Geometry of Interaction of the Histidine Ring with Other Planar and Basic Residues

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · JUNE 2003

Impact Factor: 4.25 · DOI: 10.1021/pr025584d · Source: PubMed

4 AUTHORS, INCLUDING:

Rudra Saha
Adamas University
15 PUBLICATIONS **189** CITATIONS

SEE PROFILE

research articles Journal of proteome research

Geometry of Interaction of the Histidine Ring with Other Planar and Basic Residues

Rajasri Bhattacharyya,[†,‡] Rudra Prasad Saha,[†,‡] Uttamkumar Samanta,[§] and Pinak Chakrabarti*[,†]

Department of Biochemistry and Bioinformatics Centre, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700 054, India

Received November 13, 2002

Among the aromatic residues in protein structures, histidine (His) is unique, as it can exist in the neutral or positively charged form at the physiological pH. As such, it can interact with other aromatic residues as well as form hydrogen bonds with polar and charged (both negative and positive) residues. We have analyzed the geometry of interaction of His residues with nine other planar side chains containing aromatic (residues Phe, Tyr, Trp, and His), carboxylate (Asp and Glu), carboxamide (Asn and Gln) and guanidinium (Arg) groups in 432 polypeptide chains. With the exception of the aspartic (Asp) and glutamic (Glu) acid side-chains, all other residues prefer to interact in a face-to-face or offset-face-stacked orientation with the His ring. Such a geometry is different from the edge-to-face relative orientation normally associated with the aromatic−aromatic interaction. His−His pair prefers to interact in a face-to-face orientation; however, when both the residues bind the same metal ion, the interplanar angle is close to 90°. The occurrence of different interactions (including the nonconventional N−H···π and C−H···π hydrogen bonds) have been correlated with the relative orientations between the interacting residues. Several structural motifs, mostly involved in binding metal ions, have been identified by considering the cases where His residues are in contact with four other planar moieties. About 10% of His residues used here are also found in sequence patterns in PROSITE database. There are examples of the amino end of the Lys side chain interacting with His residues in such a way that it is located on an arc around a ring nitrogen atom.

Keywords: interaction geometry • hydrogen bonding • aromatic−aromatic interaction • structural motifs • metal binding • histadine

Introduction

A specific interaction manifests itself with the two interacting groups having a fixed geometry or orientation relative to each other. The most conspicuous example of this is hydrogen bonding.[1,2] However, other weaker interactions, such as those between the aromatic rings[3,4] or the one between the sulfide group of methionine and an aromatic ring,[5] also maintain a rather well-defined geometry. These and other interactions,[6−8] as they are found in large number in proteins, can stabilize the native fold.[9]

Of all the residues, histidine with an imidazole ring, owing to its pK_a being close to the neutral pH, is quite conspicuous at the active site of proteins,[10] and because of this one generally tends to overlook any contribution that it might have toward protein stability. This, coupled with its rather low abundance,[11] (2.3%, vis-à-vis 3.9% for Phe, for example) meant that the residue is left out in most of the studies dealing with aromatic−aromatic interactions.[3,4,12] With the increase in the database, His is now included in analysis, but usually merged with other aromatic residues.[13] However, when His is considered separately it is found that it has a preference to have an interplanar angle less than 30° with other aromatic residues.[14] Along the same lines, a recent analysis of the interaction of the largest planar ring, Trp with individual aromatic residues showed that His is quite distinct in that it prefers to have a face-to-face orientation with Trp.[15] Thus, it would be of interest to analyze the geometry of interaction of His with aromatic residues (Phe, Tyr, Trp, and His) taken individually, as well as other planar side groups, such as those of Asp, Glu, Asn, Gln, and Arg. Singh and Thornton[16] published an atlas of statistical data for the interplanar angle between interacting side chains for all the possible pairs of residues. However, the database was fairly small and did not exclude all homologues. Moreover, a comparative analysis involving pairs of residues of distinct chemical properties was not attempted. In another study, Mitchell et al.[17] have shown that there is nonrandomness in the packing of side chains. However, they considered only one parameter, viz., the interplanar angle (*P*), which, however, is not enough to completely describe the relative orientation. For

* To whom correspondence should be addressed. E-mail: pinak@boseinst.ernet.in. Fax: (91)-33-2334-3886.
† Department of Biochemistry.
‡ RB and RPS contributed equally to this work.
§ Bioinformatics Centre.

10.1021/pr025584d CCC: $25.00

Published on Web 02/15/2003

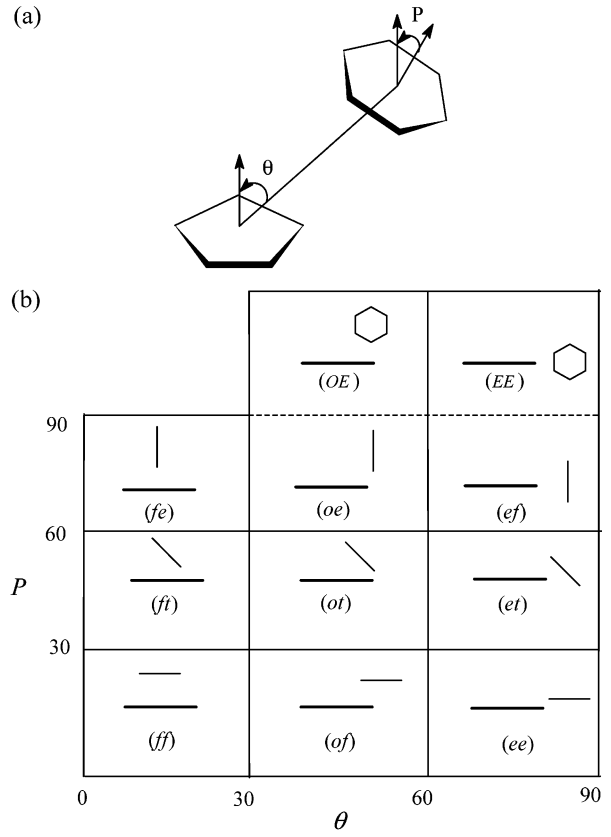

Figure 1. (a) Parameters (P and θ) describing the orientation of a planar group (for example, a six-membered ring) relative to the five-membered His ring. The normal vectors passing through the centroids are shown. The interplanar angle (P) is the angle between the two vectors; θ is the angle between the normal to His plane and the center-center direction. (b) Schematic representation and their nomenclature (see text for details) for ring orientations corresponding to various combinations of P and θ values (in °). Lines signify planes of the side chains (the longer one for His and the shorter one for the partner) perpendicular to the paper; in two cases; however, the planar group, shown as a hexagon, is in the plane of the paper. The two orientations given on top-right-hand are the distinct alternatives of the orientations given in the grid elements below, having the same range of P and θ.

example, an angle of 0° may result from the two rings being stacked one over the other, or placed side by side in a coplanar fashion (and numerous other orientations between these two extremes). Though the precise description of the location of a planar moiety relative to another would need more parameters, Samanta et al.[15] have shown that an additional polar angle (θ) provides an adequate representation to visualize the relative orientation in two dimensions (Figure 1). Here, we use such a procedure to analyze the geometry of interaction of His with nine other planar groups.

Though acidic and basic residues are found in close contact in protein structures,[18] it is not uncommon to find carboxyl−carboxyl interactions.[19,20] At the other end of the spectrum, Arg and Lys are two basic residues that carry a formal positive charge at neutral pH, while depending on the environment, His with a pK_a of 6.5,[10] may or may not be protonated. Like the His−Arg pairs, it is also of interest to identify His−Lys interactions and examine if the terminal $-NH_3^+$ group of Lys has any preferred mode of binding to His. Finally, all the interacting residue pairs are analyzed on the basis of their location in secondary structural elements, solvent accessibility, functional significance, and as components of larger structural motifs.

Materials and Methods

Atom coordinates were obtained from the Protein Data Bank (PDB) at the Research Collaboratory for Structural Bioinformatics (RCSB).[21] 432 chains (in 418 files) were selected using PDB_SELECT[22] from PDB files (as of March 2000) with an R-factor ≤ 20%, and resolution of ≤ 2.0 Å and sequence identity less than 25%. The list of the PDB files used is given in Bhattacharyya et al.[23] To restrict the analysis to well-ordered residues, all polypeptide chains with more than 40% of atoms with temperature factor (B-factor) > 30 Å^2 were excluded. All planar side chains (Phe, Tyr, Trp, His, Asp, Asn, Glu, Gln, and Arg) in contact with the side chain atom(s) of a given His (the central residue) were identified. Only the histidines not having more than 40% of atoms with a B-factor > 30 Å^2 were considered as central residues. Similarly, to qualify as a partner, the atoms in contact with His had to have an occupancy factor 1.0 and the B-factor ≤ 30 Å^2. A distance of ≤ 4.0 Å was used to locate residues in contact; however, any residue having an unreasonable short contact (≤ 2.0 Å) was not considered as a partner. For His−His pair, each residue in turn was considered as the central residue. However, as the criterion for being selected as the central residue is slightly different from that for the partner, in a few pair one of the residues got excluded (and consequently, an odd number of His residues is reported in the Results).

To calculate the geometry the centroids (the center of the five- or six-membered ring for His, Phe, and Tyr, the midpoint of CD2 and CE2 atoms for Trp, CG for Asp and Asn, CD for Glu, and Gln and CZ for Arg) of the planes were first calculated. A molecular axial system was defined with the origin at the centroid of the central His and the z-axis along the normal to the plane. The interplanar angle, P, and the angle, θ, made between the z-axis and the line joining the centroid of the partner to the origin were computed (Figure 1a). For each interaction the geometry was placed in one of the 9 grids (each spanning a range of 30° along P and θ). We have also used the same schematic representation as given in Samanta et al.[15] to indicate the relative orientation of the partner with respect to His at the 9 grid elements (Figure 1b). Each relative orientation is designated by a two-letter code (ff, ot, ef, etc.). The first letter indicates if the His residue in a pair is interacting with its face (f), edge (e) or has the centroid of the partner plane in an intermediate (offset or *o*) position. The second letter indicates if the partner is tilted (t) with respect to His or has its face (f) or edge (e) pointing toward His. When the two rings are at 90°, a rotation about the vertical axis of the partner ring does not change the values of P and θ, although the relative orientations may be altered. As a result, two further limiting orientations are shown in two extra boxes at the top of Figure 1b. It is not possible to resolve this ambiguity using the two parameters. However, in two dimensions (with no shift possible between the centroids in the perpendicular direction), it is possible to distinguish between the two limiting geometries (for example, ef vs EE) by looking at the relative orientation from the opposite perspective (His relative to the partner). This would convert an ideal ef geometry into fe, but EE would remain EE. Considering this, as well as visually examining on a graphics terminal a large number of interacting residues occupying the

two grid elements, we could find out the geometry (ef or EE, and oe or OE) with the maximum number of cases (for a given partner).

If O_{ij} is the observed frequency of occurrence in the grid element corresponding to the i^{th} row and j^{th} column (i and j varying from 1 to 3), the corresponding expected value (E_{ij}) can be calculated as the product of the sum of the observed numbers in the elements in the i^{th} row and that in the j^{th} column, divided by the total number of observations in all of the nine grid elements.[15,24] The χ^2 test was used to determine the statistical significance in the difference in the observed and expected distributions (with degrees of freedom, df = 4)[24] for each His-partner pair.[11]

The propensities of the nine planar residues to interact with His were calculated as the ratio of the observed number of residues in contact to the expected number, the latter being obtained by distributing the total number of observations (His-partner pairs) according to the percentage occurrence of the partner residues in protein structures.

For the His−Lys interaction, all Lys residues with the NZ atom within a distance 4.0 Å from any atom of a His ring were found out. In the Xray crystal structure analysis it is not always possible to unambiguously decide between the two possible orientations (related by 180° rotation about the CB−CG bond) of the imidazole ring of His. We were interested in identifying the His atom in contact with NZ and as this depended on the orientation of the ring, we first decided on the most likely orientation using the suit of programs developed by Word et al.[25] Of the two flipped orientations, the one selected had the minimum number of bad contacts (shown as red dots) and the maximum number of hydrogen bonds.

The secondary structural elements were determined using the algorithm (DSSP) of Kabsch and Sander.[26] The solvent accessible surface area (ASA) was computed using the program ACCESS,[27] which is an implementation of the Lee and Richards algorithm.[28] Molecular diagrams were generated using MOLSCRIPT.[29] Hydrogen bond interactions between His and its partners were found out using the program HBPLUS.[30] The scatterplot for His−Lys interactions was made using SURFNET.[31] In the text the PDB files are identified with the four-lettered code (in small letters), followed by the subunit identifier (in capital), if any.

Results

Planar Groups in Contact with the His Ring and the Accessible Surface Areas. A total of 1115 histidines in 432 polypeptide chains were selected (as central residues), and other planar side chains in contact (partner residues) were identified. The counts of His residues with different numbers of partners are as follows, 1:566, 2:328, 3:145, 4:47, 5:26, 6:2, and 7:1. In Table 1 the numbers of partner residues and their propensities to be in contact with His are given. In analyses involving residue contacts in protein structures one normally associates His with Lys and Arg for having high affinity to bind negatively charged residues.[18] Contrary to this, the propensity of Asp and Glu to interact with His is close to the average value of 1.0, whereas the highest propensity values are observed for three other aromatic residues. Along the expected lines, Arg has a low propensity value of 0.70, which however is marginally higher than the values of neutral planar side chains of Asn and Gln. Additionally, it is found that rather than having a single His−His contact, the propensity of a His residue to have multiple His residues around it is greater—his is likely due to the occurrence of metal binding sites with multiple His ligands (discussed latter).

Table 1. Numbers and Propensities of Planar Residues to Be in Contact with His

	central His with			
	a single contact[a]		any no. of contacts[a]	
planar residues	no. of cases	propensities	no. of cases	propensities
Phe	87	1.52	276	1.41
Tyr	95	1.74	285	1.51
Trp	33	1.47	131	1.70
His	26[b]	0.76	139[b]	1.18
Arg	45	0.66	162	0.70
Asp	97	1.10	341	1.13
Glu	83	0.94	253	0.84
Asn	45	0.64	159	0.65
Gln	30	0.52	110	0.56

[a] With the planar residue under consideration. [b] For His−His contacts, only one His in a pair has been considered.

Table 2. Mean Accessible Surface Areas (ASA) and Relative Accessibilities (RA) of the Side Chains of His Residues and the Corresponding Partners

	ASA (Å^2)		RA	
partner residue	His	partner	His	partner
Phe	38(32)	23(22)	26(21)	14(14)
Tyr	41(32)	31(32)	28(22)	18(18)
Trp	36(31)	33(36)	24(21)	16(17)
Arg	42(33)	55(42)	28(23)	27(21)
Asp	43(34)	30(25)	30(23)	29(25)
Glu	52(37)	35(24)	35(25)	26(18)
Asn	41(35)	29(32)	28(24)	26(28)
Gln	36(35)	40(32)	24(24)	28(23)

Considering the cases with only one contact with a given planar residue. RA is the percentage ASA of His in the structure as compared to its ASA in an extended Ala−His−Ala tripeptide. Standard deviations are in parentheses.

Table 2 compares the accessible surface areas (ASA) and relative accessibilities (RA) of His residues in contact with different planar side chains. It can be seen that His is more exposed than its aromatic partner; but with other residues both are exposed almost to the same extent; with Glu, His seems to have a larger surface area accessible to solvent.

Geometry of Interaction. The geometry of interaction of His ring with other planar groups has been characterized using three sets of ranges of two parameters, P and θ, giving a total of 9 relative orientations, typical representations of which are shown in Figure 1b. The number of occurrences in each of these orientations for all the partners is provided in Figure 2, along with the expected number and the χ^2 values representing the difference between the observed and the expected distributions. The probability that the observed distribution is random (and P and θ independent) is < 0.5% when χ^2= 14.9 (df = 4).[24] Except for Asp and Gln, all χ^2 values are larger than this. Looking at the grid elements where the observed number exceeds the expected number one can discern a pattern that except for Asp and Glu all other residues prefer to interact in a face-to-face (ff) or offset-face-stacked (of) mode. Although for some of these residues Mitchell et al.[17] observed an interplanar angle in the range 0−30°, we find that even in the range of P only two orientations are favored, whereas the third orientation (ee) is normally avoided by all residues (except Asp).

Phe (276)
$\chi^2 = 17.3\ (7.5)$

	OE	ef
43 (36.4)	39 (50.6)	45 (40.0) [1]
15 (24.3)	38 (33.9)	32 (26.8)
21 (18.3)	33 (25.5)	10 (20.2)

Tyr (285)
$\chi^2 = 40.2\ (11.6)$

	OE	EE[a]
17 (24.4)	34 (40.7) [7]	65 (50.9) [21]
10 (17.7)	29 (29.5) [1]	45 (36.8) [14]
33 (17.9)	37 (29.8)	15 (37.3) [6]

Trp (131)
$\chi^2 = 15.4\ (4.9)$

	OE	ef
13 (11.1)	16 (19.7) [2]	18 (16.1)
4 (9.0)	14 (16.0)	20 (13.1) [2]
14 (10.9)	25 (19.3)	7 (15.8) [1]

His (277)
$\chi^2 = 32.0\ (7.7)$

	OE	ef
9 (18.3)	69 (57.0) [3]	71 (73.7) [6]
10 (10.4)	24 (32.5) [1]	51 (42.0) [12]
15 (5.3)	13 (16.5)	15 (21.3) [3]

Arg (162)
$\chi^2 = 49.5\ (8.3)$

	oe	EE
6 (12.4)	25 (24.0) [3]	28 (22.6) [12]
2 (11.8)	27 (22.8)	27 (21.4) [14]
26 (9.9)	14 (19.1)	7 (18.0) [3]

Asp (341)
$\chi^2 = 13.6\ (2.5)$

	oe	EE
1 (5.2)	19 (15.6) [4]	128 (127.2) [77]
5 (4.8)	14 (14.3) [1]	117 (116.9) [69]
6 (2.0)	3 (6.0)	48 (49.0) [29]

Glu (253)
$\chi^2 = 19.2\ (4.3)$

	OE	EE
0 (2.0)	19 (16.4) [4]	82 (82.6) [56]
1 (2.1)	11 (17.2)	94 (86.7) [59]
4 (0.9)	11 (7.5)	31 (37.6) [19]

Asn (159)
$\chi^2 = 40.0\ (8.2)$

	oe	EE
3 (8.9)	14 (15.7) [1]	32 (24.4) [12]
6 (11.7)	18 (20.5)	40 (31.8) [16]
20 (8.4)	19 (14.7)	7 (22.9) [4]

Gln (110)
$\chi^2 = 5.4\ (2.5)$

	OE	EE
3 (6.9) [2]	10 (10.1) [1]	23 (19.0) [7]
8 (7.4)	11 (11.0)	20 (20.6) [6]
10 (6.7)	10 (9.9)	15 (18.4) [10]

Figure 2. Observed and expected (in parentheses) numbers of interactions at nine boxes (corresponding to the geometry shown in (Figure 1b) for each planar side chain. The total number of interactions, χ^2 value (indicating the difference between the observed and expected distributions) and (in parentheses) the root-mean-square-differences (RMSD) between the observed and expected numbers are given on top of each diagram. Grid elements in which the observed number exceeds the expected number by the RMSD value are shaded, and those where the number is less are enclosed in perforated lines—in either case the observed or the expected number has also to be ≥10. The number of hydrogen bonding (from among the total number of contacts), if it exists in a grid element, is given within square brackets. The range of P and θ in the two elements in the top row can lead to two extreme orientations (oe or OE, ef or EE), and the one observed in the majority of the cases is indicated against each element. In 41 out of 65 cases in this box, the hydroxyl oxygen of Tyr is involved in the shortest contact with a His side-chain atom (either N−H···O or C−H···O hydrogen bond). In such cases, the geometry is *EE*; when there is no contact with the oxygen atom, the *ef* geometry is usually observed.

Asp and Glu, because they mostly interact with His through hydrogen bonding show a different behavior, largely populating the grids oe and et, respectively. Examples drawn from assorted protein structures to indicate various geometries of interactions are shown in Figure 3.

For His−His interactions, in addition to the expected ff geometry, one also observes the perpendicular (or nearly so) orientations oe and et. As can be seen in Figure 4a, when the residues simultaneously bind a cation the geometry mostly occupies the three grids at the top-right-hand corner. For the remaining His−His pairs (which contain 11 cases of His residues binding two different cations, 20 cases where only one of the residues bind a cation and 113 cases where none binds a metal) the observed number is greater than the expected

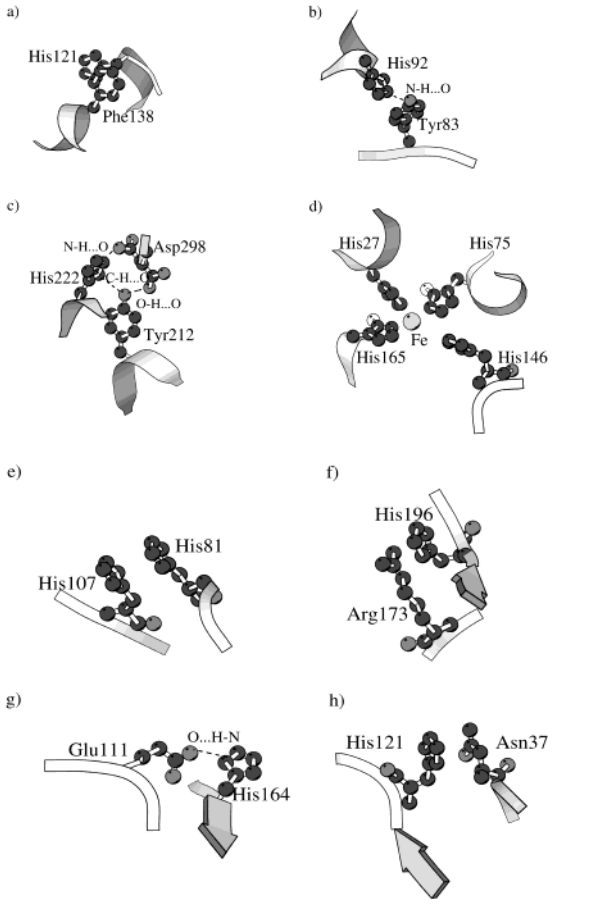

Figure 3. Some typical examples of the geometry of interaction of His with other planer groups; the residues and part of the secondary structures they are located in are shown. (a) His−Phe (in ribonuclease MC1, PDB file: 1bk7A), geometry: of; (b) His−Tyr (with N−H···O hydrogen bond between the rings) (phosphotriesterase homology protein, 1bf6A), EE; (c) His−Tyr (with C−H···O hydrogen bond between the rings) (chloroperoxidase, 1cpo), EE; (d) His−His (superoxide dismutase, 1bsmA) (where residues 27, 75, and 165 are metal ligands; His146 is not a ligand, but is in contact with His75. Hydrogen atoms at the CD2 positions of His75 and His165, placed close to the π electron cloud of His27 are shown), ef (between His75−His27), oe (between His27−His165), ef (between His75−His146); (e) His−His (adenosine kinase, 1bx4A) (where none of the His residues are metal ligands), ff; (f) His−Arg (methylene tetrahydrofolate dehydrogenase/cyclohydrolase, 1a4iB), ff; (g) His−Glu (cysteine protease, 1cv8), et; (h) His−Asn (enterotoxin B, 3seb), ff.

(a)

	71	61
	29	55
		40

(b)

9 (13.5)	20 (19.8)	28 (23.8)
10 (11.8)	17 (17.4)	23 (20.8)
15 (8.7)	13 (12.9)	9 (15.4)

Figure 4. Of the 277 cases of His−His interactions, in 133 cases both the His residues bind the same metal ion. The percentage of such residues relative to the number observed in each grid (for His, Figure 2) is shown in (a); the grid elements where there is no occurrence are left blank. (b) For the remaining 144 cases, the observed and expected numbers are shown (as in Figure 2); $\chi^2 = 9.9$ and RMSD = 3.8.

number for the ff and ef orientations (Figure 4b). Thus, the geometry of two His residues in contact can be different depending on whether the interaction is mediated by a cation.

Table 3. Hydrogen Bonding between His and Its Partner Residue

partner	no. of partners	HB	FRAC
Phe	276		
Tyr	285	49	0.17
Trp	131	5	0.04
His	277	25	0.09
Arg	162	32	0.20
Asp	341	180	0.53
Glu	253	138	0.55
Asn	159	33	0.21
Gln	110	26	0.24

HB is the number of hydrogen bonds considering only the side chain of the interacting pair. The proportion of interacting residue pairs that are hydrogen-bonded is given by FRAC.

To find out if the relative orientation between His and a planar partner in a metal-binding site changes in the absence of metal, we calculated the geometry in cases where the apo structure is also known (for the protein of the same species). Only in five proteins (myoglobin, 1a6m; galactose oxidase, 1gof; bacterial luciferase, 1lucA; arabinose operon regulatory protein, 2arcB; carboxypeptidase A, 2ctc) the structure is available both in the presence and the absence of the cation. In all cases, His remains in contact with the same planar group and the magnitude of the changes in P and θ are 10.7 $(\pm 8.6)^0$ and 8.7 $(\pm 6.6)^0$, respectively, such that the overall geometry remains in the same grid element or moves to an adjacent one (Figure 1b). Thus, it appears that the metal-binding site is essentially preformed and can accommodate the cation with only minor structural adjustments.

For two grid elements with P about 90° it is not possible to uniquely represent the orientations and two extreme possibilities are shown in Figure 1b. As discussed in Methods in the majority of the cases it is possible to associate the observed orientation with one of the representations in each grid by looking at the interacting residues on a graphics terminal and/or calculating the geometry of the His ring relative to the partner plane. Figure 2 shows that the orientation that is usually observed depends on the type of the partner. With another aromatic residue the two grids prefer the geometries of OE and ef, which place a C−H (or N−H) group of one residue pointing toward the π electron cloud of the other. The top-right-corner grid of Tyr is an exception in that EE rather than ef that is preferred, and this is caused by the occurrence of hydrogen bonding (N−H···O or C−H···O) involving the hydroxyl oxygen of Tyr. For all the remaining residues the preferred geometry in this grid is EE.

Hydrogen Bonding. Results presented in Table 3 show that about 55% of the side-chain contacts with Asp and Glu are hydrogen bonds. Less than 10% of His−His and His−Trp pairs are hydrogen bonds, whereas with the remaining residues the proportions that are hydrogen bonded are around 20%. The percentages given here are considerably higher than the values of Mitchell et al.,[17] especially for Arg, Asp, and Gln.

The geometry of the side chains engaged in hydrogen bonding show a clear pattern (Figure 2). Most of such interactions have a θ in the range 60−90°. Moreover, the interplanar angle is usually > 30°, with only a few occurrences of nearly coplanar geometry (ee). Ippolito et al.,[2] whereas studying the hydrogen bond stereochemistry relative to all the donor and acceptor groups in proteins, found that the atom hydrogen bonded to NE2 or ND1 atom of His prefers to be in the plane

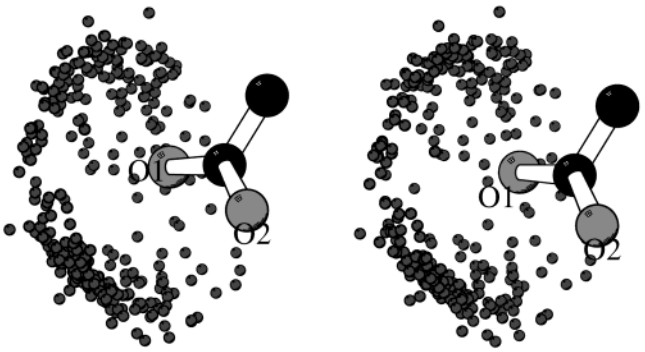

Figure 5. Scatterplot of superimposed positions of His nitrogen atoms relative to the carboxylate group of Asp and Glu, in stereo. Of the two oxygen atoms the one with the shorter distance to His is labeled O1.

of the imidazole ring. However, they did not consider if the interacting planes (when the hydrogen-bonded atom is also a part of a planar group) have any preferential geometry. What we find here is the clear tendency of carboxylate plane to interact with a P value in the range 30−90°.

It would be of interest to consider if the geometry of interaction between His and Asp residues in the catalytic triad of Ser−His−Asp located in serine proteinases and lipases is any different from the general trend observed here. For this, we randomly selected (based on lower R-factor and higher resolution) two or three structures from different groups of enzymes having the triad.[32] In a sample of 26 structures, we found the ee orientation in 8 cases, et in 15 and EE in 3. Thus, et seems to be the preferred geometry in enzyme structures, which is in agreement with the general trend.

On the basis of hydrogen bonding features we can assign the protonation state to a few His residues. In 190 out of 1115 residues, both the ring N atoms act as hydrogen bond donors (mostly to carboxylate side-chains and main-chain oxygen atoms), and these are the definite cases of protonated His. In 44 cases, one N atom acts as donor and the other acceptor (NE2 as donor in 23 cases and acceptor in 21); in 27 cases, His acts as acceptor only, and in 130 cases, it is bound to metal ion—these categories together represent neutral His residues. In the remaining examples (no hydrogen bonding, or one or both N atoms hydrogen bonded to one or more water or hydroxyl side-chains, which can act both as acceptor and donor), the protonation state cannot be ascertained unambiguously.

Geometry of His Nitrogen Atom Relative to the Carboxylate Plane. As the hydrogen bond interactions between His and Asp/Glu was found to occur at a rather well-defined orientation of the carboxylate group relative to the His plane, we wanted to see if there was also a specific location of the nitrogen atom of His when looked from the perspective of the carboxylate plane. Figure 5 shows that there indeed exists a definite geometry [the average O···N distribution, 2.8(2) Å], with the nitrogen atoms interacting along the directions of carboxylate-oxygen sp^2 lone electron pairs (angle C−O···N around 120°, and very few atoms along the extension of the C−O direction), as has been observed with other proton donor atoms also.[1,2] Although mostly close to the carboxylate plane, the nitrogen atoms can be distributed up to an angle of 45° from it. Ippolito et al.[2] also considered the syn/anti stereochemistry of the proton donor atoms relative to the carboxylate group (syn corresponds to a projected position within the two C−O

directions and anti, outside) and found a significant excess of syn-oriented interaction (syn/anti ratio of 57%:43%) for glutamate, but not for aspartate (51%:49%). On the contrary, we find a greater syn/anti discrimination for aspartate (65%:35%) than glutamate (54%:46%) when the interaction is with a His residue.

His-Lys Interaction. The basic residues Arg and His are found to have a preference to stack against another His residue. Hence, it is of interest to see if the $-NH_3^+$ group (atom NZ) of the Lys side chain also has any inclination to be in contact with the His side chain and if so, whether the interaction is with the face of the ring or its edge and also the ring atoms involved.

We first identified the His ring atom having the shortest contact with the NZ atom of a Lys residue. Then, as discussed in the Methods section, both the orientations—one corresponding to the coordinates in the PDB file and the other obtained by flipping—were investigated for the existence of short contacts with all the neighboring atoms, and the orientation with no or lesser number of bad contacts was selected. It was found that out of 54 cases of His−Lys interaction, in 26 cases where the NZ atom showed contact with a ring C-atom the ring had to be flipped and the contact occurred with ND1 or NE2 atom of His (flipping about the C^β—C^γ bond interchanges the positions of ND1 and CD2, and CE1 and NE2 atoms). Of the 54 His−Lys pair, the shortest contact, at an average distance of 3.4(3) Å, is with NE2 (25 cases), ND1 (24), and ring carbon (CG, CE1, CD2) atoms (5).

The nonionized imidazole ring can exist as tautomers, with the hydrogen atom on either ND1 or NE2 atom.[10,33] From NMR studies on model peptides, it has been shown that the hydrogen atom is usually predominantly on the NE2 atom, which has a pK_a value of about 0.6 pH unit higher than that of ND1. However, the relative affinities of the two nitrogen atoms for protons can vary with conditions in the local environment. Assuming that the interaction is between the $-NH_3^+$ group and the unprotonated nitrogen atom of His, it can be seen that both the tautomeric forms are found in almost equal numbers.

Figure 6 shows that the NZ atoms of Lys are distributed on an arc around the two His nitrogen atoms. If we calculate the angle θ (Figure 1a) for the NZ atoms, with the origin on the nitrogen atom in contact (and not the ring centroid), and distinguish between two types of interactions—one with the face ($0° \le \theta \le 45°$) and another with the edge ($\theta > 45°$) of the His ring[6]—then the respective numbers are 23 and 31. This suggests that there may not be much energetic difference between the NZ atom interacting with the sp^2 lone-pair of electrons on ring N atom (a conventional N−H···N hydrogen bonding) or its interaction with the π electron on the nitrogen (N−H···π hydrogen bonding). Ab initio calculations of a heteroaromatic ring interacting with different model ligands (water, ammonia, methane, and benzene) in different relative orientations have shown that for some ligands there can still be considerable interaction energy when the geometry is changed from a conventional hydrogen bond to an interaction with the face of the heteroaromatic ring.[34] Analysis of non-hydrogen bond interactions has also revealed cases where methionine sulfur atom is found on the face of His ring.[5] Thus, His−Lys interaction, because of its small number of occurrences, may not contribute to the stability of protein structures in general, but it is interesting to note that when it occurs, the geometry is centered around a N-atom of the His ring, and because of this specificity, it may even provide local stability to the structure.

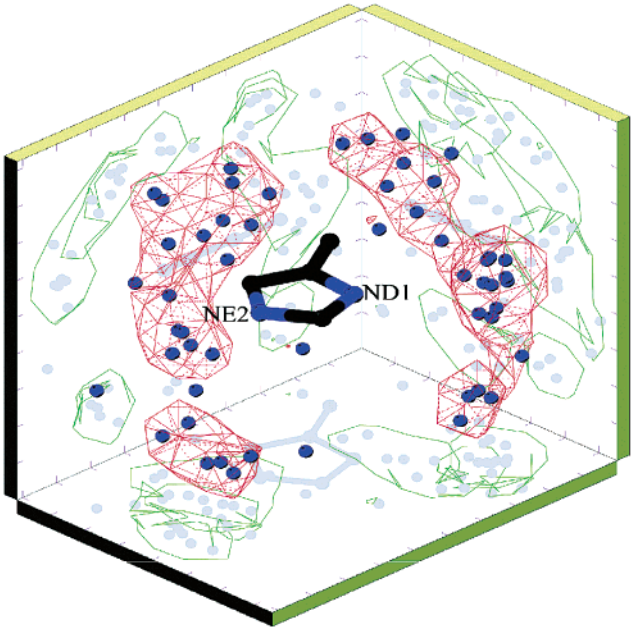

Figure 6. Scatterplot of superimposed positions of NZ atoms of Lys relative to the imidazole ring of His. The contours surround the regions with the maximum clustering of NZ atoms. To have a stereo perspective all the atoms are projected in three directions.

Table 4. Number of Cases with a Given Sequence Difference, Δ, between His and Its Partner

residue pair	Δ								
	4	3	2	1	−1	−2	−3	−4	rest
His−Phe	3	4	7	9	8	8	2	16	219
His−Tyr	5	3(1)	12(3)	7	7	3	4	13	231(45)
His−Trp	2(1)	5	3	5	4	6	1	8	97(4)
His−His	12(1)	2	10(1)	7	7	11	2	12(2)	214(21)
His−Arg	9(1)	12(5)	1(1)	4	9(2)	1	3	1	122(23)
His−Asp	8(2)	3(1)	13(4)	13(5)	8(2)	9(5)	9(4)	8(2)	270(155)
His−Glu	7(4)	4(2)	7(5)	7	5	15(10)	9(2)	9(6)	190(109)
His−Asn	3	0	6	9(2)	3(2)	10(1)	5(1)	5	118(27)
His−Gln	5	2	3(1)	3	7	7(3)	4	2	77(22)

Δ = (partner-His) residue numbers. The number of hydrogen bond at each Δ is given in parentheses.

Secondary Structures and Some Motifs Containing His. When the absolute value of the sequence difference, Δ, between His and its partner is 4 or less, the interaction can be assumed to be local, taking place within the same secondary structure, whereas a number beyond 4 indicates a longer range, tertiary interaction, and the latter is found to predominate (Table 4). The number of cases with $\Delta = -4$ exceeds that for $\Delta = 4$, when the partner is an aromatic residue, but the opposite is true for Arg. When $|\Delta|$ is 4, the two residues are spatial neighbours on the same side of the helix, and as observed here, a His following another aromatic residue, rather than in the reverse order, has been found to occur significantly as it places a C−H or N−H group of His interacting with the π electron cloud of the aromatic partner, a stabilizing interaction when the His ring is protonated.[23]

In 9 out of 12 cases, when an Arg follows a His by three residues both the residues are in the same helix and in four of

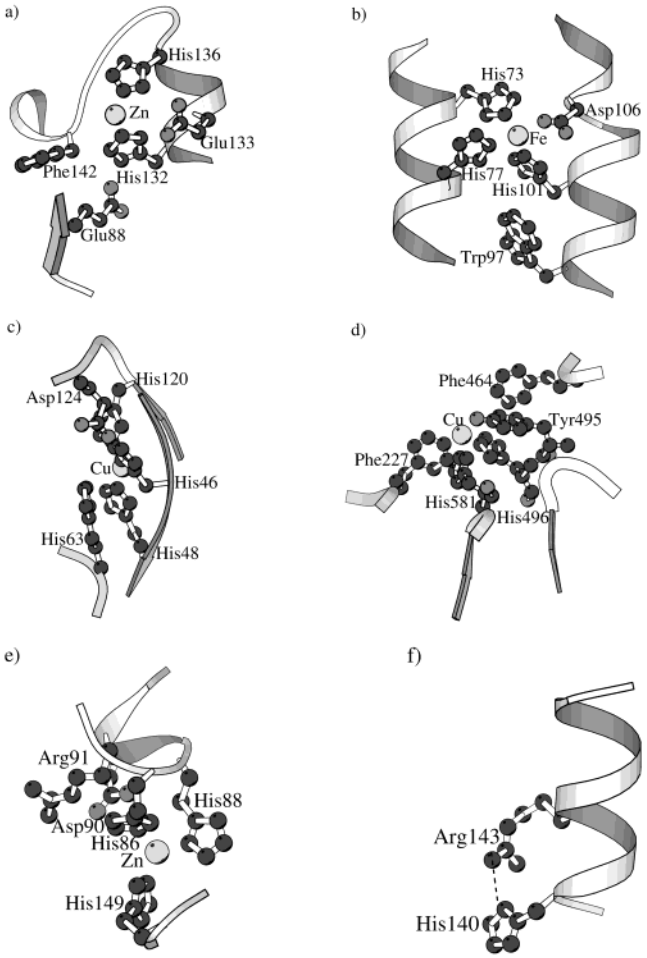

Figure 7. Representative examples of motifs (a−e) enumerated in Table 5. The proteins (and the PDB files) are (a) peptide deformylase (1bs4A), (b) hemerythrin (2hmzA), (c) zinc superoxide dismutase (1mfmA), (d) galactose oxidase (1gof), and (e) metallo beta-lactamase II (2bc2A) (details are in the table). In (f) is shown a pair of hydrogen-bonded Arg and His (with sequence difference 3, and the hydrogen-bond length of 2.93 Å between NH2 of Arg and ND1 of His) in an α-helix (endonuclease III, 2abk).

them the geometry (EE) is optimum for a hydrogen bond, such that the positively charged guanidinium group donates a proton to the neutral His side chain (Figure 7f). When $\Delta = 4$ there are examples of N−H···π interactions involving the side chain of Arg and the π electron cloud of His.

The majority of His−Glu pairs with $|\Delta| = 4$ are hydrogen bonded, which is not the case for His−Asp pair. On the contrary, with $|\Delta| = 1$, only the latter can interact through hydrogen bonding, but not the former. This indicates that because of the differences in chain lengths (His vs Asp or Glu), the sequence difference that can position the side chains for being linked by hydrogen bond is not the same for Asp and Glu.

Using the knowledge of the secondary structures of the residues, we looked for spatial and sequential patterns involving a constellation of four planar residues in contact with a common His. A number of structural motifs could be identified, given in Table 5 and illustrated in Figure 7, which occur in proteins with no apparent relationship between them. In fact, within a given a category examples are found belonging to different protein classes. Although there are two motifs (a and

Table 5. Some Motifs Involving His with Four Planar Partners

(a) metal-binding motif with at least three helical residues
H (*i*, H), E/D/F (*i*+1, H), *H*/F/E (*i*+4, H), possibly Ar (∼*i*+10) and Pl (r). D at *i*+1 can form a hydrogen bond with H (*i*). Pl (*r*) forms a hydrogen bond with H (*i*) or is a metal−ligand. examples: 1bs4A (α/*β*, Zn): *H132*, E133, *H136*, F142, E88. 1iab (α/*β*, Co): *H92*, E93, *H96*, *H102*, *Y149*. 1vns (all-α): H62, D63, F66, F74, R333. 2hmzA (all-α, Fe): *H54*, F55, *E58*, E24, H25 (E58 also binds to a second Fe).
(b) Helix bundle
all (or all but one) helical residues with at least two groups at *i* and *i*+4. 2hmzA (all-α, Fe): *H73*, *H77*, W97, *H101*, *D106*. 1dpsD (all-α): H51__D, F59__D, H63__D, W160__G, F161__G (residues are from two different subunits, which are specified after '__'). 1vns (all-α): H404, D292, Y304, N486, R490 (H404 and R490 are involved in binding a sulfate anion).
(c) Metal-binding motif with at least three histidines in two antiparallel *β*-strands (with one having a pair of ligands at *i* and *i*+2 positions)
1guqA (α/*β*, Fe): *H296*, *H298*, *H281*, *E182*, W293 (E182 also binds to a second Fe). 1mfmA (*β*, Cu): *H46*, *H48*, H63, *H120*, D124 (H63 binds to another cation, Zn).
(d) (At least four) residues in turn regions
1gof (*β*, Cu): *H496*, F227, F464, Y495, *H581*. 1qh5A (α/*β*, Zn): *H54*, H55, *H56*, *H110*, Y145. 1qh8A (α/*β*): H449, D226, D232, H272, F298.
(e) (At least four) residues in sequential proximity
2bc2A (α/*β*, Zn): *H86*, *H88*, D90, R91, *H149*. 1bqcA (α/*β*): H34, Y31, Q33, Q36, F272.

The table contains the name/description of the motifs, some comments, followed by typical examples. One-letter amino acid code is used, Ar stands for an aromatic and Pl, any planar residue. The relative position (r indicates a remote residue) and the secondary structure, if it is same across different examples, are given in parentheses (H = helix). PDB file names (with protein class and the bound metal ion, if any), followed by residue names/numbers are given as examples. Metal−ligand residues are in italics.

c) that have been exclusively termed as metal-binding, other categories also include examples where cations are held. Although included in the metal-binding motif (a), 1vns does not contain any metal. However, comparison with other members suggests that mutating two Phe residues by His may possibly create a cationic site in this protein. Motif (b) is interesting because it is not only found in the tertiary structure of a molecule, it can also be located at the interface between two protein subunits, thus showing the usefulness of His in stabilizing intersubunit contacts. In addition to binding cation, a His residue can also be a ligand to anions (exemplified in 1vns) and thus the same motif (b) is used in binding cation in one structure and anion in another.

His in PROSITE Patterns. For proteins with a common function, the amino acid sequences usually contain clusters of residue types, termed as patterns, motifs or fingerprints, which may occur even in distantly related proteins in which sequence similarities are hardly discernible.[35] Such patterns (with possible combinations of amino acids) have been derived from analyses of sequences of proteins with distinct functions

and organized in the form of a database, PROSITE.[36,37] We wanted to see if the His residue in our analysis occur in PROSITE patterns and if so, whether the corresponding partners are also located in the same pattern so that one can establish a possible role for the interaction of His with its partner in the context of the conservation of residue types within the pattern.

It is found that only about 10% of His residues (99 out of 1115) are located in 77 patterns (of which 57 are unique) in PROSITE. On average, these residues have about 2 partners and about half of them have a partner in the same pattern. The frequently used partners (whether located in the same pattern as the His residue or outside) are His and Asp. The most common pattern involving His has the role of binding metal ions; there are 22 unique patterns of metal binding. We observe that not all His residues with bound cation in PDB are shown as being part of metal-binding patterns in PROSITE. Only 10 His residues in our analysis are designated as active site residues in 9 unique patterns in PROSITE.

Discussion

The imidazole ring of His is unique among amino acid residues. As a heteroaromatic moiety, it can interact with other aromatic and nonpolar groups, whereas the heteroatoms can also participate in hydrogen bonds. Depending on the protonation state, it can also be involved in salt-bridges with acidic groups. We studied if His has any specific geometry in its interactions with all these diverse moieties, and if the preference for any particular geometry is the same irrespective of whether His is in the active site. Although weaker than a hydrogen bond, interactions such as N−H···π or C−H···π are known to affect the conformation and intermolecular packing in small molecule crystal structures[38,39] and stabilize the fold and substrate binding in protein structures.[6−9,40] When His interacts with another planar residue, it would be of interest to see if there is any preference between the stacked orientation and a near perpendicular orientation that will facilitate the occurrence of N−H···π or C−H···π interactions.

Of the nine planar side chains (Phe, Tyr, Trp, His, Asp, Asn, Glu, Gln, and Arg) a histidine residue has the highest propensities to interact with aromatic groups, its tendency to associate with a negatively charged residue (Asp, Glu) being average (propensity value of about 1) (Table 1). When interacting with another aromatic residue, histidine has its surface more exposed to solvent than the partner (Table 2).

Geometry of Interaction. We have used two parameters, P and θ, to specify the relative orientation between two interacting planes. Nine grid elements, each spanning of range of 30° along the two parameters, and their idealized representations and names are given in Figure 1b. The number of observed points in each grid for a given partner was compared to the expected number and the statistical significance of the difference found out (Figure 2). With the exception of Asp and Glu, the preferred grid element (or one of the preferred elements) for all other residues is face-to-face (ff) or offset-face-stacked (of). For Trp, the et geometry with the edge of His interacting with the face of Trp in a tilted fashion is also preferred; other aromatic residues also have a higher-than-expected number of occurrence in this grid element(though not reaching the significant level).

Studies dealing with the aromatic−aromatic interactions have mostly suggested an electrostatically favorable edge-to-face (ef or fe) or T-shaped configuration and an unfavorable face-to-face orientation.[3,4,12] However, a recent work has found that the parallel-displaced structure is 0.5−0.75 kcal/mol more stable than a T-shaped structure for the Phe−Phe dimers.[13] Here, we find that the interactions of Phe and Trp with His have a clear preference for the of orientation (Figure 3a), whereas with Tyr the neighboring grid element, ff, is preferred. Another grid preferentially occupied (especially for Trp) is et, in which the edge of His ring points toward the face of the partner, a geometry which has been found to be stabilizing when His is protonated.[23] In addition to ff, the EE geometry is also preferred for the His−Tyr contacts. These are usually the cases with N−H···O or C−H···O hydrogen bonds between a ring atom of His and the hydroxyl oxygen atom of Tyr (Figure 3b,c).

A parallel stacking arrangement of the guanidinium group of Arg directly over the center of Phe, Tyr, and Trp rings has been observed much more frequently than could be expected by chance.[41] Here, we find that even with His as the aromatic residue the geometry is the same (Figure 3f). Mitchell et al.[42] and Samanta et al.[6] observed that the amide side chains of Asn and Gln tend to form stacked interactions with the aromatic rings of Phe, Tyr, and Trp, these geometries outnumbering amino···aromatic unconventional hydrogen bonds (or N−H···π interactions). Asn−His and Gln−His interactions follow the same pattern (Figure 3h).

'Free' vs Constrained His−His Pair. About 50% of all His−His pairs have both the residues binding the same cation, with P and θ angles closer to 90° (Figures 4 and 3d). For the rest, the face-to-face parallel geometry is found more than expected (Figure 3e). When two residues are found close to each other, it could be due to the combination of two factors—the existence of some stabilizing interaction between them and the optimization of interactions that the two residues have with other chemical groups in the local environment. The former would result in a well-defined geometry of interaction. In its interaction with another His ring, a His residue shows, as in a similar interaction with another basic residue, Arg, a ff geometry. However, two His rings can also be brought together by being bound to the same metal ion and under this constraint the geometrical preference changes to a near perpendicular orientation. It needs to be pointed out that residues brought together by such extraneous factors would also affect the calculation of parameters related to affinity between residues based on interresidue distance measures.[18] Thus, considering the cases when His has a single planar partner, the propensity of His to interact with another His residue is only 0.76 (Table 1), but the value increases to 1.18 when cases with multiple partners (and these include metal-binding histidines) are also considered. In an analogous situation, by virtue of their common role as metal ligands two carboxylate groups can also come to within a distance of 3.0 Å.[20]

Various Interactions. About 55% of Asp and Glu residues in contact with His interact through hydrogen bonding (Table 3), for which there is a need for the relative geometry to have a $\theta > 60°$ (Figure 2). The et geometry is the most preferred one (Figure 3g), which is also found to be the case in His−Asp pairs in the catalytic triad of serine proteinases and lipases. Thus, the geometry of a His−carboxylate ion pair is the same irrespective of whether it is part of an active site or not. Like other proton donors[1,2] the nitrogen atom of His is located close to the direction of the sp^2 lone pair of electrons on the oxygen atom of the carboxylate group, with the syn oriented interaction

outnumbering the anti orientation in the ratio 60:40 (Figure 5).

17% of Tyr residues in contact with His are engaged through hydrogen bonding (Table 3). The EE geometry is particularly favorable for conventional N−H···O hydrogen bonding, as well as C−H···O interaction that may occur between the His ring and the hydroxyl oxygen atom of Tyr (Figure 3b,c). Though one could not normally associate a hydrogen bond between two positively charged histidines, or between His and Arg residues, when (one) His is neutral there can be N−H···N hydrogen bond. In fact, when an Arg follows a His with a sequence difference of 3 in an α-helix (Table 4), a hydrogen bond can link them in an EE geometry (Figure 7f). Likewise, there can be N−H···N hydrogen bond between the Lys and His side-chains. However, these bonds seem to have a softer characteristics as the $-NH_3^+$ group of Lys is not restricted to being on the plane of the His ring and can also be located nearly on the face of the ring (Figure 6).

As already discussed, His residues in metal sites have a typical geometry. For example, in 1bsmA (Figure 3d), His75 and His165 have both an ef geometry to His27, such that CD2-H of the former two rings are positioned close to the top of NE2 atom of His27 and there would be C−H···π interactions. Most metal binding sites with a pair of His residues have a relative orientation that generates C−H···π interactions, which may provide extra stability and affect the redox properties of the metal centers.

Sequence and Structural Motifs. Only about 10% of His residues analyzed here are found in sequence patterns in the PROSITE database[36] and nearly half of them have a planar partner in the same pattern. The three-dimensional protein structures are now being annotated in terms of the existence of residue clusters.[43] Here, we have also looked at clusters of planar residues centered on His and by combining with the secondary structural features of the residues we have been able to identify a few structural motifs (Table 5 and Figure 7). It is interesting to note that a motif like (a) has a structure (1vns) that unlike the other members of the motif is not involved in binding metal. As the relative disposition of the groups and their secondary structure are very similar, a change of two Phe groups to His residues in this structure may generate a new metal binding site. As such, an example like this can be a candidate to be used in modeling of new functionality in protein structures, to be analyzed using programs such as DEZYMER.[44] Similarly, the motif (b) has both metal as well as anion-binding sites and the structures are amenable for alteration of their binding properties (cation-binding to anion-binding and vice versa) through mutation of some residues. The identification of spatial motifs is not very easy unless they constitute the active site residues or metal-binding sites.[45] Here, we have found some patterns in the relative sequence numbers and secondary structures in the constellation of residues centered on His, which can occur within a polypeptide chain or between different subunits (for example, motif b).

In conclusion, in this paper, we have looked into the geometry of interaction of His with other planar residues and the basic side chain of Lys and identified the preferred relative orientations. Various types of interactions and their geometry will help in modeling protein structures and their interactions with other molecules. Knowledge of the structural motifs identified here can be an aid in protein engineering experiments to design such sites in other protein molecules.

Acknowledgment. We thank the Department of Biotechnology for computational facilities and the Council of Scientific and Industrial Research for fellowships to R.B. and R.P.S.

References

(1) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97.
(2) Ippolito, J. A.; Alexander, R. S.; Christianson, D. W. *J. Mol. Biol.* **1990**, *215*, 457.
(3) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23.
(4) Singh, J.; Thornton, J. M. *FEBS Lett.* **1985**, *191*, 1.
(5) Pal, D.; Chakrabarti, P. *J. Biomol. Struct. Dynam.* **2001**, *19*, 115.
(6) Samanta, U.; Pal, D.; Chakrabarti, P. *Proteins Struct. Funct. Genet.* **2000**, *38*, 288.
(7) Steiner, T.; Koellner, G. *J. Mol. Biol.* **2001**, *305*, 535.
(8) Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. *J. Mol. Biol.* **2001**, *307*, 357.
(9) Burley, S. K.; Petsko, G. A. *Adv. Prot. Chem.* **1988**, *39*, 125.
(10) Creighton, T. E. *W. H.*; Freeman and Company: New York ,1993.
(11) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1.
(12) Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837.
(13) McGaughey, G. B.; Gagne, M.; Rappe, A. K. *J. Biol. Chem.* **1998**, *273*, 15458.
(14) Brocchieri, L.; Karlin, S. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9297.
(15) Samanta, U.; Pal, D.; Chakrabarti, P. *Acta Crystallogr.* **1999**, *D55*, 1421.
(16) Singh, J.; Thornton, J. M. Atlas of Protein Side-Chain Interactions; Vols. I and II. Oxford, UK: IRL press, 1992.
(17) Mitchell, J. B. O.; Laskowski, R. A.; Thornton, J. M. *Proteins* **1997**, *29*, 370.
(18) Karlin, S.; Zuker, M.; Brocchieri, L. *J. Mol. Biol.* **1994**, *239*, 227.
(19) Sawyer, L.; James, M. N. G. *Nature* **1982**, *295*, 79.
(20) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1995**, *254*, 96.
(21) 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.
(22) Hobohm, U.; Sander, C. *Protein Sci.* **1994**, *3*, 522.
(23) Bhattacharyya, R.; Samanta, U.; Chakrabarti, P. *Protein Engng.* **2002**, *15*, 91.
(24) Rao, C. R. *Linear Statistical Inference and Its Applications*; John Wiley: New York, 1973.
(25) Word, J. M.; Lovell, S. C.; LaBean, T. H.; Taylor, H. C.; Zalis, M. E.; Presley, B. K.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1711.
(26) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(27) Hubbard, S. ACCESS: A program for calculating accessibilities. Department of Biochemistry and Molecular Biology. University College of London; 1992.
(28) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(29) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946.
(30) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777.
(31) Laskowski, R. A. *J. Mol. Graph.* **1995**, *13*, 323.
(32) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Sci.* **1996**, *5*, 1001.
(33) Chakrabarti, P. *Protein Eng.* **1990**, *4*, 57.
(34) Samanta, U.; Chakrabarti, P.; Chandrasekhar, J. *J. Phys. Chem. A.* **1998**, *102*, 8964.
(35) Bork, P.; Koonin, E. V. *Curr. Opin. Struct. Biol.* **1996**, *6*, 366.
(36) Bairoch, A. *Nucleic. Acids Res.* **1993**, *21*, 3097.
(37) Bairoch, A.; Bucher, P.; Hofmann, K. *Nucleic. Acids Res.* **1997**, *25*, 217.
(38) Nishio, M.; Hirota, M.; Umezawa, Y. *The CH/π Interaction. Evidence, Nature, and Consequences*; Wiley-VCH: New York, 1998.
(39) Desiraju, G.; Steiner, T. *The Weak Hydrogen Bond in Structural Chemistry and Biology;* Oxford University Press: Oxford, 1999.
(40) Chakrabarti, P.; Samanta, U. *J. Mol. Biol.* **1995**, *251*, 9.
(41) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709.
(42) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price. S. L. *J. Mol. Biol.* **1994**, *239*, 315.
(43) Karlin, S.; Zhu, Z.-Y. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8344.
(44) Hellinga, H. W.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 763.
(45) Kleywegt, G. J. *J. Mol. Biol.* **1999**, *285*, 1887.

PR025584D