

## Halogen-Ionic Bridges: Do They Exist in the Biomolecular World?

Peng Zhou,<sup>†</sup> Yanrong Ren,<sup>§</sup> Feifei Tian,<sup>‡,#</sup> Jianwei Zou,<sup>¶</sup> and Zhicai Shang<sup>\*,†</sup>

*Department of Chemistry, Zhejiang University, Hangzhou 310027, China, Department of Biological and Chemical Engineering, Chongqing Education College, Chongqing 400067, China, College of Bioengineering, Chongqing University, Chongqing 400044, China, Key Laboratory for Molecular Design and Nutrition Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China, and Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, Florida 32611*

Received March 27, 2010

**Abstract:** If considering that the pronouncedly charged halide anions are ubiquitous in the biological world, then it is interesting to ask whether the halogen-ionic bridges—this term is named by us to describe the interaction motif of a nonbonded halogen ion with two or more electrophiles simultaneously—commonly exist in biomolecules and how they contribute to the stability and specificity of biomolecular folding and binding? To address these problems, we herein present a particularly systematic investigation on the geometrical profile and the energy landscape of halogen ions interacting with and bridging between polar and charged molecular moieties in small model systems and real crystal structures, by means of ab initio calculation, database survey, continuum electrostatic analysis, and hybrid quantum mechanics/molecular mechanics examination. All of these unequivocally demonstrate that this putative halide motif is broadly distributed in biomolecular systems ( $>6000$ ) and can confer a substantial stabilization for the architecture of proteins and their complexes with nucleic acids and small ligands. This stabilization energy is estimated to be generally more than  $100 \text{ kcal}\cdot\text{mol}^{-1}$  for gas-phase states or about  $20 \text{ kcal}\cdot\text{mol}^{-1}$  for solution conditions, which is much greater than that found in sophisticated water-mediated ( $<10 \text{ kcal}\cdot\text{mol}^{-1}$ ) and salt ( $\sim 3.66 \text{ kcal}\cdot\text{mol}^{-1}$ ) bridges. In this respect, we would expect that the proposed halogen-ionic bridge, which has long been unrecognized in the arena of biological repertoires, could be appreciated in chemistry and biology communities and might be exploited as a new and versatile tool for rational drug design and bioengineering.

### 1. Introduction

Specific ion effects play an essential role in many physicochemical and biological processes. Such effects exhibit a reoccurring trend called the Hofmeister series.<sup>1</sup> Originally,

it was thought that an ion's influence on macromolecular properties was caused at least in part by ‘making’ or ‘breaking’ bulk water structures.<sup>2</sup> Recent time-resolved and thermodynamic studies of water molecules in salt solutions, however, shed light on that, instead of remodeling water structures through ions, direct macromolecule–ion interactions as well as the interactions with water molecules that are bound to the macromolecules seem to be more responsible for the Hofmeister effect.<sup>3</sup>

In fact, the metal cation–protein/nucleic acid interactions, which are commonly known as coordinate bonding, have been well characterized in chemistry and biology communities. In contrast, the ubiquitous anions in biologi-

\* Corresponding author. Telephone: +86-0571-87952379. Fax: +86-0571-87951895. E-mail: shangzc@zju.edu.cn.

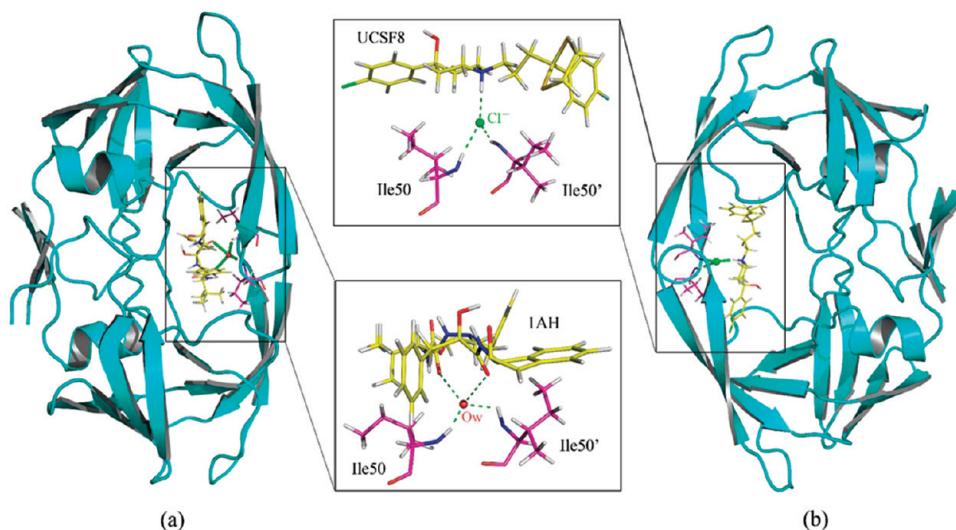
<sup>†</sup> Department of Chemistry, Zhejiang University.

<sup>§</sup> Department of Biological & Chemical Engineering, Chongqing Education College.

<sup>‡</sup> College of Bioengineering, Chongqing University.

<sup>#</sup> Department of Chemistry, University of Florida.

<sup>¶</sup> Ningbo Institute of Technology, Zhejiang University.



**Figure 1.** Crystallographic evidence showing that the role of a water molecule in mediating the hydrogen-bond network in biomolecules can be functionally replaced by a halogen ion. Usually, a conserved water molecule (Ow) is located in the active pocket of HIV-1 protease to mediate the hydrogen-bond network with its cognate substrates and noncognate inhibitors (PDB: 2cej) (a). However, there has an exception that a chlorine ion ( $\text{Cl}^-$ ) is observed at the water's position in the complex of HIV-1 protease with its nonpeptide inhibitor UCSF8 (PDB: 1aid) (b).

cal systems, such as  $\text{X}^-$  (where  $\text{X}^- = \text{F}^-, \text{Cl}^-, \text{Br}^-, \text{I}^-$ ,  $\text{SO}_4^{2-}$ , and  $\text{H}_2\text{PO}_4^-$ ), are traditionally recognized as counterions, and their interactions with biomolecules as well as the effect of these interactions on biomolecular functions have long been underappreciated in the field of biology. Theoretically, anions, especially the small, hard halogen ions, are expected to serve as a good hydrogen-bond acceptor and a friendly ion-pairing partner to specifically and nonspecifically interact with the polar hydrogen atoms and the basic groups of biomolecules. This point has been preliminarily rationalized by experimental and theoretical studies of simple molecular complexes in gas phase or vacuum conditions (see reviews).<sup>4,5</sup> For example, the dissociation enthalpies of  $\text{X}^- \cdots \text{H}_2\text{O}$  and  $\text{CH}_3(\text{CH}_2)_n\text{OH} \cdots \text{X}^-$  adducts were early determined to be  $\sim 10\text{--}30 \text{ kcal} \cdot \text{mol}^{-1}$  by using high-pressure mass spectrometry,<sup>6–9</sup> which are even six-fold greater than the stabilization energy of a water dimer ( $\sim 5 \text{ kcal} \cdot \text{mol}^{-1}$ ).<sup>10</sup> The experimentally measured energies were later systematized by means of ab initio calculations using model systems.<sup>11–15</sup>

Close contacts between the nonbonded halogen ions and the hydrogen atoms were observed in crystal structures of amino acids, peptides, and related molecules as early as 30 years ago.<sup>16</sup> In the past two decades, the interactions of halogen ions with macromolecular systems, including colloids, polymers, and proteins, were investigated intensively via nuclear magnetic resonance (NMR),<sup>17</sup> aqueous gel sieving chromatography,<sup>18</sup> and vibrational sum frequency spectroscopy<sup>19</sup> as well as molecular dynamics simulation.<sup>20</sup> Particularly, it was found that specific ion effects on protein stability could be explained by incorporating the ionic dispersion potentials into classical double-layer theory<sup>21</sup> and that small anions, such as  $\text{F}^-$ , are prone to pair with charged groups, while larger anions, such as  $\text{I}^-$ , are more likely to be bound on hydrophobic patches of protein surfaces.<sup>22</sup> Very recently and also very intriguingly, Heyda et al. have presented computational evidence for the ion-specific inter-

actions between biological entities and halides. By employing both nonpolarizable and polarizable force fields to simulate the dynamic behavior of amino acid–ion systems, they attained several clear trajectory pictures showing obvious congregations of halogen anions around the positively charged hydrogen atoms of basic amino acids.<sup>23</sup>

It is known that water molecule can serve as mediator to “glue” adjacent polar groups together through hydrogen bonds and hydrophilic forces. Traditionally, these water-participating interactions are referred to as a water-mediated hydrogen-bond bridge<sup>24</sup> and a water-induced hydrophilic interaction.<sup>25</sup> Given that the halide anions, as mentioned above, are shown to be effective in interaction with biomolecules, a question would be raised naturally, that is, whether the halogen ions can bridge between the spatially vicinal moieties in biomolecules, just like what the water molecules do? In other words, do the putative halogen ion-participating interaction motifs, that we named halogen-ionic bridges to stress their shared similarities with water-mediated bridges, exist in the biomolecular world? Actually, there has been at least one crystallographic report clearly showing a  $\text{Cl}^-$  bound functionally between the residues Ile50/Ile50' of HIV-1 protease and the protonated tertiary amine of its nonpeptide inhibitor UCSF8, a haloperidol derivative which strongly inhibits both wild-type and mutant HIV-1 proteases (Figure 1b).<sup>26</sup> As we know, however, this  $\text{Cl}^-$  position is usually occupied by a conserved water molecule (Figure 1a).<sup>27</sup>

To address these open questions related to the existence and significance of halogen-ionic bridges in biological context, in the present work we launch a systematic investigation on this putative halide motif through various theoretical and computational approaches. First, high-level quantum mechanical (QM) calculations were carried out for a series of small model systems to elucidate the geometrical preference and the energy landscape of halogen ions interacting with model molecules which mimic polar and charged

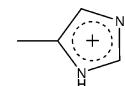
biomolecular groups. The resulting geometrical and energetic features of such interactions subsequently recurred in an exhaustive survey of the high-resolution crystal structures of all the biomolecules deposited in the Protein Data Bank (PDB),<sup>28</sup> including proteins, nucleic acids, and their complexes with small ligands. In particular, the electrostatic property and structural basis of halogen-ionic bridges in real biomacromolecular systems and their contributions toward the stability and specificity of protein architecture and protein–ligand recognition were analyzed in detail with the Poisson–Boltzmann model and a two-layer quantum mechanics/molecular mechanics (QM/MM) scheme. This study would provide solid evidence for the halogen-ionic bridges existing in and functionalizing to biomolecules and might give a new view to support the notion that direct ion–macromolecule interactions, rather than indirect water structures making and breaking by ions, are more responsible for the specific ion effects on biological systems.

## 2. Methods and Materials

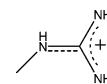
**2.1. Quantum-Mechanical (QM) Calculation.** The simplest model systems, i.e., water molecule ( $\text{H}_2\text{O}$ ) and ammonium ion ( $\text{NH}_4^+$ ) in complex with four kinds of halogen ions ( $\text{F}^-$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{I}^-$ ) and  $\text{H}_2\text{O}$  (serving as the neutral counterpart of halogen ions), were used to perform a detailed examination of geometrical, energetic, and electronic properties associated with the interactions of halogen ions with polar and charged groups, respectively. Potential energy surface scans were carried out at the Møller–Plesset second-order perturbation theory level,<sup>29</sup> in conjunction with the Dunning's augmented correlation consistent basis set, MP2/aug-cc-pVDZ. The equilibrium structures as well as corresponding parameters of atoms in molecules (AIM)<sup>30</sup> and natural bond orbitals (NBO)<sup>31</sup> for these complexes were further obtained at the MP2/aug-cc-pVTZ level of theory; the more accurate intermolecular potentials for the equilibrium structures were evaluated using the coupled cluster with single, double, and noniterative triple excitations correction term, CCSD(T)/aug-cc-pVTZ. The supermolecule approach was employed to obtain intermolecular potentials (viz.  $\Delta E_{\text{int}} = E_{\text{complex}} - E_{\text{monomer1}} - E_{\text{monomer2}}$ ),<sup>32</sup> and the associated basis set superposition error (BSSE) was eliminated by the standard counterpoise method of Boys and Bernardi.<sup>33</sup> The ideal and real intermolecular Coulombic energies ( $\Delta E_{\text{coul}}^{\text{ideal}}$  and  $\Delta E_{\text{coul}}^{\text{real}}$ ) were calculated in terms of the classical Coulomb's law using the natural charges derived from NBO analysis of the complex members in isolated and in complexed states, respectively. Since Dunning's basis set series is unavailable for iodine, the Lanl2DZ basis set, augmented by a set of *d* and *f* polarization functions (exponents 0.292 and 0.441, respectively) and *s* and *p* diffuse functions (exponents 0.0569 and 0.0330, respectively), abbr. Lanl2DZ+(df), was used for  $\text{I}^-$ . This large version of a valence electron orbit seems to be necessary for reliably describing the outer electronic structure of diffuse anions, and previous theoretical calculations which used this modified effective core potential (ECP) basis set have been shown to give reasonably good results for the  $\text{I}^-$ -participating  $\text{S}_{\text{N}}2$  reactions<sup>34</sup> and the

$\text{OCS}\cdots\text{I}^-$  van der Waals complexes.<sup>35</sup> Because no physical meaning can be ascribed to regions of Cartesian space delimited by zero-flux surfaces derived from the valence electron densities,<sup>36</sup> the wave functions generated from the all electron basis set DGDZVP, but not the valence basis set Lanl2DZ+(df), were used for AIM analysis of iodine-containing systems.<sup>37</sup>

To inspect the interaction profile of halogen ions with the protein moieties of interest, a thorough search for all the low-lying energy structures of  $\text{Cl}^-$  binding to the electrophilic hydrogen atoms of six protein groups, respectively, modeled by methanol ( $\text{CH}_3\text{OH}$ ) (for hydroxyl group), *N*-methylacetamide ( $\text{CH}_3\text{CONHCH}_3$ ) (for main chain's amide), acetamide ( $\text{CH}_3\text{CONH}_2$ ) (for side chain's amide), methylammonium ( $\text{CH}_3\text{NH}_3^+$ ) (for lysine's ammonium), 4-methylimidazolium



(for histidine's imidazolium), and *N*-methylguanidinium

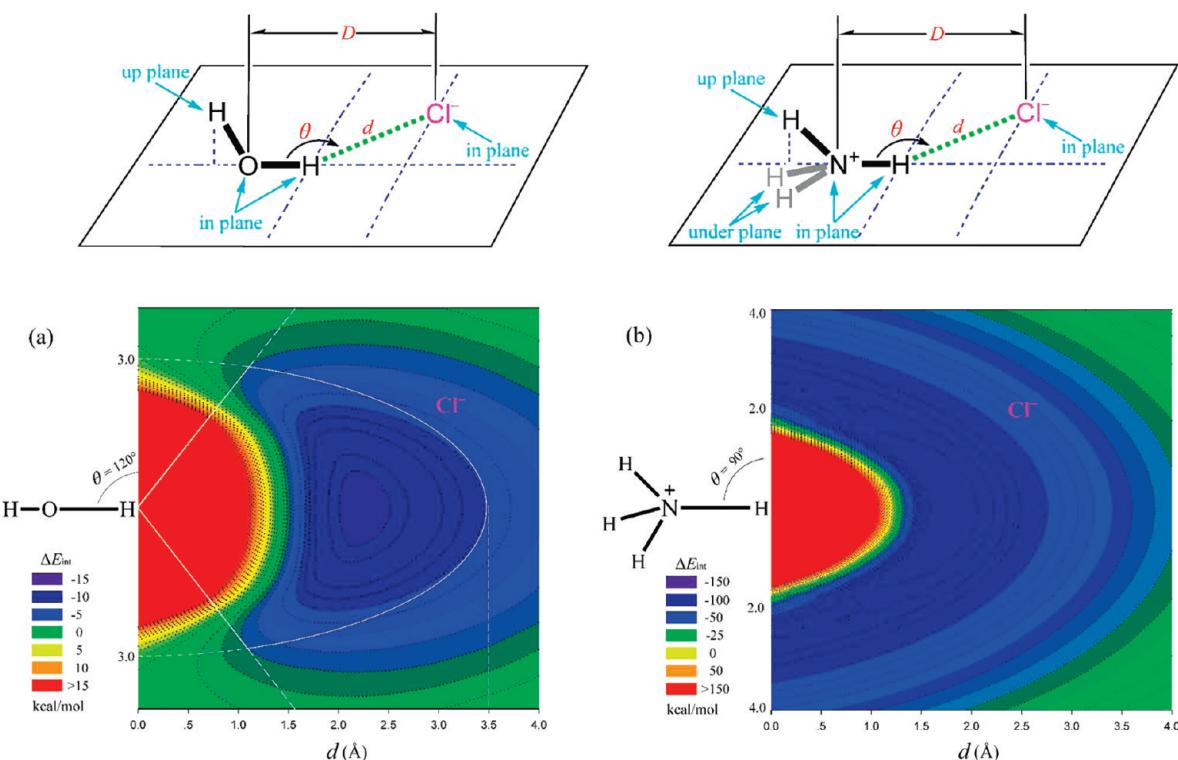


(for arginine's guanidinium), has been done at the MP2/aug-cc-pVDZ level. No symmetries were constrained in optimization procedures, and the stability of optimized structures was confirmed in the following vibrational frequency analysis.

A two-layer ONIOM-based QM/MM scheme<sup>38</sup> was adopted to fully optimize and energetically analyze the protein–ligand interactions through halogen-ionic bridges. The central halogen ion and the corresponding protein residues and ligand that are directly bound to the halogen ion were included in the QM layer and treated at a high level of density functional theory (B3LYP/6-31+G\*), while the rest atoms were in the MM layer and treated at a low level of molecular force field (AMBER parm96).<sup>39</sup> In the MM layer, water molecules were described by the TIP3P model,<sup>40</sup> and the restricted electrostatic potential (RESP) fitting procedure<sup>41</sup> was employed to assign partial atomic charges for small ligands and nonstandard amino acid atoms. Parameters that were not found in standard AMBER force fields were defined using the generalized amber force field (GAFF).<sup>42</sup> Recently, we have successfully employed this ONIOM protocol (only slight modification) to investigate the halogen–water–hydrogen bridges<sup>43</sup> and the fluorine bonds<sup>44</sup> in protein structures and, therefore, expect that this hybrid QM/MM methodology could be used for the halogen-ionic bridge systems as well.

Structure optimizations, energy evaluations, and ONIOM calculations were carried out with the help of the GAUSSIAN 03 suite of programs.<sup>45</sup> AIM and NBO analyses were implemented in AIM2000<sup>46</sup> and NBO5.0,<sup>47</sup> respectively.

**2.2. Database Survey. Pretreatment of PDB Files.** Up to January, 2010, there were 3391 protein records and 133 nucleic acid entries (solved at 3 Å or better) deposited in the PDB in which at least one nonbonded halogen ion is contained. These structures were extracted and treated with following procedure: (i) removing water molecules, metal



**Figure 2.** Potential energy surface scans for the complexes  $\text{OH}_2 \cdots \text{Cl}^-$  (a) and  $\text{NH}_4^+ \cdots \text{Cl}^-$  (b) at the MP2/aug-cc-pVDZ level of theory. In this procedure, the positions of  $\text{H}_2\text{O}$  and  $\text{NH}_4^+$  with standard geometries were fixed, and then the probe  $\text{Cl}^-$  was employed to detect its interaction potentials with the fixed  $\text{H}_2\text{O}$  and  $\text{NH}_4^+$  when it ranges over the plane space as shown in top schematic diagrams. As a result, two two-dimensional contours intuitively displaying the intermolecular potential  $\Delta E_{\text{int}}$  as a function of the geometrical arrangement of the complexes were presented. The step sizes for  $d$  and  $\theta$  in the scans were set to 0.2 Å (from 1.0 to 4.0 Å) and 10° (from 90 to 180°), respectively.

ions, and other cofactors, except halogen ions and small organic ligands, (ii) repairing missing side chains of protein residues, using the newly released SCWRL4 program,<sup>48</sup> (iii) assigning secondary structure class for protein residues, according to the dictionary secondary structure of proteins (DSSP) protocol,<sup>49</sup> (iv) adding hydrogen atoms for all protein and nucleic acid heavy atoms, using the REDUCE program<sup>50</sup> (REDUCE was adopted here because this program was tested in our previous study to be capable of precisely reproducing the neutron diffraction-determined hydrogen's positions),<sup>51</sup> (v) defining protonation state of all charged residues at pH 7.0, using the PROPKA 2.0 program,<sup>52</sup> and (vi) interpreting structural information of small ligands, which are marked by header 'HETATM' in the PDB files, using the I-INTERPRET program.<sup>53</sup> This program reads an assembly of ligands in standard PDB format and writes a MOL2 file in which the atomic states, connection manners, and neutral/charged hydrogen's positions are assigned in a considerable accuracy for these ligands.

**Definition of Geometrical Constraints.** In order to determine the appropriate geometrical constraints used for screening effective interactions between halogen ions and biomolecules in these treated PDB structures, the biomolecular groups that perform as potential halogen ion-acceptors were roughly classified into two types as polar and charged, which were respectively modeled using the  $\text{H}_2\text{O}$  and  $\text{NH}_4^+$ , and then, potential energy surface scans for the complexes  $\text{OH}_2 \cdots \text{Cl}^-$  and  $\text{NH}_4^+ \cdots \text{Cl}^-$  with systematically varying in distance  $d_{\text{H} \cdots \text{Cl}^-}$  and angle  $\theta_{\angle(\text{O/N}-\text{H} \cdots \text{Cl}^-)}$  have been done at

the MP2/aug-cc-pVDZ level. As a result, two two-dimensional contours intuitively displaying the intermolecular potential  $\Delta E_{\text{int}}$  as a function of the geometrical arrangement of the complexes were presented (Figure 2). A significant difference between these two potential landscapes can be seen. For the complex  $\text{NH}_4^+ \cdots \text{Cl}^-$ , a strong repulsion appears at the region (red) nearby the interacting H atom, which is surrounded by a prominently attractive area (navy blue) and, farther out, a weak interaction domain (green) (Figure 2b); for the  $\text{OH}_2 \cdots \text{Cl}^-$ , however, the intermolecular potential is anisotropically distributed around the hydrogen (H) atom, a strong attractive potential in the "head on" orientation (navy blue) and a weak repulsive force in the "side on" direction (green) (Figure 2a). According to this finding, together with the conclusions arisen from our other investigations, the following criteria were defined to describe the effective biological interactions involving halogen ions: (i) For a uncharged polar group, an ellipsoid with its center at the polar H atom and its semi-minor/semi-major axis of 3.0/3.5 Å was constructed. Only those halogen ions occurring within the ellipsoidal space and with the forming angle  $\theta > 120^\circ$  were considered (i.e., the region encompassed by white solid line in Figure 2a); and (ii) For a charged basic group, the halogen ions with their distances,  $D$ , to any one of the heavy atoms in the group less than 4.5 Å were considered. In this way, a halogen-ionic bridge can be readily defined as the entity in which a halogen ion effectively interacts with two or more biomolecular groups simultaneously; the number

**Table 1.** Experimentally Measured Values of Hydration Enthalpy  $\Delta H^\circ_{\text{hydr}}$ , Hydration Entropy  $\Delta S^\circ_{\text{hydr}}$ , and Hydration Free Energy  $\Delta G^\circ_{\text{hydr}}$  for Halogen Ions ( $T = 298.15 \text{ K}$ )

halogen ion	$\Delta G^\circ_{\text{hydr}}$ (kcal·mol <sup>-1</sup> ) <sup>a</sup>	$\Delta H^\circ_{\text{hydr}}$ (kcal·mol <sup>-1</sup> ) <sup>a</sup>	$-\bar{T}\Delta S^\circ_{\text{hydr}}$ (kcal·mol <sup>-1</sup> ) <sup>b</sup>
F <sup>-</sup>	-101.9	-111.5	9.6
Cl <sup>-</sup>	-73.9	-79.5	5.6
Br <sup>-</sup>	-70.6	-76.1	5.5
I <sup>-</sup>	-59.5	-62.3	2.8

<sup>a</sup> From ref 59. <sup>b</sup>  $-\bar{T}\Delta S^\circ_{\text{hydr}}$  is obtained by subtracting  $\Delta H^\circ_{\text{hydr}}$  from  $\Delta G^\circ_{\text{hydr}}$ .

of the groups participating in bridging was called the branch degree of this halogen-ionic bridge.

It is worth noting that, although the geometrical criteria presented here were derived on the basis of chlorine ion (the most abundant halogen ion found in biomolecules), this conclusion is also applicable for three other halogen ions.

**2.3. Structural and Energetic Properties of Halogen Ions in Halogen-Ionic Bridges.** Solvent accessible surface area (SASA<sub>brd</sub>) and packing density (PD<sub>brd</sub>) of the bridging halogen ions in protein context can be solved numerically using the, respectively, Sanner's and Voronoi Cell algorithms implemented in the MSMS program<sup>54</sup> and VORONOIA python package<sup>55</sup> with the ProtOr radii (for protein atoms),<sup>56</sup> Shannon effective ionic radii (for halogen ions),<sup>57</sup> and 1.4 Å radii (for water probe). Furthermore, the changes in an ion's hydration enthalpy ( $\Delta\Delta H^\circ_{\text{hydr}}$ ), hydration entropy ( $\Delta\Delta S^\circ_{\text{hydr}}$ ), and hydration free energy ( $\Delta\Delta G^\circ_{\text{hydr}}$ ) due to it transfers from solvent (water) to protein interior and fixed in a halogen-ionic bridge were estimated using the additive models of Ooi et al.<sup>58</sup>

$$\Delta\Delta H^\circ_{\text{hydr}} = \Delta H^\circ_{\text{hydr}} \cdot \left( \frac{\text{SASA}_{\text{brd}} - \text{SASA}_{\text{free}}}{\text{SASA}_{\text{free}}} \right) \quad (1)$$

$$\Delta\Delta S^\circ_{\text{hydr}} = \Delta S^\circ_{\text{hydr}} \cdot \left( \frac{\text{SASA}_{\text{brd}} - \text{SASA}_{\text{free}}}{\text{SASA}_{\text{free}}} \right) \quad (2)$$

$$\Delta\Delta G^\circ_{\text{hydr}} = \Delta G^\circ_{\text{hydr}} \cdot \left( \frac{\text{SASA}_{\text{brd}} - \text{SASA}_{\text{free}}}{\text{SASA}_{\text{free}}} \right) \quad (3)$$

where SASA<sub>brd</sub> and SASA<sub>free</sub> are the SASA of the studied halogen ion in bridging and free states, respectively, and  $\Delta H^\circ_{\text{hydr}}$ ,  $\Delta S^\circ_{\text{hydr}}$ , and  $\Delta G^\circ_{\text{hydr}}$  are the standard hydration enthalpy, hydration entropy, and hydration free energy of the halogen ion, i.e., the changes in enthalpy, entropy, and free energy when it moves from the gas phase to a solvent at the standard conditions (1 atm and 298.15 K). The experimentally measured values of  $\Delta H^\circ_{\text{hydr}}$ ,  $\Delta S^\circ_{\text{hydr}}$ , and  $\Delta G^\circ_{\text{hydr}}$  for the four kinds of halogen ions are compiled in Table 1.

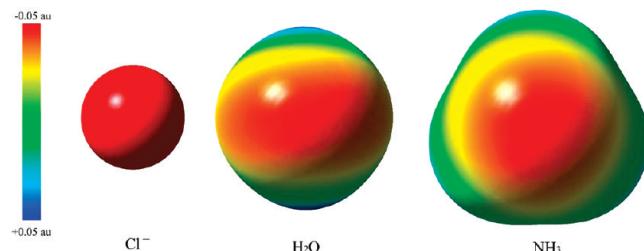
**2.4. Continuum Electrostatic Analysis.** The electrostatic contribution of halogen-ionic bridges to protein stability was ascertained via continuum electrostatic analysis by solving Poisson–Boltzmann (PB) equation, which was implemented in the DELPHI program<sup>60</sup> with probe radii 1.4 Å, temperature 298.15 K, ionic strength 0.145 M, and dielectric constants 4 for protein and 80 for solvent. A grid spacing of 0.833 Å per grid, in which the longest linear dimension of

the protein occupied 60% of the lattice, was used to determine the size of the cubic lattice, and the Debye–Hückel (full Coulombic) boundary conditions were applied. The PARSE set<sup>61</sup> of partial atomic charges and atomic radii was used for protein atoms, and the formal charge -1 and Shannon effective ionic radii (F<sup>-</sup> 1.33, Cl<sup>-</sup> 1.81, Br<sup>-</sup> 1.96, and I<sup>-</sup> 2.20 Å)<sup>57</sup> were assigned for halogen ions.

The total electrostatic contribution ( $\Delta\Delta G_{\text{tot}}$ ) to a halogen-ionic bridge's stability was decomposed into three terms: (i) Bridging energy ( $\Delta\Delta G_{\text{brd}}$ ), which arises from the Coulombic interactions between the halogen-ionic bridge's members (including central halogen ion and its interacting groups) in the folded state of the protein. The  $\Delta\Delta G_{\text{brd}}$  can be further divided into two parts:  $\Delta\Delta G_{\text{brd}}^{\text{grp}}$ , the generally repulsive interaction energy between protein groups in the bridge, and  $\Delta\Delta G_{\text{brd}}^{\text{hal}}$ , the always attractive interaction energy between the protein groups and halogen ion. (ii) Desolvation cost ( $\Delta\Delta G_{\text{dslv}}$ ), which represents the desolvation penalties incurred by the halogen ion and its interacting partners transferring from a high-dielectric water solvent in the unfolded state to the low-dielectric protein interior in the folded state of the protein.  $\Delta\Delta G_{\text{dslv}}$  can also be regarded as the sum of two aspects:  $\Delta\Delta G_{\text{dslv}}^{\text{hal}}$  and  $\Delta\Delta G_{\text{dslv}}^{\text{grp}}$ , the desolvation energies of halogen ions and protein groups, respectively. (iii) Additional effect ( $\Delta\Delta G_{\text{add}}$ ), which accounts for all the Coulombic interactions of the studied halogen-ionic bridge with the charges in rest of the protein in the folded state of the protein. Similarly,  $\Delta\Delta G_{\text{add}}$  is broken down into  $\Delta\Delta G_{\text{add}}^{\text{hal}}$  and  $\Delta\Delta G_{\text{add}}^{\text{grp}}$ . These three terms can be readily computed using a strategy proposed by Hendsch and Tidor,<sup>62</sup> who, and later Kumar et al.,<sup>63</sup> had successfully applied this method to investigate protein salt bridges. Briefly, a thermodynamic cycle was performed to trace the changes in Coulombic and reaction field energies of halogen-ionic bridge's members upon the bridge formation during protein folding. In this procedure, electrostatic contribution to free energy change was calculated relative to a mutation of its members to their hydrophobic isosteres. The hydrophobic isosteres were identical with those in the halogen-ionic bridge, except that their partial atomic charges were set to 0. A detailed description of this procedure can be found in refs 62 and 63. The protein moieties, which were considered in the continuum electrostatic calculation as well as their PARSE parameters,<sup>61</sup> employed in this calculation are provided in Supporting Information, Figure S1.

### 3. Results and Discussion

**3.1. Small Model System. I. Electrostatic Potentials.** Electrostatic potentials (ESPs),<sup>64</sup> the most intuitive physical quantity characterizing an electronic distribution state around a molecule or ion, are mapped on the same electronic isodensity surfaces of Cl<sup>-</sup>, H<sub>2</sub>O, and NH<sub>3</sub>, as shown in Figure 3. The latter two have been widely used as the model of lone-pair donors to study geometrical and energetic features of canonical hydrogen bonding. As might be anticipated, the Cl<sup>-</sup> performs the typical behavior of hard base with a small radii and a high charge density, which should act as a strong Brønsted base to accept protons donated from the hydrogen-

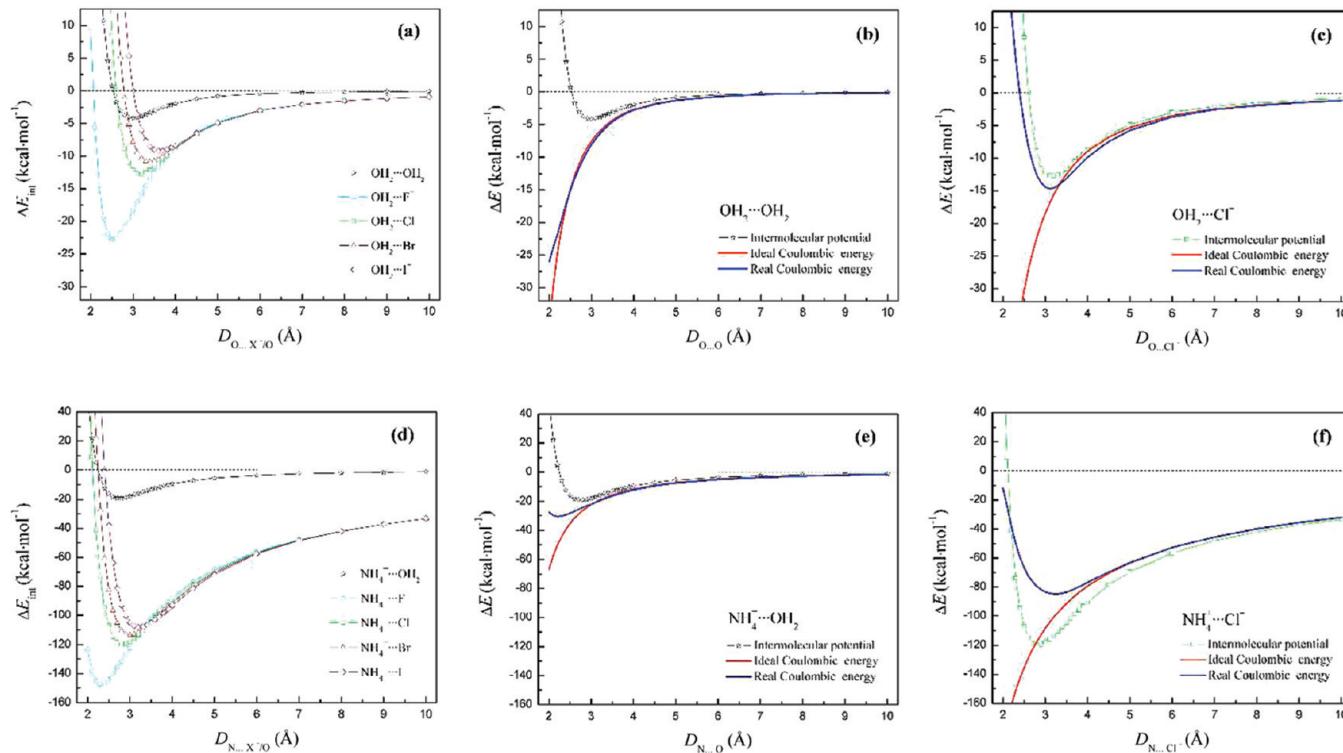


**Figure 3.** Electrostatic potential distribution, in Hartrees, at the 0.0004 electrons Bohr-3 isodensity surfaces of  $\text{Cl}^-$ ,  $\text{H}_2\text{O}$ , and  $\text{NH}_3$ .

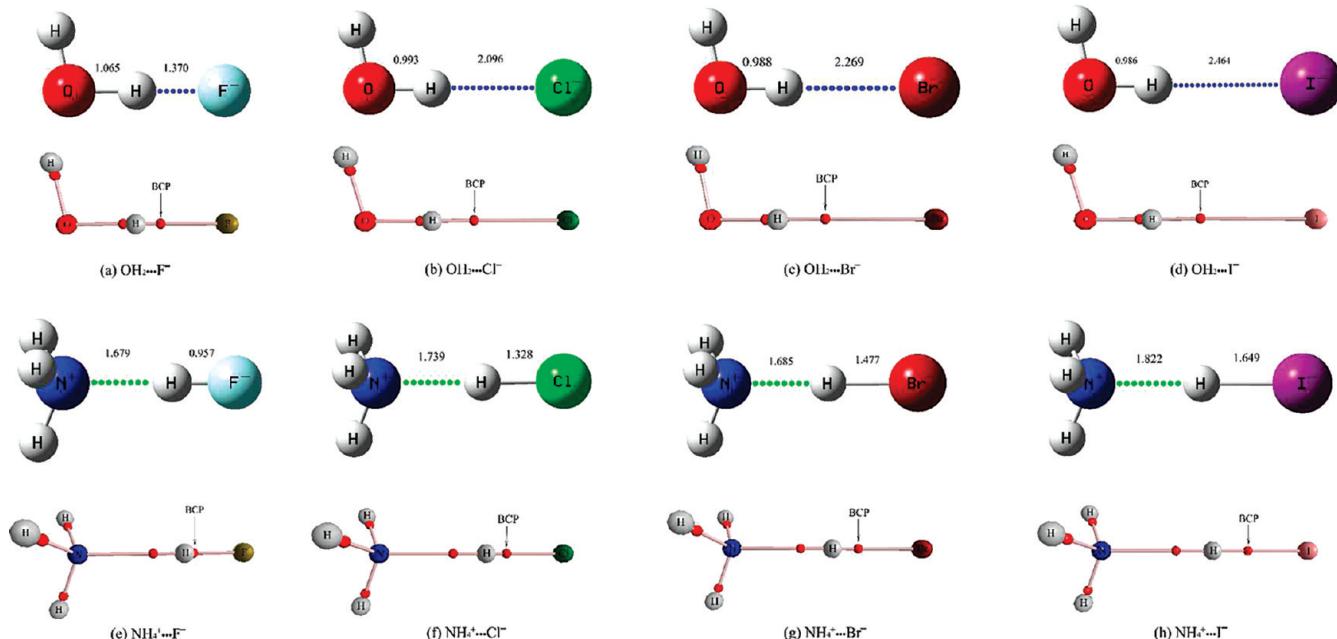
bond donors, such as serine hydroxyl and lysine ammonium. By contrast, these conventional Lewis bases,  $\text{H}_2\text{O}$  and  $\text{NH}_3$ , are apparently more ‘soft’ than the  $\text{Cl}^-$ , given by the less negative electrostatic potential at their lone-pair sites. In this respect, the  $\text{Cl}^-$ , and also the other three halogen ions, expectantly can form strong (ionic) hydrogen bonds, as compared to covalently bonded O and N atoms and can even pair to positively charged biomolecular moieties through ionic bonds.

**3.2. Small Model System. II. Intermolecular Interactions.** To quantitatively characterize the interaction profile of different halogen ions with polar and charged hydrogen atoms, Figure 4 shows the intermolecular potentials and the Coulombic energies for the complex model systems of  $\text{H}_2\text{O}$  and  $\text{NH}_4^+$  with  $\text{F}^-$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{I}^-$ , as a function of the

intermolecular  $\text{O}/\text{N}\cdots\text{X}^-$  distances. Also plotted are, for comparison purposes, the potential and Coulombic curves of  $\text{H}_2\text{O}$  and  $\text{NH}_4^+$  interacting with the oxygen atom of  $\text{H}_2\text{O}$ , the neutral counterpart of halogen ions. At a first glance, the interactions of halogen ions with  $\text{H}_2\text{O}$  and  $\text{NH}_4^+$  are, as that inferred from ESPs, markedly stronger than that of  $\text{H}_2\text{O}$  lone-pairs with the same hydrogen donors. For example, the well depth of the  $\text{OH}_2\cdots\text{I}^-$  potential curve, the most weak complex in the  $\text{OH}_2\cdots\text{X}^-$  series, was predicted to be 9.20 kcal $\cdot\text{mol}^{-1}$ , which is more than two-fold of the optimal dissociation energy (4.19 kcal $\cdot\text{mol}^{-1}$ ) of the water dimer (Figure 4a). This difference would increase to six-fold when the hydrogen donor  $\text{H}_2\text{O}$  is replaced by  $\text{NH}_4^+$  ( $\sim$ 120 kcal $\cdot\text{mol}^{-1}$  for  $\text{NH}_4^+\cdots\text{X}^-$  vs  $\sim$ 20 kcal $\cdot\text{mol}^{-1}$  for  $\text{NH}_4^+\cdots\text{OH}_2$ ) (Figure 4d). Comparison of intermolecular potentials to ideal and real Coulombic energies for both  $\text{OH}_2\cdots\text{OH}_2$  (Figure 4b) and  $\text{OH}_2\cdots\text{Cl}^-$  (Figure 4c) adducts preliminarily sheds light on the physical nature of these interactions;  $\text{OH}_2\cdots\text{Cl}^-$  bonding is dominated by an electrostatic force, given by the good agreement of its potential curve with the corresponding real Coulombic curve. The charge transfers (CTs) seem to be significant in the  $\text{OH}_2\cdots\text{Cl}^-$  system, which is implied by the large deviation of the real Coulombic curve from the corresponding ideal one (the former was calculated based on the real charge distribution in the complex system, whereas the latter based on the atomic charges of isolated complex members). In



**Figure 4.** (a) Intermolecular potentials for the complexes of  $\text{H}_2\text{O}$  with  $\text{F}^-$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$ ,  $\text{I}^-$ , and  $\text{H}_2\text{O}$ . (b) Comparison of intermolecular potential to ideal and real Coulombic energies for complex  $\text{OH}_2\cdots\text{OH}_2$ . (c) Comparison of intermolecular potential to ideal and real Coulombic energies for complex  $\text{OH}_2\cdots\text{Cl}^-$ . (d) Intermolecular potentials for the complexes of  $\text{NH}_4^+$  with  $\text{F}^-$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$ ,  $\text{I}^-$ , and  $\text{H}_2\text{O}$ . (e) Comparison of intermolecular potential to ideal and real Coulombic energies for complex  $\text{NH}_4^+\cdots\text{OH}_2$ . (f) Comparison of intermolecular potential to ideal and real Coulombic energies for complex  $\text{NH}_4^+\cdots\text{Cl}^-$ . All of energetic data plotted here were determined at the MP2/aug-cc-pVDZ (or MP2/Lanl2DZ+(df) for iodine) level of theory. Limited by space, comparisons of intermolecular potentials to Coulombic energies for the complexes of  $\text{H}_2\text{O}$  and  $\text{NH}_4^+$  with  $\text{F}^-$ ,  $\text{Br}^-$ , and  $\text{I}^-$  are provided in the Supporting Information, Figures S2 and S3.



**Figure 5.** Equilibrium structures and corresponding molecular graphs for complex model systems  $\text{OH}_2 \cdots \text{F}^-$  (a),  $\text{OH}_2 \cdots \text{Cl}^-$  (b),  $\text{OH}_2 \cdots \text{Br}^-$  (c),  $\text{OH}_2 \cdots \text{I}^-$  (d),  $\text{NH}_4^+ \cdots \text{F}^-$  (e),  $\text{NH}_4^+ \cdots \text{Cl}^-$  (f),  $\text{NH}_4^+ \cdots \text{Br}^-$  (g), and  $\text{NH}_4^+ \cdots \text{I}^-$  (h). In the equilibrium structures, the shorter one of the bonding between O/N and H and between X and H is plotted as a solid line, while the longer one is plotted as a dotted line. Bond lengths are shown in Å. Optimizations were performed at the MP2/aug-cc-pVTZ (or MP2/Lanl2DZ+(df) for iodine) level with the angle  $\theta_{\angle(\text{O}/\text{N}-\text{H}\cdots\text{X}^-)}$  frozen in  $180^\circ$  to avoid secondary interactions between halogen ions and other hydrogen atoms in these systems. Molecular graphs were generated by AIM analysis of corresponding MP2-optimized geometries.

contrast, the real Coulombic curve of the  $\text{OH}_2 \cdots \text{OH}_2$  dimer is coincidentally well with the ideal, indicating only a slight CT accompanied with the formation of this complex. These phenomena basically recurred when we investigated the complex systems of  $\text{NH}_4^+ \cdots \text{OH}_2$  (Figure 4e) and  $\text{NH}_4^+ \cdots \text{Cl}^-$  (Figure 4f), albeit not completely recurred. Most unexpectedly, the real Coulombic curve of the doubly charged  $\text{NH}_4^+ \cdots \text{Cl}^-$ , unlike that of  $\text{OH}_2 \cdots \text{Cl}^-$ , is almost on the above of and deviated considerably from the intermolecular potential curve, this phenomenon is quite obvious in the region nearby curve minimum. It means that, besides the electrostatic effect, there also exists other chemical force(s) to contribute the  $\text{NH}_4^+ \cdots \text{Cl}^-$  attraction, but what is (are) it? As we know, the  $\text{NH}_4^+ \cdots \text{Cl}^-$  complex can also be viewed as formed by  $\text{NH}_3$  with  $\text{HCl}$ , if considering that the proton bound between  $\text{N}^+$  and  $\text{Cl}^-$  is always mobile. In this regard, the  $\text{H} \cdots \text{Cl}^-$  interaction (maybe written as  $\text{H}^{\delta+} - \text{Cl}^{\delta-}$  bond is more suitable) is, to some extent, imparted with a covalent component and thereby tagged with an additional stabilization energy to enhance its bonding strength. This hypothesis will be proved by topological analysis of electron density and chemical bonding in these model systems (vide post).

**3.3. Small Model System. III. Geometrical, Energetic, And Electronic Analyses.** The complex model systems of  $\text{H}_2\text{O}$  and  $\text{NH}_4^+$  with  $\text{F}^-$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{I}^-$  were further fully optimized at the MP2/aug-cc-pVTZ (or MP2/Lanl2DZ+(df) for iodine) level with the angle  $\theta_{\angle(\text{O}/\text{N}-\text{H}\cdots\text{X}^-)}$  frozen in  $180^\circ$  to avoid secondary interactions between halogen ions and other hydrogen atoms in these systems. The equilibrium structures and corresponding molecular graphs, which were generated by AIM analysis of the MP2 optimized geometries,

are shown in Figure 5 and parametrized as that listed in Table 2. It is evident that the halogen ions in all complexes have a pronounced interaction with the hydrogen donors, rationalized by the presence of bond paths linking the nuclei of  $\text{X}^-$  and H. In addition, the interatomic separations between halogen and hydrogen are always longer than corresponding O–H bonds in  $\text{OH}_2 \cdots \text{X}^-$ , but this is converse in the  $\text{NH}_4^+ \cdots \text{X}^-$  series (due to the proton transfers). Specifically, the  $\text{H} \cdots \text{F}$  distance in equilibrium  $\text{NH}_4^+ \cdots \text{F}^-$  structure is only 0.957 Å, which follows the typical feature of open-shell (shared) interactions. Hence, rather than the nonbonded intermolecular force, it would better be recognized as a polar covalent or ionic bond. Qualitative graphic conclusions could be substantiated by quantitative examination of the geometrical, energetic, and electronic parameters associated with these interactions (Table 2). The first and most straightforward evidence is the particularly high interaction energies (up to 175 kcal·mol<sup>-1</sup>) attached to the  $\text{NH}_4^+ \cdots \text{X}^-$  associations, these values fall within the normal range of the bond energies of covalent and ionic bonds. By contrast, the intermolecular potentials of  $\text{OH}_2 \cdots \text{X}^-$  are only a tenth of that found in corresponding  $\text{NH}_4^+$ -involved adducts, satisfying the definition of ionic hydrogen bonds by Meot-Ner.<sup>65</sup> Second, the unshared attribute for the  $\text{OH}_2 \cdots \text{X}^-$  and the shared character for the  $\text{NH}_4^+ \cdots \text{X}^-$  can be clearly characterized by the electronic topological parameters (including electron density  $\rho_b$ , Laplacian of the electron density  $\nabla^2 \rho_b$ , ellipticity  $\varepsilon_b$ , and electronic energy density  $H_b$ ) at the bond critical points (BCPs) of  $\text{H} \cdots \text{X}^-$  bond paths (as marked in Figure 5). For example,  $\rho_b$  and  $\nabla^2 \rho_b$  of  $\text{OH}_2 \cdots \text{X}^-$  were predicted to be in the range of 0.006–0.096 and 0.037–0.121 au, respectively, which are basically consistent with or

**Table 2.** Geometrical, Energetic, And Electronic Parameters for the Complexes of Halogen Ions with H<sub>2</sub>O and NH<sub>4</sub><sup>+</sup> Serving As Hydrogen Donor<sup>a</sup>

Complex	$d_{\text{H} \cdots \text{X}}^{-b}$	$D_{\text{Y} \cdots \text{X}}^{-c}$	$\Delta E_{\text{int}}^{d\prime}$	$\rho_0^e$	AIM analysis $\nabla^2 \rho_b^f$	$H_b^g$	Wiberg <sup>h</sup>	NLMO/NPA <sup>i</sup>	$\Delta q_{\text{X}}^{-\kappa}$	NBO analysis $\Delta E_{\text{coul}}^{\text{real}} /$ $\Delta E_{\text{SE}}^m$	$\Delta E^2 n$
OH <sub>2</sub> ...F <sup>-</sup>	1.370	2.435	-26.61 (-23.32) <sup>q</sup>	0.0958	0.1207	0.0002	-0.0514	0.1699	0.0883	0.1095	-32.02
OH <sub>2</sub> ...Cl <sup>-</sup>	2.096	3.089	-14.11 (-14.71) <sup>r</sup>	0.0328	0.0585	0.0014	-0.0062	0.0701	0.0356	0.0458	-16.42
OH <sub>2</sub> ...Br <sup>-</sup>	2.269	3.257	-11.99 (-11.71) <sup>r</sup>	0.0277	0.0480	0.0012	-0.0041	0.0615	0.0314	0.0404	-14.22
OH <sub>2</sub> ...I <sup>-</sup>	2.464	3.449	-9.80 (-10.30) <sup>r</sup>	0.0056	0.0371	0.0008	-0.0036	0.0603	0.0304	0.0398	-11.76
NH <sub>4</sub> <sup>+</sup> ...F <sup>-</sup>	0.957	2.636	-175.56	0.3186	-2.5734	0.0002	-0.7423	0.5716	0.3489	0.3600	-119.17
NH <sub>4</sub> <sup>+</sup> ...Cl <sup>-</sup>	1.328	3.067	-131.98	0.2219	-0.7107	0.0001	-0.2302	0.8106	0.6107	0.6135	-26.96
NH <sub>4</sub> <sup>+</sup> ...Br <sup>-</sup>	1.477	3.162	-120.63	0.1808	-0.4494	0.0000	-0.1584	0.8245	0.6512	0.6552	-17.47
NH <sub>4</sub> <sup>+</sup> ...I <sup>-</sup>	1.649	3.471	-113.17	0.1466	-0.1506	0.0000	-0.1113	0.9129	0.8151	0.8198	-2.90

<sup>a</sup> The equilibrium structures as well as AIM and NBO parameters of all complexes were obtained at the MP2/aug-cc-pVTZ (or MP2/Lanl2DZ+(df) for iodine) for iodine for structure optimization and NBO analysis or at MP2/DGDDZVP for iodine for AIM analysis level of theory. Interaction energies  $\Delta E_{\text{int}}$  between halogen ions and their interacting partners were calculated at a higher level of CCSD(T)/aug-cc-pVTZ (or CCSD(T)/Lanl2DZ+(df) for iodine) using the MP2 optimized geometries. <sup>b</sup>  $d_{\text{H} \cdots \text{X}}^{-c}$  (Å), interatomic distance between X<sup>-</sup> and H (for OH<sub>2</sub>...X<sup>-</sup> and NH<sub>4</sub><sup>+</sup>...X<sup>-</sup>), where X<sup>-</sup> is halogen ion. <sup>c</sup>  $D_{\text{Y} \cdots \text{X}}^{-c}$  (Å), interatomic distance between Y and X<sup>-</sup>, where X<sup>-</sup> is halogen ion and Y is O or N in H<sub>2</sub>O or NH<sub>4</sub><sup>+</sup>, respectively. <sup>d</sup>  $\Delta E_{\text{int}}$  (kcal·mol<sup>-1</sup>), calculated and, if exist, experimental (in parentheses) interaction energies. <sup>e</sup>  $\rho_0$  (au), electron density at BCPs. <sup>f</sup>  $\nabla^2 \rho_b$ (au), Laplacian of the electron density at BCPs. <sup>g</sup>  $\varepsilon_b$  (au), ellipticity at BCPs. <sup>h</sup>  $H_b$  (au), Wiberg bond index associated with the interactions. <sup>i</sup> NLMO/NPA, atom–atom net linear NLMO/NPA bond order associated with the interactions. <sup>j</sup>  $\Delta q_{\text{X}}^{-\kappa}$ , changes in charge number of halogen ion due to the interactions. <sup>k</sup>  $\Delta E_{\text{coul}}^{\text{real}}$  (kcal·mol<sup>-1</sup>), real Coulombic energy between donor and acceptor, obtained via Coulomb's analysis. <sup>m</sup>  $\Delta E_{\text{SE}}$  (kcal·mol<sup>-1</sup>), steric exchange energy between halogen ion and its interacting partners. <sup>n</sup>  $\Delta E^2$  (kcal·mol<sup>-1</sup>), second-order perturbation stabilization energy between the lone pair of halogen ion and the antibonding orbital  $\sigma$  of O–H or N–H bond in H<sub>2</sub>O or NH<sub>4</sub><sup>+</sup>, respectively. <sup>o</sup> Not applicable. <sup>p</sup> From gas-phase equilibrium measurements by high-pressure mass spectrometry.<sup>8</sup> <sup>q</sup> From gas-phase equilibrium measurements by high-pressure mass spectrometry.<sup>8</sup> <sup>r</sup> From gas-phase equilibrium measurements by high-pressure mass spectrometry.

slightly larger than that proposed for (nonionic) hydrogen bonds (i.e., 0.002–0.035 and 0.024–0.139 au, respectively),<sup>66</sup> but these quantities are noticeable in the  $\text{NH}_4^+ \cdots \text{X}^-$  repertoire. In addition, the negative values of  $H_b$  for the both  $\text{OH}_2 \cdots \text{X}^-$  and  $\text{NH}_4^+ \cdots \text{X}^-$  suggest that these two kinds of interactions are stronger than conventional hydrogen bonding, which usually has a positive  $H_b$ .<sup>66</sup> The last but most importantly, NBO population analysis of the natural atomic and bond orbits in these complexes depicted a clearer profile for the chemical nature of interactions involving halogen ions. As listed in Table 2, the Wiberg<sup>67</sup> and natural localized molecular orbital/natural population analysis (NLMO/NPA)<sup>68</sup> bond indices of the  $\text{NH}_4^+ \cdots \text{X}^-$  interactions are quite significant with respect to  $\text{Cl}^-$ ,  $\text{Br}^-$ , and, particularly,  $\text{I}^-$  (close or greater than 0.6) but relatively lower in  $\text{F}^-$ -containing system. This is inconsistent with that reflected in interaction energies  $\Delta E_{\text{int}}$ , which enhances in the order of  $\text{NH}_4^+ \cdots \text{I}^- < \text{NH}_4^+ \cdots \text{Br}^- < \text{NH}_4^+ \cdots \text{Cl}^- < \text{NH}_4^+ \cdots \text{F}^-$ . It must be recalled here that most of bond orders (BOs), such as the Wiberg and NLMO/NPA discussed here, generally underestimate for polar covalent bonds, since the ionic bond component in these polar covalent bonds is almost ignored by BOs. Compared to those found in  $\text{NH}_4^+ \cdots \text{Br}^- / \text{Cl}^- / \text{I}^-$  complexes, the relatively small degree of charge transfers ( $\Delta q_F^- = 0.36$ ) and the dominant Coulombic effect ( $\Delta E_{\text{coul}}^{\text{real}} = -119.17 \text{ kcal} \cdot \text{mol}^{-1}$ ) unravel a marked ionic bond character associated with the  $\text{NH}_4^+ \cdots \text{F}^-$  interaction, because these electrostatic properties are always related to the ionic bonding.<sup>69</sup> Furthermore, it must be reminded here that complicated biological context would undermine the “idea fashion” (as that in small model systems) of halogen ions approaching charged hydrogen atoms, giving rise to a more important role of the nondirectional ionic bonding than the directional covalent bonding in the interactions involving not only  $\text{F}^-$  but also other three halogen ions.

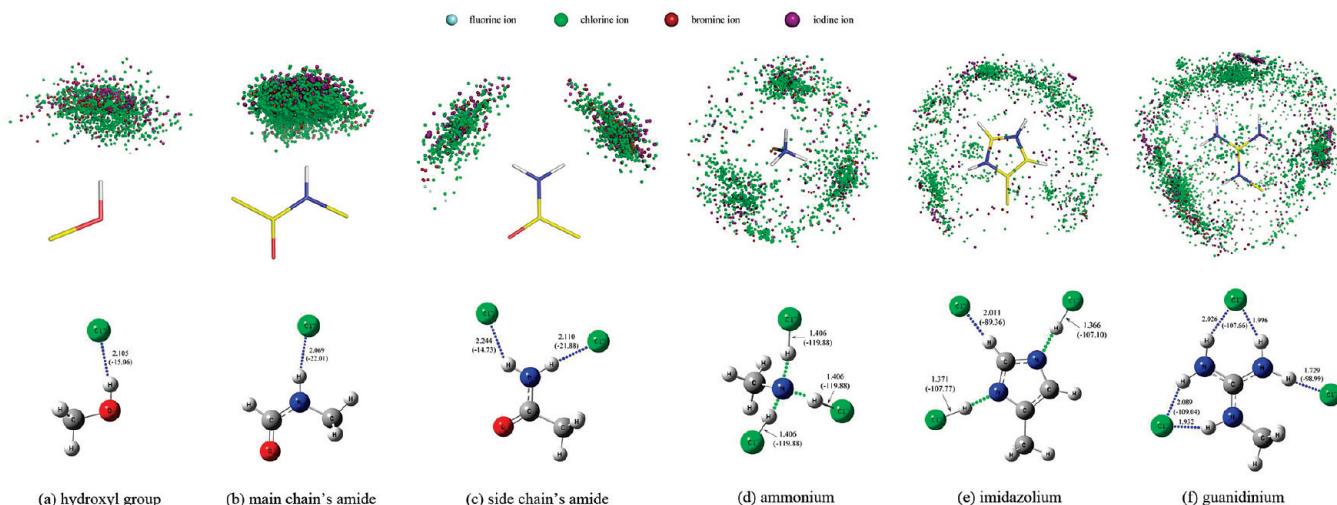
Overall, halogen ions in physiological environment are presumed to adopt three types of interactions to bridge between biomolecular moieties: (i) ionic hydrogen bonding with polar hydrogen atoms, such as those in amide and hydroxyl group; (ii) ionic bonding with positively charged species, such as ammonium, imidazolium, and guanidinium; and (iii) covalent bonding with the hydrogen atoms of proton donors (this interaction type must have a “good” geometrical arrangement as compared to ionic bonding). It should be noted here that this division is not absolute and most of the real cases must be compatible simultaneously with two or even three of these interaction types.

**3.4. Real Biomolecular System. I. PDB Survey of Halogen-Ionic Bonding.** The PDB (January, 2010 release) contains 3391 and 133 entries of X-ray crystal structures (at resolutions of 3.0 Å or better) of proteins and nucleic acids showing 11 852 and 1345 nonbonded halogen ions, respectively (total 37  $\text{F}^-$ , 9966  $\text{Cl}^-$ , 1065  $\text{Br}^-$ , and 2129  $\text{I}^-$ ). The pronouncedly unbalanced numbers of different halogen ions found in biomolecules mirror the variance in chemical activity and the natural abundance of these halogens. Fluorine is the most active element in halogens and usually exists in combined state. Hence, only a few of free fluorine ions were observed in the survey. However, the also chemically active

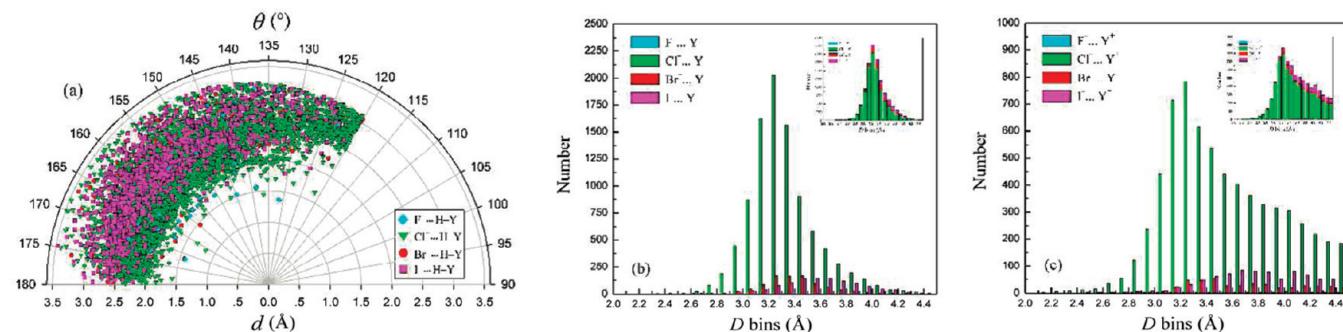
chlorine ions were found to be quite abundant in biological systems. This could be attributed to the important physiological function of the chlorine ions in keeping, for example, electrical neutrality, acid–base balance, and correct pressure of cell and body. Using the criteria described in Section 2.2, we selected the halogen ions in effective interaction (for convenience, termed as halogen-ionic bonding) with the polar hydrogen atoms and positively charged groups of biomolecules to define a reliable set of solid biological contacts involving halogen ions, which consists of 20 826, 793, and 174 halogen-ionic bonding with proteins, nucleic acids, and small ligands, respectively. Considering the prominent magnitude of halogen-ionic bonding with protein moieties, we herein give a detailed inspection on the geometrical characteristics and distribution of this kind of interactions.

Relative halophilicity of different protein moieties was assessed using their contact rates (CRs) with halogen ions, which were simply defined as the ratio of those in effective interaction with halogen ions to all presented in our data set. As might be expected, three charged moieties, i.e. ammonium, imidazolium, and guanidinium, have the highest propensity to pair with halogen ions, with their CRs of 2.34, 4.32, and 5.47%, respectively. In the remaining polar moieties, the side chain’s amide performs as well in the halophilicity (CR = 1.62%), whereas the main chain’s amide and hydroxyl group exhibit a relatively halophobic feature (CRs < 1%). The difference in halophilicity between polar and charged protein moieties well echoes the bonding strength variation among small model systems and also reflects the electrostatic nature of biological halogen-ionic bonding. These conclusions could be further visualized by superposing halogen ions around their common protein moieties and by comparing the distribution states of these superposed halogen ions to corresponding MP2-determined low-lying energy structures, as shown in Figure 6. It can be seen that the main chain’s amide and hydroxyl group hold only one hydrogen site to accommodate halogen ions, hence, their halophilicity is lower than the congeneric side chain’s amide, which provides two hydrogen sites for halogen ions. The polyatomic imidazolium and guanidinium have a large surface to contact with surrounding halogen ions and thus show the highest halophilicity, while the smaller ammonium can only possess a moderately strong halophilicity. Besides, the distribution preferences of halogen ions around different protein moieties are compatible with the  $\text{Cl}^-$  locations in corresponding low-lying energy structures. For example, the arrangement patterns of  $\text{Cl}^-$  in interaction with three charged moieties in low-lying energy structures are clearly mirrored as the presence of local halogen ion-dense regions in the statistical distribution plots, although the contacting angle of halogen ions with these charged moieties has been entirely ignored when we performed the PDB survey.

Despite the significant unbalancedness of four kinds of halogen ions occurring in a biological environment, the contacting behavior of different halogen ions with protein moieties could also be analyzed in terms of the frequency distributions of geometrical parameters of halogen-ionic bonds (HIBs) retrieved from the PDB. It is shown that the geometrical profile of polar HIBs (bonding between halogen



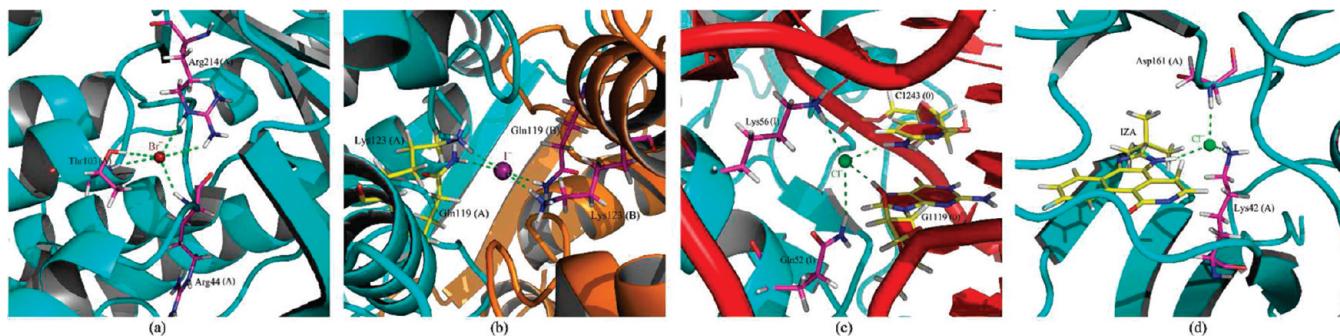
**Figure 6.** The first row showing the distribution states of halogen ions around different protein moieties retrieved from the PDB. The second row showing the low-lying energy structures of  $\text{Cl}^-$  in complex with corresponding protein moieties obtained by a thorough MP2/aug-cc-pVDZ search (to render this figure more readable, multiple low-lying energy sites of  $\text{Cl}^-$  in complex with the same moiety are artificially resettled in a subplot).



**Figure 7.** (a) Polar scatter plot of  $\theta$  vs  $d$  for halogen-ionic bonding with polar hydrogen atoms of proteins retrieved from the PDB. Histogram distributions of  $D$  (in 0.1 bins) for halogen-ionic bonding with: (b) polar and (c) charged moieties of proteins retrieved from the PDB.  $\theta$  is the angle  $\angle(\text{O}-\text{H}\cdots\text{X}^-)$  for hydroxyl groups or the angle  $\angle(\text{N}-\text{H}\cdots\text{X}^-)$  for amides;  $d$  is the interatomic distance between X and polar H in polar moieties;  $D$  is the interatomic distance between X and heavy atom Y, where Y is the antecedent of the interacting polar H in polar moieties or the nearest heavy atom in charged moieties.

ions and polar protein groups) is in agreement with those received from statistical examination of water–halide ion interactions in the CSD<sup>70</sup> and routine hydrogen bonds in protein crystals,<sup>71</sup> albeit the  $d$  values of HIBs seem to be slightly larger than that of hydrogen bonds (Figure 7a). This is not unexpected if considering that these polar HIBs are natural of ionic hydrogen bonds, and the longer interatomic separations between H and  $\text{X}^-$  in polar HIBs than those between H and O/N in routine hydrogen bonds are apparently owing to the larger radii of halogen ions (except fluorine ion) relative to oxygen and nitrogen atoms. Moreover, from the frequency distributions of bond lengths  $D$  derived from polar and charged HIBs (Figure 7bc), it can be readily appreciated that the  $D$  values also increase with the size of halogen ions, i.e.  $\text{F}^-\cdots\text{Y}/\text{Y}^+ < \text{Cl}^-\cdots\text{Y}/\text{Y}^+ < \text{Br}^-\cdots\text{Y}/\text{Y}^+ < \text{I}^-\cdots\text{Y}/\text{Y}^+$ . Interestingly, the peak positions of  $D$  distributions for three polar HIBs ( $\text{Cl}^-\cdots\text{Y}$ ,  $\text{Br}^-\cdots\text{Y}$ , and  $\text{I}^-\cdots\text{Y}$ ) and for three charged HIBs ( $\text{Cl}^-\cdots\text{Y}^+$ ,  $\text{Br}^-\cdots\text{Y}^+$ , and  $\text{I}^-\cdots\text{Y}^+$ ) are completely consistent, as both series located at the 3.25, 3.45, 3.65 Å bins, respectively ( $\text{F}^-\cdots\text{Y}$  and  $\text{F}^-\cdots\text{Y}^+$  were not considered here because their numbers found in the PDB are too small to generate

statistically significant conclusions), and these peak locations are uniformly accompanied with a red-shift of about 0.2 Å relative to equilibrium distances in corresponding small model complexes (see Table 2). The elongating of HIBs in biomolecules could be ascribed to steric hindrance and constraint in complicated biological context. Furthermore, although polar and charged HIBs have a coherency in their  $D$  peak locations, the whole profile of their  $D$  distributions is solidly distinct, particularly in the regions to the right of the peak positions (Figure 7b and c). Compared to polar HIBs, charged are more long-range and hold a considerable number with bond lengths  $D > 4.0$  Å. This can be reflected in the intermolecular potential curves derived from small model systems (Figure 4a and d). At the 4.5 Å separation, for example, intermolecular potentials for  $\text{OH}_2\cdots\text{X}^-$  models are only about  $-7 \text{ kcal}\cdot\text{mol}^{-1}$ , while for  $\text{NH}_4^+\cdots\text{X}^-$ , models are of striking values as more than  $-80 \text{ kcal}\cdot\text{mol}^{-1}$ . The fundamental difference between polar and charged HIBs in long-range interaction behavior reveals and substantiates their natures of ionic hydrogen bonding and ionic bonding, respectively.



**Figure 8.** Some examples of halogen-ionic bridges in biomolecules. (a) Bromine-ionic bridge in protein interior (PDB: 1doc). (b) Iodine-ionic bridge at protein–protein interface (PDB: 2vgz). (c) Chlorine-ionic bridge at protein–rRNA interface (PDB: 1jj2). (d) Chlorine-ionic bridge at protein–small ligand interface (PDB: 2j90).

**3.5. Real Biomolecular System. II. PDB Survey of Halogen-Ionic Bridges.** On the basis of the HIBs retrieved from the PDB, we have defined a distinct data set consisting of 6406 biological halogen-ionic bridges; each of them is composed of a halogen ion and several biomolecular moieties that are directly bound to the halogen ion through HIBs. This substantial magnitude of halogen-ionic bridges found here is remarkably more than the numbers of halogen bonds (113)<sup>72</sup> and halogen–water–hydrogen bridges (106)<sup>43</sup> surveyed in previous works, implying an non-negligible role of the halogen-ionic bridges in biological systems. We further classified these halogen-ionic bridges into four groups according to their locations, i.e., 4910 in protein interiors, 1132 at protein–protein interfaces, 273 at protein–nucleic acid interfaces, and 91 at protein–small ligand interfaces. A classification list of these halogen-ionic bridges is tabulated in Supporting Information, Tables S1–S4.

**Halogen-Ionic Bridges in Protein Interiors (Figure 8a).** A considerable number of halogen-ionic bridges stand in protein interiors and are thought to be functional in figuring the advanced structure of protein architectures. Formation of halogen-ionic bridges in the low dielectric environment due to protein folding causes a large desolvation penalty, which can be, more or less, compensated by the favorable electrostatic interactions both between the halogen ions and their oppositely charged partners and between the halogen-ionic bridges and their protein surroundings. In the next section, we will provide computational evidence supporting that halogen-ionic bridges are generally stabilizing toward protein architectures, and this stabilization tendency is quite significant as compared to the diverse salt bridges described previously. In addition, most of the halogen-ionic bridges in proteins are formed between sequentially farther residues in comparison with salt bridges, which usually pair within the vicinal residue blocks.<sup>63</sup> The average number of amino acids separating the halogen-ionic bridging residues is 24.01, this is far beyond that to be considered in the hierarchical model of protein folding.<sup>73</sup> Hence, halogen-ionic bridges in proteins are most likely to be formed later than the “molten globule” phase of the folding.

**Halogen-Ionic Bridges at Protein–Protein Interfaces (Figure 8b).** Owing to the chemically natural similarity between protein–protein interface and protein interior,<sup>74</sup> halogen-ionic bridges at the interfaces are supposed to confer stability and specificity for protein binding as much as that for protein

folding. In fact, halogen-ionic bridges seem to be more effective in contributing to protein binding rather than to folding because the binding does not need too much of a degree of packing the halogen-ionic bridging groups from the already structured protein monomers to complex, thus leading to a lesser desolvation penalty.

**Halogen-Ionic Bridges at Protein–Nucleic Acid Interfaces (Figure 8c).** Almost all of the halogen-ionic bridges across protein–nucleic acid interfaces were found in huge, complicated ribosomes. Since ribosome growth is an exhaustive process which involves a series of molecular operation steps, such as protein/rRNA splicing, folding, and packing,<sup>75</sup> the role played by halogen-ionic bridges in the ribosomes should be different to those in protein complexes, which are normally formed by direct, rigid protein–protein association. Visual inspection of these ribosomal halogen-ionic bridges found that they are usually located at the regions where protein and rRNA atoms are fully buried but not in sufficient contact with each other, the halogen ions occupy at the cavities embedded within these atoms and interact with vicinal polar groups. On this point, the halogen-ionic bridges at protein–nucleic acid (rRNA) interfaces could be regarded as structural fillers to refine the shape complementarity and to tune the local conformation of the interface structures.

**Halogen-Ionic Bridges at Protein–Small Ligand Interfaces (Figure 8d).** Although limited numbers of halogen-ionic bridges were observed at the protein–small ligand interfaces, we will demonstrate that they do play an important role in inhibitor recognition and binding, at least by HIV-1 protease and PDK1 kinase, using the QM/MM scheme (see Section 3.7). Statistical analysis unraveled that halogen-ionic bridges at protein–ligand interfaces share a similar solvent accessibility (measured by SASA) and fluctuation rate (measured by isotropic B-factor) with those ligand-bound water molecules,<sup>76</sup> but they usually stand in multifurcated form (measured by branch degree) and are packed tightly by surrounding protein and ligand atoms (measured by packing density). In general, the halogen-ionic bridge, if it exists, is shown to be essential in assisting the ligand positioning in the protein active pocket because only the accurate ligand location can result in the bridge with optimal geometry and thus highest stability (this will be rationalized by QM/MM procedure).

The mean statistics of structural and energetic parameters for the 6042 halogen-ionic bridges found in proteins (includ-

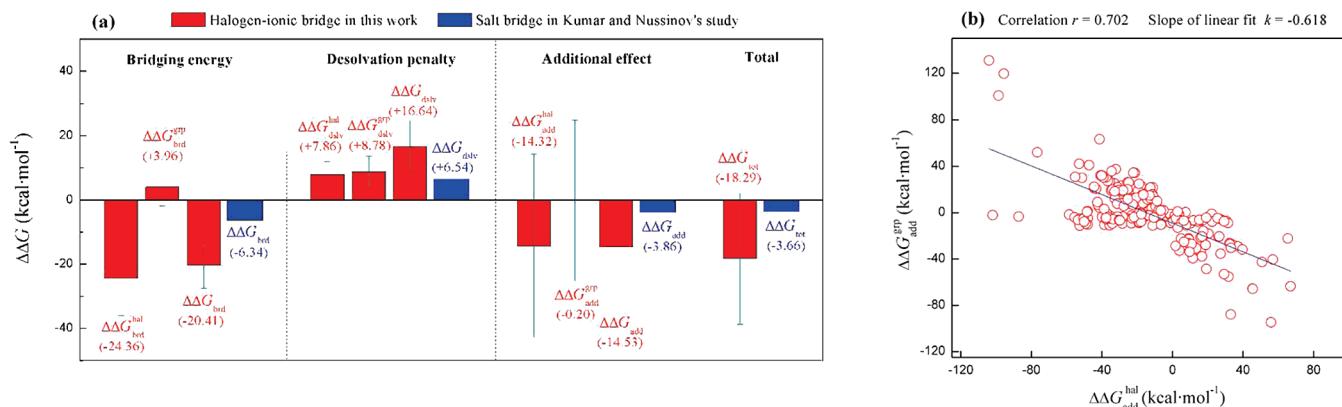
ing protein interiors and interfaces) are compiled in Table 3. Nearly 35% of halogen-ionic bridging amino acids are located at a protein helix region, and the remainders ( $\sim 65\%$ ) are approximately equivalently distributed in strand, turn, and loop. This assignment agrees to the abundance of these secondary structure classes observed in native proteins.<sup>77</sup> Comparison of the average *B*-factors between bridging and nonbridging halogen ions suggested that halogen ions are generally less mobile when they are bound in halogen-ion bridges than when they are out of the bridges, given by  $\sim 20\%$  difference in their average *B*-factor values. This phenomenon has also been observed for the water-mediated bridges in protein crystals.<sup>76</sup> In addition, the fluorine-ionic bridge has a large branch degree and packing density as compared to the other three; this could be reflected in the significant desolvation effect accompanied with the fluorine-ionic bridge formation. As can be seen in Table 3, the average percentage of reduction in the SASA when the  $F^-$  transfer from solvent to bridges is 87.7%; this, coupled with its prominent thermodynamic effect of hydration, gives rise to the  $F^-$  with a noticeable value in  $\Delta\Delta G_{\text{hydr}}^\circ$ ,  $\Delta\Delta H_{\text{hydr}}$ , and  $-T\Delta\Delta S_{\text{hydr}}$ . In contrast, the desolvation penalties of bridging  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{I}^-$  have only about two-thirds of that with the bridging  $F^-$ . In this regard, the desolvation profile of formation of different halogen-ionic bridges is compatible with the specific ion effects observed experimentally, i.e.,  $F^-$  is referred as kosmotrope, whereas  $\text{Cl}^-$ ,  $\text{Br}^-$ , and  $\text{I}^-$  are called chaotropes.<sup>3</sup>

**3.6. Real Biomolecular System. III. Continuum Electrostatic Analysis.** Based on conventional biochemical intuition, one would expect halogen-ionic bridges to be stabilizing toward the folded conformations of proteins. However, we should be cautious of this notion, recalling that some salt bridges have been demonstrated to be destabilizing for protein structures since their desolvation penalties, due to the burial of ionizable salt-bridging groups in the low dielectric protein interior during protein folding, are not fully recovered by favorable electrostatic interactions in the folded state.<sup>78–80</sup> So, we herein performed continuum electrostatic calculations on a panel of high-quality halogen-ionic bridges derived from monomeric protein crystal structures to answer questions like do the halogen-ionic bridges generally stabilize protein architectures and whether they confer more stability for proteins than traditional salt bridges? Since all the calculations are essentially based upon the atomic coordinates provided in protein PDB files, the accuracy of the results gained from the continuum electrostatic analysis is highly dependent on the quality of the protein structures in which the studied halogen-ionic bridges are contained. Therefore, we used a set of monomeric protein structures with high resolution ( $\leq 1.8 \text{ \AA}$ ) and low homology (sequence identity  $< 30\%$  between any two proteins) taken from the current (October 14, 2009 updated) PDB-REPRDB list of structures.<sup>81</sup> From this list, we culled the halogen-ionic bridges satisfying the criteria as defined in Section 2.2, and also which the positions of the central halogen ions possess a high precision (occupancy = 1 and *B*-factor  $< 30 \text{ \AA}^2$ ). Consequently, 241 high-quality halogen-ionic bridges distributed in 189 monomeric proteins were selected for

**Table 3.** Mean Statistics of Structural and Energetic Parameters for Halogen-Ionic Bridges in Proteins<sup>a</sup>

ion	num.	SSC <sup>b</sup>	BF <sub>brd</sub> <sup>c</sup>	BD <sub>brd</sub> <sup>d</sup>	PD <sub>brd</sub> <sup>e</sup>	SASA <sub>brd</sub> <sup>f</sup>	$\Delta\text{SASA}_{\text{brd}}^g\%$	$\Delta\Delta G_{\text{hydr}}^\circ h$	$\Delta\Delta H_{\text{hydr}}^i$	$\Delta\Delta S_{\text{hydr}}^j$
$F^-$	18	46.5% H, 13.5% S, 17.7% T, 22.3% L	34.47(13.52)	3.22(1.48)	0.74(0.04)	11.49(12.79)	87.7%	89.40(13.91)	97.82(15.22)	-8.42(1.31)
$\text{Cl}^-$	5074	37.1% H, 19.9% S, 21.5% T, 21.4% L	37.63(19.04)	2.56(0.78)	0.40(0.12)	20.80(20.18)	83.9%	62.03(11.52)	66.73(12.39)	-4.70(0.87)
$\text{Br}^-$	296	30.8% H, 32.0% S, 23.6% T, 13.7% L	32.87(19.93)	2.41(0.66)	0.41(0.11)	25.01(19.63)	82.4%	58.15(9.77)	62.69(10.53)	-4.53(0.76)
$\text{I}^-$	553	30.1% H, 19.8% S, 27.2% T, 22.9% L	46.30(23.52)	2.29(0.53)	0.40(0.10)	33.57(25.99)	79.4%	47.23(9.50)	49.46(9.94)	-2.22(0.45)
All	6042	34.1% H, 23.5% S, 21.3% T, 21.1% L	38.09(19.72)	2.53(0.76)	0.40(0.11)	22.22(21.06)	83.3%	60.50(12.14)	64.98(13.19)	-4.47(1.12)

<sup>a</sup> The numbers in parentheses are corresponding standard deviations. <sup>b</sup> SSC, secondary structure class assignment for the amino acids that are directly bound to the bridging halogen ions. <sup>c</sup> helix ( $\alpha$ ,  $\pi$ , and 3/10-helix); S, strand (isolated  $\beta$ -strand and multiple  $\beta$ -ladder); T, turn (3, 4, 5 turns and bend); L, loop and others (loop, coil, etc.). <sup>c</sup> BF<sub>brd</sub> ( $\text{\AA}^2$ ), mean *B*-factor (or Debye-Waller factor) of the bridging halogen ions. <sup>d</sup> BD<sub>brd</sub>, mean branch degree of the halogen-ionic bridges. <sup>e</sup> PD<sub>brd</sub>, mean pack density of the bridging halogen ions. <sup>f</sup> SASA<sub>brd</sub> ( $\text{\AA}^2$ ), mean SASA of the bridging halogen ions. <sup>g</sup>  $\Delta\text{SASA}_{\text{brd}}^g\%$ , average percentage of changes in the SASA when the halogen ions transfer from solvent (free state) to halogen-ionic bridges (bridging state). <sup>h</sup>  $\Delta\Delta G_{\text{hydr}}^\circ$  (kcal·mol<sup>-1</sup>), mean value of changes in hydration free energy when the halogen ions transfer from solvent to halogen-ionic bridges. <sup>i</sup>  $\Delta\Delta H_{\text{hydr}}^i$  (kcal·mol<sup>-1</sup>), mean value of changes in hydration entropy when the halogen ions transfer from solvent to halogen-ionic bridges. <sup>j</sup>  $-T\Delta\Delta S_{\text{hydr}}^j$  (kcal·mol<sup>-1</sup>), mean value of changes in hydration entropy when the halogen ions transfer from solvent to halogen-ionic bridges.



**Figure 9.** (a) Average energy terms in 241 halogen-ionic bridges formed in monomeric proteins and their comparisons with that in 222 salt bridges. The error bar ranges within  $\pm$  standard deviation. (b) Negative linear correlation with a correlation coefficient  $r = 0.702$  between  $\Delta\Delta G_{\text{add}}^{\text{hal}}$  and  $\Delta\Delta G_{\text{add}}^{\text{grp}}$ .

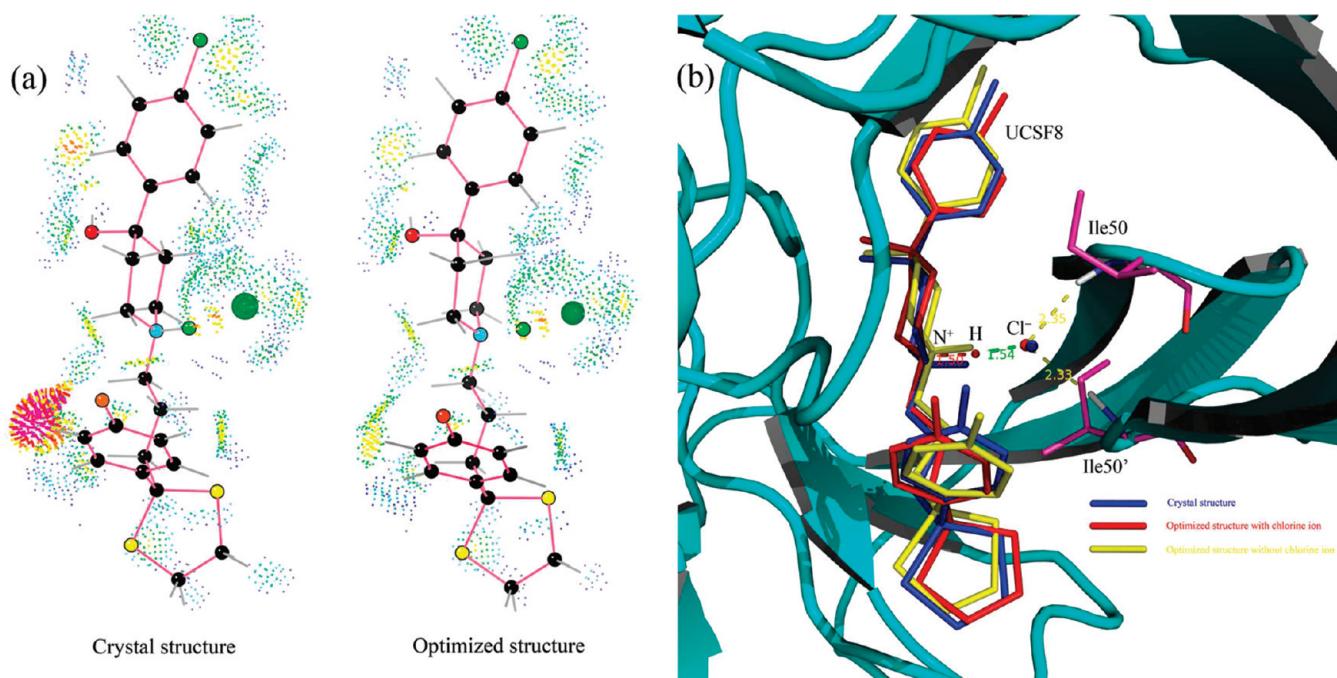
continuum electrostatic analysis. Only monomeric proteins are considered here because our primary interest is the effect of halogen-ionic bridges on protein folding rather than on protein binding. The detailed information about these selected halogen-ionic bridges and their electrostatic energy terms are collected in Supporting Information, Table S5.

A total of 204 (84.6%) out of the 241 halogen-ionic bridges in our data set are stabilizing ( $\Delta\Delta G_{\text{tot}} < 0$ ). This stabilization energy  $\Delta\Delta G_{\text{tot}}$ , as shown in Figure 9a, is a compromise between the favorable electrostatic interaction within the bridge ( $\Delta\Delta G_{\text{brd}} < 0$ ) as well as the interaction of the bridge with the charges in the rest of the protein ( $\Delta\Delta G_{\text{add}} < 0$ ) and the unfavorable desolvation penalty ( $\Delta\Delta G_{\text{dsv}} > 0$ ) of the polar/charged bridge due to its burial in low-dielectric protein interior. On average, halogen-ionic bridge formations incur a desolvation penalty  $\Delta\Delta G_{\text{dsv}}$  of  $+16.54 \text{ kcal}\cdot\text{mol}^{-1}$ . This penalty is over paid by the bridging energy term  $\Delta\Delta G_{\text{brd}}$  of  $-20.41 \text{ kcal}\cdot\text{mol}^{-1}$ . The electrostatic interactions of halogen-ionic bridges with the rest of the proteins are generally attractive with an average  $\Delta\Delta G_{\text{add}}$  of  $-14.53 \text{ kcal}\cdot\text{mol}^{-1}$ , which assists the bridging energy term to ultimately overcome the desolvation energy penalty, making the halogen-ionic bridge stabilizing. This free energy profile presented here for halogen-ionic bridge is coincident with that proposed previously for salt bridge,<sup>63</sup> but each energy term in the halogen-ionic bridge is much greater than that in the salt bridge (Figure 9a). Consequently, the stabilization effect of halogen-ionic bridge appears to be quite significant in comparison to that of salt bridge, given by the substantial difference in their total stabilization energies  $\Delta\Delta G_{\text{tot}}$  ( $-18.29 \text{ kcal}\cdot\text{mol}^{-1}$  for halogen-ionic bridge vs  $-3.66 \text{ kcal}\cdot\text{mol}^{-1}$  for salt bridge).

Breaking down of these energy terms in halogen-ionic bridge into two parts separately associated with the halogen ion and the bridging protein groups could provide a further insight into the energy composition of the bridge. From Figure 9a, it is seen that: (i) the bridging energy,  $\Delta\Delta G_{\text{brd}}$ , is made up of a dominant attraction term of halogen ion with its interacting groups ( $\Delta\Delta G_{\text{brd}}^{\text{hal}} = -24.36 \text{ kcal}\cdot\text{mol}^{-1}$ ) and a marginal repulsion term among these groups ( $\Delta\Delta G_{\text{brd}}^{\text{grp}} = +3.96 \text{ kcal}\cdot\text{mol}^{-1}$ ); (ii) the desolvation penalty,  $\Delta\Delta G_{\text{dsv}}$ , is the sum of two nearly equivalent terms as  $+7.86 \text{ kcal}\cdot\text{mol}^{-1}$

accounting for the desolvation of halogen ion ( $\Delta\Delta G_{\text{dsv}}^{\text{hal}}$ ) and  $+8.78 \text{ kcal}\cdot\text{mol}^{-1}$  for desolvating protein groups ( $\Delta\Delta G_{\text{dsv}}^{\text{grp}}$ ). As can be seen, the value of desolvation penalty for bridging halogen ions calculated using the continuum electrostatic approach is significantly lesser than those obtained from the empirical additive model, as described earlier (see Table 3). This is because the additive model only gives consideration in the “net” hydration effect of halogen ions. It equals the move of halogen ions from a solvent to a gas-phase condition, regardless of the fact that the dielectric constant in protein interiors is actually greater than 1 and the folded proteins are not of infinite extent, so the halogen ions in bridging state can also interact with solvent;<sup>62</sup> and (iii) the additional effect,  $\Delta\Delta G_{\text{add}}$ , seems to have arisen from a strong electrostatic attraction between the halogen ion and the rest of the protein ( $\Delta\Delta G_{\text{add}}^{\text{hal}} = -14.32 \text{ kcal}\cdot\text{mol}^{-1}$ ) and from a quite weak term of the bridging groups interacting with the rest of the protein ( $\Delta\Delta G_{\text{add}}^{\text{grp}} = -0.20 \text{ kcal}\cdot\text{mol}^{-1}$ ). However, one should beware of this statement, considering that these two mean quantities are accompanied with large standard deviations, indicating a significant variance within the sample scatters. In fact, most of  $\Delta\Delta G_{\text{add}}^{\text{hal}}$  and  $\Delta\Delta G_{\text{add}}^{\text{grp}}$  in our data set fall into a wide scope ranging from  $-80$  to  $+80 \text{ kcal}\cdot\text{mol}^{-1}$ , with few even getting more than  $+120 \text{ kcal}\cdot\text{mol}^{-1}$ . Moreover, there exists a negative linear correlation between  $\Delta\Delta G_{\text{add}}^{\text{hal}}$  and  $\Delta\Delta G_{\text{add}}^{\text{grp}}$  (Figure 9b), well reflecting the oppositely charged feature of halogen ion and its bridging groups in interaction with the same charges in the protein region out of the bridge.

**3.7. Real Biomolecular System. IV. QM/MM Calculation.** Recently, several specific intermolecular forces involved in ligand recognition and binding by protein receptors have been investigated in detail by means of the hybrid QM/MM methodology.<sup>82–84</sup> These works manifested that, if reasonably collocated with a MM context, it is possible to apply the expensive QM method to treat the nonbonding interactions of interest in the whole biomacromolecular framework. In order to give quantitative insight into the role and significance of halogen-ionic bridges in protein–ligand recognition and to explore their relevance to rational drug design, we herein addressed an ONIOM-based QM/MM study on two paradigms of chlorine-ionic bridges function-

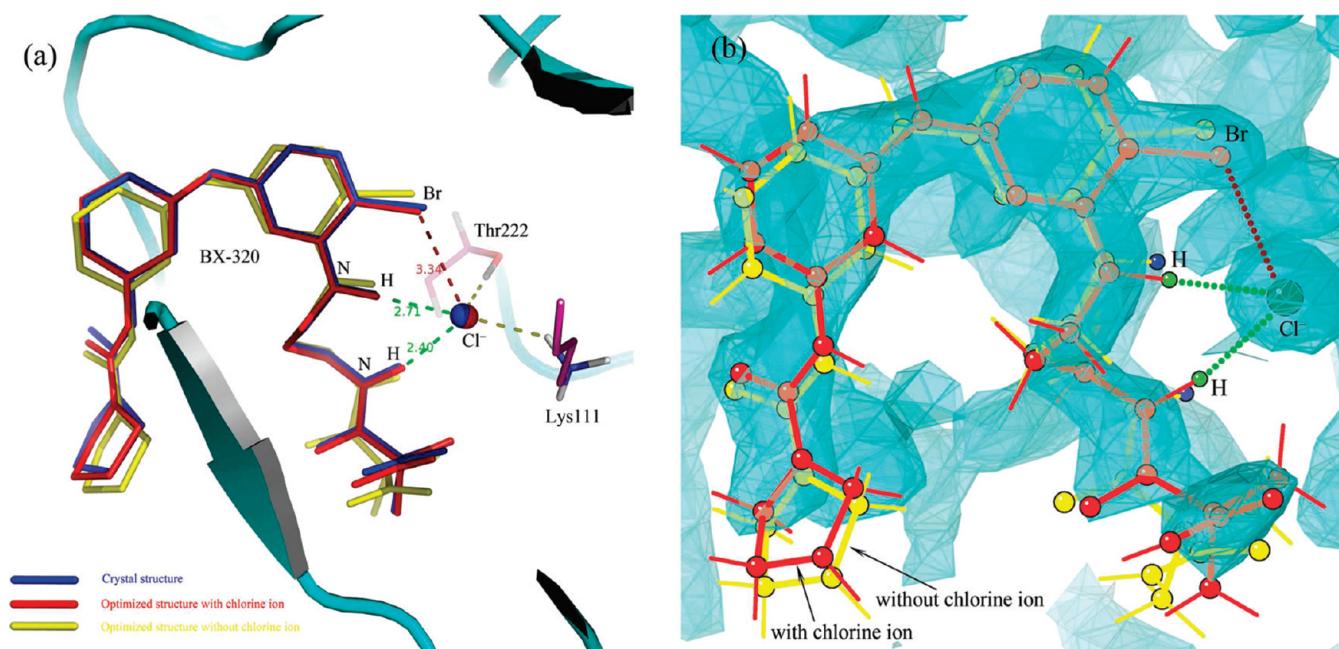


**Figure 10.** Stereoview of ONIOM-optimized UCSF8 structures in HIV-1 PR Q7K binding pocket. (a) Visualizing vdW clashes between UCSF8 and Q7K in crystal and optimized complex structures. Blue dots for wide contacts ( $>0.25\text{ \AA}$ ), green or yellow dots for good contacts (green for close contacts and yellow for slight overlaps,  $<0.2\text{ \AA}$ ), and red spikes for bad overlaps ( $\geq0.4\text{ \AA}$ ). (b) Superposition of the chlorine ion-containing (red) and chlorine ion-removed (yellow) structures to crystal structure (blue).

alizing to the binding of inhibitors by their targets. The first one is in the earlier mentioned complex of HIV-1 protease (HIV-1 PR) with a nonpeptide inhibitor UCSF8. Here, we used the crystal structure of its mutant, i.e. HIV-1 PR Q7K-UCSF8 complex, which has a similar inhibition profile ( $K_i = 15\text{ }\mu\text{M}$ ), similar kinetic parameters but higher resolution level (solved at  $1.9\text{ \AA}$ ) when compared to the wild type, as template to perform QM/MM analysis (PDB: 2aid).<sup>26</sup> The second is located at the binding interface of 3-phosphoinositide-dependent kinase 1 (PDK1) with its selective inhibitor BX-320, an aminopyrimidine derivative which can specifically bind to the catalytic domain of PDK1 at a nmol level of affinity ( $\text{IC}_{50} = 39\text{ nM}$ ) (PDB: 1z5m).<sup>85</sup> These two complex structures were submitted to an ONIOM minimization procedure, as described in Section 2.1, followed by single-point energy analyses of the optimized model layers using the rigorous MP2/aug-cc-pVDZ theory. For the purpose of comparison, halide anion-removed versions of these two complexes were also analyzed in the same way.

**HIV-1 PR Q7K-UCSF8 Complex.** Although the crystal structure of this complex was solved at a higher resolution level ( $1.9\text{ \AA}$ ), the larger thermal *B*-factors for atoms of UCSF8 ( $<B \geq 68.6\text{ \AA}^2$ ) versus all atoms in the protein ( $<B \geq 30.3\text{ \AA}^2$ ) suggest that the inhibitor position was not determined quite precisely.<sup>26</sup> This point can be validated by detecting van der Waals (vdW) clashes between Q7K and UCSF8 in both the complex crystal structure and the ONIOM-optimized structure. The small probe technique implemented in the PROBE program<sup>86</sup> was employed to fulfill this purpose, and the result is a graphic diagram visualizing the distribution of collisions around the UCSF8. From Figure 10a, it should be appreciated here that ONIOM

optimization can give a substantial refinement for the active region of the complex crystal structure, as shown by the fact that most of the bad overlaps at the crystal interface were eliminated after the ONIOM minimization procedure. The optimized UCSF8 conformations, with or without chlorine ion, are superposed on the crystal structure (Figure 10b), from which the root-mean-square deviations (RMSDs) of chlorine ion-containing and chlorine ion-removed ligand structures relative to the crystal one were computed to be, respectively,  $1.18$  and  $1.97\text{ \AA}$ . The former is far below the X-ray diffraction resolution of the studied crystal, whereas the latter is above of the resolution, indicating that the optimized model structures should be reliable and the absence of chlorine ion would throw a considerable effect on the ligand arrangement in Q7K binding pocket. Noteworthy, in the optimized structure, the N-H bond of UCSF8 amine moiety is elongated remarkably as much as to  $1.50\text{ \AA}$ , manifesting that there exists a significant trend of proton transfer toward the chlorine ion, which is consistent with that found earlier in small model systems. Energy analysis further revealed a noticeable effect of the proton transfer contributing to the complex stabilization. Interaction energy between the chlorine ion and UCSF8 was predicted to be  $-121.10\text{ kcal}\cdot\text{mol}^{-1}$ , this value is far more than that when a water molecule is placed at the same position of the chlorine ion ( $-15.97\text{ kcal}\cdot\text{mol}^{-1}$ ). In addition, the chlorine ion is shown to be also effective in interaction with the residues Ile50 and Ile50' of Q7K, on account of the strong intermolecular potential of  $-40.03\text{ kcal}\cdot\text{mol}^{-1}$ . In conclusion, the chlorine-ionic bridge should be important in assisting the specific binding of Q7K by UCSF8 and in maintaining the complex conformation and stability.



**Figure 11.** Stereoview of ONIOM-optimized BX-320 structures in PDK1 binding pocket. (a) Superposition of the chlorine ion-containing (red) and chlorine ion-removed (yellow) structures to crystal structure (blue). (b) Superposition of chlorine ion-containing (red) with chlorine ion-removed (yellow) structures in 2Fo–Fc electron-density map (shown contoured at  $2.2\sigma$ ).

**PDK1-BX-320 Complex.** ONIOM optimization of the huge BX-320 molecule in the PDK1 binding pocket was an exhaustive process, but the resulting BX-320 structure had only a little conformational change relative to original crystal structure ( $\text{RMSD} = 0.73 \text{ \AA}$ ). As shown in Figure 11a, the optimized BX-320 structure is nearly perfectly superposed on the crystal, with slight fluctuations over whole molecular heavy atoms. Each of the two N–H bonds in the flexible chain of BX-320 forms a typical ionic hydrogen bond with the chlorine ion that is bound to the residues Lys111 and Thr222 of PDK1. Intriguingly, the bromine atom on the BX-320 pyrimidine ring seems to be in weak interaction with the chlorine ion through a nonlinear halogen bond,<sup>87–89</sup> as claimed by their interatomic distance being shorter than the sum of respective van der Waals radii ( $3.34 < 3.66 \text{ \AA}$ )<sup>57,90</sup> and their angle  $\angle(\text{C–Br}\cdots\text{Cl}^-)$  meeting the criterion defined by Auffinger et al. for biological halogen bonds ( $121.3^\circ > 120^\circ$ ).<sup>72</sup> Topological analysis of the electron density in the optimized model layer confirmed the existence of a bond path linking the nuclei of Br and Cl<sup>-</sup>. However, the value (0.0132 au) of electron density  $\rho_b$  at the BCP for this interaction was predicted to be much smaller than that for two N–H $\cdots$ Cl<sup>-</sup> interactions (0.0365 and 0.0313 au), implying a very weak halogen bond in comparison to the strong ionic hydrogen bonds in this model. The Br $\cdots$ Cl<sup>-</sup> halogen bond can be, therefore, considered as secondary interaction contributions to the H $\cdots$ Cl<sup>-</sup> hydrogen bonds that conduct the formation of the chlorine-ionic bridge. At the MP2/aug-cc-pVDZ level, the interaction energies of the chlorine ion with BX-320 and PDK1 were calculated to be  $-34.02$  and  $-131.62 \text{ kcal}\cdot\text{mol}^{-1}$ , respectively. This noticeable energy level involved in the chlorine-ionic bridge should contribute considerably to the binding affinity of BX-320, though the desolvation penalty is not deducted from the QM-calculated interaction energies. The functionality of the chlorine ion in

the PDK1 active site can be intuitively characterized by comparing the optimized conformations of chlorine ion-containing and chlorine ion-removed complexes. It can be seen from Figure 11b that BX-320 structure exhibits an obvious motion when the chlorine ion is taken off from the complex, leading to a large RMSD ( $1.94 \text{ \AA}$ ) relative to that optimized with the chlorine ion. In addition, the chlorine ion-containing structure can be fitted into the 2Fo–Fc electron-density map fairly well, but the chlorine ion-removed one, especially at its two ends, departs from the  $2.2\sigma$  contour appreciably, indicating a nonignorable effect of the chlorine ion on the native architecture of this system.

#### 4. Conclusions

The main aim of this study is to prove the existence and significance of the putative halogen-ionic bridges in biomolecular systems. To achieve this, we present a comprehensive investigation on the geometrical profile and energy landscape of biological interactions involving halide anions. High-level ab initio calculations on small model systems preliminarily unveil the noticeable stabilization and the typical bonding character of halogen ion complexes with polar and charged groups. Database surveys of massive crystal structures deposited in the PDB further reveal a considerable number of geometrically preferential contacts between the nonbonded halogen ions and the electrophilic moieties of proteins, nucleic acids, and small ligands; these contacts are used to define a distinct data set consisting of 6406 biological halogen-ionic bridges. Continuum electrostatic analyses and hybrid quantum mechanics/molecular mechanics (QM/MM) examinations ultimately give a quantitative pronouncement for the important role of halogen-ionic bridges in conferring stability and specificity for protein folding and protein–ligand binding. All of these forebode that the halogen-ionic bridges should widely exist in

and solidly functionalize to biomolecules. We, therefore, expect that this newly proposed halide motif, with respect to its substantial magnitude and marked stabilization in biological context, could be exploited as a novel and versatile tool for rational drug design and bioengineering.

**Acknowledgment.** This work was supported by the State Key Laboratory of Trauma, Burns, and Combined Injury Foundation (no. SKLKF200904).

**Supporting Information Available:** A classification list of the 6406 biological halogen-ionic bridges retrieved from the current PDB database (January, 2010 release) and the calculated electrostatic energy terms for the 241 high-quality halogen-ionic bridges are tabulated. The six protein moieties used in continuum electrostatic calculations and their PARSE parameters as well as comparison of intermolecular potentials to ideal and real Coulombic energies for some small model complexes are shown. Tables S1–S5; Figures S1–S3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Hofmeister, F. *Arch. Exp. Pathol. Pharmakol.* **1888**, *24*, 247–260.
- (2) Marcus, Y. *Chem. Rev.* **2009**, *109*, 1346–1370.
- (3) Zhang, Y.; Cremer, P. S. *Curr. Opin. Chem. Biol.* **2006**, *10*, 658–663.
- (4) Takashima, K.; Riveros, J. M. *Mass Spectrom. Rev.* **1998**, *17*, 409–430, and references therein.
- (5) Kovács, A.; Varga, Z. *Coord. Chem. Rev.* **2006**, *250*, 710–727, and references therein.
- (6) Arshadi, M.; Yamdagni, R.; Kebarle, P. *J. Phys. Chem.* **1970**, *74*, 1475–1482.
- (7) Caldwell, G.; Kebarle, P. *J. Am. Chem. Soc.* **1984**, *106*, 967–969.
- (8) Hiraoka, K.; Mizuse, S.; Yamabe, S. *J. Phys. Chem.* **1988**, *92*, 3943–3957.
- (9) Bogdanov, B.; Peschke, M.; Tonner, D. S.; Szulejko, J. E.; McMahon, T. B. *Int. J. Mass Spectrom.* **1999**, *187*, 707–725.
- (10) Klopper, W.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227–2234.
- (11) Xantheas, S. S.; Dunning, T. H., Jr. *J. Phys. Chem.* **1994**, *98*, 13489–13497.
- (12) Johnson, M. S.; Kuwata, K. T.; Wong, C. K.; Okumura, M. *Chem. Phys. Lett.* **1996**, *260*, 551–557.
- (13) Choi, J. H.; Kuwata, K. T.; Cao, Y. B.; Okumura, M. *J. Phys. Chem. A* **1998**, *102*, 503–507.
- (14) Kim, J.; Lee, H. M.; Suh, S. B.; Majumdar, D.; Kim, K. S. *J. Chem. Phys.* **2000**, *113*, 5259–5272.
- (15) Bogdanov, B.; McMahon, T. B. *J. Phys. Chem. A* **2000**, *104*, 7871–7880.
- (16) Vinogradov, S. N. *Int. J. Pept. Protein Res.* **1979**, *14*, 281–289.
- (17) Zhang, W.; Piculell, L.; Nilsson, S. *Macromolecules* **1992**, *25*, 6165–6172.
- (18) Washabaugh, M. W.; Collins, K. D. *J. Biol. Chem.* **1986**, *261*, 12477–12485.
- (19) Chen, X.; Yang, T.; Kataoka, S.; Cremer, P. S. *J. Am. Chem. Soc.* **2007**, *129*, 12272–12279.
- (20) Lund, M.; Vrbka, L.; Jungwirth, P. *J. Am. Chem. Soc.* **2008**, *130*, 11582–11583.
- (21) (a) Ninham, B. W.; Yaminsky, V. *Langmuir* **1997**, *13*, 2097–2108. (b) Böstrom, M.; Tavares, F. W.; Bratko, D.; Ninham, B. W. *J. Phys. Chem. B* **2005**, *109*, 24489–24494. (c) Boström, M.; Lonetti, B.; Fratini, E.; Baglioni, P.; Ninham, B. W. *J. Phys. Chem. B* **2006**, *110*, 7563–7566.
- (22) (a) Lund, M.; Jungwirth, P.; Woodward, C. E. *Phys. Rev. Lett.* **2008**, *100*, 258105. (b) Lund, M.; Vácha, R.; Jungwirth, P. *Langmuir* **2008**, *24*, 3387–3391.
- (23) Heyda, J.; Hrobárik, T.; Jungwirth, P. *J. Phys. Chem. A* **2009**, *113*, 1969–1975.
- (24) Meyer, E. *Protein Sci.* **1992**, *1*, 1543–1562.
- (25) Sun, Y.; Kollman, P. *J. Phys. Chem.* **1996**, *100*, 6760–6763.
- (26) Rutenber, E.; Fauman, E. B.; Keenan, R. J.; Fong, S.; Furth, P. S.; Ortiz de Montellano, P. R.; Meng, E.; Kuntz, I. D.; DeCamp, D. L.; Saltoll, R.; Rosb, J. R.; Craik, C. S.; Stroud, R. M. *J. Biol. Chem.* **1993**, *268*, 15343–15346.
- (27) Suresh, C. H.; Vargheese, A. M.; Vijayalakshmi, K. P.; Mohan, N.; Koga, N. *J. Comput. Chem.* **2008**, *29*, 1840–1849.
- (28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (29) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.
- (30) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
- (31) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 889–926.
- (32) Chesnut, D. B.; Moseley, R. W. *Theor. Chim. Acta* **1969**, *13*, 230–248.
- (33) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (34) Glukhovtsev, M. N.; Pross, A.; Radom, L. *J. Am. Chem. Soc.* **1995**, *117*, 2024–2032.
- (35) Sanov, A.; Faeder, J.; Parson, R.; Lineberger, W. C. *Chem. Phys. Lett.* **1999**, *313*, 812–819.
- (36) Cioslowski, J.; Piskorz, P. *Chem. Phys. Lett.* **1996**, *255*, 315–319.
- (37) Asaduzzaman, A. M.; Schreckenbach, G. *Theor. Chem. Acc.* **2009**, *122*, 119–125.
- (38) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- (39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (41) Comell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.
- (42) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

- (43) Zhou, P.; Lv, J.; Zou, J.; Tian, F.; Shang, Z. *J. Struct. Biol.* **2010**, *169*, 172–182.
- (44) Zhou, P.; Zou, J.; Tian, F.; Shang, Z. *J. Chem. Inf. Model.* **2009**, *49*, 2344–2355.
- (45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2003.
- (46) Biegler-Konig, F.; Schoenbohm, J. *AIM2000*, 2nd ed.; Buro fur Innovative Software: Bielefeld, Germany, 2002.
- (47) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpente, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. *NBO 5.0*; Theoretical Chemistry Institute, University of Wisconsin: Madison, WI, 2001.
- (48) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr *Proteins* **2009**, *77*, 778–795.
- (49) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
- (50) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735–1747.
- (51) Zhou, P.; Tian, F.; Lv, F.; Shang, Z. *Proteins* **2009**, *76*, 151–163.
- (52) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins* **2008**, *73*, 765–783.
- (53) Zhao, Y.; Cheng, T.; Wang, R. *J. Chem. Inf. Model.* **2007**, *47*, 1379–1385.
- (54) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. *Biopolymers* **1996**, *38*, 305–320.
- (55) Rother, K.; Hildebrand, P. W.; Goede, A.; Gruening, B.; Preissner, R. *Nucleic Acids Res.* **2009**, *37*, D393–D935.
- (56) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. *J. Mol. Biol.* **1999**, *290*, 253–266.
- (57) Shannon, R. D. *Acta Crystallogr.* **1976**, *A32*, 751–767.
- (58) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086–3090.
- (59) Klots, C. E. *J. Phys. Chem.* **1981**, *85*, 3585–3588.
- (60) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem.* **2001**, *105*, 6507–6514.
- (61) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.
- (62) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211–226.
- (63) Kumar, S.; Nussinov, R. *J. Mol. Biol.* **1999**, *293*, 1241–1255.
- (64) Naray-Szabo, G.; Ferenczy, G. G. *Chem. Rev.* **1995**, *95*, 829–847.
- (65) Meot-Ner, M. *Chem. Rev.* **2005**, *105*, 213–284.
- (66) Nakanishi, W.; Hayashi, S.; Narahara, K. *J. Phys. Chem. A* **2008**, *112*, 13593–13599.
- (67) Wiberg, K. B. *Tetrahedron* **1968**, *24*, 1083–1096.
- (68) Reed, A. E.; Schleyer, P. *Inorg. Chem.* **1988**, *27*, 3969–3987.
- (69) Chesnut, D. B. *J. Chem. Theory Comput.* **2008**, *4*, 1637–1642.
- (70) Prabakaran, P.; Singarayan, M. G. *Chem. Lett.* **2004**, *33*, 1640–1641.
- (71) Rajagopal, S.; Vishveshwara, S. *FEBS J.* **2005**, *272*, 1819–1832.
- (72) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16789–16794.
- (73) Baldwin, R. L.; Rose, G. D. *Trends Biochem. Sci.* **1999**, *24*, 77–84.
- (74) Tsai, C. J.; Xu, D.; Nussinov, R. *Folding Des.* **1998**, *3*, R71–R80.
- (75) Osawa, S. *Annu. Rev. Biochem.* **1968**, *37*, 109–130.
- (76) Lu, Y.; Wang, R.; Yang, C. Y.; Wang, S. *J. Chem. Inf. Model.* **2007**, *47*, 668–675.
- (77) Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14258–14263.
- (78) Sun, D. P.; Sauer, U.; Nicholson, H.; Matthews, B. W. *Biochemistry* **1991**, *30*, 7142–7153.
- (79) Dao-pin, S.; Anderson, D. E.; Baase, W. A.; Dahlquist, F. W.; Matthews, B. W. *Biochemistry* **1991**, *30*, 11521–11529.
- (80) Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. *Nat. Struct. Biol.* **1995**, *2*, 122–128.
- (81) Noguchi, T.; Matsuda, H.; Akiyama, Y. *Nucleic Acids Res.* **2001**, *29*, 219–220.
- (82) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. *J. Med. Chem.* **2009**, *52*, 2854–2862.
- (83) Lu, Y.; Wang, Y.; Xu, Z.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. *J. Phys. Chem. B* **2009**, *113*, 12615–12621.
- (84) Alzate-Morales, J. H.; Caballero, J.; Jague, A. V.; Nilo, F. D. G. *J. Chem. Inf. Model.* **2009**, *49*, 886–899.
- (85) Feldman, R. I.; Wu, J. M.; Polokoff, M. A.; Kochanny, M. J.; Dinter, H.; Zhu, D.; Biroc, S. L.; Aliche, B.; Bryant, J.; Yuan, S.; Buckman, B. O.; Lentz, D.; Ferrer, M.; Whitlow, M.; Adler, M.; Finster, S.; Chang, Z.; Arnaiz, D. O. *J. Biol. Chem.* **2005**, *280*, 19867–19874.
- (86) Word, J. M.; Lovell, S. C.; LaBean, T. H.; Taylor, H. C.; Zalis, M. E.; Presley, B. K.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1711–1733.
- (87) Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y.; Murray, J. S. *J. Mol. Model.* **2007**, *13*, 305–311.
- (88) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. *J. Mol. Model.* **2007**, *13*, 291–296.
- (89) Murray, J. S.; Lane, P.; Politzer, P. *J. Mol. Model.* **2009**, *15*, 723–729.
- (90) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.