shown in Figure 6, with the increase of dielectric constant, the binding energy between the pairing and acidic residues ranges from -154 to -22 kcal/mol. For comparing the binding energy with that in Arg–Asp/Glu, a complex formed by a methylguanidinium and an acetate anion was also optimized. The binding energy in Arg–Asp/Glu (Figure 6) ranges from -122 to -18 kcal/mol in different solvents, about 25% higher than that between the methylguanidinium pair and acetate anion. Therefore, it is clear the carboxylate group has a significant stabilization effect on the Arg–Arg pairing (Figures 5 and 6). For individual Arg–Arg pairing in water ($\epsilon = 78.4$), the binding energies are around -1 kcal/mol, while the binding energies of the (Arg–Arg)–Asp/Glu clusters are about -25 kcal/mol. In the lower dielectric constant solvents, the binding between Arg–Arg pairing and Asp/Glu is even stronger (Figures 6 and 5a), suggesting again that the acidic residues could tremendously stabilize the Arg–Arg pairings that are located inside protein. EDA calculations were performed for the

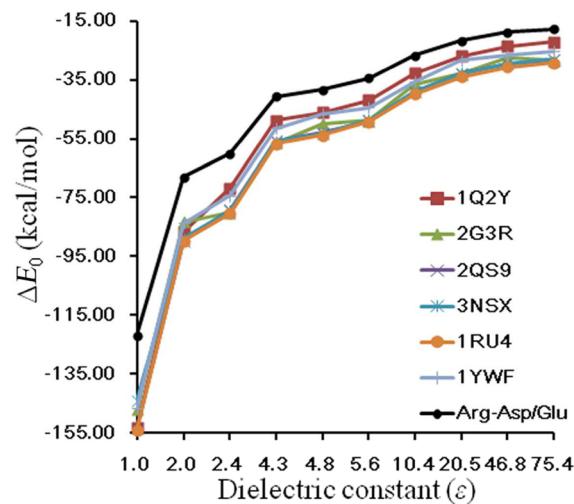


Figure 6. Binding energies of Arg–Asp/Glu and (Arg–Arg)–Asp/Glu clusters, optimized with B97-D/6-311++G(d,p) in a vacuum ($\epsilon = 1.0$) and a series of solvents (cyclohexane, $\epsilon = 2.0$; toluene, $\epsilon = 2.4$; diethyl ether, $\epsilon = 4.3$; chloroform, $\epsilon = 4.8$; chlorobenzene, $\epsilon = 5.6$; dichloroethane, $\epsilon = 10.4$; acetone, $\epsilon = 20.5$; DMSO, $\epsilon = 46.8$; water, $\epsilon = 78.4$).

Arg–Arg–Asp/Glu cluster using ADF (Table S5, Supporting Information). Some selected results calculated in the solvents acetone ($\epsilon = 20.5$), DMSO ($\epsilon = 46.8$), and water ($\epsilon = 78.4$) were shown in Table 3. It was found that electrostatic interaction is the dominant contribution to the attractions between the Arg–Arg and Asp/Glu moieties, with the contribution more than 75% while the other 25% comes from the orbital term (ΔE_{orb}), which is different from the binding in Arg–Arg pairings that solvation energy is the dominant contribution.

CONCLUSIONS

Database survey in this study revealed that about one-third of the protein structures deposited in the PDB contain at least one Arg–Arg pairing in the whole PDB and about one-fourth of the nonredundant structures have at least one Arg–Arg pairing. It is found that all the Arg–Arg pairings have a polar counterpart like water molecules, acidic residues, and polar or polarizable moieties. Furthermore, most of the Arg–Arg pairings are either solvent exposed or stabilized by acidic residues to form Arg–Arg–Asp/Glu clusters. The structures of Arg–Arg pairings and Arg–Arg–Asp/Glu clusters were optimized in a series of solvents with dielectric constants ranging from 1 to 78.4, and the results suggested that the perpendicular Arg–Arg pairings are favorable in low dielectric constant solvents ($\epsilon \leq 20.5$) as metastable structures, whereas the parallel Arg–Arg pairings are favorable in high dielectric constant solvents ($\epsilon \geq 46.8$) as thermodynamically stable structures. Arg–Arg pairings become attractive when the dielectric constant increases to 46.8 (DMSO). In a strong polar environment, the binding energies of Arg–Arg pairing are calculated to be -1 kcal/mol, while the binding energies of (Arg–Arg)–Asp/Glu clusters amount to about -25 kcal/mol. In addition, the binding energy decomposition suggested that the solvation effect plays an essential role in stabilizing the Arg–Arg pairing structures, while the electrostatic interaction contributes to attraction between Arg–Arg pairing and Asp/Glu moieties (more than 75%). The data of the Arg–Arg interaction between ligand and

Table 3. Energy Components in kcal/mol Calculated Using the PBE Method at the DZP Basis Set of B97-D/6-311++G(d,p) Optimized Arg–Arg–Asp/Glu Clusters in Acetone ($\epsilon = 20.5$), DMSO ($\epsilon = 46.8$), and Water ($\epsilon = 78.4$)

PDB code	dielectric constant	ΔE_{int}	ΔE_{Pauli}	ΔE_{elstat}	ΔE_{orb}	ΔE_{sol}
1Q2Y	$\epsilon = 20.5$	-26.3	50.8	-166.7 (80.8) ^a	-39.5 (19.2)	129.1
	$\epsilon = 46.8$	-22.4	48.2	-164.4 (81.5)	-37.4 (18.5)	131.2
	$\epsilon = 78.4$	-20.4	49.6	-165.0 (81.0)	-38.6 (19.0)	133.7
2G3R	$\epsilon = 20.5$	-34.5	51.7	-179.0 (81.4)	-40.8 (18.6)	133.7
	$\epsilon = 46.8$	-27.1	53.3	-166.3 (78.7)	-45.0 (21.3)	130.9
	$\epsilon = 78.4$	-25.2	50.6	-171.0 (80.5)	-41.4 (19.5)	136.6
2QS9	$\epsilon = 20.5$	-33.6	51.5	-174.6 (80.5)	-42.2 (19.5)	131.7
	$\epsilon = 46.8$	-29.6	50.9	-173.9 (80.7)	-41.6 (19.3)	135.0
	$\epsilon = 78.4$	-27.0	50.7	-174.2 (80.9)	-41.1 (19.1)	137.6
3NSX	$\epsilon = 20.5$	-34.5	50.2	-179.8 (82.0)	-39.4 (18.0)	134.4
	$\epsilon = 46.8$	-30.3	50.1	-178.3 (81.8)	-39.8 (18.2)	137.6
	$\epsilon = 78.4$	-27.7	50.0	-178.5 (82.0)	-39.3 (18.0)	140.1
1RU4	$\epsilon = 20.5$	-33.0	52.8	-173.8 (80.0)	-43.5 (20.0)	131.4
	$\epsilon = 46.8$	-29.0	52.3	-173.2 (80.1)	-43.0 (19.9)	134.7
	$\epsilon = 78.4$	-26.5	52.1	-173.5 (80.4)	-42.4 (19.6)	137.3
1YWF	$\epsilon = 20.5$	-28.2	50.4	-159.4 (78.9)	-42.7 (21.1)	123.5
	$\epsilon = 46.8$	-25.4	49.9	-159.7 (79.0)	-42.5 (21.0)	126.9
	$\epsilon = 78.4$	-23.1	49.6	-159.9 (79.3)	-41.8 (20.7)	128.9

^aThe values in parentheses are the percentage contribution to the total attractive interactions $\Delta E_{\text{orb}} + \Delta E_{\text{elstat}}$.

receptor presented in this study should be very useful in drug discovery and design.

ASSOCIATED CONTENT

Supporting Information

Twelve optimized Arg–Arg pairing structures in chloroform ($\epsilon = 4.8$) and water ($\epsilon = 78.4$). Arg–Arg–Asp/Glu optimized structures in vacuum ($\epsilon = 1.0$), chloroform ($\epsilon = 4.8$), and chlorobenzene ($\epsilon = 10.4$). Geometrical parameters and binding energies of Arg–Arg dimers, optimized with B97-D/6-311++G(d,p) in a series of nine solvents (cyclohexane, $\epsilon = 2.0$; toluene, $\epsilon = 2.4$; diethyl ether, $\epsilon = 4.3$; chloroform, $\epsilon = 4.8$; chlorobenzene, $\epsilon = 5.6$; dichloroethane, $\epsilon = 10.4$; acetone, $\epsilon = 20.5$; DMSO, $\epsilon = 46.8$; water, $\epsilon = 78.4$). Binding energies of six selected Arg–Arg structures, optimized with B97-D at 6-31G(d,p), 6-311++G(d,p), cc-pVTZ, and aug-cc-pVTZ basis sets in a series of solvents (diethyl ether, $\epsilon = 4.3$; chloroform, $\epsilon = 4.8$; chlorobenzene, $\epsilon = 5.6$; dichloroethane, $\epsilon = 10.4$; acetone, $\epsilon = 20.5$; DMSO, $\epsilon = 46.8$; water, $\epsilon = 78.4$). Energy components calculated using the PBE method at the DZP basis set in a series of solvents (chloroform, $\epsilon = 4.8$; chlorobenzene, $\epsilon = 5.6$; dichloroethane, $\epsilon = 10.4$; acetone, $\epsilon = 20.5$; DMSO, $\epsilon = 46.8$; water, $\epsilon = 78.4$) with the B97-D/6-311++G(d,p) optimized structures. Root mean square derivations (RMSDs) to the initial structure of Arg–Arg–Asp/Glu clusters and binding energies, optimized with B97-D/6-311++G(d,p) in a vacuum ($\epsilon = 1.0$) and nine solvents. Energy components of (Arg–Arg)–Asp/Glu clusters calculated using the PBE method at the DZP basis set in a vacuum ($\epsilon = 1.0$) and nine solvents with the B97-D/6-311++G(d,p) optimized structures. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

AUTHOR INFORMATION

Corresponding Author

*Phone: +86-512-65882230 (S.L.); +86-21-64252767 (Y.L.); +86-21-50805020 (W.Z.). E-mail: shujinli@suda.edu.cn (S.L.); yxlu@ecust.edu.cn (Y.L.); wlzhu@mail.shcnc.ac.cn (W.Z.). Fax: +86-21-50807088 (W.Z.).

Author Contributions

[¶]Z.Z., Z.X.: These authors contributed equally to the work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the Ministry of Science and Technology (2012AA01A305), National Major Project (2012ZX09301001-004), National Natural Science Foundation (81273435, 20977064, 21103047, 91013008, and 21021063) and the priority academic program development of Jiangsu higher education institutions.

REFERENCES

- (1) Sapse, A. M.; Massa, L. J. Guanidinium Ion: SCF Calculations. *J. Org. Chem.* **1980**, *45* (4), 719–721.
- (2) Magalhaes, A.; Maigret, B.; Hoflack, J.; Gomes, J. N.; Scheraga, H. A. Contribution of Unusual Arginine–Arginine Short-Range Interactions to Stabilization and Recognition in Proteins. *J. Protein Chem.* **1994**, *13* (2), 195–215.
- (3) Pednekar, D.; Tendulkar, A.; Durani, S. Electrostatics-Defying Interaction Between Arginine Termini as a Thermodynamic Driving Force in Protein-Protein Interaction. *Proteins* **2009**, *74* (1), 155–63.
- (4) Singh, J.; Thornton, J. M.; Snarey, M.; Campbell, S. F. The Geometries of Interacting Arginine-Carboxyls in Proteins. *FEBS Lett.* **1987**, *224* (1), 161–171.
- (5) Sumowski, C. V.; Schmitt, B. B. T.; Schweizer, S.; Ochsenfeld, C. Quantum-Chemical and Combined Quantum-Chemical/Molecular-Mechanical Studies on the Stabilization of a Twin Arginine Pair in Adenovirus Ad11. *Angew. Chem., Int. Ed.* **2010**, *49* (51), 9951–9955.
- (6) Vazdar, M.; Vymetal, J.; Heyda, J.; Vondrasek, J.; Jungwirth, P. Like-Charge Guanidinium Pairing from Molecular Dynamics and Ab Initio Calculations. *J. Phys. Chem. A* **2011**, *111* 11193–11201.
- (7) No, K. T.; Nam, K. Y.; Scheraga, H. A. Stability of Like and Oppositely Charged Organic Ion Pairs in Aqueous Solution. *J. Am. Chem. Soc.* **1997**, *119* (52), 12917–12922.
- (8) Neves, M. A.; Yeager, M.; Abagyan, R. Inter- and Intramolecular Dispersion Interactions. *J. Comput. Chem.* **2010**, *32*, 1117–1127.
- (9) Vondrasek, J.; Mason, P. E.; Heyda, J.; Collins, K. D.; Jungwirth, P. The Molecular Origin of Like-Charge Arginine–Arginine Pairing in Water. *J. Phys. Chem. B* **2009**, *113* (27), 9041–9045.

- (10) Vazdar, M.; Uhlig, F.; Jungwirth, P. Like-Charge Ion Pairing in Water: An Ab Initio Molecular Dynamics Study of Aqueous Guanidinium Cations. *J. Phys. Chem. Lett.* **2012**, *3* (15), 2021–2024.
- (11) Wernersson, E.; Heyda, J.; Vazdar, M.; Mikael Lund, P. E. M.; Jungwirth, P. Orientational Dependence of the Affinity of Guanidinium Ions to the Water Surface. *J. Phys. Chem. B* **2011**, *115* (43), 12521–12526.
- (12) Kubíčková, A.; Křížek, T. s.; Coufal, P.; Wernersson, E.; Heyda, J.; Jungwirth, P. Guanidinium Cations Pair with Positively Charged Arginine Side Chains in Water. *J. Phys. Chem. Lett.* **2011**, *2* (12), 1387–1389.
- (13) Karp, D. A.; Gittis, A. G.; Stahley, M. R.; Fitch, C. A.; Stites, W. E.; García-Moreno, E. B. High Apparent Dielectric Constant Inside a Protein Reflects Structural Reorganization Coupled to the Ionization of an Internal Asp. *Biophys. J.* **2007**, *92* (6), 2041–2053.
- (14) Patargias, G. N.; Harris, S. A.; Harding, J. H. A Demonstration of the Inhomogeneity of the Local Dielectric Response of Proteins by Molecular Dynamics Simulations. *J. Chem. Phys.* **2010**, *132* (23), 235103–1–235103–8.
- (15) Simonson, T.; Brooks, C. L. Charge Screening and the Dielectric Constant of Proteins: Insights from Molecular Dynamics. *J. Am. Chem. Soc.* **1996**, *118* (35), 8452–8458.
- (16) Guest, W. C.; Cashman, N. R.; Plotkin, S. S. A Theory for the Anisotropic and Inhomogeneous Dielectric Properties of Proteins. *Phys. Chem. Chem. Phys.* **2011**, *13* (13), 6286–6295.
- (17) Mellor, B. L.; Cruz Cortés, E. n.; Busath, D. D.; Mazzeo, B. A. Method for Estimating the Internal Permittivity of Proteins Using Dielectric Spectroscopy. *J. Phys. Chem. B* **2011**, *115* (10), 2205–2213.
- (18) Nakamura, H. A Theoretical Study of the Dielectric Constant of Protein. *Protein Eng.* **1988**, *2* (3), 177–183.
- (19) Lamm, G.; Pack, G. R. Calculation of Dielectric Constants near Polyelectrolytes in Solution. *J. Phys. Chem. B* **1997**, *101* (6), 959–965.
- (20) Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; García-Moreno, E. B. High Apparent Dielectric Constants in the Interior of a Protein Reflect Water Penetration. *Biophys. J.* **2000**, *79* (3), 1610–1620.
- (21) Mitra, M.; Manna, P.; Seth, S. K.; Das, A.; Meredith, J.; Helliwell, M.; Bauzá, A.; Choudhury, S. R.; Frontera, A.; Mukhopadhyay, S. Salt-Bridge- π (sb- π) Interactions at Work: Associative Interactions of sb- π , π - π and anion- π in Cu(ii)-Malonate-2-Aminopyridine-Hexafluoridophosphate Ternary System. *CrystEngComm* **2013**, *15* (4), 686–696.
- (22) Wang, G.; Dunbrack, R. L. PISCES: A Protein Sequence Culling Server. *Bioinformatics* **2003**, *19* (12), 1589–1591.
- (23) Soetens, J.-C.; Millot, C.; Chipot, C.; Jansen, G.; Ángyán, J. G.; Maigret, B. Effect of Polarizability on the Potential of Mean Force of Two Cations. The Guanidinium–Guanidinium Ion Pair in Water. *J. Phys. Chem. B* **1997**, *101* (50), 10910–10917.
- (24) Grimme, S. Density Functional Theory with London Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2011**, *1* (2), 211–228.
- (25) Mackie, I. D.; Dilabio, G. A. Accurate Dispersion Interactions from Standard Density-Functional Theory Methods with Small Basis Sets. *Phys. Chem. Chem. Phys.* **2010**, *12* (23), 6092–6098.
- (26) Grimme, S. Semiempirical GGA-type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27* (15), 1787–1799.
- (27) Yamashita, H.; Yumura, T. The Role of Weak Bonding in Determining the Structure of Thiophene Oligomers inside Carbon Nanotubes. *J. Phys. Chem. C* **2012**, *116* (17), 9681–9690.
- (28) Zhao, Y.; Truhlar, D. G. Density Functional Theory for Reaction Energies: Test of Meta and Hybrid Meta Functionals, Range-Separated Functionals, and Other High-Performance Functionals. *J. Chem. Theory Comput.* **2011**, *7* (3), 669–676.
- (29) 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.; et al. *Gaussian 09*, revision C.01; Gaussian, Inc.: Wallingford, CT, 2009.
- (30) Gao, W.; Feng, H.; Xuan, X.; Chen, L. The Assessment and Application of an Approach to Noncovalent Interactions: The Energy Decomposition Analysis (EDA) in Combination with DFT of Revised Dispersion Correction (DFT-D3) with Slater-type Orbital (STO) Basis Set. *J. Mol. Model.* **2012**, *18* (10), 4577–4589.
- (31) Liu, Z. Chemical Bonding in Silicon–Carbene Complexes. *J. Phys. Chem. A* **2009**, *113* (22), 6410–6414.
- (32) Wu, Q.; Ayers, P. W.; Zhang, Y. Density-Based Energy Decomposition Analysis for Intermolecular Interactions with Variationally Determined Intermediate State Energies. *J. Chem. Phys.* **2009**, *131* (16), 164112–1–164112–8.
- (33) Velde, G. T.; Bickelhaupt, F. M.; Baerends, E. J.; Guerra, C. F.; Gisbergen, S. J. A. V.; Snijders, J. G.; Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22* (9), 931–967.
- (34) Bayat, M.; von Hopffgarten, M.; Salehzadeh, S.; Frenking, G. Energy Decomposition Analysis of the Metal–Oxime Bond in [M{RC(NOH)C(NO)R}2] (M = Ni(II), Pd(II), Pt(II), R = CH₃, H, F, Cl, Br, Ph, CF₃). *J. Organomet. Chem.* **2011**, *696* (18), 2976–2984.
- (35) Szatyłowicz, H.; Sadlej-Sosnowska, N. Characterizing the Strength of Individual Hydrogen Bonds in DNA Base Pairs. *J. Chem. Inf. Model.* **2010**, *50* (12), 2151–2161.
- (36) Grant, D. J.; Weng, Z.; Jouffret, L. J.; Burns, P. C.; Gagliardi, L. Synthesis of a Uranyl Persulfide Complex and Quantum Chemical Studies of Formation and Topologies of Hypothetical Uranyl Persulfide Cage Clusters. *Inorg. Chem.* **2012**, *51* (14), 7801–7809.
- (37) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. Towards an Order-N DFT Method. *Theor. Chem. Acc.* **1998**, *99* (6), 391–403.
- (38) McGaughey, G. B.; Gagné, M.; Rappé, A. K. π -Stacking Interactions Alive and Well in Proteins. *J. Biol. Chem.* **1998**, *273* (25), 15458–15463.
- (39) Neves, M. A. C.; Yeager, M.; Abagyan, R. Unusual Arginine Formations in Protein Function and Assembly: Rings, Strings, and Stacks. *J. Phys. Chem. B* **2012**, *116* (23), 7006–7013.