Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 27 (2008) 20–33
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Self-contacts in Asx and Glx residues of high-resolution protein structures: Role of local environment and tertiary interactions

Tuhin Kumar Pal, Ramasubbu Sankararamakrishnan *

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

Received 17 October 2007; received in revised form 2 February 2008; accepted 5 February 2008
Available online 12 February 2008

Abstract

In protein structures, side-chains of asparagine and aspartic acid (Asx) and glutamine and glutamic acid (Glx) can approach their own backbone nitrogen or carbonyl group. We have systematically analyzed intra-residue contacts in Asx and Glx residues and their secondary structure preferences in two different datasets consisting of 500 and 1506 high-resolution structures. Intra-residue contact in an Asx/Glx residue between the heavy atoms of side-chain and main-chain functional groups of the same residue was investigated irrespective of whether such contacts are due to hydrogen bonding or not. Our search yielded 563 and 1462 cases of self-contacting Asx and Glx residues from the two datasets. Two important observations have been made in this analysis. First, self-contacts involving side-chain oxygen and backbone nitrogen atoms in majority of Asx residues are not due to hydrogen bonds. In the second instance, surprisingly, side-chain and backbone carbonyl oxygens of a significant number of Asx and Glx residues approach each other. For a wide-range of accessible surface areas, self-contacting residues are surrounded by less number of polar groups compared to all other Asx/Glx residues. In buried and partially buried regions, side-chain and main-chain functional groups of these residues together participate in simultaneous interactions with the available polar groups or water molecules. Asx/Glx residues with self-contacts are rarely observed in the middle of an α-helix or a β-strand. Asx/Glx side-chain having contact with its own backbone nitrogen shows different capping preferences compared to those having contact with its backbone oxygen. Examples of proteins with multiple self-contacting Asx/Glx residues are found. We speculate that mutation of a self-contacting residue in the buried or partially buried region of a protein will destabilize the structure. The results of this analysis will help in engineering protein structures and site-directed mutagenesis experiments.
© 2008 Elsevier Inc. All rights reserved.

Keywords: Accessible surface area; Intra-residue hydrogen bond; Local hydrophobicity; Secondary structure preferences; Multiple interactions; Carbonyl contacts

1. Introduction

The amino acid sequence of a polypeptide chain specifies its characteristic three-dimensional structure and determines all its chemical and biological properties. Each residue's conformational preference and its ability to interact with other residues (and the solvent) contribute to the folding and stability of a protein. The conformational role of several individual residues in different secondary structures has been established in many experimental and computational studies [1–13]. Specific interactions involving the side-chains of certain residues seem to influence the protein structure and its stability. Hydrogen-bond interactions between side-chains of Asx/Glx residues with the backbone amine groups of the same residue have been observed in earlier studies [14–16]. Interactions between Asp–Glu pairs of residues involving their carbonyl oxygen atoms have been investigated using a non-redundant set of protein structures [17]. Asn and to a lesser extent Asp have high preference for partially allowed regions of the Ramachandran map and Deane et al. [18] have proposed that this could be due to a stabilizing interaction between the side-chain carbonyl of these residues and the backbone carbonyl of either the same residue in question or the previous one. It is possible for side-chains of some of the residues to approach their own backbone nitrogen or oxygen. This could happen as a result of intra-residue hydrogen bonding or due to some tertiary constraints. In this paper, we have analyzed Asp and Asn (Asx) residues whose side-chain functional groups approach their own backbone CO or NH groups. Similar study was also carried out for Glu and Gln (Glx) residues and the results are compared with that

* Corresponding author. Tel.: +91 512 2594014; fax: +91 512 2594010.
E-mail address: rsankar@iitk.ac.in (R. Sankararamakrishnan).

1093-3263/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2008.02.001

obtained for Asx residues. The following points distinguish our study from those carried out previously. (1) To our knowledge, a systematic study of Asx intra-residue interactions and their preference for specific secondary structure has not been carried out. (2) Here we have not restricted our analysis only to hydrogen-bonded interactions. All Asx and Glx residues are considered if side-chain carbonyl oxygen or amide nitrogen comes in close contact with either the backbone amine nitrogen or backbone carbonyl oxygen *irrespective of whether a hydrogen bond is formed or not*. (3) We have not restricted this analysis to any particular secondary structure or the capping positions. (4) The present study was carried out on two different data sets of protein structures. Our results show that in a significant number of Asx/Glx cases, close contact occurs between the oxygens of two C=O groups from side-chain and the backbone. Self-contacting Asx/Glx residues participate in fewer tertiary polar interactions compared to the global set consisting of all Asx/Glx residues and hence they exist in an environment that is relatively more hydrophobic. Hence, the self-contacts could be due to hydrogen bond or due to maximization of the polar interactions with the available polar groups. Such interactions are likely to be stronger in hydrophobic environment. A large number of Asx/Glx residues with intra-residue contacts is observed in several proteins. If mutated, the polar interactions due to the self-contacting residue in the hydrophobic environment will be disrupted and this could destabilize the protein structure.

2. Materials and methods

2.1. Datasets

We have carried out the analysis on two different datasets of protein structures and the datasets are described in detail below.

2.1.1. Dataset #1

The first dataset (http://kinemage.biochem.duke.edu/databases/top500.php) contained 500 high-resolution protein structures [19]. This dataset was constructed from three resources: database of 240 structures with resolution 1. 7 Å or better [20], structures from PDBSELECT list [21] released in February 2000 with 30% homology cut-off and resolution $\leq$1.8 Å and structures from Protein Data Bank [22] with resolution $\leq$1.5 Å released between February and May 2000. In this dataset, if multiple structures of the same polypeptide chain were present in any PDB file, the first chain was chosen, unless the file header contained information that another was better ordered. If there were multiple non-homologous chains, they were split if each formed a separate compact unit. All the structures in this dataset have clashscore (for atoms with *B*-factor <40) less than 22 per 1000 atoms [23]. Hydrogen atoms have been built using the REDUCE [24] program.

2.1.2. Dataset #2

The second set is the representative list of protein chains as given in the PDBSELECT [21] released in March 2006. To begin with, this dataset contained 3080 polypeptide chains from 25% threshold list. The protein structures in this set have resolution 3.0 Å or better and *R*-factor less than 30%. This dataset, however, contains structures determined using many different experimental techniques including X-ray, NMR, fiber diffraction and cryo-electron microscopy. The selection of PDB structures is an important step for the type of analysis reported in this paper and the quality of structures should be uniformly high. Hence, we eliminated nearly half of the structures that were determined using NMR and other experimental techniques. We also excluded 38 chains for which only Cα coordinates are available. We selected only the X-ray crystallography structures with resolution 2.5 Å or better. At the end, this dataset contained 1556 chains from 1506 structures and these structures were used for our analysis.

2.2. Determining the intra-residue contacts

All Asx and Glx residues from the two datasets were considered if their atoms do not have alternate location. Distances between the side-chain atoms of Asx/Glx residues [Oδ1, Nδ2 (Asn), Oδ1, Oδ2 (Asp), Oε1, Nε2 (Gln) and Oε1, Oε2 (Glu)] and their own backbone atoms N and O were calculated. If any of these distances is $\leq$2.75 Å, then we considered them as residues having intra-residue contacts. We also refer such residues as ''self-contacting'' or ''self-interacting'' residues. This cut-off is stringent since we wanted to pick up only those residues in which side-chain and main-chain atoms are in close contacts. The distance cut-off 2.75 Å was chosen because this is close to the normal limit for the oxygen and nitrogen atoms used in calculating the Ramachandran map [25]. This distance is also the lower limit for the hydrogen-bond contact distance calculated from the potentials of mean force compiled from 833 highly resolved protein structures [26].

2.3. Assignment of Asn and Gln side-chains

The next step essentially investigated the exact nature of intra-residue contacts in Asx and Glx residues. An unambiguous assignment of atoms is required to characterize the self-contacts. The identity of oxygen and nitrogen atoms in the amide groups of Asn and Gln is not clearly resolved in the X-ray structures and it needs to be established for proper investigation of self-contacts. The amide rotamers of Asn and Gln have already been examined and corrected in Dataset #1 (personal communication from Dr. Michael Word and Prof. Jane Richardson). However, the same cannot be said for these residues in Dataset #2. So we used three different methods to determine whether a ''flip'' is needed for a particular Asn or Gln residue that is being considered from Dataset #2. The programs REDUCE [24], NQ-Flipper [26] and HBPLUS [27] were used for this purpose. The program REDUCE adds hydrogens to the PDB structure and uses a numerical score to decide whether to flip the orientation of the amide group [24]. Weichenberger and Sippl have derived knowledge-based self-consistent potential functions to automatically detect and correct unfavorable amide rotamers in Asn or Gln and NQ-Flipper is a web-based service

[28]. In the third method, the program HBPLUS [27] adds polar hydrogens in the structures and also checks whether a flipping is required for the nitrogen and oxygen atoms of Asn and Gln amide groups. We have examined all the self-contacting Asn and Gln residues using the above three methods and using a consensus approach we flipped the amide groups of Asn or Gln residues if at least two methods supported it.

2.4. Definition of hydrogen bonds and short hydrogen bonds

A hydrogen bond in HBPLUS [27,29] is defined using two distances (D···A and H···A) and three angles (D–H···A, D···A–AA and H···A–AA) involving the donor (D), hydrogen (H), acceptor (A) and acceptor-antecedent (AA) atoms. The geometric criteria used to define hydrogen bond in the present study are: (i) D···A and H···A distances must be $\leq$3.5 and 2.5 Å, respectively, and (ii) all the three angles should be $\geq 90^{\circ}$. The criteria applied in HBPLUS are slightly more generous compared to the recent observation in protein structures by Baker and coworkers [30]. It should be noted that the distance between the donor and acceptor atoms in the current study is $\leq$2.75 Å and hence one of the distance criteria is satisfied in all the cases. To check the hydrogen bonding interactions in self-contacting Asx/Glx residues, all the angle and distance criteria have to be verified. Short hydrogen bonds are defined with distance D···A < 2.7 Å and angle D–H···A $\geq 150^{\circ}$ [31].

2.5. Accessible surface area calculations

The program NACCESS [32] was used to calculate the accessible surface area (ASA) of all self-contacting residues. For comparison purpose, we also calculated the ASA values of all Asx/Glx residues in the entire datasets. This program calculates the atomic surface accessibilities defined by rolling a probe of given size around a van der Waals surface and it is an implementation of Lee and Richards method [33]. The probe size was 1.4 Å and the values for van der Waals radii were taken from Chothia [34]. By default, the program ignores heteroatoms, hydrogens and crystal waters in the PDB file records.

2.6. Determination of polar contacts involving Asx/Glx residues

We also investigated the number of residues surrounding the self-contacting residues that make polar vdw contacts with them. Polar contacts were determined by considering only the nitrogen and oxygen atoms. Since electronegativity of sulphur is about the same as carbon, Sγ atom of Cys was not included in this analysis. To define a polar contact, we applied the criteria used by Chothia [35] in the analysis of inter-helical atomic contacts. If the distance between two polar atoms is less than the sum of their vdw radii + 0.6 Å, then the two atoms are said to have polar contact. When polar contacts from surrounding residues were determined for each self-contacting residue, we also considered all the neighboring polypeptide chains present in the crystal, although these polypeptides chains were not part of the dataset. For comparison purpose, polar contacts of all Asx/Glx residues that do not have intra-residue contacts were also determined using the same approach.

2.7. Secondary structure preferences and propensities

Secondary structures of all the proteins were determined using the Kabsch and Sander algorithm [36] as implemented in InsightII package [InsightII (Version 2000). Accelrys, San Diego, CA]. The occurrence of self-contacting Asx/Glx residues in different secondary structures was investigated. Preference for such Asx/Glx residues to occur in helix caps, helix middle, strand caps, strand middle and turn/random coil structures was found out. A minimum helix length of 6 and strand length of 4 were considered for this analysis. Secondary structures of shorter lengths were considered as turns/random coil regions. The helix capping region as defined by Aurora and Rose [37] is used in our analysis. For helices and β-strands, nomenclature for helix/strand and their flanking residues is as follows:

$$\cdots -\mathrm{N''}-\mathrm{N'}-\mathrm{Ncap}-\mathrm{N1}-\mathrm{N2}-\mathrm{N3}-\cdots-\mathrm{C3}-\mathrm{C2}-\mathrm{C1}-\mathrm{Ccap}-\mathrm{C'}-\mathrm{C''}-\cdots$$

Residues from N1 to C1 refer, respectively, to the beginning and end of the helix/strand as defined by the Kabsch and Sander algorithm [36]. Ncap and Ccap are the first and last residues to occur in non-helical conformations in helical secondary structures. Six residues in the N-terminus (N″, N′, Ncap, N1, N2 and N3) and six residues in the C-terminus (C3, C2, C1, Ccap, C′ and C″) were combined to, respectively, define the N-cap and C-cap regions in helices. Similarly, five residues at the N-terminus (N″, N′, Ncap, N1 and N2) and five residues at the C-terminus (C2, C1, Ccap, C′ and C″) were combined to define the strand N-cap and C-cap regions, respectively. The residues that occur between the two capping regions (N4 to C4 for helices and N3 to C3 for strands) were considered to be part of the middle region. The propensity P_i of self-contacting Asx/Glx residues to occur in the specific region of helix/strand is determined as a ratio of two frequencies and is given below:

$$P_i = \frac{f(\mathrm{X}_i^{\mathrm{intra}})}{f(\mathrm{X}_{\mathrm{ssec}}^{\mathrm{intra}})} \quad (1)$$

$$f(\mathrm{X}_i^{\mathrm{intra}}) = \frac{N(\mathrm{X}_i^{\mathrm{intra}})}{N(\mathrm{X}_i)} \quad (2)$$

$$f(\mathrm{X}_{\mathrm{ssec}}^{\mathrm{intra}}) = \frac{N(\mathrm{X}_{\mathrm{ssec}}^{\mathrm{intra}})}{N(\mathrm{X}_{\mathrm{ssec}})} \quad (3)$$

where $N(\mathrm{X}_{\mathrm{ssec}})$ is the number of times the amino acid X occurs in a particular secondary structure and $N(\mathrm{X}_{\mathrm{ssec}}^{\mathrm{intra}})$ represents the number of times the amino acid X occurs in the same secondary structure with intra-residue contacts. ''ssec'' includes all three regions of a secondary structure (Ncap region, helix/strand proper and Ccap region) that is being considered. $N(\mathrm{X}_i)$ is the number of times the amino acid X occurs in the region i of a particular secondary structure ''ssec''. The symbol $N(\mathrm{X}_i^{\mathrm{intra}})$ represents the number of self-contacting Xs at the same

region *i*. Here, the region *i* could be one of the three regions (helix/strand proper, Ncap or Ccap region) of the secondary structure ''ssec''.

3. Results

All the Asx/Glx residues were analyzed in the two datasets of protein structures to find out whether the side-chain oxygen/nitrogen of these residues makes close contact with the backbone nitrogen or carbonyl oxygen of the same residue. In the present context, four types of intra-residue contacts are possible between the side-chain and main-chain atoms of the same residue: side-chain oxygen–main-chain nitrogen [SC(O)···MC(N)], side-chain oxygen–main-chain oxygen [SC(O)···MC(O)], side-chain nitrogen–main-chain nitrogen [SC(N)···MC(N)] and side-chain nitrogen–main-chain oxygen [SC(N)···MC(O)]. The latter two contacts are possible only for Asn and Gln residues.

3.1. Self-contacts in Asx/Glx residues

All the 500 high-resolution crystal structures in Dataset #1 were analyzed and the result of this analysis is given in Table 1. We found 563 cases (2.6% of total Asx and Glx residues) in which the side-chains of Asx/Glx residues are in close contact with their own backbone oxygen or nitrogen atoms. In Dataset #2, we found 1462 examples of self-contacting Asx/Glx residues from 792 proteins (Table 1) and this is about 2.8% of the total Asx/Glx residues present in the dataset. The fraction of Asx/Glx examples with self-contacts found in this dataset is almost similar to that found in the Dataset #1 (2.8% vs. 2.6%). Histograms of the contact distances (Fig. 1) for both datasets show that 90 and 84% of the distances is between 2.55 and 2.75 Å in Datasets #1 and #2, respectively. About 20% of the distances in Datasets #1 and #2 falls below the extreme limit (2.6 Å) for the O···O, O···N and N···N atom pairs [25] indicating that there could be steric clash in such intra-residue contacts.

3.2. Amide rotamers in Asn/Gln residues of Dataset #2

Self-contacts in Asx/Glx residues could be due to intra-residue hydrogen bonds or electrostatic interactions. Such contacts could also arise due to some tertiary constraints in protein structures. The nature of self-contacts needs to be characterized to find out the factors that bring the side-chain and backbone functional groups close together. This requires unambiguous identification of atoms in the crystal structures for the residues being considered. For the Asp and Glu residues, there will be no ambiguity in the assignment of side-chain atoms (both are carbonyl oxygen atoms). For the amide side-chains of Asn and Gln, distinguishing the N and O atoms of the side-chain amide from the electron density map is not possible except at extremely high resolution. An unambiguous assignment of these atoms is required to characterize the interactions arising from the intra-residue contacts. In a recent analysis of more than one million side-chain Asn and Gln amides from 54,189 polypeptide chains derived from 22,574 protein structures, it was found that even in structures with resolution better than 1 Å and NMR structures, the average error rate in amide rotamers exceeded above 10 and 20%, respectively [26]. So in order to identify the exact nature of intra-residue interactions, it is important to (i) generate all the hydrogen atom positions in the protein structures and (ii) correctly identify the amide rotamers for Asn and Gln residues. Examination of PDB records from Dataset #1 indicates that Asn/Gln flips have been corrected in these structures after adding hydrogen atoms [19,24] (Dr. Michael Word and Prof. Jane Richardson, personal communication). However, the contacts formed by the side-chain Asn and Gln residues in Dataset #2 could not be ambiguously defined. The self-contacting Asn and Gln residues from the Dataset #2 were evaluated using REDUCE [24], NQ-Flipper [26] and HBPLUS [27] programs (see Section 2). Among 395 self-contacting Asn and Gln residues found in Dataset #2 (Table 1), 281 residues did not require correction of their amide rotamers. In 114 residues, a flip in the orientation of the amide group was made and 96 of them were recommended by both REDUCE and NQ-Flipper. Analysis of the self-contacting Asn and Gln residues was carried out after the flip was made in 114 of the 395 residues.

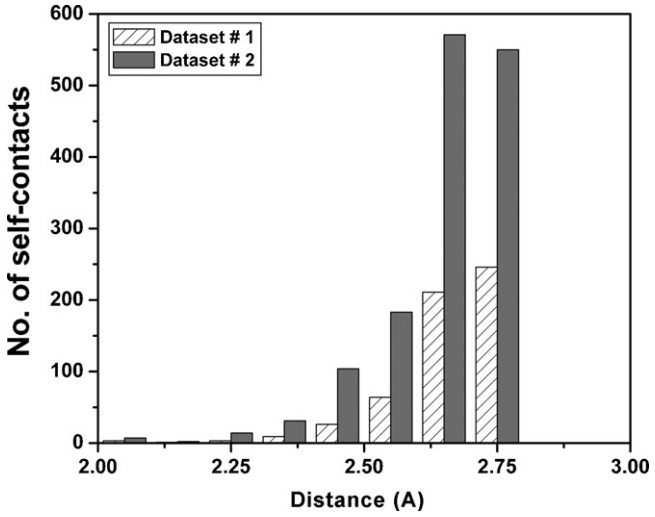

Fig. 1. Histograms showing the distribution of Asx/Glx contact distances between the SC(O)/SC(N) and MC(O)/MC(N) atoms for Datasets #1 and #2. All distances less than or equal to 2.0 Å were combined together and are shown in the same bin.

Table 1
SC(O)/SC(N)···MC(O)/MC(N) intra-residue contacts in Asx/Glx residues analyzed in the structures of Datasets #1 and #2

Intra-residue interactions from Asx/Glx residues	Number of examples	
	Dataset#1	Dataset #2
Asp/Glu Oδ1/Oδ2/Oε1/Oε2···O	30	146
Asp/Glu Oδ1/Oδ2/Oε1/Oε2···N	368	921
Asn/Gln Oδ1/Oε1···O	22	72
Asn/Gln Oδ1/Oε1···N	133	281
Asn/Gln Nδ2/Nε2···O	7	31
Asn/Gln Nδ2/Nε2···N	3	11
Total Asx/Glx residues with i → i contacts	563	1462

3.3. Nature of self-contacts in Asx/Glx residues

Hydrogen atoms are already added in the PDB structures of Dataset #1. Hence characterization of self-contacts in Dataset #1 was straightforward. In structures of Dataset #2, polar hydrogen atoms were added using the program HBPLUS [27,29]. The same program was also used to find out whether residues in Dataset #2 observed with self-contacts are actually due to intra-residue ''hydrogen-bond'' interactions. In Asx/Glx residues with intra-residue contacts, we also checked whether the side-chain of these residues participate in hydrogen-bond interactions with rest of the protein or water molecules or with heteroatoms. Similarly, we have investigated whether the backbone atom (amine nitrogen or carbonyl oxygen) involved in intra-residue contact can take part in hydrogen bonding activity with other regions of protein or solvent molecules. In the following sections, we discuss the predominantly observed contacts within the same residues in both datasets and their possible intra- and inter-residue interactions. Table 2 gives the list of self-contacting Asx/Glx residues from Dataset #2 that forms intra-residue hydrogen bonding and/or hydrogen bonds with rest of the protein or water molecules or heteroatoms.

3.3.1. SC(O)···MC(N) contacts

About 89% of the observed intra-residue contacts in Asx and Glx residues in Dataset #1 are of the type SC(O)···MC(N) (Table 1). Such contacts could possibly be due to the hydrogen-bond interactions involving the backbone amine NH and the side-chain carbonyl C=O groups. According to the criteria used by HBPLUS, the SC(O)···MC(N) self-contacts of two-thirds of Asx residues are not hydrogen bonds. In contrast, ∼90% of the Glx residues with this kind of contacts can be considered as hydrogen bonds in Dataset #1. Among the hydrogen bonds found in Glx residues, significant number of Glu (35 out of 116) and Gln (10 out of 40) actually form ''short hydrogen bonds'' [31] between the backbone NH and their own side-chain carbonyl oxygen atoms. Such short hydrogen bonds could be due to stronger electrostatic interactions between the groups involved. Surprisingly, intra-residue hydrogen bonds of this class observed in almost all Asp and Asn residues do not satisfy the criterion for short hydrogen bonds.

The fraction of SC(O)···MC(N) self-contacts in Asx/Glx residues of Dataset #2 (83%) is slightly less than that of Dataset #1 (89%) (Table 1). The side-chain dihedral angles were analyzed in these residues and it is found that in both Asx and Glx residues $\pm 60^{\circ}$ is preferred for $\chi 1$ and in Glx residues $\chi 2$ also prefers one of the gauche conformations (Supplemental Figure S1). The angle and the distance criteria used in HBPLUS that define a hydrogen bond are satisfied only in 532 out of 921 Asp/Glu and 125 out of 281 Asn/Gln residues with SC(O)···MC(N) self-contacts (Table 2). As observed in Dataset #1, large number of Glu and Gln (>80%) from this group form intra-residue hydrogen bonds which is much higher than that of Asn (∼25%) or Asp (∼40%) with similar self-contacts. Majority of Asx residues with SC(O)···MC(N) self-contacts do not satisfy all the angle criteria for a hydrogen bond as defined by HBPLUS (Supplemental Figure S2). At least one of the three angles involving donor, hydrogen, acceptor and acceptor-antecedent atoms is less than 90° in these cases. In Glx residues, the longer side-chain with an additional methylene group might help orienting the side-chain acceptor atom for optimal intra-residue hydrogen-bond interactions.

An example of SC(O)···MC(N) self-contact in an Asp residue that is not a hydrogen bond is shown in Fig. 2A. In the structure of human kappa class glutathione transferase (PDB ID: 1YZX; resolution: 1.93 Å) the residue Asp201 from chain A has SC(O)···MC(N) self-contact (Fig. 2A) and this occurs in the interface of this dimeric enzyme. This contact is not due to the formation of a hydrogen bond according to HBPLUS criteria. However, both SC(O) and MC(N) groups together participate in multiple interactions with water molecules, heteroatom and residues from the other monomer.

3.3.2. SC(O)···MC(O) contacts

In Dataset #1, a small but a significant number of examples (9.3%) are found in which side-chain oxygen atom of an Asx or Glx residue is in close contact with the main-chain oxygen of the same residue (Table 1). We have observed 52 examples in

Table 2
Hydrogen-bond interactions observed in self-contacting Asx/Glx residues of Dataset #2

Type of intra-residue contact	Only HB_{self}[a]	HB_{self}[a] + $HB_{tertiary}$[b]	Only $HB_{tertiary}$[b]	No HB[c]	Total
Asp H–N···Oδ1/Oδ2	34	195	260	84	573
Glu H–N···Oε1/Oε2	68	235	25	20	348
Asn H–N···Oδ1	6	40	105	36	187
Gln H–N···Oε1	20	59	10	5	94
Asp C=O···Oδ1/Oδ2	–	–	87	27	114
Glu C=O···Oε1/Oε2	–	–	24	8	32
Asn C=O···Oδ1	–	–	45	16	61
Gln C=O···Oε1	–	–	8	3	11
Asn H–N···Nδ2	–	–	8	2	10
Gln H–N···Nε2	–	–	1	–	1
Asn C=O···Nδ2	4	11	7	3	25
Gln C=O···Nε2	2	2	1	1	6

[a] HB_{self} indicates that the intra-residue contacts between the side-chain and main-chain of Asx and Glx residues can be described as hydrogen bonds.
[b] $HB_{tertiary}$ indicates that the side-chain of self-contacting Asx/Glx residues and/or the backbone atom involved in intra-residue contact take part in hydrogen-bond interactions with the rest of the protein or with water molecules/heteroatoms.
[c] No HB implies that neither HB_{self} nor $HB_{tertiary}$ could be detected for the self-contacting residues. Such residues are likely to interact with the solvent molecules.

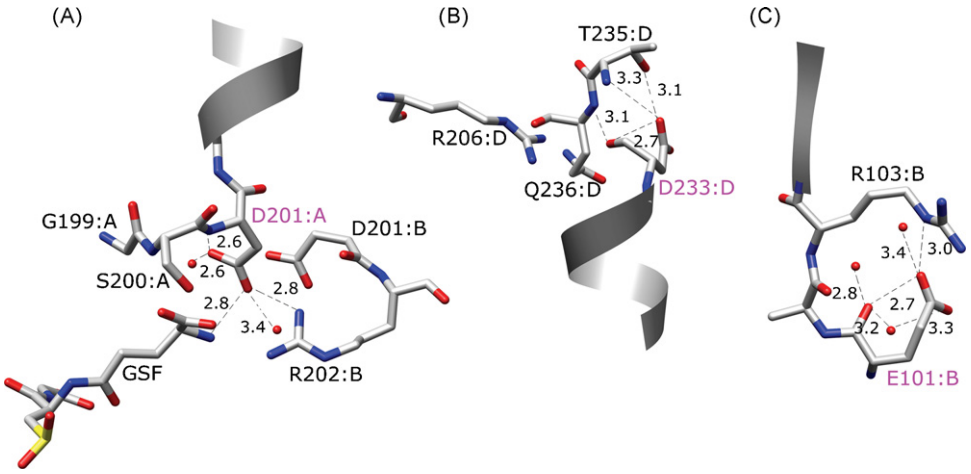

Fig. 2. (A) Asp201 from chain A of human kappa class glutathione transferase (PDB ID: 1YZX; resolution 1.93 Å) has self-contact of the type SC(O)···MC(N). This contact is not due to intra-residue hydrogen bond according to the HBPLUS criteria and the residue occurs at the N′ position of an α-helix (shown as ribbon). Both the main-chain nitrogen and side-chain oxygen atoms take part in multiple interactions. (B) Asp233 (D-chain) from the DNA-binding domain of KorB protein (PDB ID: 1R71; resolution 2.2 Å). This residue occurs at the C-cap position of an α-helix, shown as ribbon. The self-contacting carbonyl groups interact with residues at C″ and C‴ positions. An additional interaction is also seen with the side-chain of Arg206 from a neighboring loop. (C) Glu101 residue from B chain of (+)-epi-biotin bound to streptavidin (PDB ID: 2F01; resolution 0.85 Å) occurs at the N′ position of a β-strand (shown as ribbon). Self-contacting backbone and side-chain carbonyl oxygen atoms interact with water molecules. Glu side-chain also forms a hydrogen bond with Arg103 at the N1 position of the strand. In all the figures, only heavy atoms are shown and nitrogen and oxygen atoms are displayed in blue and red, respectively. Water molecules are shown as red spheres. Only part of the secondary structures is displayed. The self-contacting residue and other interacting residues are labeled as follows—one-letter code residue number: chain ID. Distances within 4.0 Å from the self-contacting backbone heavy atom and the side-chain functional groups are indicated in dotted lines. All molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco [43].

this category. At the outset, it appears that close contact of carbonyl oxygen atoms do not seem to be electrostatically favorable. However, a more detailed analysis of protein structures indicates that the two carbonyl groups are surrounded by water molecules or residues from other parts of the proteins. Thus the unfavorable contact between the two carbonyl oxygen atoms is offset by the favorable interactions of residues surrounding these two functional groups. In Table 3, all the Asp and Glu residues from Dataset #1 having SC(O)···MC(O) self-contacts and their tertiary interactions with other residues and/or water molecules are shown. In 23 out of 30 cases, the Asp/Glu residues participate in hydrogen bonds with other regions of the protein or water molecules. In many cases, we observe that both the carbonyl groups are part of a network of interactions.

Structures from Dataset #2 show a slightly higher percentage of Asx/Glx residues (15%) with intra-residue contacts of SC(O)···MC(O) type (Table 1). When such contacts occur, side-chain dihedral angle $\chi 1$ of Asx and Glx residues adopts either a *trans* or a *gauche*+ conformation (Supplemental Figure S1). Globally, *gauche*− is the preferred conformation for the Asx/Glx residues [38]. The $\chi 2$ of Glx in these circumstances assumes one of the gauche conformations. Our analysis indicates that nearly 75% of Asx/Glx residues with such contacts take part in tertiary hydrogen-bond interactions (Table 2) and in many cases both the contacting carbonyl oxygen atoms seem to take part *together* in additional hydrogen bonding or other tertiary interactions. As in Dataset #1, a closer look at the Asp/Glu residues with SC(O)···MC(O) self-contacts indicate that more than half of observed 146 residues have water molecules within 4.0 Å of the side-chain functional group or MC(O). Residues from other parts of the proteins and/or water molecules are involved in hydrogen bonds in nearly 70% of the Asp/Glu residues with SC(O)···MC(O) self-contacts. One such example is shown in the structure of DNA-binding domain of KorB protein (PDB ID: 1R71; resolution 2.2 Å; Fig. 2B). Asp-233 from D chain occurs at the C-terminal end of an α-helix and forms an intra-residue contact [d(Oδ1···O) = 2.74 Å]. Structural analysis indicates that both the carbonyl oxygens are involved in multiple interactions simultaneously with the backbone and side-chain atoms of adjacent residues. An additional interaction is also observed with Arg206 from an adjacent loop. In these cases, it is also possible that the tertiary interaction of one carbonyl oxygen and the environment can force the other carbonyl oxygen to adopt an orientation that can bring these two oxygens of the same residue together. For example, in the high-resolution structure of (+)-epi-biotin bound to streptavidin (PDB ID: 2F01; resolution: 0.85 Å), the interaction of self-contacting backbone and one of the side-chain carbonyl oxygens to the same water molecule is likely to be the factor that brings the two C=O groups of the same residue within 2.75 Å (Fig. 2C).

3.3.3. SC(N)···MC(N)/MC(O) contacts

We have observed very few examples of Asn/Gln from the structures of Dataset #1 in which the side-chain amide nitrogen closely approaches either the main-chain amine nitrogen or the carbonyl oxygen (Table 1). Only in 10 cases, SC(N)···MC(O)/MC(N) contact occurs and this is less than 2% of all self-contacting Asx/Glx residues in Dataset#1. A similar situation is observed in Dataset #2 also. Only about 3% of Asx/Glx residues with intra-residue contacts are due to the close

Table 3
Asp/Glu residues from Dataset #1 with SC(O)···MC(O) self-contacts and their polar interactions with other residues

PDB ID[a]	Residue	Self-contact[b]	Tertiary contacts[c]
1A3A/D	Asp95	2.67	Ile97
1B0Y/A	Asp52	2.71	HOH130, HOH196, HOH209, HOH277, HOH301
1C3P/A	Asp181	2.55	Glu210, Gly213, HOH200, HOH211
1C3P/A	Asp334	2.4	HOH63
1DBG/A	Asp188	2.47	Ser190
1DOZ/A	Asp42	2.73	Lys44, Asp45, Arg46, HOH185
1FLM/B	Asp24	2.29	HOH1067
1FLM/B	Asp40	2.52	Arg43
1HXN	Asp241	2.62	His243, Gly244, Ala245, Tyr307
1MOQ	Asp451	2.57	Tyr257, Glu455, His586, HOH176, HOH339
1NZY/B	Asp72	2.63	
1PMI	Asp133	2.71	Lys116, Ala120, HOH808, HOH853, HOH884
1QE3/A	Asp345	2.49	Tyr347
1QGQ/A	Asp208	2.69	HOH98, HOH244
1QH4/A	Asp326	2.30	
1QK5/A	Asp43	2.66	Asn28, Ala29
1QSG/G	Asp58	2.59	Val60, HOH656
1SLU/B	Asp178	2.68	Met180, Lys230, Asn233
1TTB/A	Asp39	2.56	HOH549
2TPS/A	Asp64	2.47	Leu66, HOH3060
3SEB	Asp9	2.42	Lys7, Leu11, HOH1115
1BYI	Glu208	2.69	Pro210
1C24/A	Glu12	2.67	Met14, Arg15, Val16
1FXD	Glu27	2.70	Asp29
1PMI	Glu226	2.69	Thr228, Asp229, Arg230
1QJD/A	Glu481	2.60	Pro483, HOH1118, HOH1128, HOH1230
1RHS	Glu148	2.68	Arg29, Gly88, Ala150, HOH325, HOH437, HOH491, HOH501, HOH691
1TTB/A	Glu7	2.72	HOH298, HOH596
1YGE	Glu236	2.65	Lys260, Gln544, HOH1464
2CUA/A	Glu51	2.43	Val132

[a] If chain ID is available, they are also given along with the PDB ID. For example, 1A3A/D indicates that '1A3A' is PDB ID and 'D' is the chain ID.
[b] The minimum distance between the side-chain carbonyl and backbone carbonyl oxygen atoms.
[c] The preceding and succeeding residues of self-contacting Asp/Glu are not considered for polar tertiary contacts. If the side-chain/main-chain oxygen or nitrogen atoms are within 4.0 Å from the self-contacting residues, then they are considered to have tertiary interactions.

approach of SC(N) and MC(O)/MC(N) atoms (Table 1). It may be that in some cases, all the three methods (REDUCE, NQ-Flipper and HBPLUS) used in this study have failed to correctly orient the amide groups. It is also possible that SC(N) prefers to be exposed to the solvent or some tertiary interactions involving SC(N) occur without making a close contact with MC(O)/MC(N).

3.4. Secondary structure preferences

We have examined the secondary structures in which self-contacting Asx/Glx residues occur irrespective of whether the contacts are due to hydrogen bonds or other tertiary interactions. Majority of Asx/Glx amino acids with self-contacts occur in the helix or strand cap regions. In Dataset #1, only 15 examples are found in the middle of an α-helix or a β-strand (see Supplemental Figure S3). Rest of them is observed in the random coil/turn regions. Although it is clear that self-contacting Asx/Glx residues prefer ends of secondary structures or random coil/turn regions, the number of examples found in Dataset #1 is too small to derive any statistically meaningful conclusions. Additionally, preferences for the capping regions would be more meaningful to analyze from a larger sample. Hence we have analyzed the bigger dataset of protein structures with more than 1500 polypeptide chains (Dataset #2). Analysis of this larger dataset is presented in detail below.

While making intra-residue contact, side-chains of Asx and Glx residues could come close to either the backbone amine nitrogen or carbonyl oxygen. So we have divided the self-contacting Asx/Glx residues of Dataset #2 into two groups, namely those that contact the main-chain nitrogen and those that approach carbonyl oxygen of the backbone. Secondary structures in which they occur were investigated and majority of self-contacting Asx/Glx residues occur in the helical or strand regions. About one third of these residues are observed in the turn/random coil region (Supplemental Table S1). Fig. 3 summarizes the preferences of self-contacting residues to occur in different regions of the secondary structure.

We have calculated the propensity values for each region of the helix and strand structures as described in Section 2. The method used to calculate the propensities is similar to the one used by Aurora and Rose [37] and Doig et al. [39]. The propensity values for different helix and strand regions are shown in Fig. 3. Propensity value 1 at a specific region indicates that self-contacting Asx/Glx residues can be as frequently

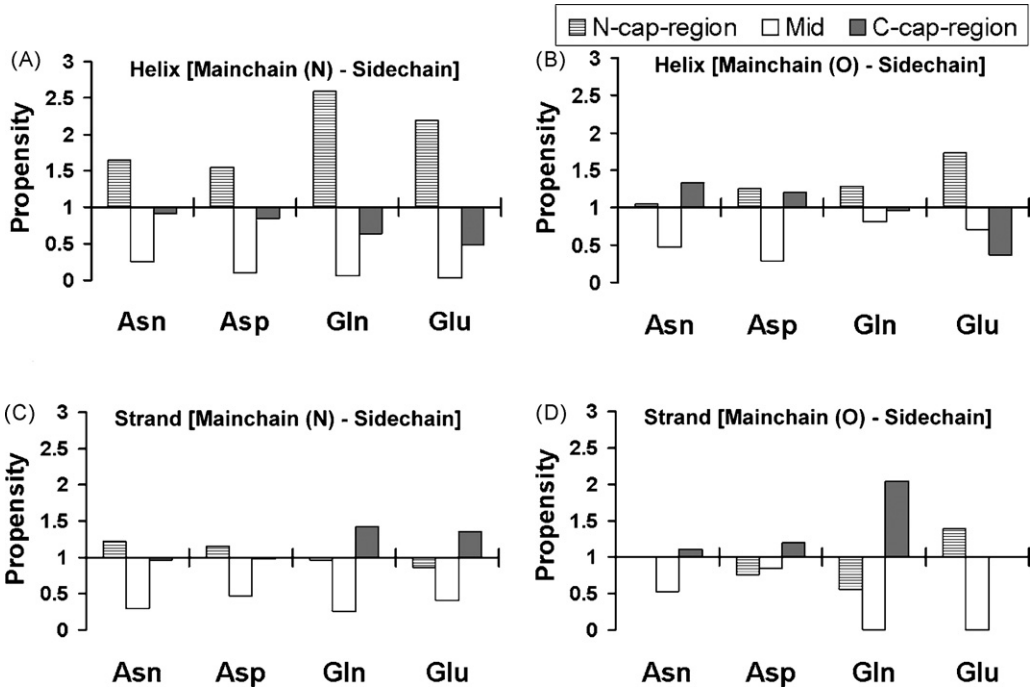

Fig. 3. Propensities of self-contacting Asx and Glx residues in different regions of the secondary structure. Preferences of Asx/Glx residues having self-contacts with MC(N) are shown for (A) α-helical and (C) β-strand regions. (B) and (D) display the propensities for Asx/Glx residues having intra-residue contacts with MC(O) in the α-helical and β-strand regions, respectively. For helices, N-cap and C-cap regions constitute six residues at the N-terminus (N″, N′, Ncap, N1, N2 and N3) and six residues at the C-terminus (C3, C2, C1, Ccap, C′ and C″), respectively. Similarly, five residues at the N-terminus (N″, N′, Ncap, N1 and N2) and five residues at the C-terminus (C2, C1, Ccap, C′ and C″) were considered, respectively, as N-cap and C-cap regions for β-strands. The middle region is defined by all residues that lie between N3 and C3 for helices and N2 and C2 for strands.

found as any other Asx/Glx residues in that region. A value greater than 1 means Asx/Glx residues are more likely to contact their own backbone atom when it occurs at that particular region. As observed in the Dataset #1, Asx/Glx residues with intra-residue contacts are rarely observed in the middle of an α-helix or a β-strand. Fig. 3A shows that N-Cap region of an α-helix is clearly preferred for Asx/Glx residues with self-contacts involving MC(N). Such self-contacts are least likely to occur in the middle or C-cap region of an α-helix. The backbone nitrogen atoms in the N-cap region are not involved in intra-helical hydrogen bonds and hence they are available for interactions. As a result, side-chain carbonyls of self-contacting Asx/Glx residues at these positions approach their own backbone nitrogen atoms to participate in intra-residue hydrogen bonding or both functional groups could together participate in interactions with rest of the protein (Fig. 2A). The N-cap region of a β-strand is only slightly more preferable for Asx residues that have intra-residue contact with the backbone nitrogen (Fig. 3C). Although the middle of the strand is obviously not preferred, Glx residues having self-contacts with MC(N) seem to have slightly higher preference to occur in the C-cap region of a β-strand.

Asx/Glx residues with self-contacts involving their side-chain and their own main-chain C=O atoms are more frequently found in the N-cap region of an α-helix (Fig. 3B). When Asx side-chains approach backbone oxygen of the same residue, positions in the C-terminal side of the helix seem to be more favored. Since backbone C=O in the C-cap positions do not take part in intra-helical hydrogen bonds, the side-chain functional group of Asx can come close together with the MC(O) and participate in multiple tertiary interactions simultaneously (Fig. 2B). As far as the β-strand is concerned, Asx and Gln residues have higher preference to occur in the C-cap region (Fig. 3D). Here again, residues with such self-contacts are involved in multiple tertiary interactions or intra-residue hydrogen bond. Glu having self-contacts with MC(O) also show marginal preference for the N-cap region of the strand (Fig. 2C).

Since the majority of the self-contacting residues occur at the ends of secondary structures or loop regions, we also examined the *B*-factors of these residues. The atomic temperature factor or *B*-factor, determined in an X-ray diffraction study of a protein crystal, can be used to quantify the atomic motions in proteins. If electron density in the loop regions or the ends of secondary structures is less well defined, then the *B*-factor of self-contacting residues occurring in these regions would be very high. Our analysis shows that ∼80 and ∼70% of the self-contacting residues in Datasets #1 and #2 have *B*-factor ≤40 indicating that thermal motion or positional disorder is less in the regions where these residues are observed.

In summary, Asx/Glx residues that have intra-residue contacts with backbone nitrogen clearly have higher preference to occur in the N-cap region of an α-helix and this is also seen to some extent for Asx residues in β-strand structures. C-Cap region of a helix is completely avoided for such residues. For Asx and Gln residues with self-contacts involving their MC(O)

atoms, the C-cap regions are mostly preferred in the helical or strand structures. However, such residues also seem to occur in the N-cap regions of helical structures.

3.5. Factors causing self-contacts in Asx/Glx residues

Although formation of intra-residue hydrogen bonds could be a driving force for the observed intra-residue contacts, the SC(O)···MC(N) self-contacts in two-thirds of Asp/Asn residues do not satisfy the criteria for hydrogen bonds. Moreover, we have examples of Asx/Glx residues in which the carbonyl oxygen atoms from side-chain and main-chain groups approach closely. Hence in addition to hydrogen bonds, there must be other factors that could bring the two functional groups of a same residue close together. To deduce these factors, understanding of the environment of Asx/Glx residues with intra-residue contacts is necessary. We have analyzed the accessible surface area and the amino acids surrounding these self-contacting residues. Results of this analysis are compared with all Asx/Glx residues that do not have intra-residue contacts. Similar results were obtained for both Datasets #1 and #2 and the results are presented for Dataset #2.

Accessible surface areas of self-contacting Asx/Glx residues have been compared with that of all Asx/Glx residues (Table 4). Asx/Glx residues with intra-residue contacts are in general less frequently observed in the interior of the protein (ASA < 40 Å^2). More examples of Asx/Glx residues with self-contacts are found on the surface of the protein. This difference is especially pronounced for those Asx/Glx residues that have intra-residue contacts involving their backbone carbonyl oxygen atoms. In majority of the cases, self-contacts with MC(O) are made by SC(O) atoms and close contacts of two carbonyl oxygen atoms in the more hydrophobic interior of a protein are energetically not favorable. Hence, we see less instances of Asx/Glx residues with SC(O)···MC(O) contacts in the interior of a protein. Under these circumstances, both C=O groups together participate in favorable interactions to offset the repulsive interactions due to the close approach of the carbonyl oxygen atoms (see above). When exposed, favorable interactions of both contacting carbonyl oxygen atoms with water molecules are likely to dominate and hence they are more likely to be observed on the surface of a protein.

Comparison with all Asx/Glx residues indicate that the average number of amino acids that participate in polar contacts is less for self-contacting residues than that observed for the global set (Fig. 4). Except for the highly exposed residues (ASA > 120 Å^2), a two-sample *t*-test shows that the difference in the average number of residues that take part in polar contacts between the self-contacting set and the global set is statistically highly significant ($P < 0.001$). This indicates that the self-contacting residues in general encounter less number of polar groups from other regions of a protein for a wide-range of accessible surface areas. Hence, in the interior and moderately buried regions of a protein, the environment of self-contacting residue is much more hydrophobic than that of all other Asx/Glx residues. In such cases, the exposure of polar side-chain of Asx/Glx residues in a hydrophobic environment is energetically not favorable. Instead, they tend to come close to their backbone functional groups to maximize the favorable electrostatic interactions. This is achieved by forming intra-residue hydrogen bonds or by interactions with other available polar groups in the vicinity. For example, we analyzed the interactions of 18 buried (ASA < 20 Å^2) Asx residues with SC(O)···MC(O) self-contacts (data not shown). Our analysis shows that in majority of the examples both the contacting carbonyl groups together participate in interactions with the same polar atom (backbone nitrogen/side-chain functional group of another residue). Such favorable interactions offset the energetically unfavorable contacts between the carbonyl oxygen atoms. In relatively exposed regions, the self-contacting residues could interact with water molecules even when enough polar groups are not available for interactions. It is also interesting to note that ~8.4% of self-contacting Asx/Glx residues have polar contacts with residues from neighboring polypeptide chains in the crystal (for an example, see Fig. 2A). We speculate that the polar interactions arising from these self-contacting residues in the relatively higher hydrophobic environment will be stronger and hence any disruption of these interactions is likely to make the structure of the protein energetically unstable.

Table 4
Comparison of accessible surface areas of self-contacting and all Asx/Glx residues

Residues	Accessible surface area (Å^2)[a]				Total
	<40	40–80	80–120	>120	
Asx self-contacting with MC(N)	211 (27)	252 (33)	219 (28)	88 (11)	770
Asx self-contacting with MC(O)	38 (19)	54 (27)	69 (35)	37 (19)	198
All self-contacting Asx	249 (26)	306 (32)	288 (30)	125 (13)	968
All Asx[b]	8712 (35)	8504 (34)	5784 (23)	2067 (8)	25,067
Glx self-contacting with MC(N)	93 (21)	108 (24)	151 (34)	88 (20)	440
Glx self-contacting with MC(O)	2 (4)	12 (24)	24 (49)	11 (22)	49
All self-contacting Glx	95 (19)	120 (24)	175 (36)	99 (20)	489
All Glx[b]	5987 (24)	6956 (27)	7903 (31)	4560 (18)	25,406

[a] For each range of accessible surface area, number of Asx/Glx residues are shown. The percentage within each category is given within brackets.
[b] Self-contacting residues are not included.

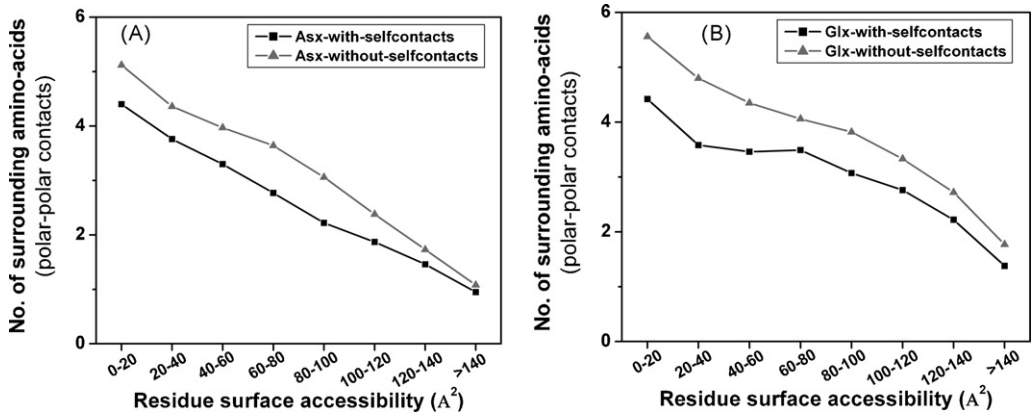

Fig. 4. Residue accessible surface area (Å^2) is plotted against the number of amino acids that make polar contacts with the self-contacting (A) Asx and (B) Glx residues. For comparison, data is also shown for all Asx and Glx residues in the respective plots.

3.6. Proteins with multiple self-contacting Asx/Glx residues

Structures of more than 280 proteins from Dataset #1 have at least one Asx/Glx residue with intra-residue contact. While 142 proteins have each one self-contacting Asx/Glx residue, half that number is found with two such Asx/Glx residues (Fig. 5). Eighteen proteins have five or more Asx/Glx residues with self-contact. The structure of histone deacetylase catalytic core (1C3P) has maximum number of 10 Asx/Glx residues with intra-residue contacts. Salivary alpha-amylase (1SMD), 3-isopropylmalate dehydrogenase from *Salmonella typhimurium* (1CNZ) and benzoylformate decarboxylase (1BFD) have, respectively, nine, eight and seven self-contacting Asx and Glx residues.

In Dataset #2, about 54% of 792 proteins have just one Asx/Glx residue with intra-residue contact (Fig. 5). Thirty-nine proteins in this dataset have five or more self-contacting Asx/Glx residues. Structure of a serine kinase from *Saccharomyces cerevisiae* (1Q97) and crystal structure of a heme oxygenase from cyanobacterium (1WE1) each contain eight self-contacting Asx/Glx residues and this is the highest number observed in any protein from this dataset. The other structures with seven Asx/Glx residues with intra-residue contacts include rhamnose isomerase from *Escherichia coli* (1DE5), RNA-dependent RNA polymerase from double-stranded RNA bacteriophage (1HI8), endoglucanase from termite (1KS8) and calcium-bound protease core of calpain I from rat (1KXR).

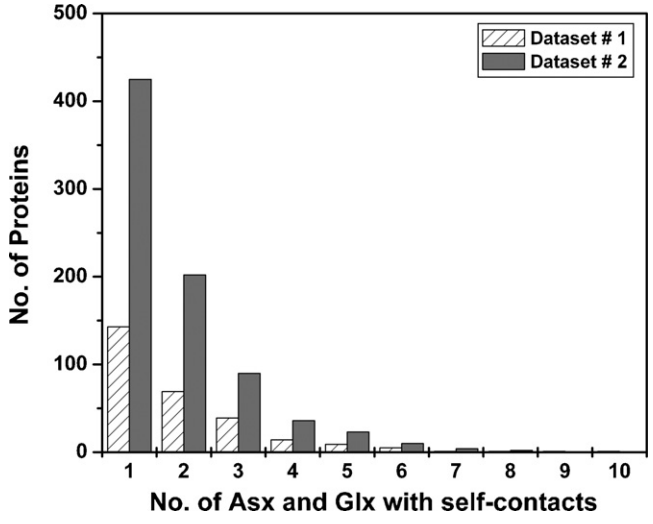

Fig. 5. The number of Asx/Glx residues with intra-residue contacts in different proteins is shown for Datasets #1 and #2.

A representative example of a protein with multiple self-contacting Asx/Glx residues is shown for each dataset (Figs. 6 and 7). The structure of histone deacetylase catalytic core (PDB ID: 1C3P) from Dataset #1 determined at resolution 1.8 Å has six Asp and four Glu residues that have intra-residue contacts (Fig. 6). Eight of them have SC(O)···MC(N) self-contacts and one has SC(O)···MC(O) self-contact. The SC(O) of Asp334 has close contacts with both MC(N) and MC(O) of the same residue. The SC(O)···MC(N) self-contacts of only three residues are due to intra-residue hydrogen bonding. All the self-contacting residues occur in the N- or C-terminal ends of helices or loop regions. Nine of the 10 self-contacting Asp and Glu residues interact with at least one water molecule and/or with residues from other parts of the protein (Fig. 6). The residues Asp173, Asp258 and Glu262 take part in salt-bridge interactions with His132, His170 and Arg356, respectively. The interaction pattern of self-contacting residues was compared with other Asx/Glx residues of the same protein having similar ASA. It clearly indicates that majority of the self-contacting residues has less polar tertiary interactions compared to the 71 other Asx/Glx residues in the same protein that do not have self-contacts. The formation of self-contacts results in optimum favorable interactions. Interactions of Asx/Glx residues with intra-residue contacts are likely to be stronger since they are in a relatively higher hydrophobic environment. Their presence in helix ends and loops indicate that the interactions due to them can help to fix the orientations of the helices or reduce the flexibility of the loop regions. The source of the present protein is a hyperthermophilic bacterium. The stronger interactions involving the self-contacting residues could help to increase the stability of this protein at higher temperatures.

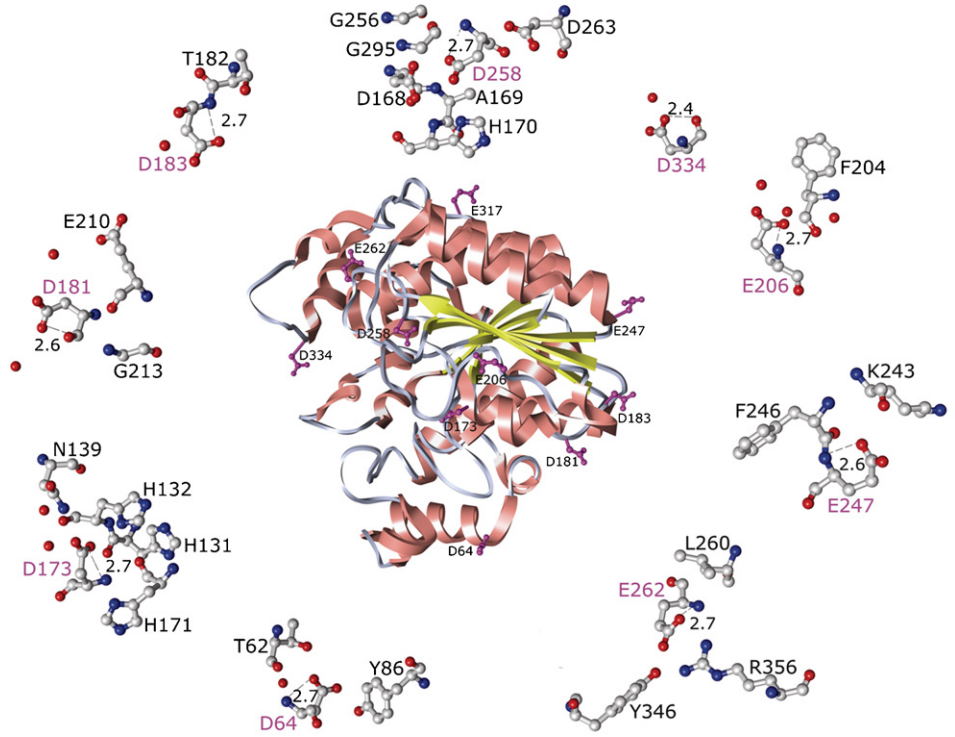

Fig. 6. Structure of histone deacetylase catalytic core derived from a hyperthermophilic bacterium from Dataset #1 (PDB ID: 1C3P; Chain A; resolution: 1.8 Å) has maximum number of 10 self-contacting Asx/Glx residues. A ribbon representation of the entire protein along with the Asx/Glx residues with self-contacts (magenta) is shown. Nine out of 10 self-contacting Asx/Glx residues interact with other residues and/or water molecules. These residues are shown separately with interacting residues and water molecules that are within 4.0 Å in ball-and-stick representation. Distances between the atoms that are responsible for self-contacts are shown. Water molecules are shown as red spheres.

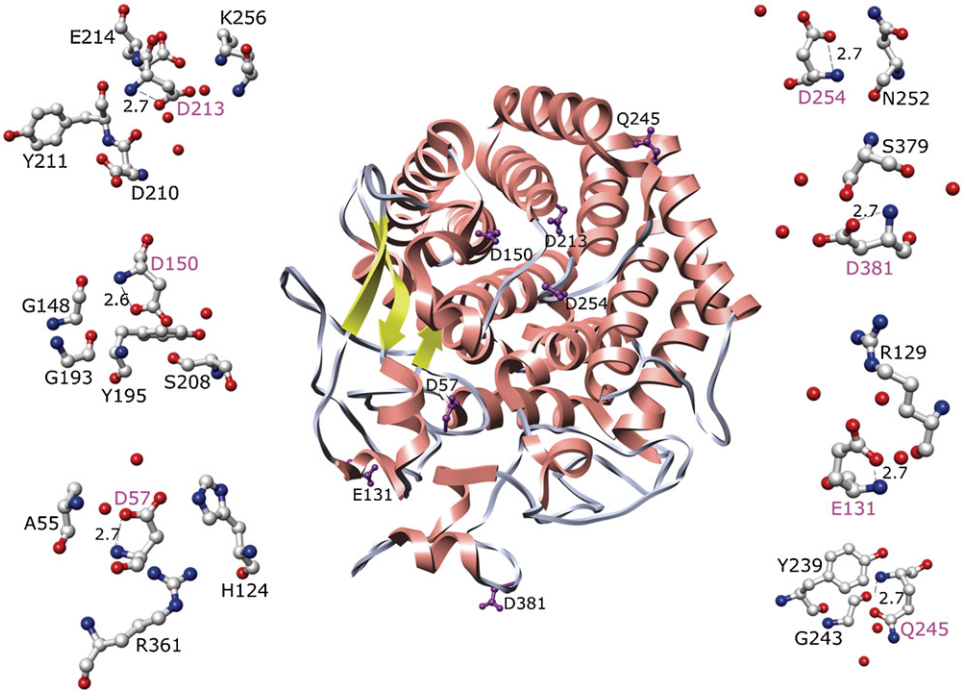

Fig. 7. Structure of endoglucanase from a termite from Dataset #2 (PDB ID: 1KS8; Chain A; resolution 1.4 Å) has seven Asx/Glx self-contacting residues. A ribbon representation of the entire protein along with the Asx/Glx residues with intra-residue contacts (magenta) is shown. All self-contacting Asx/Glx residues interact with residues from other regions of the protein and/or water molecules. Self-contacting Asx/Glx residues are shown separately with interacting residues and water molecules that lie within 4.0 Å in ball-and-stick representation. Distances between the atoms that give rise to self-contacts are shown. Water molecules are shown as red spheres.

The structure of an endoglucanase (PDB ID: 1KS8) has seven self-contacting residues, five of them are Asp, one Gln and one Glu residue (Fig. 7). This is a high-resolution structure (resolution 1.4 Å) from Dataset #2 with large number of self-contacting Asx/Glx residues. All of them occur at the N- or C-terminal ends of helices. All the self-contacts in this protein are of SC(O)···MC(N) type, but only three of them are due to hydrogen bonds. All the self-contacting residues are interacting with water molecules and residues that come from other regions of the protein. Every self-contacting residue forms at least one hydrogen bond with water molecule and at least one with residues that are distant at the primary sequence level. There are 90 other Glx/Asx residues in this protein that do not have intra-residue contacts. Their tertiary interactions when compared with that of self-contacting residues with similar ASA again exhibit the same trend, i.e. the number of hydrophilic interactions with other residues is less in self-contacting residues. The strength of the interactions increases in a relatively higher hydrophobic environment and hence interactions involving such self-contacting residues could be crucial. As in the structure of histone deacetylase catalytic core, these stronger interactions occur at the ends of secondary structures or in the loops and hence could possibly help in the orientation of helices or fixing the loops.

4. Discussion

Amino acid side-chains participate in important short- and long-range interactions and contribute to the stability of protein structures. In this paper, we have identified all the Asx/Glx residues with intra-residue contacts occurring between the side-chain and main-chain functional groups. Hydrogen bonding criteria was *not* used to find the close approach of side-chain and main-chain atoms of Asx/Glx residues and a short distance of 2.75 Å was used in the analysis. Our analysis was carried out on two different datasets, one with 500 X-ray structures with resolution $\leq$ 1.8 Å and the structures in this dataset were determined in the year 2000 or before. The other dataset contained more than 1500 structures with 2.5 Å resolution or better and many of these structures were determined more recently (PDBSELECT list, 25% homology cut-off released in March 2006). It is interesting to observe that the results obtained from these two different datasets are almost identical, validating our strategy of using two different datasets to strengthen our conclusions.

Previous studies have focused on intra-residue hydrogen bonds in Glx/Asx residues between the side-chain and the backbone nitrogen [14–16]. Our results show that significant number of contacts is *not* due to intra-residue hydrogen bonds. For example, majority of Asx residues with SC(O)···MC(N) self-contact do not participate in intra-residue hydrogen bonding interactions. It should be noted that this observation is made in spite of the more generous criteria used by HBPLUS to identify hydrogen bonds compared to the recent observation by Baker and coworkers [30]. We have also observed that the intra-residue contacts can occur between the side-chain and the main-chain carbonyl oxygen atoms and this analysis has not been carried out earlier. At the outset, it appears that the SC(O) and MC(O) cannot come close together, since they cannot have favorable electrostatic interactions. However, we have found many examples in which oxygen atoms of both C=O groups do approach each other with distances between the oxygen atoms $\leq$ 2.75 Å. Closer examination of such cases clearly indicates that both the C=O groups together participate in multiple interactions simultaneously. A similar observation was made in the case of Asx residues with self-contacts of the type SC(O)···MC(N). Even though, these contacts do not satisfy the hydrogen-bond criteria, most of them participate in a network of interactions.

The self-contact between the side-chain and the main-chain C=O groups reported in this study is different from the interactions of C=O groups observed in earlier studies [6,18,40–42]. In these studies, the partial positive charge of carbonyl carbon from one C=O group has a favorable electrostatic interaction with the partial negative charge of carbonyl oxygen from another C=O group. Here the closer approach of two carbonyl oxygens with partial negative charges will result in repulsion between the like charges. However, this is offset by multiple favorable interactions occurring between the C=O groups and water molecule(s)/residues from other parts of the protein.

We have also investigated the probable factors that are responsible for such self-contacts in Asx/Glx residues. These residues are surrounded by less number of polar groups compared to all other Asx/Glx residues. In such cases, they exist in a relatively more hydrophobic environment. Fully exposing the polar side-chains under these circumstances is not likely to be energetically favorable. Instead, if the main-chain and side-chain functional groups are brought together, an intra-residue hydrogen bond can be formed and/or polar interactions can occur with both the contacting groups simultaneously. Such interactions in a more hydrophobic environment will also be stronger especially if the residue is partially or fully buried. Mutation of such self-contacting residue will result in disruption of these interactions and hence it will be detrimental to the protein structure. We have recently carried out molecular dynamics simulations on an enzyme carboxypeptidase (PDB ID: 1ES5) to test this hypothesis (Pal and Sankararamakrishnan, manuscript in preparation). When a single self-contacting Asp residue in this protein is mutated, the structure is completely destabilized and the single residue mutation affected regions that are farther from the site of mutation.

In this study, we have used a stringent cut-off of 2.75 Å to determine the self-contacts in Asx/Glx residues. Using these criteria, we have observed 2.6 and 2.8% of self-contacting Asx/Glx residues in Datasets #1 and #2, respectively. If we relax this distance cut-off to 2.85 Å, then the percentage of self-contacting Asx/Glx residues increases to 5.5 and 5.7% of total Asx/Glx residues, respectively. Thus the present analysis has focused only on a small fraction of unambiguously self-contacting Asx/Glx residues. We have also observed that several proteins contain large number of self-contacting residues. Mutation of self-contacting residues in such proteins would result in the disruption of stronger polar interactions in a

relatively higher hydrophobic environment and this could seriously affect the stable nature of the protein. On the other hand, homologous proteins can be engineered so that these residues can be introduced at appropriate positions. This will result in stronger polar interaction networks and as a result, the stability of a protein can be enhanced.

5. Conclusions

We have systematically identified and analyzed all self-contacting Asx and Glx residues in two different datasets of protein structures. The first dataset contained 500 high-resolution structures and the second set consisted of X-ray structures of more than 1500 polypeptide chains. The distance cut-off for intra-residue contact was chosen close to the normal limit of O–O, O–N and N–N pairs so that only close contacts/interactions can be considered. In the previous studies, hydrogen-bond interactions between the side-chains of Asx/Glx residues with the main-chain atoms of the same residue or nearby residues have been analyzed. In this study, we have first extracted all the close contacts involving the backbone and side-chain functional groups of the same residue. Such self-contacts need not be due to hydrogen-bond interactions. These residues with intra-residue contacts were further examined to find out the nature of interactions they are involved and the secondary structures in which they occur. The main conclusions we have derived from this study are summarized below.

When contacting the backbone nitrogen, majority of Asx side-chains do not form hydrogen bonds. These self-contacting Asx residues together participate in extensive interactions with other side-chains and/or water molecules. On the contrary, Glx residues with similar self-contacts are due to hydrogen bonds and many of them can be described as short hydrogen bonds. It may be that the longer Glx side-chain with one additional methylene group can correctly position its side-chain to form such interactions. Close contacts of main-chain C=O and side-chain C=O have been observed in Asx/Glx residues in both the datasets. At the outset, such interactions seem to be counter-intuitive. But when individual structures were examined, many examples were found in which both carbonyl oxygens together take part in multiple interactions with other regions of the proteins and/or water molecules. Self-contacting Asx/Glx residues are rarely found in the middle of an α-helix or a β-strand. Majority of them prefer to be in capping positions or in random coils/turns. In secondary structures, positional preferences of Asx/Glx side-chains contacting the backbone nitrogen seem to be different from those having self-contacts with main-chain carbonyl oxygen. Self-contacting Asx/Glx residues have fewer polar vdw contacts compared to the global set containing all Asx/Glx residues for a wide-range of accessible surface areas. Hence, the polar interactions involving these residues are likely to be stronger if the residues are not completely exposed, since they exist in a relatively more hydrophobic environment. A large number of proteins with many self-contacting Asx/Glx residues have been observed. We postulate that Asx/Glx residues with self-contacts in such proteins have significant role in the stability of these proteins.

Acknowledgements

We gratefully acknowledge Drs. Roman Laskowski and Simon Hubbard for providing the programs HBPLUS and NACCESS. Drs. Christian Weichenberger and Michael Word are acknowledged for their help and valuable suggestions in using NQ-Flipper and REDUCE programs. We thank Avik Sil, Alok Jain and Tarini Shankar Ghosh for their help during the course of this research project. An anonymous referee is gratefully acknowledged for the comments that greatly improved the manuscript. This research is supported by a research grant awarded to RS by the Ministry of Human Resources Development, Government of India. RS is a Joy Gill Chair Professor at IIT-Kanpur. TKP acknowledges IIT-Kanpur for a fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2008.02.001.

References

[1] L. Serrano, A.R. Fersht, Capping and α-helix stability, Nature 342 (1989) 296–299.
[2] D.A.E. Cochran, S. Penel, A.J. Doig, Effect of the N1 residue on the stability of the α-helix for all 20 amino acids, Protein Sci. 10 (2001) 463–470.
[3] K. Gunasekaran, C. Ramakrishnan, P. Balaram, β-Hairpins in proteins revisited: lessons for de novo design, Protein Eng. 10 (1997) 1131–1141.
[4] T.P. Creamer, G.D. Rose, Side-chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 5937–5941.
[5] R. Aurora, R. Srinivasan, G.D. Rose, Rules for alpha-helix termination by glycine, Science 264 (1994) 1126–1130.
[6] P. Chakrabarti, D. Pal, The interrelationships of side-chain and main-chain conformations in proteins, Prog. Biophys. Mol. Biol. 76 (2001) 1–102.
[7] J.F. Espinosa, F.A. Syud, S.H. Gellman, Analysis of the factors that stabilize a designed two stranded antiparallel β-sheet, Protein Sci. 11 (2002) 1492–1505.
[8] H.M. Fooks, A.C.R. Martin, D.N. Woolfson, R.B. Sessions, E.G. Hutchinson, Amino acid pairing preferences in parallel β-sheets in proteins, J. Mol. Biol. 356 (2006) 32–44.
[9] J.S. Merkel, J.M. Sturtevant, L. Regan, Sidechain interactions in parallel β-sheets: the energetics of cross-strand pairings, Structure 7 (1999) 1333–1343.
[10] M.B. Ulmschneider, M.S.P. Sansom, Amino acid distributions in integral membrane protein structures, Biochim. Biophys. Acta 1512 (2001) 1–14.
[11] W.J. Duddy, J.W.M. Nissink, F.H. Allen, E.J. Milner-White, Mimicry by asx- and ST-turns of the four main types of beta-turn in proteins, Protein Sci. 13 (2004) 3051–3055.
[12] W. Wai-Yan, E.J. Milner-White, A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at α-helical N termini and in other situations, J. Mol. Biol. 286 (1999) 1651–1662.
[13] R. Sankararamakrishnan, S. Vishveshwara, Geometry of proline-containing alpha-helices, Int. J. Pept. Prot. Res. 39 (1992) 356–363.
[14] N. Eswar, C. Ramakrishnan, Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures, Protein Eng. 13 (2000) 227–238.
[15] W. Wai-Yan, E.J. Milner-White, A natural grouping motif with an aspartate or asparagine residue forming two hydrogen bonds to residues ahead in sequence: their occurrence at α-helical N termini and other situations, J. Mol. Biol. 286 (1999) 1633–1649.

[16] S. Penel, E. Hughes, A.J. Doig, Side-chain structures in the first turn of the α-helix, J. Mol. Biol. 287 (1999) 127–143.
[17] I.Y. Torshin, R.W. Harrison, I.T. Weber, Close pairs of carboxylates: a possibility of multicenter hydrogen bonds in proteins, Protein Eng. 16 (2003) 201–207.
[18] C.M. Deane, F.H. Allen, R. Taylor, T.L. Blundell, Carbonyl–carbonyl interactions stabilize the partially allowed Ramachandran conformations of asparagine and aspartic acid, Protein Eng. 12 (1999) 1025–1028.
[19] S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Cα geometry: ϕ, ψ and Cβ deviation, Proteins: Struct. Funct. Genet. 50 (2003) 437–450.
[20] S.C. Lovell, J.M. Word, J.S. Richardson, D.C. Richardson, The penultimate rotamer library, Proteins: Struct. Funct. Genet. 40 (2000) 389–408.
[21] U. Hobohm, C. Sander, Enlarged representative set of protein structures, Protein Sci. 3 (1994) 522–524.
[22] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliand, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.
[23] J.M. Word, S.C. Lovell, T.H. LaBean, H.C. Taylor, M.E. Zalis, B.K. Presley, J.S. Richardson, D.C. Richardson, Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms, J. Mol. Biol. 285 (1999) 1711–1733.
[24] J.M. Word, S.C. Lovell, J.S. Richardson, D.C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation, J. Mol. Biol. 285 (1999) 1735–1747.
[25] G.N. Ramachandran, V. Sasisekharan, Conformation of polypeptides and proteins, Adv. Protein Chem. 23 (1968) 283–438.
[26] C.X. Weichenberger, M.J. Sippl, Self-consistent assignment of asparagine and glutamine amide rotamers in protein crystal structures, Structure 14 (2006) 967–972.
[27] I.K. McDonald, J.M. Thornton, The application of hydrogen bonding analysis in X-ray crystallography to help orient asparagine, glutamine and histidine side chains, Protein Eng. 8 (1994) 217–224.
[28] C.X. Weichenberger, M.J. Sippl, NQ-Flipper: validation and correction of asparagine/glutamine amide rotamers in protein crystal structures, Bioinformatics 22 (2006) 1397–1398.
[29] I.K. McDonald, J.M. Thornton, Satisfying hydrogen bonding potential in proteins, J. Mol. Biol. 238 (1994) 777–793.
[30] A.V. Morozov, T. Kortemme, K. Tsemekhman, D. Baker, Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 6946–6951.
[31] S. Rajagopal, S. Vishveshwara, Short hydrogen bonds in proteins, FEBS J. 272 (2005) 1819–1832.
[32] S.J. Hubbard, J.M. Thronton, 'NACCESS', Computer program, Department of Biochemistry and Molecular Biology, UCL (1993).
[33] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J. Mol. Biol. 55 (1971) 379–400.
[34] C. Chothia, The nature of accessible and buried surfaces in proteins, J. Mol. Biol. 105 (1976) 1–14.
[35] C. Chothia, Structural invariants in protein folding, Nature 254 (1975) 304–308.
[36] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[37] R. Aurora, G.D. Rose, Helix capping, Protein Sci. 7 (1998) 21–38.
[38] J. Janin, S. Wodak, M. Levitt, B. Maigret, Conformation of amino acid side-chains in proteins, J. Mol. Biol. 125 (1978) 357–386.
[39] A.J. Doig, M.W. MacArthur, B.J. Stapley, J.M. Thornton, Structures of N-termini of helices in proteins, Protein Sci. 6 (1997) 147–155.
[40] F.H. Allen, C.A. Baalham, J.P.M. Lommerse, P.R. Raithby, Carbonyl–carbonyl interactions can be competitive with hydrogen bonds, Acta Cryst. B 54 (1998) 320–329.
[41] P.H. Maccallum, R. Poet, E.J. Milner-White, Coulombic interactions between partially charged main-chain atoms not hydrogen-bonded to each other influence the conformations of α-helices and antiparallel β-sheet. A new method for analyzing the forces between hydrogen bonding groups in proteins includes all the Coulombic interactions, J. Mol. Biol. 248 (1995) 361–373.
[42] P.H. Maccallum, R. Poet, E.J. Milner-White, Coulombic attractions between partially charged main-chain atoms stabilise the right-handed twist found in most β-strands, J. Mol. Biol. 248 (1995) 374–384.
[43] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.