

# Quantum Chemical Investigations on Intraresidue Carbonyl–Carbonyl Contacts in Aspartates of High-Resolution Protein Structures

Tuhin Kumar Pal and Ramasubbu Sankararamakrishnan\*

Department of Biological Sciences and Bioengineering, Indian Institute of Technology, Kanpur, Kanpur 208016, India

Received: September 29, 2009; Revised Manuscript Received: December 2, 2009

The folding and stability of a polypeptide chain are due to many different and simultaneous noncovalent interactions. Recent studies have observed several novel and counterintuitive contacts in protein structures, and the nature of interactions due to such contacts is yet to be fully elucidated. We have identified carbonyl–carbonyl intraresidue contacts in 102 Asp residues from a data set of high-resolution protein structures. At the outset, it appears that such close approach of two carbonyl oxygen atoms is energetically not favorable. We have carried out ab initio quantum chemical calculations on 10 representative examples of self-contacting Asp residues from different regions of the Ramachandran map. Potential energy scan using three levels of theory (HF, B3LYP, and MP2) and two basis sets (6-31+G\* and 6-31++G\*\*) was performed by varying the side-chain dihedral angle  $\chi_1$  while keeping all other parameters corresponding to that observed in the protein structures. We also calculated interaction energies by considering the surrounding interacting residues and water molecules. Our results show that the energy difference between a self-contacting Asp residue from the crystal structures and the minimum energy conformations is about 10–15 kcal/mol. This small energy difference is compensated by its interactions with the surrounding residues and water molecules as observed in the interaction energy analysis. The results are independent of the levels of theory used. The contacting carbonyl–carbonyl groups adopt a sheared parallel motif orientation which helps to expose both the backbone and side-chain carbonyl oxygen atoms and enable them to participate in tertiary interactions. Natural bond orbital calculations indicate that carbonyl–carbonyl groups in self-contacting Asp residues interact through  $n \rightarrow \pi^*$  electron delocalization. The geometry analysis and nature of chemical interactions together explain the rationale for the existence of such Asp residues in protein structures and their importance in the protein stability.

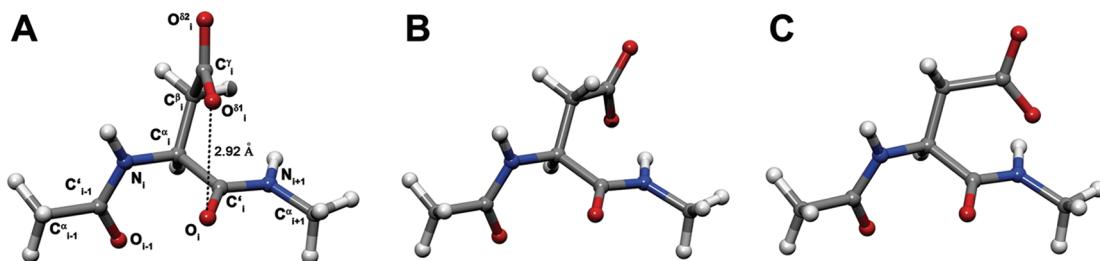
## Introduction

Interactions of many of the naturally occurring 20 amino acids have been shown to play a significant role in the structure and function of proteins. Side-chains of several amino acids participate in conventional as well as counterintuitive interactions. For example, the  $\pi$ -electron cloud of aromatic residues has been shown to take part in interactions with the lone pair electrons of oxygen or sulfur atoms.<sup>1–6</sup> Similarly,  $\pi$ -electrons have been shown to interact with the Cl groups in the crystal structures of certain protein–ligand complexes.<sup>7</sup> Although acidic and basic residues are prominently known to be involved in salt-bridge or hydrogen bond interactions, it has been recently observed that the nature of contacts between their side-chains in some cases is difficult to comprehend. For example, a recent analysis showed contacts between arginine residues in an analysis of 266 protein dimers.<sup>8</sup> It has been speculated that such clustering among arginines could produce polarization responses and thus could be a driving force for bringing protein subunits together. Similarly, in the high-resolution structures of 3150 polypeptide chains, close pairs of carboxylates were identified in acidic residues.<sup>9</sup> Such pairs of negatively charged residues are often found to be tightly packed with low B-factors, indicating possible attractive interactions. It is suggested that water molecules, metal ions, a positive residue, or, in the absence

of all three entities, a proton could provide stabilization to this seemingly repulsive interaction. In another study, analysis of aspartic and asparagine residues from the left-handed helical region of the Ramachandran map revealed stacking of the side-chain carbonyl over its own backbone carbonyl or the backbone carbonyl of the preceding residue.<sup>10</sup> Such carbonyl–carbonyl interaction presumably gives rise to favorable dipole–dipole interaction, and it has been argued that this stabilization energy is likely to be responsible for aspartic/asparagine to occur in the left-handed or partially allowed regions of the Ramachandran map. The evidence for the existence of carbonyl–carbonyl interactions was provided by the analysis of small molecule crystal structures available in Cambridge Structural Database.<sup>11</sup> The favorable nature of dipolar interactions between carbonyl groups was demonstrated using ab initio quantum chemical calculations.<sup>12</sup> Recently, the orthogonal dipolar interactions between amide carbonyl bond dipoles were quantified experimentally by employing a model system.<sup>13</sup> The computed and experimental interaction energies are comparable to those of hydrogen bonds<sup>12</sup> or aromatic–aromatic C–H $\cdots\pi$  and  $\pi\cdots\pi$  interactions.<sup>14–18</sup> Such interactions have been proposed to play a significant role in the stability of protein secondary structural elements.<sup>19,20</sup>

In an analysis of Asx (Asp/Asn) and Glx (Glu/Gln) residues with intraresidue contacts, we have recently identified Asx and Glx residues with carbonyl–carbonyl contact that occur between the side-chain C=O and main-chain C=O of the same residue.<sup>21</sup>

\* To whom correspondence should be addressed. E-mail: rsankar@iitk.ac.in. Phone: +91 512 2594014. Fax: +91 512 2594010.



**Figure 1.** (A) Example of a self-contacting Asp residue (D129) from the protein epidermolytic toxin A (PDB ID: 1AGJ). The ( $\varphi$ ,  $\psi$ ) angle of this residue falls in the extended region in the Ramachandran map, and it belongs to cluster II. The N- and C-terminal ends were capped by acetyl and *N*-methylamide groups to mimic the peptide linkages from preceding and succeeding residues. Individual atoms are labeled with the self-contacting residue as the *i*th residue. This labeling is used throughout the paper. In parts B and C, the minimum ( $E_{\text{min}}$ ) and maximum energy ( $E_{\text{max}}$ ) conformations obtained from the potential energy scan calculations of this self-contacting Asp residue using the B3LYP/6-31++G\*\* level of theory are shown.

Such self-contacts were observed in the high-resolution protein structures of two data sets and are not restricted to any specific region within the Ramachandran map. They presumably occur in a relatively more hydrophobic environment to maximize the number of polar contacts with the available polar groups in the surroundings. The distance criterion used in that study<sup>21</sup> to define self-contact between the two carbonyl oxygen atoms is stringent (2.75 Å) and is close to the normal limit for two oxygen atoms.<sup>22</sup> The partial negative charges from both the carbonyl oxygens in such a close distance will result in a highly unfavorable interaction between the two functional groups. It is intriguing to find out how these acidic residues with such seemingly unfavorable carbonyl–carbonyl contacts are accommodated within protein structures.

Several research groups have carried out quantum chemical calculations on different aspartic acid compounds at various levels of theory. Conformational preferences of *N*-acetyl-L-aspartic acid-*N'*-methylamide,<sup>23–25</sup> *N*-formyl-L-aspartic acid amide, and *N*-formyl-L-aspartatamide<sup>26</sup> have been studied at the Hartree-Fock (HF) level or density functional theory (DFT). Similarly, cationic, zwitterionic, and anionic species of L-aspartic acid were also investigated.<sup>27</sup> Both backbone and side-chain conformations were explored in these studies. Results from such calculations focused on relative energies, and the most stable conformations mostly involved one or two side-chain–backbone or backbone–backbone hydrogen bonds.

In this study, we have first identified Asp residues with intraresidue carbonyl–carbonyl contacts from high-resolution protein structures. Quantum chemical calculations were carried out at three different levels of theory by varying the side-chain dihedral angle,  $\chi_1$ , for a given backbone conformation observed from the protein structures. The relative energies of self-contacting aspartic acid are compared with those of the most and least stable conformations. The calculations were performed for 10 aspartates with different backbone conformations. In all 10 cases, interaction energies were also calculated between each self-contacting aspartate and its surrounding interacting residues using quantum chemical methods. Geometry characterization and natural bond orbital (NBO) calculations on the contacting carbonyl–carbonyl group explained the rationale for occurrence of such moieties in protein structures.

## Methods

**Database Search.** Two data sets of high-resolution protein structures were merged to identify the aspartic acid residues that have intraresidue contacts between the side-chain and main-chain carbonyl oxygen atoms of the same residue. The first data set contained 500 high-resolution structures carefully culled by

Jane Richardson and her colleagues (<http://kinemage.biochem.duke.edu/databases/top500.php>).<sup>28</sup> The second data set is from PDBSELECT<sup>29</sup> released in March 2006 which contains a representative list of protein chains. From both data sets, only crystal structures with a resolution of 2.0 Å or better and an *R*-factor of 20% or better were considered. In the combined data set, if two protein structures showed 25% or more sequence identity, then only one was considered. In this nonredundant set of high-resolution protein structures, all Asp residues whose average *B*-factor was  $\leq 40$  and occupancy of all of the atoms was equal to 1.0 were examined to identify self-contacts. An aspartic acid residue is defined as self-contacting if the distance between backbone and side-chain carbonyl oxygens is  $\leq 3.0$  Å. This is a slightly relaxed criterion compared to our previous studies<sup>21</sup> where a distance of 2.75 Å was used to find self-contacting Asp residues. The cutoff value 3.0 Å is just below twice the van der Waals radius of oxygen.<sup>30</sup>

**Quantum Chemical Calculations.** Two types of quantum chemical calculations were performed, namely, (i) potential energy surface (PES) scan of self-contacting Asp residues and (ii) interaction energy analysis between self-contacting Asp and surrounding residues and water molecules in the protein crystal structures. All calculations were performed using the Gaussian 03 program package.<sup>31</sup> The MOLDEN v4.6 program<sup>32</sup> was utilized to scrutinize the geometries.

**(i) Potential Energy Surface Scan.** A potential energy surface (PES) scan was carried out on Asp residues which have intraresidue contacts between the carbonyl oxygen atoms and were observed in different regions of the Ramachandran map.<sup>22</sup> Ten representative examples from right-handed  $\alpha$ -helical, extended, and left-handed  $\alpha$ -helical regions were considered for the calculations. The calculations were carried out on the *N*-acetyl-L-aspartate-*N'*-methylamide molecule, and the coordinates of the heavy atoms were taken directly from the corresponding protein structures. Hydrogen coordinates were generated using the Builder module of the InsightII molecular modeling package (InsightII v2000, Accelrys Inc., San Diego, CA). While all of the parameters including the backbone dihedral angles were fixed corresponding to the X-ray structure, the side-chain dihedral angle  $\chi_1$  ( $\text{N}-\text{C}^{\alpha_i}-\text{C}^{\beta_i}-\text{C}^{\gamma_i}$ ; for atom labels, see Figure 1A) was systematically varied in steps of 5° starting from the experimentally determined value. Thus, single point calculations were carried out using the 6-31+G\* and 6-31++G\*\* basis sets on 72 conformations for each of the 10 examples to generate the PES. Three distinct levels of theories were used in the calculations, namely, Hartree–Fock (HF), density functional theory (DFT), and Møller–Plesset perturbation theory to second order (MP2). For each self-contacting Asp

residue, a total of six combinations (two basis sets and three distinct theories) were considered. In total, we carried out 4320 single point calculations. Relative energies with respect to the experimentally observed self-contacting Asp residue were analyzed and compared with the most and least favorable conformations.

In order to understand and characterize the stereoelectronic effect of carbonyl–carbonyl interactions in self-contacting Asp residues, natural bond orbital (NBO)<sup>33,34</sup> calculations were carried out on the same 10 representative examples obtained from different allowed regions of the Ramachandran map. NBO analyses were performed on the *N*-acetyl-L-aspartate-*N'*-methylamide molecule at the B3LYP/6-31++G\*\* level using NBO 3.1<sup>35</sup> as implemented in Gaussian 03.<sup>31</sup>

(ii) **Interaction Energy Analysis.** The representative examples of self-contacting Asp residues from different regions of the Ramachandran map considered for potential energy scan have at least two polar contacts from backbone nitrogen atoms of surrounding interacting residues. In the majority of cases (9 out of 10), they also interact with at least one water molecule. We computed the interaction energies between these self-contacting Asp residues and the interacting polar groups/water molecules. Polar contact was defined, if the distance between the two polar atoms is less than the sum of van der Waals radii + 0.6 Å.<sup>36</sup> The nature of chemical interactions can be understood only if the hydrogen positions are clearly defined. Since hydrogen atom positions are not available in most of the protein structures, the system for quantum mechanical calculations was constructed as follows. First, hydrogen atoms were built for the entire protein using two programs, namely, GROMACS<sup>37</sup> with the OPLS force field<sup>38</sup> and the Builder module of the InsightII molecular modeling package (InsightII v2000, Accelrys, San Diego, CA). If the polar contact with the self-contacting Asp was due to a backbone atom, then that residue was substituted by a Gly residue. If a side-chain participates in polar contact, then an equivalent model system was considered in its place. Side-chains of Arg, Lys, His, and Thr were found to take part in polar contacts in the 10 examples considered in this study. They were respectively substituted by guanidinium, ammonium, imidazole, and ethanol moieties. If both side-chain and backbone atoms are involved in contacts, then the entire residue was retained. In the place of self-contacting Asp, we considered the *N*-acetyl-L-aspartate-*N'*-methylamide molecule as in potential energy scan calculations. The coordinates of the heavy atoms were directly taken from the crystal structures. With the two programs to construct hydrogen atom positions, we have considered 20 systems for ab initio calculations.

We have used the 6-31++G\*\* basis set, and the HF, DFT (B3LYP), and MP2 levels of theory were used. Our purpose here is to calculate the interaction energy between self-contacting Asp and the residues/water molecules making polar contacts with this residue. Hence, we treated the system as a bimolecular system (*AB*) in which *A* was considered as self-contacting Asp and the interacting residues/water molecules were considered as *B*. The scheme used to calculate the interaction energy ( $\Delta E$ ) is as follows.

$$\Delta E = E_{AB(AB)} - (E_{A(B)} + E_{B(A)}) \quad (1)$$

where  $E_{AB(AB)}$  is the energy of the entire system for the given basis set (*AB*),  $E_{A(B)}$  is the energy of *A* given basis set (*B*) and  $E_{B(A)}$  is the energy of *B* given basis set (*A*). The size of the systems varied from 46 to 70 atoms, and BSSE correction was fully taken into account. In both the potential energy scan and

interaction energy analysis, single point calculations were carried out without optimizing the systems. This is due to the fact that the systems under consideration are present in a protein environment where tertiary interactions from the surrounding residues impose constraints and any optimization of simplified systems from such compactly folded structure is likely to move away from the actual environment. Results of calculations of optimized systems in these cases are not likely to reflect the real picture present in the protein environment, and hence, the present approach of performing the single point calculations was adopted.

## Results

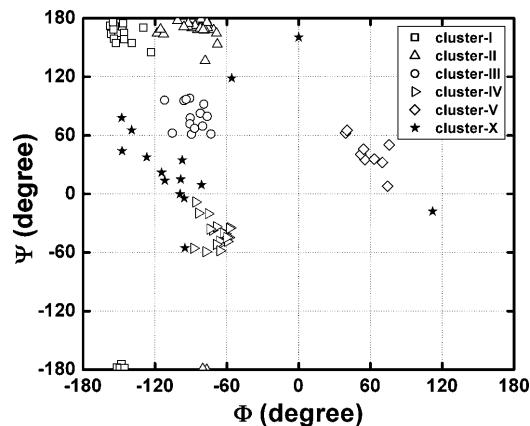
**Crystal Structure Analysis.** Our database search of high-resolution protein structures yielded 102 examples in which side-chain carbonyl oxygen of an Asp residue is within 3.0 Å distance from the backbone carbonyl oxygen atom of the same residue (Table 1). An example of a self-contacting Asp residue (D129) from the protein epidermolytic toxin A (PDB ID: 1AGJ) is shown in Figure 1A. Analysis of their backbone dihedral angles indicates that the self-contacting Asp residues are not restricted to any particular region of the Ramachandran map (Figure 2). They are spread over all of the allowed regions of ( $\varphi$ ,  $\psi$ ) map.<sup>22</sup> We have calculated several stereochemical and geometrical parameters involving the self-contacting Asp residues. The side-chain dihedral angle  $\chi_1$  ( $N_i-C^{\alpha}_i-C^{\beta}_i-C^{\gamma}_i$ ) shows a distribution around  $\pm 180^\circ$  (trans) and  $+60^\circ$  (gauche<sup>+</sup> or  $g^+$ ) (see Figure S1 in the Supporting Information). Surprisingly, the region around  $-60^\circ$  (gauche<sup>-</sup> or  $g^-$ ) is completely avoided. In general,  $g^-$  is more preferred than  $g^+$  for the  $\chi_1$  dihedral angle of Asp residues.<sup>39</sup> Similarly, the distribution of dihedral angles  $\chi_{21}$  and  $\chi_{22}$  ( $C^{\alpha}_i-C^{\beta}_i-C^{\gamma}_i-O^{\delta 1}/C^{\alpha}_i-C^{\beta}_i-C^{\gamma}_i-O^{\delta 2}_i$ ) shows two clearly distinct peaks in the regions of 0 and  $\pm 180^\circ$  (see Figure S1 in the Supporting Information). Due to the delocalization of electrons in the side-chain carbonyl group, a planar arrangement is preferred for this dihedral angle.

Potential energy surface scan of an Asp residue requires varying several parameters that include the backbone and side-chain dihedral angles. Such a calculation requires an enormous amount of computational time. Thus, we have grouped the observed examples into six different clusters in the Ramachandran map (Figure 2). The first five clusters represent different allowed regions and constitute ~85% of the observed examples. Clusters I–III correspond to three distinct groupings within the extended region. Clusters IV and V fall in the right- and left-handed  $\alpha$ -helical regions, respectively. The sixth cluster (cluster X) corresponds to the remaining examples not included in the previous five clusters. We have selected two examples from each of the first five clusters for the potential energy surface calculations, and they are shown in bold in Table 1 (see also Tables 2 and 3). In each case, the self-contacting Asp residue has polar contacts with two backbone nitrogen atoms and one oxygen atom from the neighboring interacting residues (Table 4). Additionally, at least one water molecule also interacted with the Asp residue in 9 out of 10 examples. During the PES scan, all geometric parameters including the backbone dihedral angles were fixed. The side-chain dihedral angle  $\chi_{21}/\chi_{22}$  of Asp was also the same as that of the crystal structure, i.e., close to 0 or  $180^\circ$  to keep the planar arrangement of the side-chain carboxyl group. Only the side-chain dihedral angle  $\chi_1$  of self-contacting Asp was varied as described in the Methods section. For each  $\chi_1$ , single point calculation was carried out and the energy was compared to that of the self-contacting Asp residue. The

TABLE 1: Example of Aspartates with Carbonyl–Carbonyl Self-Contacts

PDB <sup>a</sup>	residue	$D_{O\cdots O}^b$	PDB <sup>a</sup>	residue	$D_{O\cdots O}^b$	PDB <sup>a</sup>	residue	$D_{O\cdots O}^b$
1A3A/D	D95	2.67	1M15/A	D263	2.75	1TU9/A	D117	2.96
<b>1AGJ/A</b>	<b>D129</b>	<b>2.92</b>	1MNN/A	D103	2.65	1TZV/A	D40	2.99
1AGJ/A	D218	2.95	1MNN/A	D88	2.34	1TZV/A	D97	2.94
1AJS/A	D169	2.96	1MRO/C	D206	2.96	1U02/A	D25	2.83
<b>1AMF</b>	<b>D59</b>	<b>2.83</b>	1MUG/A	D98	2.95	1U4G/A	D243	2.96
1B0U/A	D143	2.87	1MUW/A	D100	2.98	1UPQ/A	D65	2.95
1BG6	D333	2.99	1O2D/A	D152	2.96	1USH	D456	2.97
<b>1BS0/A</b>	<b>D154</b>	<b>2.99</b>	1O4Y/A	D279	2.91	1USH	D477	2.95
1BU7/A	D232	2.81	<b>1O9R/A</b>	<b>D31</b>	<b>2.96</b>	1VJF/A	D51	2.98
1BW9/A	D267	2.80	1OD3/A	D142	2.96	1VMH/A	D15	2.83
1CS1/A	D126	2.99	1P4C/A	D284	2.99	1W0H/A	D230	2.88
1CVR/A	D103	2.95	1P4C/A	D332	2.90	1W0N/A	D125	2.95
1D3V/A	D100	2.90	1PCF/A	D91	2.89	1WCW/A	D117	2.81
1D3V/A	D234	2.98	1PFV/A	D255	2.99	1X9D/A	D357	2.95
1DBG/A	D188	2.47	1PFV/A	D351	2.72	1X9D/A	D675	2.93
<b>1DOZ/A</b>	<b>D42</b>	<b>2.73</b>	1PFV/A	D535	2.69	1Y0E/A	D94	2.88
1DQS/A	D44	2.94	1PHN/A	D13	2.95	1Y43/B	D39	2.65
1E29/A	D35	2.82	1PMI	D133	2.71	1YGE	D607	2.95
1EB6/A	D121	2.97	1PMI	D398	2.98	1Z3E/B	D257	2.95
1EUV/A	D412	2.92	1QGQ/A	D208	2.69	1ZD7/A	D34	2.96
1F1E/A	D130	2.87	1QKS/A	D9	2.98	2BBK/L	D17	2.79
1F60/B	D1191	2.99	1QOP/B	D176	2.84	2BJF/A	D63	2.87
1F8E/A	D111	2.87	1QQF/A	D1245	2.96	<b>2BKQ/C</b>	<b>D79</b>	<b>2.95</b>
1F8E/A	D293	2.88	1QTN/B	D454	2.87	2BKV/B	D22	2.48
1G61/A	D2033	2.97	1R5L/A	D108	2.96	2BW4/A	D53	2.98
1GK9/B	D183	2.99	<b>1RWI/A</b>	<b>D39</b>	<b>2.81</b>	2CCM/A	D127	2.97
1H4X/A	D22	2.58	1RY9/A	D26	2.76	2CWQ/A	D53	2.92
1HZT/A	D108	2.70	1S3Z/B	D115	2.94	2ERB/A	D118	2.99
1JFB/A	D260	2.95	1S3Z/B	D51	2.89	2EX4/A	D128	2.83
1JH6/A	D16	2.52	1SQ9/A	D145	2.95	2NAC/A	D23	2.83
1LAM	D332	2.78	<b>1TIF</b>	<b>D20</b>	<b>2.99</b>	2POR	D294	2.92
1LAM	D412	2.99	1TL2/A	D96	2.99	3SEB	D55	2.90
1LTZ/A	D11	2.86	1TP6/A	D112	2.90	<b>7A3H/A</b>	<b>D181</b>	<b>2.98</b>
1LTZ/A	D213	2.92	<b>1TP6/A</b>	<b>D27</b>	<b>2.97</b>	7ATJ/A	D43	2.98

<sup>a</sup> Four letter Protein Data Bank<sup>48</sup> ID/Chain ID. <sup>b</sup> Minimum of the two distances in angstroms  $d(O_i\cdots O^{\delta 1}_i)$  and  $d(O_i\cdots O^{\delta 2}_i)$ . PDB examples selected for QM calculations are shown in bold. For atom labels, see Figure 1A.



**Figure 2.** The  $(\varphi, \psi)$  angles of 102 self-contacting Asp residues are shown in the Ramachandran map. On the basis of the distribution, the self-contacting Asp residues are grouped into five major clusters and the corresponding  $(\varphi, \psi)$  values are shown in different symbols. The sixth cluster is comprised of all of the examples that are not included in the first five clusters.

calculations were performed for all 10 examples considered in this study, and the results are presented below.

**Potential Energy Surface Scan.** The potential energy profiles obtained from single point calculations as a function of the side-chain dihedral angle  $\chi_1$  are plotted for all 10 examples in Figure 3. Results from all three levels of theory are shown. Distance profiles for the two carbonyl oxygen atoms from the same residue are also plotted in the same graph. For each example,

the energy profiles are plotted relative to the energy of the self-contacting Asp residue from the X-ray structure. All three levels of theory (MP2, DFT, and HF) have resulted in similar profiles for all 10 examples. It is immediately clear that the energy of the self-contacting Asp conformation ( $E_{xray}$ ) is well separated from the energetically unfavorable conformations that can be recognized from the peaks ( $E_{max}$ ) in the potential energy profiles. In examples belonging to clusters I, IV, and V, there are clearly two peaks. We have presented results for conformations from peaks that give rise to the largest energy difference with respect to the self-contacting Asp residues (Table 2 and Table S1 in the Supporting Information). In the majority of the cases, the potential energy difference between the conformations corresponding to  $E_{xray}$  and  $E_{max}$  is more than 100 kcal/mol. Only in 1DOZ and 1O9R which are representative examples for the right-handed  $\alpha$ -helical region, this energy difference is smaller (22–27 kcal/mol). The distances between the backbone and side-chain carbonyl oxygen atoms for energetically unfavorable conformations representing  $E_{max}$  are greater than 3.0 Å for the majority of the examples studied. However, closer examination of these conformations revealed that there are severe steric clashes between the side-chain carbonyl oxygen atoms and backbone heavy atoms.

We have also compared the relative energies of the self-contacting Asp residue,  $E_{xray}$ , and the minimum energy ( $E_{min}$ ) conformations. The energy difference between these two conformations varies between 4.0 and 22 kcal/mol (Table 2 and Table S1 in the Supporting Information). However, in the

**TABLE 2: Relative Energies and Other Parameters of Self-Contacting Asp Residues from X-ray Structures and Minimum and Maximum Energy Conformations**

cluster	PDB	Asp <sup>a</sup>	<i>E</i> <sub>xray</sub>		<i>E</i> <sub>min</sub>			<i>E</i> <sub>max</sub>		
			$\chi_1^b$	<i>D</i> <sub>O...O<sup>c</sup></sub>	$\chi_1^b$	<i>D</i> <sub>O...O<sup>c</sup></sub>	$\Delta E^d$	$\chi_1^b$	<i>D</i> <sub>O...O<sup>c</sup></sub>	$\Delta E^d$
I	1AMF	D59	66.3	2.83	-163.6	4.41	-17.1	-58.6	5.10	60.2
	1RW1	D39	69.2	2.81	-160.8	4.36	-13.4	154.2	3.37	106.0
II	1AGJ	D129	68.4	2.92	-161.6	4.51	-12.8	143.4	3.41	169.3
	1TIF	D20	84.4	2.99	-150.6	4.44	-13.6	159.4	3.45	162.7
III	1BS0	D154	-177.0	2.99	47.9	3.14	-4.6	112.9	1.53	140.2
	7A3H	D181	-172.3	2.98	37.6	3.51	-9.4	112.6	1.42	189.7
IV	1DOZ	D42	170.2	2.73	35.2	4.42	-21.4	85.2	3.25	26.1
	1O9R	D31	-176.5	2.96	48.5	4.41	-19.0	98.5	3.34	22.2
V	1TP6	D27	-174.3	2.97	-64.3	4.92	-10.7	20.6	4.95	178.1
	2BKQ	D79	-174.8	2.95	-69.8	5.03	-10.7	120.2	1.37	184.5

<sup>a</sup> Self-contacting Asp residue; for polypeptide chain ID, see Table 1. <sup>b</sup>  $\chi_1$ : Side-chain dihedral  $N_i-C^{\alpha}_i-C^{\beta}_i-C^{\gamma}_i$  (in degrees). <sup>c</sup> Minimum of the two distances in angstroms  $d(O_i \cdots O^{31}_i)$  and  $d(O_i \cdots O^{32}_i)$  determined for conformations corresponding to  $E_{xray}$ . For atom labels, see Figure 1A. <sup>d</sup>  $\Delta E$ : Relative energy calculated at the B3LYP/6-31++G\*\* level. For each example, the energies (in kcal/mol) for  $E_{min}$  and  $E_{max}$  are calculated with respect to the energy of  $E_{xray}$  conformation.

**TABLE 3: Relative energies<sup>a</sup> (in kcal/mol) of All 30 Structures at the B3LYP/6-31++G\*\* Level Obtained from the Potential Energy Surface Scan Calculations of Different Clusters**

PDB	<i>E</i> <sub>xray</sub>	<i>E</i> <sub>min</sub>	<i>E</i> <sub>max</sub>
1AMF	-18.0	-35.2	42.1
1RW1	-1.7	-15.1	104.3
1AGJ	-15.2	-28.1	154.0
1TIF	-13.6	-27.2	149.1
1BS0	-19.4	-24.0	120.9
7A3H	-16.6	-26.0	173.1
1DOZ	-8.5	-30.0	17.6
1O9R	-4.0	-23.0	18.2
1TP6	0.0	-10.7	178.1
2BKQ	-2.0	-12.7	182.5

<sup>a</sup> The energies of all 30 structures are relative to the X-ray structure of 1TP6.

majority of the examples, this difference is between 10 and 15 kcal/mol. It is clear that the energy of the self-contacting Asp residue in all of the clusters lies closer to that of minimum energy conformations in the respective clusters. In conformations corresponding to  $E_{min}$ , the side-chain carbonyl oxygen is involved in hydrogen bonding interaction with the backbone nitrogen of Asp ( $N_i$ ) or the nitrogen atom ( $N_{i+1}$ ) belonging to the *N*'-methylamide part of the molecule and this gives additional stability to the Asp residue.  $E_{min}$  and  $E_{max}$  conformations of the self-contacting Asp residue, D129, from epidermolytic toxin A (PDB ID: 1AGJ) are shown in Figure 1B and 1C, respectively.

We have also considered the conformations and energies of all 10 self-contacting Asp residues and their corresponding minimum and maximum energy conformations together. In total, 30 different conformations and their energies were analyzed and all of the energies were relative to the self-contacting Asp from the crystal structure 1TP6 of cluster V. Only results obtained from the DFT level of theory will be discussed. MP2 and HF level calculations produced qualitatively similar results (data not shown). It is interesting to note that some of the conformations of the self-contacting Asp residues from specific clusters are more stable than others by 10–19 kcal/mol (Table 3 and Figure S2 in the Supporting Information). The self-contacting Asp residue corresponding to 1BS0 (cluster III) has the most favorable energy. It is surprising that, even among different self-contacting Asp residues, such a wide range of stability is seen although in all of the cases the distance between the two carbonyl oxygen atoms is  $\leq 3.0$  Å. Other parameters including

the backbone dihedral angles and the interactions between other pairs of atoms definitely play a role in influencing the potential energy of self-contacting Asp residues.

Comparison of the relative energies of self-contacting Asp residues and those of minimum energy conformations reveals that some of the self-contacting residues from particular clusters have energetically more favorable conformations than the minimum energy conformations belonging to other clusters. For example, energies of self-contacting Asp residues representing the X-ray structures 1AMF (cluster I), 1AGJ (cluster II), and 1BS0 and 7A3H (cluster III) are more favorable than those of the minimum energy conformations corresponding to 1RW1 (cluster I) and 1TP6 and 2BKQ (cluster V). If we consider all of 10 self-contacting Asp residues and all 10 minimum energy conformations, the maximum energy difference occurs between the self-contacting Asp residue of 1TP6 and the minimum energy conformation of 1AMF (35.2 kcal/mol). The minimum energy difference of 0.1 kcal/mol is observed between the self-contacting Asp residue of 1AGJ and the minimum energy conformation of 1RW1. The relative energies of minimum energy conformations vary between -11 and -35 kcal/mol, and the majority of them fall between -23 and -30 kcal/mol.

We also carried out a full geometry optimization followed by frequency calculation on conformations corresponding to  $E_{xray}$ ,  $E_{min}$ , and  $E_{max}$  from all 10 representative examples. Such a procedure will help us to locate and characterize the stationary points on the potential energy surface of the *N*-acetyl-L-aspartate-*N*'-methylamide molecule and will enable us to compare how far an  $E_{xray}$  conformation is from the optimized structure. Thirty such calculations were carried out using the 6-31++G\*\* basis set at the DFT (B3LYP) level theory. Our computations using tight optimization revealed that at least one of the three starting points ( $E_{xray}$ ,  $E_{min}$ , and  $E_{max}$ ) in examples from clusters I–III converged to a structure with  $\varphi$ ,  $\psi$  values corresponding to -151.5 and +173.5°. Similarly, the energetically most favorable optimized structures belonging to cluster IV were close to the right-handed  $\alpha$ -helical region ( $\varphi = -106.5^\circ$ ;  $\psi = -54.2^\circ$ ). The relative energies of optimized conformations (calculated with respect to  $E_{xray}$  of 1TP6) were between -44 and -46 kcal/mol in clusters I–IV which is about 10 kcal/mol lower than the most favorable  $E_{min}$  structure (Table 3). Structures from cluster V optimized to conformations close to the left-handed helical region with a slightly higher energy (-38.6 kcal/mol). Thus, the energetically most favorable geometrically optimized structures from five clusters essentially gave rise to stationary points

**TABLE 4: Surrounding Residues and Water Molecules Exhibiting Polar Contacts with Self-Contacting Asp Residues (Interaction Energies Have Been Calculated with the CP Method for All 10 Examples Representing Five Clusters)**

PDB	interacting residues <sup>a</sup>	no. of atoms <sup>b</sup>	interaction energy <sup>c</sup> (kcal/mol)	BSSE correction
1AMF	S57, K61, W62, HOH309	46	-34.886	1.653
1RW1	Q41, H42, L43, HOH88	63	-38.362	2.068
1AGJ	N131, G132, V133, <b>K138</b> , HOH595, HOH851	61	-120.330	3.893
1TIF	N22, G23, D24, <b>R41</b> , HOH283	63	-95.433	2.863
1BS0	T156, H157, L158, HOH1149, HOH1472	66	-22.140	3.102
7A3H	H183, H184, HOH35, HOH126, HOH150, HOH398, HOH479	68	-30.867	4.228
1DOZ	K44, D45, <b>R46</b> , HOH185	70	-123.680	2.820
1O9R	A33, L34, V35, T91, HOH1021, HOH1022, HOH1072	68	-41.578	3.466
1TP6	L29, <b>D30</b> , A31, HOH565	58	56.770	1.624
2BKQ	P81, N82, <b>K83</b>	55	-133.195	1.691

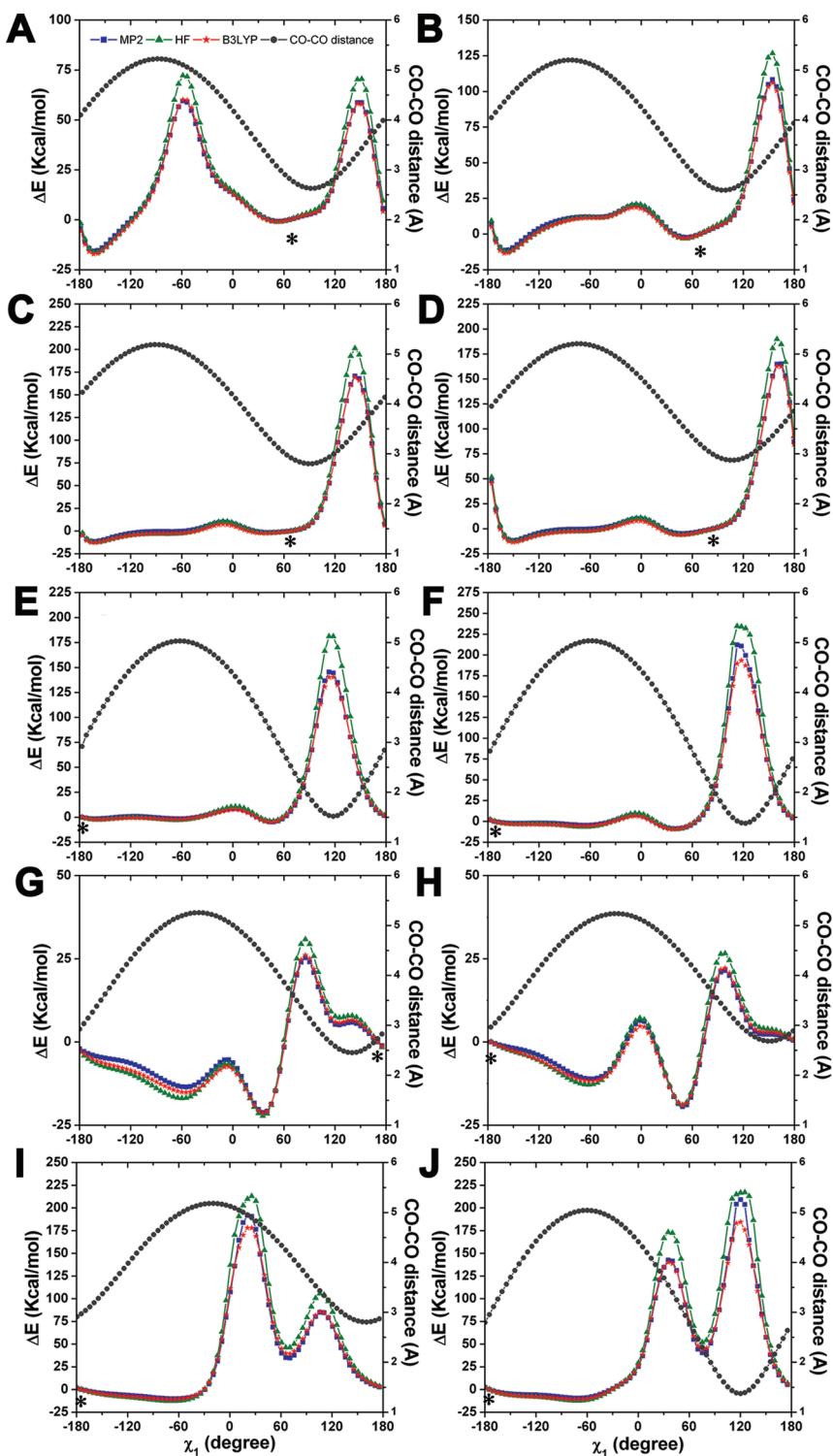
<sup>a</sup> Residue numbers of neighboring residues and water molecules with polar contacts are as given in the PDB file. Polar contacts are determined as described in the Methods section. Residues involved in salt-bridge interactions with the self-contacting Asp residue are shown in bold. The Asp residue responsible for unfavorable interaction energy in 1TP6 is shown in bold and underlined. <sup>b</sup> Interaction energies are calculated as described in the Methods section between two groups A and B. Group A has 24 atoms comprising only the self-contacting Asp residue, and the rest forms group B. The number of atoms includes atoms from both groups A and B. For example, in the case of 1AMF, the number of atoms considered for interaction energy analysis is  $24 + 22 = 46$ . <sup>c</sup> Interaction energy calculated with the CP method at the B3LYP/6-31++G\*\* level.

located in the extended, right-handed  $\alpha$ -helical, and left-handed  $\alpha$ -helical regions (Figure S3 in the Supporting Information) and this is confirmed by frequency calculations. When compared with  $E_{\text{x-ray}}$  structures, relative energies of optimized conformations are more favorable by  $-25$  (1BS0) to  $-43$  (1RW1) kcal/mol (Table 3) and this is mainly due to the absence of any close contacts and favorable hydrogen bond interactions in the optimized structures. Hence, it is clear that self-contacting Asp residues are energetically closer to the conformations belonging to the stationary points and they are well separated from  $E_{\text{max}}$  conformations by at least 100 kcal/mol in the majority of the cases. This explains why such residues are accommodated in protein structures in spite of the close intraresidue contacts between their carbonyl oxygen atoms. The small energy penalty due to this contact can be easily compensated with more favorable interactions with surrounding residues. Hence, in order to confirm this, we carried out quantum chemical calculations between self-contacting Asp residues and the surrounding interacting residues and water molecules.

**Interaction Energies.** Potential energy scan calculations show that self-contacting Asp residues are energetically closer to the conformation corresponding to  $E_{\text{min}}$  in each cluster. However, in the majority of the cases studied, their energies are still 10–15 kcal/mol higher than the  $E_{\text{min}}$  conformations. This confirms that the close approach of carbonyl oxygen atoms in the self-contacting Asp residue could contribute to the unfavorable nature of this conformation. However, when such self-contacting residues are observed in different protein structures, this unfavorable contact must be compensated by many favorable interactions from surrounding residues so that their presence within a protein structure will be justified. Such a conclusion can be arrived by considering the surrounding interacting residues and water molecules and calculating the interaction energy between them and the self-contacting Asp residue. We have included all of the residues and water molecules that are considered as interacting with the self-contacting residues using the criterion defined in the Methods section. We have used two different methods (InsightII and GROMACS<sup>37</sup> using the OPLS force field<sup>38</sup>) to build hydrogen atom positions as described in the Methods section. For each of the 10 self-contacting residues, we have considered both the InsightII-generated and GROMACS/OPLS-generated coordinates. The hydrogen atom positions generated by the two methods are almost similar except for water molecules. Differ-

ences are also observed in the protonation state of histidine residues. Interaction energies of the self-contacting Asp residues with the surrounding residues and water molecules were evaluated using the 6-31++G\*\* basis set and three levels of theory (MP2, B3LYP, and HF). We have obtained qualitatively similar results for both sets of coordinates. Hence, only the results obtained for GROMACS/OPLS-generated coordinates will be discussed. Data for InsightII-generated coordinates are available in the Supporting Information (see Tables S3 and S4). As a representative example, the system used to calculate the interaction energy of the self-contacting Asp residue, D129, from epidermolytic toxin A (PDB ID: 1AGJ) is shown in Figure 4.

In 9 out of 10 cases, the counterpoise-corrected interaction energies show that they are favorable for calculations carried out using all three levels of theory. They vary from  $-29$  to  $-136$  kcal/mol for MP2 (see Table S2 in the Supporting Information),  $-22$  to  $-133$  kcal/mol for B3LYP (Table 4), and  $-17$  to  $-135$  kcal/mol for HF (data not shown) calculations. Among the examples which showed favorable interaction energies, 1BS0 and 2BKQ showed, respectively, the least ( $-22.14$  kcal/mol) and the most ( $-133.19$  kcal/mol) favorable interaction energy (Table 4). Both structures in cluster II and one structure each in clusters IV (1DOZ) and V (2BKQ) have interaction energies close to or greater than  $-100$  kcal/mol. A close examination of the interaction pattern of self-contacting residues in these structures reveals the presence of a basic residue as one of the interacting residues, indicating a salt-bridge interaction as the main reason for the highly favorable nature of interaction energy with its magnitude exceeding 100 kcal/mol in some cases. The structures of clusters I and III show interaction energies between  $-20$  and  $-40$  kcal/mol. This is also true for one of the structures from cluster IV (1O9R). In these cases, the interactions are mostly defined by hydrogen bonds formed by the side-chain or main-chain of amino acid residues or water molecules. In the case of 1TP6, the interaction energy is observed to be positive. A closer look at the interaction energy reveals that the side-chain of another Asp residue is approaching the self-contacting Asp residue. This has resulted in unfavorable interaction energy due to the repulsion between the negatively charged side-chains of two Asp residues. In the actual system, it is possible that a water molecule will be bridging these two side-chains.<sup>9</sup> An ultrahigh-resolution X-ray structure is likely to be able to locate this water molecule, and such a finding will resolve this question.

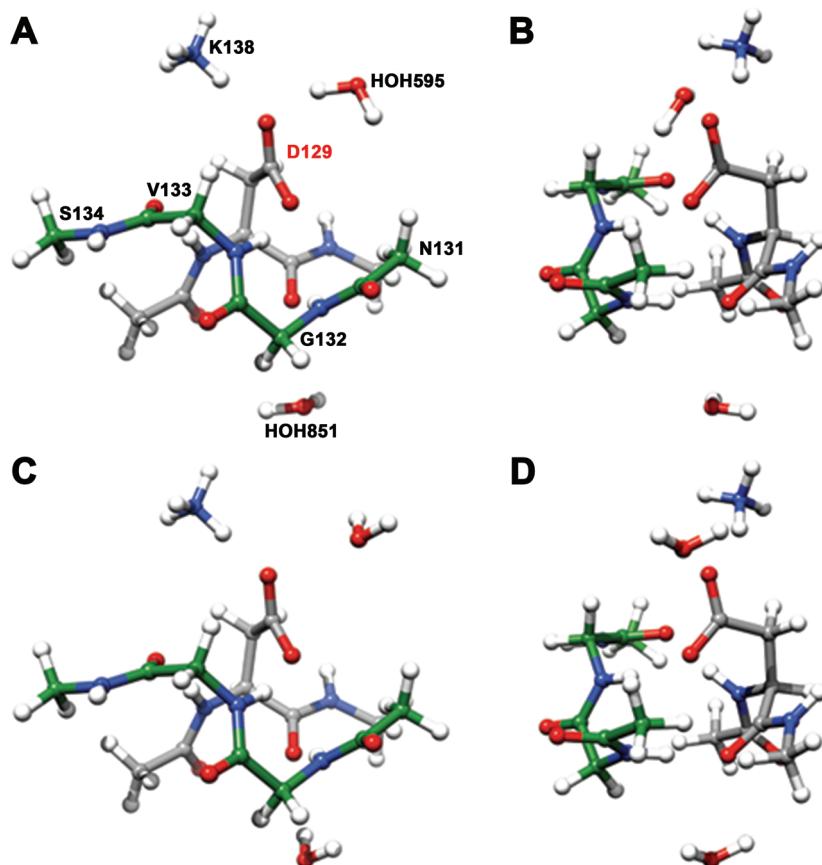


**Figure 3.** Potential energy curves obtained for 10 representative examples belonging to five different clusters are shown. The PDB IDs and the respective clusters are as follows: cluster I - (A) 1AMF, (B) 1RW1; cluster II - (C) 1AGJ, (D) 1TIF; cluster III - (E) 1BS0, (F) 7A3H; cluster IV - (G) 1DOZ, (H) 1O9R; cluster V - (I) 1TP6, (J) 2BKQ. Results of the calculations using the 6-31+G\*\* basis set and all three levels of theory (HF, B3LYP, and MP2) are displayed. Only the dihedral angle  $\chi_1$  ( $N_i-C^{\alpha}_i-C^{\beta}_i-C^{\gamma}_i$ ) was varied in steps of 5°, and all other parameters were fixed. The initial values of all of the parameters in each case correspond to that of the crystal structure (depicted with an asterisk in the plots). The distance between the two carbonyl oxygens of the Asp residue is shown as a function of  $\chi_1$ . The initial coordinates of heavy atoms were taken directly from the crystal structures. The hydrogen atom positions were built as described in the Methods section. Each energy value is relative to that of the crystal structure.

## Discussion

The noncovalent interaction as a result of unusually close contacts between two carbonyl groups observed in small molecules and biological macromolecular structures has attracted the attention of several researchers.<sup>9,10,12,13,19,20,40–42</sup> Both experiments and com-

putational studies have been carried out to find whether the nature of such contacts can be described as attractive. In the present study, we have identified 102 examples of Asp residues from high-resolution structures in which both the backbone and side-chain carbonyl oxygen atoms are within



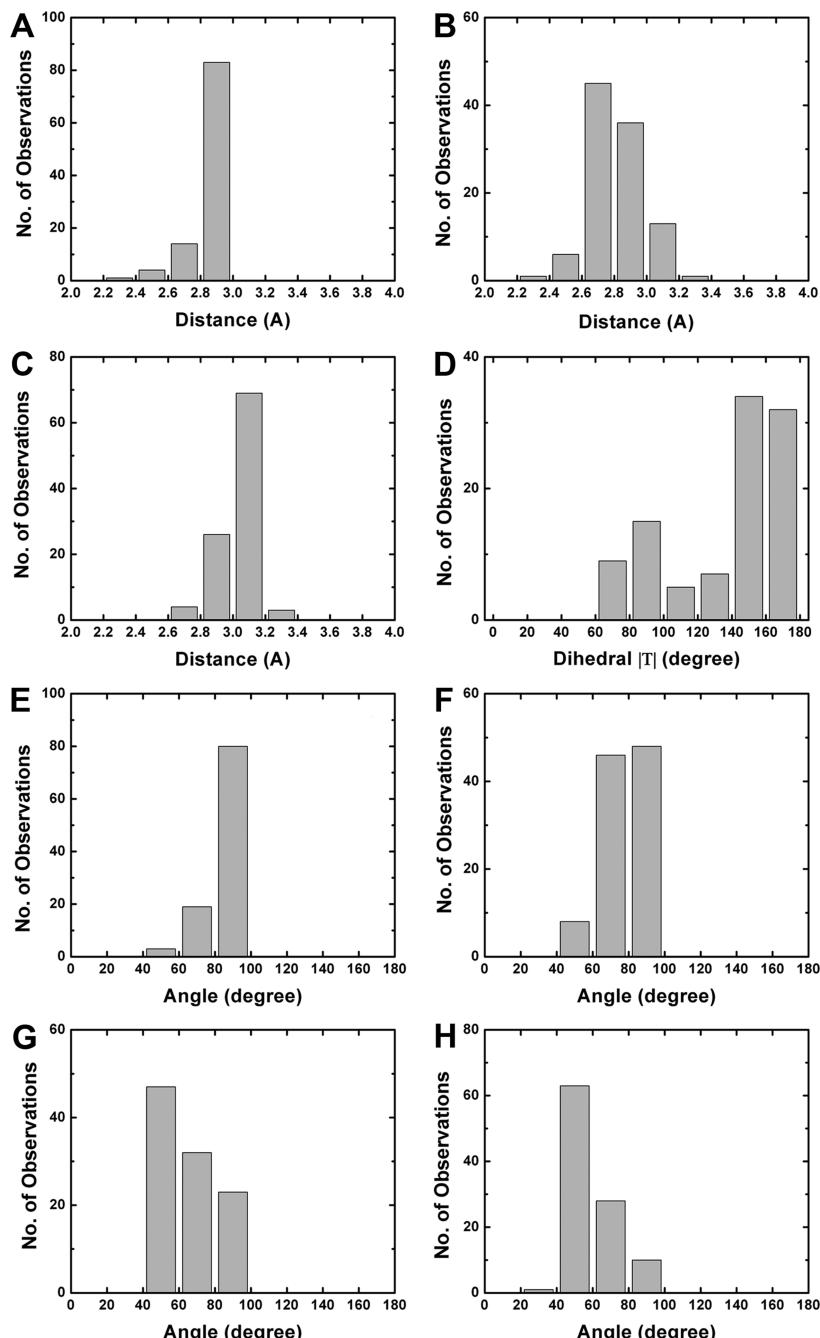
**Figure 4.** Example of a self-contacting Asp residue (labeled in red color in part A) along with the surrounding interacting residues and water molecules used in the interaction energy calculations. This example is taken from the protein epidermolytic toxin A, and the residue numbers correspond to that of PDB ID 1AGJ. Hydrogen atom positions were built either using GROMACS and the OPLS force field (A and B) or using the Builder module of the Insight II software package (C and D). For each method, two different orientations are shown. One can clearly see that the hydrogen positions generated by the two methods for the water molecules (HOH595 and HOH851) are very different. The other hydrogen positions are almost identical in both of the methods.

a distance of 3.0 Å. One can immediately suspect that these close distances could be due to the formation of a hydrogen bond between the backbone carbonyl oxygen and the hydrogen atom of the protonated carboxyl side-chain. In the absence of clearly defined hydrogen positions in the crystal structures, it will be difficult to determine the protonation state of an Asp residue and it also depends upon the local pH that is influenced by the environment. In order to rule out the possibility that the pH due to crystallization conditions could have resulted in a self-contacting Asp residue in the protonated state, we have checked the pH in which the structures are determined. The majority of protein structures are determined at physiological pH (see Figure S4 in the Supporting Information), indicating that the side-chains of self-contacting Asp residues are most likely to exist in the ionized state. Although the surrounding protein environment greatly influences the protonation state of an Asp residue, one can at least conclude that crystallization conditions could not have resulted in a protonated self-contacting Asp residue in the majority of the proteins. Interestingly, we could find only 25 cases of Glu residues in which the backbone and side-chain carbonyl oxygens approach each other within 3.0 Å. It could be that, due to the additional methylene group in Glu, it would be entropically unfavorable to constrain the side-chain and main-chain functional groups by bringing them close to each other.

Studies to understand the nature of carbonyl–carbonyl interactions have produced two different interpretations. By analyzing small molecule crystal structures and ab initio

molecular orbital calculations, Allen et al.<sup>12</sup> defined three types of interaction motifs for carbonyl–carbonyl contacts, namely, perpendicular, antiparallel, and sheared parallel motifs. Their work highlighted the importance of dipolar interactions between the carbonyl groups, and the calculated interaction energy between the carbonyl dipoles was shown to be comparable to medium strength hydrogen bonds. Raines and co-workers synthesized a series of compounds and carried out hybrid density functional theory (DFT) and natural bond orbital (NBO) calculations to investigate the carbonyl–carbonyl interactions that can occur in proteins.<sup>40,41,43</sup> They have suggested that such contacts are due to electronic delocalization in which a lone pair ( $n$ ) of the oxygen atom overlaps with the antibonding orbital ( $\pi^*$ ) of the carbonyl group. This  $n \rightarrow \pi^*$  interaction is believed to occur between the adjacent carbonyl groups in proteins, thus providing conformational stability to different secondary structures, namely,  $\alpha$ -helix and polyproline II helix. One of the two factors discussed above could stabilize the carbonyl–carbonyl self-contacts in Asp residues observed in protein structures. To find out which factor is likely to be responsible, we have examined several geometrical parameters involving the two carbonyl groups of self-contacting Asp residues that characterized the carbonyl–carbonyl interactions as reported by Allen et al.<sup>12</sup> and Raines's group.<sup>43</sup>

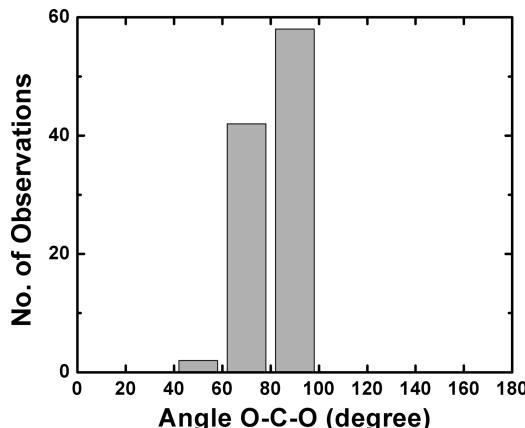
**Possibility of Favorable Dipole Interactions between Contacting Carbonyl Groups.** As per our cutoff criteria, all of the distances between the two carbonyl oxygen atoms are  $\leq 3.0$  Å (Figure 5A). However, it is surprising to note that, in



**Figure 5.** Distribution of several geometrical parameters that are used to characterize the two contacting carbonyl groups of the same Asp residue. These parameters include (A)  $\text{Min}[d(\text{O}_i \cdots \text{O}^{\delta 1}_i), d(\text{O}_i \cdots \text{O}^{\delta 2}_i)]$ ; (B)  $\text{Min}[d(\text{C}'_i \cdots \text{O}^{\delta 1}_i), d(\text{C}'_i \cdots \text{O}^{\delta 2}_i), d(\text{C}'_i \cdots \text{C}'_i)]$ ; (C)  $d(\text{C}'_i \cdots \text{C}'_i)$ ; (D) dihedral angle  $\tau(\text{C}'_i-\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i \cdots \text{C}'_i-\text{O}_i)$  or  $\tau(\text{C}'_i-\text{O}_i \cdots \text{C}'_i-\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i)$ ; (E)  $\theta(\text{C}'_i=\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i \cdots \text{C}'_i)$ ; (F)  $\theta(\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i \cdots \text{C}'_i=\text{O}_i)$ ; (G)  $\theta(\text{C}'_i=\text{O}_i \cdots \text{C}'_i)$ ; and (H)  $\theta(\text{O}_i \cdots \text{C}'_i=\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i)$ .  $d(X \cdots Y)$ : distance between atoms  $X$  and  $Y$ .  $\theta(A-B-C)$ : angle between the atoms  $A$ ,  $B$ , and  $C$ .  $\tau(A-B-C-D)$ : dihedral angle between the four atoms  $A$ ,  $B$ ,  $C$ , and  $D$ . The function “Min” is used to find out the minimum of a given set of distances. The choice of  $\text{O}^{\delta 1}_i$  or  $\text{O}^{\delta 2}_i$  in parts D, E, F, and H depends upon which of the two side-chain carbonyl oxygen atoms results in a minimum distance. The choice of one of the two dihedral angles in part D depends upon whether  $d(\text{O}_i \cdots \text{C}'_i)$  or  $d(\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i \cdots \text{C}'_i)$  is smaller. All distances are in angstroms, and angles and dihedral angles are given in degrees. For atom labels, see Figure 1A.

most of the examples, the distance between the main-chain (or side-chain) carbonyl carbon and the side-chain (or main-chain) carbonyl oxygen atoms is closer than the corresponding distance between the two carbonyl oxygen atoms (Figure 5B). This could be due to a favorable electrostatic interaction between the partial positive charge on the carbonyl carbon and the partial negative charge on the carbonyl oxygen atoms. Other geometrical parameters (Figure 5), including four angles and one dihedral angle, indicate that the contacting carbonyl groups assume an orientation corresponding to a sheared parallel motif. The average values of the four angles are  $84.6^\circ$  [ $\theta(\text{C}'_i=\text{O}_i \cdots \text{C}'_i)$ ] (Figure 5E),  $77.5^\circ$  [ $\theta(\text{O}_i \cdots \text{C}'_i=\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i)$ ] (Figure 5F),  $65.8^\circ$  [ $\theta(\text{C}'_i=\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i \cdots \text{C}'_i)$ ] (Figure 5G), and  $58.7^\circ$  [ $\theta(\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i \cdots \text{C}'_i=\text{O}_i)$ ] (Figure 5H) which are close to the idealized values  $90^\circ$ ,  $90^\circ$ ,  $55^\circ$ , and  $55^\circ$  for sheared parallel motifs.<sup>12</sup>

Similarly, the histogram of the dihedral angle  $\text{C}'_i=\text{O}_i \cdots \text{C}'_i=\text{O}^{\delta 1}_i/\text{O}^{\delta 2}_i$  shows a peak in the range  $160^\circ$ – $180^\circ$  with an average value of  $137.2^\circ$  (Figure 5D). The preference for the sheared parallel motif for the self-contacting Asp residue could also be explained from the fact that this arrangement enables both carbonyl oxygen atoms to be exposed. This facilitates the interactions of both the carbonyl oxygen atoms with other residues from the



**Figure 6.** Distribution of the angle  $\theta_{\text{BD}}$  (in degrees) describing the Bürgi–Dunitz trajectory. For each self-contacting residue, the angle  $\theta(\text{O}_i \cdots \text{C}'_i = \text{O}^{\delta 1}_i / \text{O}^{\delta 2}_i)$  or  $\theta(\text{O}^{\delta 1}_i / \text{O}^{\delta 2}_i \cdots \text{C}'_i = \text{O}_i)$  was calculated depending upon whether the distance  $d(\text{O}_i \cdots \text{C}'_i)$  or  $d(\text{O}^{\delta 1}_i / \text{O}^{\delta 2}_i \cdots \text{C}_i)$  is smaller. The choice of  $\text{O}^{\delta 1}_i$  or  $\text{O}^{\delta 2}_i$  depends upon which of the two side-chain carbonyl oxygen atoms results in a minimum distance. For atom labels, see Figure 1A.

surrounding regions of the protein. In the representative examples studied, this sheared parallel motif conformation of self-contacting Asp residue ( $E_{\text{xray}}$ ) is about 10–15 kcal/mol energetically less favorable than the minimum energy conformation ( $E_{\text{min}}$ ) and this seems to be independent of the levels of theory used in the calculations. In fact, the features of the potential energy curve obtained by using different levels of theory are almost the same and a very high level of correlation was observed between different levels of theory (see Figure S5 in the Supporting Information). The energy difference of 10–15 kcal/mol between the self-contacting Asp residue and the minimum energy conformations can be easily compensated by interactions from surrounding residues. Interaction energy analysis using the B3LYP/6-31+G\*\* level of theory revealed that neighboring residues and water molecules interact with the exposed carbonyl oxygen atoms, resulting in a favorable interaction energy ranging from –22 to –133 kcal/mol. A similar range of interaction energy is observed using the HF and MP2 levels of theory also.

**n → π\*** Interactions between the Contacting Carbonyl Groups. The close contacts between the carbonyl groups could also result in electron delocalization, as suggested by Raines and co-workers.<sup>40,41,43</sup> In such an interaction, the lone pair electrons of the carbonyl oxygen atom ( $\text{O}_i$  or  $\text{O}^{\delta 1}_i / \text{O}^{\delta 2}_i$ ; donor) overlaps with the antibonding orbital of the carbonyl carbon atom ( $\text{C}'_i$  or  $\text{C}'_i$ ; acceptor). This is in tune with the Bürgi–Dunitz trajectory<sup>44–46</sup> in which the carbonyl oxygen of one  $\text{C}=\text{O}$  group (nucleophile) is positioned proximally to the carbonyl carbon of another  $\text{C}=\text{O}$  group (electrophile). The geometric requirement of such an interaction demands the distance between the  $\text{O}_i$  and  $\text{C}'_i$  (or  $\text{O}^{\delta 1}_i / \text{O}^{\delta 2}_i$  and  $\text{C}'_i$ , whichever is the minimum; denoted as  $d_{\text{BD}}$ ) to be less than 3.2 Å and the angle  $\text{O}_i \cdots \text{C}'_i = \text{O}^{\delta 1}_i / \text{O}^{\delta 2}_i$  (or  $\text{O}^{\delta 1}_i / \text{O}^{\delta 2}_i \cdots \text{C}'_i = \text{O}_i$ ; designated as  $\theta_{\text{BD}}$ ) to be between 99 and 119°.<sup>41</sup> We have calculated these parameters for all self-contacting Asp residues and found that the overwhelming majority of the donor (carbonyl oxygen) and acceptor (carbonyl carbon) distances are less than 3.2 Å (Figure 5B). However, the distribution of the angle shows that this parameter is less than 99° for all of the cases and exhibits a peak between 60 and 100° (Figure 6). This is just below the preferred angle referred in the Bürgi–Dunitz trajectory at which the nucleophilic carbonyl oxygen approaches the electrophilic carbonyl carbon.

**TABLE 5: Second-Order Perturbation Energy for  $E_{\text{xray}}$  Showing Hyperconjugation between Donor and Acceptor NBOs**

PDB	donor <sup>a</sup> ( <i>i</i> )		acceptor <sup>b</sup> ( <i>j</i> )		<i>E</i> (2) <sup>c</sup>
	NBO	occupancy	NBO	occupancy	
1AMF	lp( $\text{O}^{\delta 1}_i$ )	1.88205	$\pi^*(\text{C}'_i - \text{O}_i)$	0.28812	0.52
1RW1	lp( $\text{O}^{\delta 1}_i$ )	1.88487	$\pi^*(\text{C}'_i - \text{O}_i)$	0.29087	0.50
1AGJ	lp( $\text{O}^{\delta 1}_i$ )	1.87093	$\pi^*(\text{C}'_i - \text{O}_i)$	0.29301	3.25
1TIF	lp( $\text{O}^{\delta 1}_i$ )	1.87629	$\pi^*(\text{C}'_i - \text{O}_i)$	0.28850	1.41
1BS0	lp( $\text{O}^{\delta 1}_i$ )	1.87793	$\pi^*(\text{C}'_i - \text{O}_i)$	0.30703	2.35
7A3H	lp( $\text{O}^{\delta 1}_i$ )	1.87815	$\pi^*(\text{C}'_i - \text{O}_i)$	0.29565	0.77
1DOZ	lp( $\text{O}^{\delta 1}_i$ )	1.87862	$\pi^*(\text{C}'_i - \text{O}_i)$	0.26758	NA
1O9R	lp( $\text{O}^{\delta 1}_i$ )	1.88234	$\pi^*(\text{C}'_i - \text{O}_i)$	0.26734	NA
1TP6	lp( $\text{O}_i$ )	1.85847	$\pi^*(\text{C}'_i - \text{O}^{\delta 2}_i)$	0.38873	0.61
2BKQ	lp( $\text{O}_i$ )	1.85551	$\pi^*(\text{C}'_i - \text{O}^{\delta 1}_i)$	0.38241	1.15

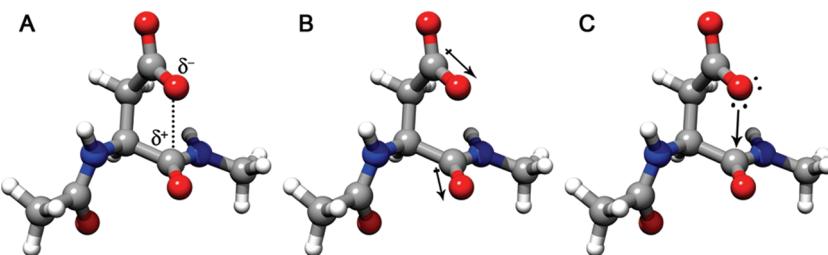
<sup>a</sup> Donor NBO for a lone pair (lp) of electrons of oxygen.

<sup>b</sup> Acceptor NBO corresponding to the antibonding ( $\pi^*$ ) orbital.

<sup>c</sup>  $E(2)$ , second-order perturbation energy (in kcal/mol); NA, not available.

In the literature,  $n \rightarrow \pi^*$  interactions have been reported to occur in hydroxyproline and chloroproline molecules with a  $\tau_{\text{BD}}$  angle less than 99°.<sup>47</sup> Hence, in order to investigate whether such interactions exist between the contacting carbonyl groups, we have carried out NBO calculations on the 10 representative examples of self-contacting Asp residues as described in the Methods section. In 8 out of 10 cases, the side-chain carbonyl oxygen is closer to the backbone carbonyl carbon. In the remaining two examples, the distance between the backbone carbonyl oxygen and the side-chain carbonyl carbon is less than 2.85 Å. Table 5 summarizes the occupancy of oxygen lone pairs, non-Lewis orbitals of  $\text{C}=\text{O}$  groups, and the second-order perturbation energies of  $n \rightarrow \pi^*$  interactions. Our results clearly show the occurrence of electron delocalization between the side-chain (backbone) carbonyl oxygen and the backbone (side-chain) carbonyl carbon atom in 8 out of 10 cases studied. The strength of this interaction can be judged from the energy values which range from 0.5 to 3.25 kcal/mol (Table 5). It must be pointed out that the magnitude of the calculated energies is small and this is similar to the experimentally estimated value for  $n \rightarrow \pi^*$  interactions in certain para-substituted phenyl esters of *N*-formyl-L-proline where electron delocalization is observed between the lone pairs of the amide carbonyl oxygen and the ester group.<sup>42</sup>

**What Stabilizes the Intraresidue Carbonyl–Carbonyl Contacts in Asp Residues? Dipole or  $n \rightarrow \pi^*$  Interactions?** Our geometric characterization indicates that the interacting carbonyl groups of self-contacting Asp residues adopt a sheared parallel motif. Intermolecular perturbation theory calculations clearly showed the significance of dipolar interactions in antiparallel motifs formed by the contacting carbonyl groups in small molecule crystal structures.<sup>12</sup> However, in the present case, the sheared parallel motif is likely to give rise to parallel dipoles and the interactions between parallel dipoles are known to be unfavorable. Hence, we feel that the dipole interactions are likely to play a destabilizing role. Then, the stability of such a motif is likely to be the result of either simple Coulombic attraction between the lone pair electrons of the carbonyl oxygen of one  $\text{C}=\text{O}$  group and the positively polarized carbonyl carbon of another  $\text{C}=\text{O}$  group of the same Asp residue. On the other hand, NBO calculations show that the closely contacting carbonyl groups have  $n \rightarrow \pi^*$  interactions and this hyperconjugative interaction is due to the approach of the lone pair electrons in the carbonyl oxygen to the electrophilic carbonyl carbon. All three scenarios for self-contacting Asp residues are



**Figure 7.** Possible interactions that could take place between the contacting carbonyl groups of a self-contacting Asp residue are depicted: (A) Coulombic, (B) dipole, and (C)  $n \rightarrow \pi^*$  electron delocalization.

illustrated in Figure 7. The fact that the angle parameter of the interacting group ( $O \cdots C=O$ ) falls just below the range (between 99 and 119°) of the Bürgi–Dunitz trajectory indicates that the strength of this electron delocalization interaction is weak and this is also reflected in the second-order perturbation energy. Hence, we conclude that the contacts between the carbonyl groups of the same Asp residue could perhaps be stabilized to some extent by both the simple Coulombic attraction and  $n \rightarrow \pi^*$  interaction due to electronic delocalization. The preference for sheared parallel motif could be due to two reasons. First, it will probably be difficult for both of the  $C=O$  groups of the same Asp residue to assume the orientations of the other two motifs (antiparallel and perpendicular) due to conformational constraints. Second, the sheared parallel motif helps to expose both of the carbonyl oxygen atoms so that they can interact with other residues from the surrounding environment of the protein. Such a motif can help to maximize the polar contacts when they occur in a relatively more hydrophobic environment<sup>21</sup> and hence play a crucial role in the structural stability of proteins.

## Conclusion

We have identified 102 examples of close intraresidue carbonyl–carbonyl contacts in Asp residues from a data set of high-resolution protein structures. This observation gave rise to two important questions: (i) How are such seemingly unfavorable contacts accommodated in protein structures? (ii) What is the chemical nature of interactions that result from such contacts? In order to answer these questions, we carried out ab initio quantum chemical calculations. Our potential energy scan on 10 representative examples from different allowed regions of the Ramachandran map indicate that self-contacting Asp residues are energetically closer to the minimum energy conformations and the energy difference is only 10–15 kcal/mol. In these calculations, only the side-chain dihedral angle  $\chi_1$  was varied and all other parameters corresponded to that of the crystal structures. Interaction energies calculated between the self-contacting Asp residues and the surrounding residues and water molecules clearly show the existence of favorable interactions and in some cases the magnitude of interaction energy exceeds 100 kcal/mol. The results obtained from quantum chemical calculations are independent of the levels of theory used and the programs to construct the hydrogen atom positions. The geometrical parameters derived for the two contacting carbonyl groups correspond to those of the sheared parallel motif. The dipole interactions arising out of such an arrangement are not likely to stabilize them. On the other hand, natural bond orbital calculations show that the two carbonyl groups can interact through  $n \rightarrow \pi^*$  electron delocalization. Coulombic interactions between the partial negative charge of carbonyl oxygen and partial positive charge of carbonyl carbon could also be an important factor as revealed by the geometry

analysis. The preference of the sheared parallel motif for the self-contacting Asp residue can be explained by the possibility that both exposed carbonyl oxygen atoms can participate in interactions with surrounding residues and water molecules.

**Acknowledgment.** We thank Alok Jain, V. Ramanathan, and S. M. Wahidur Rahman for their help in quantum chemical calculations. R.S. is holding Joy Gill Chair Professorship at IIT-Kanpur. We would like to thank all of our lab members for useful discussions.

**Supporting Information Available:** Distributions of side-chain dihedral angles of self-contacting Asp residues, distribution of potential energies corresponding to  $E_{x\text{ray}}$ ,  $E_{\min}$ , and  $E_{\max}$  conformations, molecular plot of fully geometry optimized structures, distribution of pH values that are used in the crystallization conditions of protein structures, correlation between the  $\Delta E$  values (kcal/mol) obtained for three different levels of theory in the potential energy calculations, relative energies for  $E_{x\text{ray}}$ ,  $E_{\min}$ , and  $E_{\max}$  conformations calculated using the MP2 level of theory for the 10 representative examples and interaction energies obtained using different levels of theory in which either InsightII or GROMACS/OPLS programs were used to build hydrogen atom positions. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Jain, A.; Purohit, C. S.; Verma, S.; Sankararamakrishnan, R. *J. Phys. Chem. B* **2007**, *111*, 8680.
- (2) Jain, A.; Ramanathan, V.; Sankararamakrishnan, R. *Protein Sci.* **2009**, *18*, 595.
- (3) Sarkhel, S.; Rich, A.; Egli, M. *J. Am. Chem. Soc.* **2003**, *125*, 8998.
- (4) Egli, M.; Sarkhel, S. *Acc. Chem. Res.* **2007**, *40*, 197.
- (5) Ringer, A. L.; Senenko, A.; Sherrill, C. D. *Protein Sci.* **2007**, *16*, 2216.
- (6) Scheiner, S.; Kar, T.; Pattanayak, J. *J. Am. Chem. Soc.* **2002**, *124*, 13257.
- (7) Imai, Y. N.; Inoue, Y.; Nakanishi, I.; Kitaura, K. *Protein Sci.* **2008**, *17*, 1129.
- (8) Pednekar, D.; Tendulkar, A.; Durani, S. *Proteins: Struct., Funct., Bioinf.* **2009**, *74*, 155.
- (9) Torshin, I. Y.; Harrison, R. W.; Weber, I. T. *Protein Eng.* **2003**, *16*, 201.
- (10) Deane, C. M.; Allen, F. H.; Taylor, R.; Blundell, T. L. *Protein Eng.* **1999**, *12*, 1025.
- (11) Allen, F. H. *Acta Crystallogr.* **2002**, *B58*, 380.
- (12) Allen, F. H.; Baalham, C. A.; Lommerse, J. P. M.; Raithby, P. R. *Acta Crystallogr.* **1998**, *B54*, 320.
- (13) Fischer, F. R.; Wood, P. A.; Allen, F. H.; Diederich, F. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17290.
- (14) Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R. *J. Mol. Biol.* **2001**, *307*, 357.
- (15) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210.
- (16) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *J. Am. Chem. Soc.* **2002**, *124*, 6133.
- (17) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104.

- (18) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. *J. Comput. Chem.* **2009**, *30*, 2187.
- (19) MacCallum, P. H.; Poet, R.; Milner-White, E. J. *J. Mol. Biol.* **1995**, *248*, 361.
- (20) MacCallum, P. H.; Poet, R.; Milner-White, E. J. *J. Mol. Biol.* **1995**, *248*, 374.
- (21) Pal, T. K.; Sankararamakrishnan, R. *J. Mol. Graphics Modell.* **2008**, *27*, 20.
- (22) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283.
- (23) Koo, J. C.; Chass, G. A.; Perczel, A.; Farkas, O.; Varro, A.; Torday, L. L.; Papp, J. G.; Csizmadia, I. G. *Eur. Phys. J. D* **2002**, *20*, 499.
- (24) Koo, J. C. P.; Lam, J. S. W.; Salpietro, S. J.; Chass, G. A.; Enriz, R. D.; Torday, L. L.; Varro, A.; Papp, J. G. *THEOCHEM* **2002**, *619*, 143.
- (25) Koo, J. C. P.; Chass, G. A.; Perczel, A.; Farkas, O.; Torday, L. L.; Varro, A.; Papp, J. G.; Csizmadia, I. G. *J. Phys. Chem. A* **2002**, *106*, 6999.
- (26) Salpietro, S. J.; Perczel, A.; Farkas, O.; Enriz, R. D.; Csizmadia, I. G. *THEOCHEM* **2000**, *497*, 39.
- (27) Sang-aroon, W.; Ruangpornvisuti, V. *THEOCHEM* **2006**, *758*, 181.
- (28) Lovell, S. C.; Davis, I. W.; Arendall III, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 437.
- (29) Hobohm, U.; Sander, C. *Protein Sci.* **1994**, *3*, 522.
- (30) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
- (31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.05; Gaussian, Inc.: Pittsburgh, PA, 2003.
- (32) Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 123.
- (33) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.
- (34) Weinhold, F.; Landis, C. R. *Valency and bonding: A natural bond orbital donor-acceptor perspective*; Cambridge University Press: Cambridge, U.K., 2005.
- (35) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO Version 3.1, TCI*; University of Wisconsin, Madison, 1998.
- (36) Chothia, C. *Nature* **1975**, *254*, 304.
- (37) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.
- (38) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (39) Dunbrack, J. R., L.; Karplus, M. *J. Mol. Biol.* **1993**, *230*, 543.
- (40) Choudhary, A.; Gandla, D.; Krow, G. R.; Raines, R. T. *J. Am. Chem. Soc.* **2009**, *131*, 7244.
- (41) Hinderaker, M. P.; Raines, R. T. *Protein Sci.* **2003**, *12*, 1188.
- (42) Hodges, J. A.; Raines, R. T. *Org. Lett.* **2006**, *8*, 4695.
- (43) DeRider, N. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 2497.
- (44) Burgi, H. B.; Dunitz, J. D.; Shefter, E. *J. Am. Chem. Soc.* **1973**, *95*, 5065.
- (45) Burgi, H. B.; Dunitz, J. D.; Lehn, J. M.; Wipff, G. *Tetrahedron* **1974**, *30*, 1563.
- (46) Burgi, H. B.; Dunitz, J. D.; Shefter, E. *Acta Crystallogr.* **1974**, *B30*, 1517.
- (47) Shoulders, M. D.; Guzei, I. A.; Raines, R. T. *Biopolymers* **2008**, *89*, 443.
- (48) 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.

JP909339R