Energetic Fitness of Histidine Protonation States in PDB Structures

Giorgio F. Signorini, Riccardo Chelli,* Piero Procacci, and Vincenzo Schettino

Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, and European Laboratory for Non-linear Spectroscopy (LENS), Via Nello Carrara 1, 50019 Sesto Fiorentino, Italy

Received: May 26, 2004

The energetic fitness of histidine in each of its three protonation states has been investigated for NMR-determined protein structures by using molecular mechanics calculations. The protein structures have been taken from the Protein Data Bank (PDB). For the proteins in the database, we generated all isomers, considering all combinations of protonation forms of each histidine. The energy of each isomer has been determined by conjugate gradient minimization using a well-established all-atom force field. We find that, in general, the isomer reported in the PDB is not the most stable isomer. The statistical distribution of isomer energies minus that of the PDB isomer *behaves as though* the sequence of the histidine forms reported in the PDB was chosen at random. We also show that our molecular mechanics method is a valid approach to predicting the protonation state of histidines buried in the protein core.

1. Introduction

Histidine plays a fundamental role in biochemical events, acting as ligand in metallo-proteins,[1] as a catalyst in the active site of enzymes,[2] and as a proton shuffler in glycolysis reactions.[3] Histidine is also involved in the partial unfolding of important proteins such as apomyoglobin.[4] All these biochemical functions of histidine depend, directly or indirectly, on the two nitrogen atoms of the side-chain; thus, to understand the underlying chemical mechanisms, it is of primary importance to know their protonation state. In fact, histidine can exist in three forms (see Figure 1): two electrically neutral forms with either the N_δ or N_ϵ nitrogen atom protonated (HIS-D and HIS-E in Figure 1), and one positively charged form with both N_δ and N_ϵ atoms protonated (HIS-H in Figure 1).

In principle, HIS-D, HIS-E, and HIS-H can be distinguished in nuclear Overhauser effect two-dimensional NMR spectroscopy[5−7] (NOESY) or total correlation NMR spectroscopy[5−7] (TOCSY) by both intensity and frequency peak patterns. Very recently the experimental assignment of histidine tautomers by 2D $^1H/^{13}C$ correlated spectroscopy has been discussed.[8,9] In practice, however, the assignment in two- or three-dimensional NMR spectra is often focused on the peak pattern due to the backbone atoms in order to determine (i) the torsion angles (through the Karplus relation[10]) needed for secondary structure determination and (ii) inter-residue proton−proton distances to be fed in a simulated annealing constraint minimization using empirical force fields (typically AMBER[11] or CHARMM[12]) in order to resolve the tertiary structure.

When experimental data are lacking or insufficient, hydrogen positions in histidine are assigned—with computational tools that include energy minimization, sometimes together with a certain degree of arbitrariness—in the structure determination and/or refinement procedure (usually based on popular computational packages such as X-PLOR,[13] DYANA,[14] OPAL,[15] CNS,[16] XWINNMR (from Bruker)), or in post-structural analysis tools, such as PROCHECK[17,18] or WHAT-IF.[19] To give just one example, in the program WHAT-IF the histidine structure is tested and/or assigned by evaluating the root-mean-square *z*-score on the Engh and Huber[20] structural parameters (bond distances and bending angles) or by analyzing the associated *stationary* hydrogen bond network. Most of the assignments of the protonation state of histidine in PDB structures, including those resolved by NMR, are therefore not due to *direct* observation of the H_δ and/or H_ϵ (see Figure 1) chemical shifts, but rather to post-analysis and adjustment of the resulting model structure using computational or statistical methodologies.

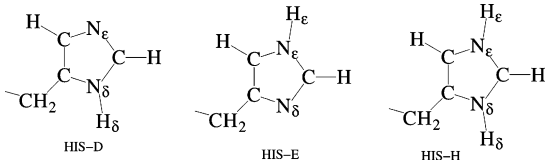

Figure 1. Protonation states of histidine.

From the modeling standpoint, an indeed important (and to our knowledge, unanswered) question is the following: how reliable are these "experimental" assignments? In fact, histidine is usually involved in hydrogen bonds, having stabilization energies comparable to the free energy difference between the folded and the unfolded states; molecular dynamics simulations or molecular mechanics calculations with the wrongly assigned protonation state of histidine could indeed lead to severely misleading conclusions. In the present study we have tested the assignment of the histidine protonation states reported in the PDB by using extensive full atomistic molecular mechanics calculations with the aid of the AMBER force field.[11] To this end, all possible histidine protonation states from selected PDB structures obtained from NMR experiments were generated and energy minimized. Computed lowest energy structures confirm the protonation state assignment from PDB for the cases corresponding to structures where chemical shifts of the H_δ and H_ϵ hydrogen atoms are available. In all other cases, the energy distribution of the protonation states reported in the PDB is similar to a random distribution of protonation states.

The article is organized as follows: in section 2 the protein database is described together with the computational method;

* To whom correspondence should be addressed. Fax: +390554573077. Phone: +390554573082. E-mail: chelli@chim.unifi.it.

10.1021/jp047716r CCC: $27.50
Published on Web 07/14/2004

TABLE 1: The Protein Database

histidines per protein[a]: N	no. of proteins:[b] $L(N)$	theoretical no. of energy minimizations:[c] $L(N) \times 3^N$	no. of energy minimizations actually performed[d]
1	159	477	477
2	107	963	547
3	61	1647	471
4	34	2754	330
5	22	5346	270
6	12	8748	172
7	2	4374	38
8	5	32805	137
9	2	39366	54
10	4	236196	142
total	408	332676	2638

[a] Number of histidines per protein (N). [b] Number of proteins (in the database) having N histidines [$L(N)$]. [c] Theoretical number of energy minimizations (protein isomers), i.e., considering 3^N isomers per protein. [d] Number of energy minimizations actually performed after application of the approximation of eq 1.

in section 3 the results are reported and discussed; the conclusions are presented in section 4.

2. Materials and Methods

2.1. Protein Database. Our study was carried out on a subset of the PDB database[21] (the PDB can be accessed from http://www.rcsb.org/pdb), restricted to non-homologue protein structures containing at least one histidine for which the positions of the H_δ and H_ϵ hydrogen atoms are experimentally known by NMR or derived with the aid of ancillary modeling tools.[13−19] This database has been further processed by eliminating all structures (i) including heteroatoms, (ii) with less than 35 amino acid residues, and (iii) with atoms species that do not match standard AMBER atomic types.[11] The reason for excluding heteroatoms from the protein database will become clear in section 3.2, where the energy distribution of superoxide dismutase, i.e., a protein where histidine acts as ligand for metal ions, is reported and discussed. The restriction on the number of residues is dictated by the fact that the solvent appears to be determinant for the stabilization of small proteins.[22] Since in our calculations we are neglecting the solvent contribution (see next section), to limit the number of cases where the solvent is sure to play a fundamental role, we have decided to exclude all small proteins. The third restriction comes only from reasons of practical order.

The resulting database includes 408 structures (the list of the PDB files of the protein database is available as Supporting Information). In this PDB subset, the number of *resolved* histidines per protein ranges from 1 to 10, the average value being 2.4 (see Table 1 for a breakdown of data). The ratio HIS-D:HIS-E:HIS-H is 0.70:0.07:0.23.

2.2. Molecular Mechanics Calculations. The assignment of the histidine protonation states given in PDB has been tested by performing a conjugated gradient energy minimization for all possible combinations of protonation states of the histidine side-chains. The adopted interaction potential is AMBER.[11] In principle, at each minimization step, using the zero-order field produced by the electric point charges, one should solve a Poisson−Boltzmann equation using a finite difference approach[23−25] in order to account for the free energy polarization of the solvent medium at room temperature. However, the computational cost of such a calculation would be prohibitive, and hence we have decided to neglect altogether the effect of the solvent in the stabilization of histidines by doing the minimization in vacuo. With this approach, we do expect a rigorous matching between minimum energy predicted isomers and PDB isomers only for those proteins that have histidines deeply buried in the protein core, where very likely only one protonation state is favored with respect to the others. We also expect matching with decreasing probability as the degree of solvent-exposure of the histidine side-chains increases.

For each protein isomer, energy minimization has been performed by adjusting only the Cartesian coordinates of all the atoms of the histidine side-chains, while all remaining residues have been left fixed at their original (PDB) site. The histidine residues are allowed to interact (by means of electrostatic and atom−atom potentials) with all other residues of the protein; that is, a cutoff radius has not been used. For each protein, the following procedure was applied. First, the minimum energy of the isomer reported in the PDB (the "PDB" isomer) was found by relaxing the histidine side-chains as stated above. Then, all possible HIS-D, HIS-E, or HIS-H forms of each histidine were considered, and for each resulting protein isomer the energy was minimized with the same method. Since the number of possible isomers per protein is 3^N, where N is the number of histidines in the protein, the computational effort can be rather challenging even for moderate values of N (see entries in column 3 of Table 1). To keep the number of calculations to a practical level, it has been assumed that, when all the nitrogen atoms (N_δ and N_ϵ) of two histidine side-chains are separated by more than 6 Å, the energy changes due to the substitution of one form with another on each histidine are independent and additive. In other terms, the energy of a protein with the forms of two histidines (X and Y) differing from those of the PDB isomer (A and B) is extrapolated from the energies of the single-substituted isomers X, B and A, Y:

$$E_{XY} = E_{AB} + (E_{XB} - E_{AB}) + (E_{AY} - E_{AB}) \qquad (1)$$

When the aforementioned independence condition of two histidine side-chains is not fulfilled, approximation (1) is not used, and all possible combinations of protonation states are taken explicitly into account. Using this methodology, we have done 2638 independent energy minimizations in vacuo (see entries in column 4 of Table 1) with an energy tolerance threshold of 10^{-6} kJ mol^{-1}. The approximation (1) has been tested on several proteins calculating E_{XY} in two ways: by a direct calculation and by using eq 1. In all cases, differences of the energies do not exceed 1 kJ mol^{-1}, well below (one or more orders of magnitude) the average energy differences calculated for different protonation states.

All calculations reported in the present work have been performed with the program ORAC.[26]

3. Results and Discussion

3.1. Distribution of Isomer Energies. For a protein with one or more histidines it is expected that, if the isomer reported in the PDB corresponds to the real structure, then its energy will be the lowest among those of all possible isomers. In other terms, the energy difference $\Delta E = E - E_{\mathrm{PDB}}$ will always be positive, independent of the isomer, and will be identically zero only for the *target* isomer, i.e., the one with the lowest energy. So, even considering the involved approximations, such as deficiencies in the potential model and lack of solvent screening for solvent-exposed histidines, one would expect that the global distribution of ΔE (that is, for all proteins) would be a function with significant nonzero values only in the positive x-axis side, with a relatively small negative-side tail. On the contrary, if the PDB structure were not systematically the lowest energy

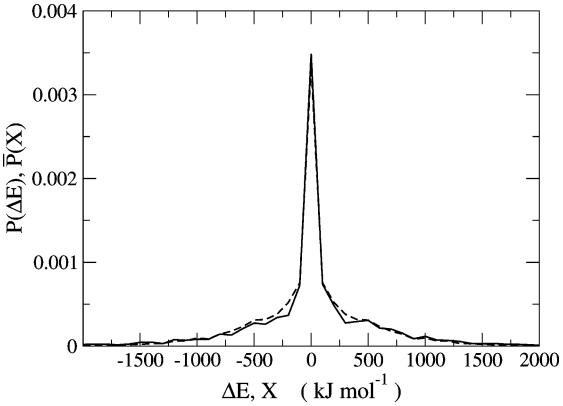

Figure 2. Solid line: global distribution function, $P(\Delta E)$, of the difference between the energy of an isomer and that of the PDB isomer of the same protein. The contributions of all proteins of the database (408 proteins) are each normalized to one and averaged. Dashed line: global distribution function, $\bar{P}(X)$, of the difference between the energy of an isomer and that of a reference isomer of the same protein, averaged over all reference isomers of each protein and over all proteins of the database (see eq 4).

structure of each protein, the distribution of ΔE would approach the appropriate distribution of a random quantity, symmetrically distributed around $\Delta E = 0$.

We found that the global distribution of ΔE is indeed of the latter type: see $P(\Delta E)$ in Figure 2. The function was built by normalizing to one the contribution of each protein, to prevent the global distribution from being dominated by proteins with a high number of isomers (high N), where statistics is poor (see entries in column 2 of Table 1). However, in this way, more weight is given to the classes of proteins that are more populated, i.e., those with low N. The distribution function has a sharp peak in the origin, and its width is about 100 kJ mol^{-1}.

It is worth noting that, even under the assumption that the PDB structure is randomly chosen among the isomers of each protein, the shape of the distribution of ΔE is not expected to be Gaussian. In fact, it is easy to recognize that the statistical distribution of the difference $\Delta E = E_n - E_n^{\mathrm{ref}}$ for a protein n, when averaged over all possible choices of E_n^{ref}, is the autocorrelation function of the single-protein energy distribution $P_n(\Delta E)$:

$$\bar{P}_n(\Delta E) = \int_{-\infty}^{\infty} P_n(\Delta E + Y)\, P_n(Y)\, \mathrm{d}Y \qquad (2)$$

and that, as such, it is an even function with maximum in 0. Now, $P_n(\Delta E)$, being the sum of 3^N independent variates, will indeed approach a Gaussian in force of the well-known *central limit theorem* (see for example Figure 3), but only for high N, and so will its autocorrelation function. Thus the global distribution function, given by the average of the L single-protein functions $\bar{P}_n(\Delta E)$,

$$\bar{P}(X) = \frac{1}{L}\sum_{n=1}^{L} \bar{P}_n(X) \qquad (3)$$

will have a Gaussian shape only if (a) high-N molecules predominate and (b) they all have approximately the same width. In general, however, it will be an even function with a pronounced maximum in the origin, which often (for example for a series of Gaussian or rectangular distributions with uniformly distributed width[27]) will assume a "witch's hat" shape like that of Figure 2, not a "bell" shape.

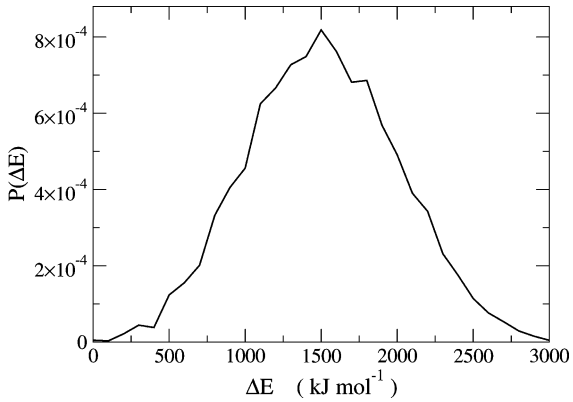

Figure 3. Distribution function, $P(\Delta E)$, of the difference between the energy of an isomer and that of the PDB isomer for superoxide dismutase. $P(\Delta E)$ has been normalized to one.

The global average distribution is clearly

$$\bar{P}(X) = \frac{1}{L}\sum_{n=1}^{L}\left[M_n^{-2}\sum_{i=1}^{M_n}\sum_{j=1}^{M_n}\delta(X - E_{ni} - E_{nj})\right] \qquad (4)$$

where δ is the Dirac function, M_n is the number of isomers of protein n, the sum on n runs over all the L proteins of the database, and E_{ni} is the energy of the ith isomer of the protein n. The distribution $\bar{P}(X)$ is reported in Figure 2 for comparison with $P(\Delta E)$. The strict resemblance between $P(\Delta E)$ and $\bar{P}(X)$ is a clear indication that the sequences of the histidine forms reported in the PDB *behave as though* they were chosen at random.

In the Supporting Information, we report the histidine protonation states for each protein of the database as obtained from our energy minimization protocol. The differences between PDB assignment and ours are also evidenced.

3.2. On the Reliability of the Method: The Cases of Superoxide Dismutase and Subtilisin BPN′. The reliability of our methodology for determining the energetic fitness of the target isomer can be convincingly tested by evaluating the energy differences between the protonation states of proteins where histidine tautomers are known with accuracy for having been determined directly through NMR measurements. As a first zero-order test we have analyzed superoxide dismutase (PDB code: *1ba9*). The structure of superoxide dismutase (SOD) in the reduced form has been determined using NMR,[1] and according to the authors, *there are spectroscopic evidences for the protonation properties* of histidine residues 43, 46, 48, 63, 71, 80, and 120. Histidines 46, 48, and 120 are coordinated to a reduced copper ion (Cu^+), and their forms are HIS-E, HIS-D, and HIS-D, respectively. Histidines 63, 71, and 80 are coordinated to a zinc ion (Zn^{2+}) and are all of type HIS-E. Histidine 43 (non metal coordinated) is of type HIS-H. In addition to these histidines of assigned type, there is another histidine in position 110 (indicated as HIS-D by the authors) for which no clear evidence exists about the protonation state. In this protein, all possible (6561) isomers have a higher energy than the PDB one (see Figure 3), with a single exception ($\Delta E = -1.2$ kJ mol^{-1}): the isomer where the seven assigned histidines have the same form as in the PDB, while histidine 110 (i.e., the one for which there is no experimental evidence of the structure) is of type HIS-E.

The case of SOD proves that our energy minimization protocol is a reliable tool for predicting the experimental protonation state of histidines that are coordinated to a metal

ion. However, this test may be seen as not particularly sensitive, since when histidine coordinates metal ions, there is in general a strong preference toward one protonation state, for both energetic and steric reasons. In fact, proteins with heteroatoms were not included in the analyzed database.

A much more severe test of the method can be performed considering proteins that have non-metal-coordinating histidines for which different protonation states might be energetically competitive. However in this case the additional factor of the solvent exposure of histidine residues should be taken into account. In fact, it may be argued that, as previously noted, because of the complete neglect of the solvation free energy, the in vacuo energy minimization procedure may be unreliable as a means of predicting the protonation state for partially or fully solvent-exposed histidines. We have then performed a calculation on a protein, subtilisin BPN′, that was found to have several histidines buried in the protein core that do not coordinate metal ions. Unfortunately the NMR structure of subtilisin BPN′ is not yet deposited in the PDB, and hence, in principle, we could not proceed to check the protonation state against the experimental assignment in solution. We then performed a mixed test: for checking the histidine protonation states we used a putative subtilisin BPN′ structure in solution starting from the X-ray PDB atomic coordinates in the solid state[28] (PDB code: *1sup*). Because hydrogen atoms are not included in the X-ray structure, we proceeded as follows: heavy atoms of the PDB structure of subtilisin BPN′ were saturated with hydrogen atoms using geometrical criteria based on the AMBER parameters. The HIS-E protonation state was arbitrarily chosen for all histidines. Then the atomic coordinates of the hydrogen atoms were energy minimized using the conjugate gradient method, by fixing the atomic coordinates of all heavy atoms to their PDB values. The resulting structure was assumed as the reference (the "PDB" reference). As discussed before, only buried histidines are considered. As found by Day et al. in a recent NMR investigation,[8] subtilisin BPN′ has four of its six histidine residues (17, 39, 67, and 226) that are neutrally charged and do not titrate. Our analysis will then focus on these four residues, by leaving the two other histidines (64 and 238) in the arbitrary protonation state (HIS-E) of the reference structure. In addition, since the protonated form, HIS-H, is not accessible experimentally, we limit the comparison to the HIS-D and HIS-E protonation states alone. The energy difference between the reference state (all HIS-E histidines) and all other possible protein isomers is reported in Table 2. The lowest energy isomer of subtilisin BPN′ has the histidines 17, 39, and 67 in the HIS-E protonation state, whereas only the histidine 226 is in the HIS-D protonation state. This was indeed what Day et al. found from their experiments.[8]

In addition to these two examples corroborating our molecular mechanics approach, results of other tests of the method are discussed in the next section. In particular we report there on calculations performed on proteins whose structure and histidine protonation states are resolved by NMR spectroscopy.

3.3. Distribution of Isomer Energies in the Restricted Database. In the analyzed database, the degree of solvent-exposure of histidine side-chains spans the entire range from deeply buried residues to superficial or fully solvated residues. For reasons discussed in several parts of the article, we built a second statistics by restricting the database only to proteins that have at least one histidine deeply buried in the core. To identify "buried" histidine side-chains in a protein, we have calculated the Voronoi polyhedron volumes[22,29,30] for each atom of the minimized PDB structure of the protein. Note that, for a generic molecule in vacuo, it is not possible to calculate the Voronoi polyhedron volumes for all the atoms, because Voronoi polyhedra cannot be defined for the atoms "on the surface" of the molecule. To meet this deficiency, we considered the protein to be embedded in a large (compared to the protein dimensions) cubic box with eight dummy particles at its vertexes. Then, for each atom of the protein, the volume of the Voronoi polyhedron has been computed, using, when needed, the dummy particles for closing the polyhedron. Large volumes indicate exposed atoms, small volumes buried ones. In our specific case, a histidine side-chain was assumed to be buried if the largest Voronoi polyhedron volume among those of N_δ, H_δ, N_ϵ, and H_ϵ did not exceed 25 Å^3 (in passing we remark that, in the case of subtilisin BPN′, this method classifies buried histidines 17, 39, 67, and 226, the same as NMR experiments).

TABLE 2: Energy of Histidine Protonation States for Subtilisin BPN′[a]

His17	His39	His67	His226	ΔE
E	E	E	D	−94.2
E	E	D	D	−68.5
E	D	E	D	−39.3
E	D	D	D	−13.6
D	E	E	D	−4.0
E	E	E	E	0.0
D	E	D	D	21.6
E	E	D	E	25.7
D	D	E	D	50.9
E	D	E	E	54.9
D	D	D	D	76.5
E	D	D	E	80.6
D	E	E	E	90.2
D	E	D	E	115.8
D	D	E	E	145.1
D	D	D	E	170.7

[a] Only the protonation state of the buried (in the protein core) histidines is reported (D ≡ HIS-D, E ≡ HIS-E). The protonation state of the solvent-exposed histidines (64 and 238) is arbitrarily assumed of type HIS-E. The energy differences (ΔE) between the reported isomer of subtilisin BPN′ and the reference isomer formed by all HIS-E are in kJ mol^{-1}.

In a restricted statistics one should also take care of the fact that, since the double-protonated form HIS-H, even though energetically favorable, may be not accessible experimentally, we have limited the comparison (as done for subtilisin BPN′) to the HIS-D and HIS-E protonation states alone.

If both these restrictions (at least one buried histidine and no HIS-H forms) are enforced, the original set is reduced to 63 proteins (the list of the PDB files of the restricted protein database is available as Supporting Information). Since in this case we are interested in buried histidines alone, the calculations have been performed by changing only their protonation state, while the protonation state of solvent-exposed histidines is kept equal to that of the PDB structure. This choice for the restricted database gives a total of 171 isomers (including the PDB isomers). The distribution of the energy differences, $P(\Delta E)$, is shown as a histogram in Figure 4. Note that the contribution of the trivial case $E = E_{\mathrm{PDB}}$ to $P(0)$ is omitted for the sake of clarity (the total number of isomers thus becomes 108). The average distribution $\bar{P}(X)$ is also shown in Figure 4, the trivial case ($E_{ni} = E_{nj}$ in eq 4) also being omitted. In the first instance, the figure shows, especially in view of the poor statistics of the sample, that ΔE is still distributed as a random variable. Second, we note that the protonation state of histidine is experimentally known for four proteins of this restricted set. Three out of these four proteins (*1cfe*, *1ji8*, and *1jwe*) were identified by cross-referencing the 63 proteins of the restricted database with the

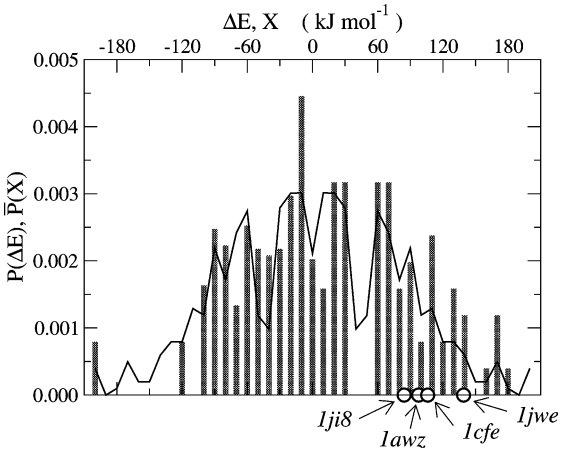

Figure 4. Histogram: global distribution function, $P(\Delta E)$, of the difference between the energy of an isomer and that of the PDB isomer of the same protein. The contributions of all proteins of the database (63 proteins) are each normalized to one and averaged. In $P(0)$, the contribution of the trivial case $E = E_{PDB}$ is not considered for the sake of clarity. Solid line: global distribution function, $\bar{P}(X)$, of the difference between the energy of an isomer and that of a reference isomer of the same protein, averaged over all reference isomers of each protein and over all proteins of the database (see eq 4). In $\bar{P}(0)$, the contribution of the trivial case $E_{ni} = E_{nj}$ is not considered for the sake of clarity. Circles: differences between the energy of the non-PDB and of the PDB isomer of the proteins *1cfe*, *1awz*, *1ji8*, and *1jwe* (see text for details).

BioMagResBank[31] (BMRB) database (the BMRB database was downloaded from http://www.bmrb.wisc.edu) by searching for HIS.HD1 or HIS.HE1 NMR chemical shifts. The fourth protein, human angiogenin (*1awz*), was identified by searching the word "histidine" in the PDB headers of the restricted protein database. In this case the protonation state of histidine 47 was determined by $^1H/^{15}N$ NMR spectroscopy.[32] Each protein contains more than one histidine residue, but all have only one buried histidine side-chain of type HIS-D (*1cfe*), HIS-D (*1awz*), HIS-E (*1ji8*), and HIS-D (*1jwe*). Thus for each protein only one isomer, in addition to the PDB one, can be built. The energy difference, $\Delta E = E - E_{PDB}$, of the isomers is marked by circles on the energy axis of Figure 4. We can see that all four isomers have a higher energy than that of the PDB isomer.

These results along with those obtained for subtilisin BPN′ all evidence that the molecular mechanics method we propose is a valid approach for assigning the protonation state for histidines buried in the protein core.

The protonation state obtained from the energy minimization protocol for the buried histidines of the restricted protein database is reported in the Supporting Information.

4. Conclusions

A study of the energetic fitness of histidine protonation states in NMR-determined structures taken from the Protein Data Bank (PDB) has been carried out by means of molecular mechanics calculations. As in metal-coordinating histidines there is a strong preference toward a well-defined protonation state, metalloproteins have not been included in the analyzed set of proteins. For each protein of the database, we have generated all isomers, considering all possible combinations of protonation forms of each histidine. Then we have performed energy minimization in vacuo for each of these isomers using the conjugate gradient method and a well-established all-atom force field. Where applicable, the energy of an isomer was obtained as the sum of independent contributions. The statistical distribution of the energy difference $P(E_n - E_n^{ref})$, where E_n is the energy of a generic minimized isomer of protein n and E_n^{ref} that of the minimized PDB isomer of the same protein, averaged on all the proteins of the database, turned out to be centered in $E_n - E_n^{ref} = 0$. An almost identical distribution has been obtained when the reference energy E_n^{ref} is averaged over all isomers of a protein, thus indicating that the sequences of the histidine forms reported in the PDB *behave as though* they were chosen at random. By contrast, in superoxide dismutase and in subtilisin BPN′ (not included in the database), where the protonation state for several histidines has been determined experimentally, our protonation state assignment agrees systematically with the experimental data.

An essentially random distribution has also been observed by restricting the statistical analysis to the HIS-E and HIS-D forms and eliminating partially or fully solvent-exposed histidines, for which the in vacuo energy minimization procedure is more questionable. By intersecting this restricted database with the BioMagResBank database of NMR chemical shifts for searching proteins with at least one histidine experimentally determined, we find that all NMR histidine assigned isomers correspond to the lowest energy isomer, thus indicating (as in the superoxide dismutase and subtilisin BPN′ cases) a good correlation between energy fitness as determined using molecular mechanics and experimental structures determined by direct observation of the relevant chemical shifts.

This study shows that, when the protonation state of histidines in a protein is not directly accessible from the experiment, our energy minimization scheme provides a viable approach for predicting histidine structure. The proposed methodology works well for histidines that are deeply buried in the protein core, while it is less reliable for histidines on the protein surface. To extend the present protocol also to solvent-exposed histidines, we are investigating the possibility of introducing in our scheme (at an acceptable computational cost) the solvent contribution using analytical (i.e., amenable to force field based energy minimization) implicit-solvent Poisson−Boltzmann[33] or generalized Born[34]-like approaches.

Acknowledgment. We are grateful to Antonio Rosato and Mario Piccioli for stimulating discussions. This work was supported by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca and by the European Union under contracts ERB-FMGE-CT950017 and HPRI-CT-1999-00111.

Supporting Information Available: (1) PDB codes of the proteins of the database; for each protein the protonation state of histidines as found in the PDB and as obtained using our computational approach is reported. (2) Notes on the theory of the statistical distribution of isomerization energies in a protein ensemble. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Banci, L.; Benedetto, M.; Bertini, I.; Del Conte, R.; Piccioli, M.; Viezzoli, M. S. *Biochemistry* **1998**, *37*, 11780−11791.

(2) Krupenko, S. A.; Vlasov, A. P.; Wagner, C. *J. Biol. Chem.* **2001**, *276*, 24030−24037.

(3) Zhang, Z. D.; Komives, E. A.; Sugio, S.; Blacklow, S. C.; Narayana, N.; Xuong, N. H.; Stock, A. M.; Petsko, G. A.; Ringe, D. *Biochemistry* **1999**, *38*, 4389−4397.

(4) Barrick, D.; Hughson, F. M.; Baldwin, R. L. *J. Mol. Biol.* **1994**, *237*, 588−601.

(5) Günther, H. *NMR−Spektroskopie*, 361; Georg Thieme Verlag: Stuttgart, 1992.

(6) Canet, D. *NMR−Konzepte und Methoden*, 184; Springer-Verlag: Berlin, 1994.

(7) Evans, J. *Biomolecular NMR Spectroscopy*, 71; Oxford University Press: Oxford, 1995.
(8) Day, R. M.; Thalhauser, C. J.; Sudmeier, J. L.; Vincent, M. P.; Torchilin, E. V.; Sanford, D. G.; Bachovchin, C. W.; Bachovchin, W. W. *Protein Sci.* **2003**, *12*, 794−810.
(9) Sudmeier, J. L.; Bradshaw, E. M.; Haddad, K. E. C.; Day, R. M.; Thalhauser, C. J.; Bullock, P. A.; Bachovchin, W. W. *J. Am. Chem. Soc.* **2003**, *125*, 8430−8431.
(10) Karplus, M. *J. Phys. Chem.* **1959**, *30*, 11−15.
(11) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(12) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(13) Brunger, A. T. *X-PLOR, Version 3.1: A system for x-ray crystallography and NMR*; Yale University Press: New Haven, 1992.
(14) Guntert, P.; Mumenthaler, C.; Wuthrich, K. *J. Mol. Biol.* **1997**, *273*, 283−298.
(15) Luginbuhl, P.; Guntert, P.; Billeter, M.; Wuthrich, K. *J. Biomol. NMR* **1996**, *8*, 136−146.
(16) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D* **1998**, *54*, 905−921.
(17) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. *PROTEINS* **1992**, *12*, 345−364.
(18) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(19) Vriend, G. *J. Mol. Graphics* **1990**, *8*, 52−56.
(20) Engh, R. H.; Huber, R. *Acta Crystallogr. A* **1991**, *47*, 392−400.
(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*, 235−242.
(22) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *PROTEINS* **2004**, *55*, 139−151.
(23) Gilson, M. K.; Honig, B. *PROTEINS* **1988**, *4*, 7−18.
(24) Gilson, M. K.; Honig, B. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 5−20.
(25) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(26) Procacci, P.; Darden, T.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848−1862.
(27) Signorini, G. F. Internal report, Chemistry Department, Università di Firenze, 2003. This report is available as Supporting Information.
(28) Gallagher, T.; Oliver, J.; Bott, R.; Betzel, C.; Gilliland, G. L. *Acta Crystallogr. D* **1996**, *52*, 1125−1135.
(29) Voronoi, G. M. *J. Reine Angew. Math.* **1908**, *134*, 198−287.
(30) Procacci, P.; Scateni, R. *Int. J. Quantum Chem.* **1992**, *42*, 1515−1528.
(31) Seavey, B. R.; Farr, E. A.; Westler, W. M.; Markley, J. L. *J. Biomol. NMR* **1991**, *1*, 217−236.
(32) Lequin, O.; Thüring, H.; Robin, M.; Lallemand, J. Y. *Eur. J. Biochem.* **1997**, *250*, 712−726.
(33) Marchi, M.; Borgis, D.; Levy, N.; Ballone, P. *J. Chem. Phys.* **2001**, *114*, 4377−4385.
(34) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606−10614.