Topological Analysis of Hydrogen Bonds and Weak Interactions in Protein Helices via Transferred Experimental Charge Density Parameters

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · AUGUST 2011

Impact Factor: 2.69 · DOI: 10.1021/jp2040289 · Source: PubMed

6 AUTHORS, INCLUDING:

Dorothee Liebschner
High Energy Accelerator Research Organizati…
20 PUBLICATIONS **116** CITATIONS
SEE PROFILE

Christian Jelsch
French National Centre for Scientific Research
127 PUBLICATIONS **1,992** CITATIONS
SEE PROFILE

Enrique Espinosa
University of Lorraine
143 PUBLICATIONS **3,947** CITATIONS
SEE PROFILE

Benoit Guillot
University of Lorraine
74 PUBLICATIONS **759** CITATIONS
SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY A

ARTICLE

pubs.acs.org/JPCA

Topological Analysis of Hydrogen Bonds and Weak Interactions in Protein Helices via Transferred Experimental Charge Density Parameters

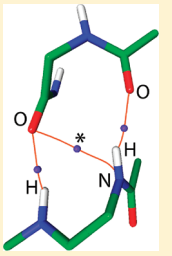
Dorothee Liebschner,[†] Christian Jelsch,[†] Enrique Espinosa,[†] Claude Lecomte,[†] Eric Chabrière,[‡] and Benoît Guillot*[,†]

[†]Crystallographie, Résonance Magnétique et Modélisations, Institut Jean Barriol, CNRS UMR 7036, Nancy-Université, Vandoeuvre-lès-Nancy, France

[‡]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, CNRS UMR 6236, Université de la Méditerranée, Marseille, France

Supporting Information

ABSTRACT: Helices represent the most abundant secondary structure motif in proteins and are often involved in various functional roles. They are stabilized by a network of hydrogen bonds between main chain carbonyl and amide groups. Several surveys scrutinized these hydrogen bonds, mostly based on the geometry of the interaction. Alternatively, the topological analysis of the electron density provides a powerful technique to investigate hydrogen bonds. For the first time, transferred experimental charge density parameters (ELMAM database) were used to carry out a topological analysis of the electron density in protein helices. New parameters for the description of the hydrogen bond geometry are proposed. Bonding contacts between the amide N and carbonyl O atoms (N···O) of helices, poorly addressed in the literature so far, were characterized from topological, geometrical, and local energetic analyses. Particularly, a geometrical criterion allowing for the discrimination between hydrogen bonds and N···O contacts is proposed.



INTRODUCTION

In proteins, the polypeptide chain adopts a specific 3-dimensional structure that is related to the properties and the function of the macromolecule.[1] Approximately 37% of the amino acids are in the helical geometry; the other residues are distributed between β-sheets (22%), β-turns (22%), and random conformation (19%).[2] The helix is, thus, the most abundant secondary structure motif that can be found in proteins. Various functional roles can be ascribed to helices, like DNA-binding, membrane spanning,[1] or electrostatic interactions due to their dipole moments.[3−5] To form a helix, the protein main chain adopts a coiled conformation (the side chains pointing outward) and is stabilized by hydrogen bonds (HBs) between backbone amide N−H and carbonyl C═O groups. Textbooks differentiate between α, 3_{10}, and π-helices. In α-helices, the HBs occur typically between N−H and C═O groups separated by four amino acids (noted $i \rightarrow i+4$), in 3_{10}-helices, the HB partners are separated by only three residues (noted $i \rightarrow i+3$). π-Helices are characterized by an $i \rightarrow i+5$ hydrogen bonding pattern. In this description, α-helices are far more common than 3_{10}-helices and they are also much longer: the mean length of α-helices is 12.6 residues, whereas 3_{10}-helices reach only a mean length of 3.3 residues.[2] π-Helices appear to be extremely rare, the large majority consists of seven-residue fragments.[6]

Various surveys scrutinized the hydrogen bonds in helices and, in particular, it has been shown that the hydrogen bond network in real protein structures is more complicated than implied by the definition given above. Thus, α-helices contain also $i \rightarrow i+3$ hydrogen bonds and in 3_{10}-helices, $i \rightarrow i+4$ HBs are observed, too,[2,7,8] implicating that some O and H atoms form more than one HB interaction. The configuration where the oxygen atom accepts two HB interactions simultaneously is called a double hydrogen bond,[2] whereas the case of one donating H atom to two O atoms is termed bifurcated hydrogen bond.[9] Helices are therefore stabilized by double and bifurcated hydrogen bonds, which succeed one after another. Several studies focused also on the geometry of such hydrogen bonds.[2,8,10] Accordingly, $i \rightarrow i+4$ bonds have shorter donor−acceptor distances and larger angles than $i \rightarrow i+3$ bonds (the average N···O distances and N···O═C angles being around 3.0 Å and 150° for the former[8,10] and around 3.1 Å and 120° for the latter[8]). The N−H and C═O groups in $i \rightarrow i+4$ hydrogen bonds are therefore more aligned, and the corresponding H···O interaction more energetic (the hydrogen bond being a directional interaction[9]). Interestingly, the study of Thomas et al.[10] pointed out that the amide N−H group is, in general, far from being located in the oxygen sp^2 plane, which contains the electron lone

Special Issue: Richard F. W. Bader Festschrift

Received: April 30, 2011
Revised: July 11, 2011
Published: August 01, 2011

Table 1. Some Properties of the Protein Models Used for this Study

	PDB code	R [%]	R_{free} [%]	resolution [Å]	number of residues	deposition date
aldose reductase[24]	n.a.[a]	9.1	9.4	0.66	316	n.a.[a]
β-glucosidase[25]	1UG6	12.2	13.0	0.99	431	2003
CTS-M-9-β-lactamase[26]	1YLJ	10.8	12.9	0.98	263	2005
xylanase[27]	1I1W	9.0	10.6	0.89	303	2001
endpopoly galacturonase[28]	1K5C	11.5	14.0	0.96	335	2001
concavanalin A[29]	1NLS	12.7	15.4	0.94	237	1997

[a] The high resolution model of aldose reductase was available in our laboratory.[24]

pairs. The hydrogen atom interacts with the oxygen atom with an average angle of 58° from the lone pairs plane.

Alternatively to the investigation of the hydrogen bond geometry, the topological analysis of the electron density provides a powerful technique to study this kind of interaction. Indeed, the topological analysis of the electron density[11,12] ρ in the H···acceptor region is a common method in small molecule crystallography.[13] This method consists in analyzing the gradient vector field (first derivative) of the electron density $\nabla\rho$ to characterize the topological features of ρ. In particular, the points where $\nabla\rho$ vanishes are called critical points (CPs). Diagonalization of the second derivatives matrix (Hessian matrix) at the CP positions yields eigenvalues and eigenvectors, which correspond to the main curvatures and axes of the electron density distribution at these particular points of the molecular structure. The volume defined by the zero flux surface of the electron density around a nucleus is termed the atomic basin. The addition of these atomic regions can be used as a partitioning scheme of the molecular or crystalline space. At critical points of covalent bonds and hydrogen bond interactions, the electron density distribution is characterized by one positive curvature, λ_3 (along the direction of the interaction pathway), and two negative ones, λ_1 and λ_2 (in the plane perpendicular to the bond path). Using a large set of experimental data comprising 83 X−H···O (X = C, N, O) hydrogen bonds, some relationships between geometrical, topological, and energetic parameters have been evidenced.[14,15] The exponential dependence of the positive curvature λ_3 on the donor−acceptor H···O distance can be noted in particular.[13] All the experimental functional dependencies observed for H···O interactions have been recently extended to other HBs in a theoretical study involving 163 H···X (X = H, C, N, O, F, S, Cl, π) interactions.[16]

In several studies, a topological analysis of the theoretical ab initio electron density in helix hydrogen bonds has been carried out:[17−20] Parthasarathi et al.[18] evidenced a linear relationship between the relative energy of the helix and the sum of the electron density at the H···O CPs. The survey of LaPointe et al.[19] proposed the existence of another interatomic interaction in helices, which occurs between the amide N atom and the carbonyl O atom of residue i-3 (noted as N···O). Furthermore, Vener et al.[20] showed the presence of hydrogen bonds between carbonyl oxygen atoms and side chain hydrogen atoms (C_β−H).

A topological analysis can nowadays be performed using either theoretical or experimental electron densities. Notably, the ELMAM library, which has been built on the basis of experimental charge density parameters of isolated amino acids and small peptides, permits to rebuild the precise electron density of proteins.[21,22] The aim of this study is to carry out a topological analysis of the electron density in helices, using transferred experimental electron densities from the ELMAM database and experimental high resolution protein structures. Additional parameters are introduced for the description of the hydrogen bond geometry.

MATERIALS AND METHODS

Choice of the Protein Structures. The topological analysis was performed on main chain interactions between N−H and C═O groups from six high resolution protein structures, most of them being retrieved from the Protein Data Bank.[23] The criteria of choice were the resolution of the diffraction data (better than 1 Å), the R-factors (R < 15% and $R_{free} - R$ < 5%) and the deposition date (no model before 1995). The proteins used for the analysis and some properties of the crystal structures are summarized in Table 1.[24−29]

To simplify the analysis, the multiple conformations in the proteins aldose reductase, CTS-M-9-β-lactamase, endopolygalacturonase I, and Xylanase have been erased with the "pdb-cure" module of the CCP4 suite.[30] Only the conformations with the highest respective occupations are then kept in the model.

The use of high resolution structures has the advantage that atomic coordinates are obtained with a higher degree of accuracy. Hence, such a resolution allows the precise positioning of hydrogen atoms either from precise C,N neighboring atom positions or on the basis of experimental electron density peaks in case of mobile H atoms.[31] However, hydrogen atoms were missing on most of the models in the PDB and, in such cases, were generated using the MOLPROBITY server.[32] As relevant hydrogen atoms for this study are bound to main chain nitrogen atoms, their positions depend only on the position of their parent nitrogen atom, as well as on the stereochemistry of the corresponding main chain peptide group. Because the heavy atom positions were obtained with subatomic resolution accuracy, no refinement steps were carried out neither before nor after hydrogen atom generation.

Topological Analysis. A formal topological analysis of the electron density can, in principle, be carried out using a spherical atom model. However, this approach does not take into account the deformation of the electron density due to shared- and closed-shell interactions. Protein crystals usually yield diffraction data whose resolution is not high enough to conduct the structure refinement with a nonspherical atom model. Therefore, a nonspherical description of the electron density has to be obtained by other means, like theoretical calculations or transfer from an electron density database.

The ELMAM experimental electron density database[21,22] was created by modeling the electron density of several subatomic resolution peptide crystal structures, via the multipolar atom formalism of Hansen and Coppens[33] and by averaging the charge density parameters of different atoms of the same chemical type.

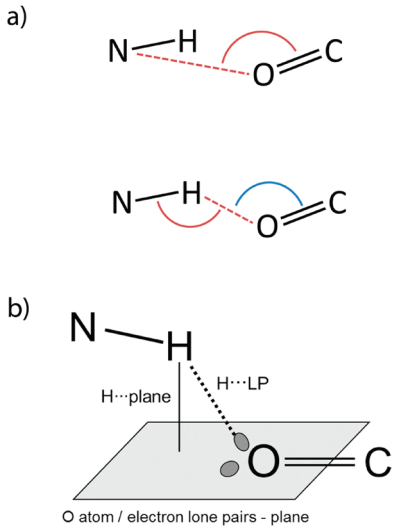


Figure 1. (a) Scheme of a hydrogen bond illustrating the geometrical parameters calculated for each interaction. Top: N···O distance and N···O=C angle. Bottom: H···O distance, N−H···O and H···O=C angles. (b) Scheme illustrating the geometrical parameters H···LP and H···plane (see text) involving the position of the oxygen electron lone pairs.

In this context, atom types are defined by the chemical species they belong to and by their neighboring and hybridization state. Therefore, experimental multipolar parameters for all 20 amino acid side and main chains are available in the ELMAM library and can be transferred to protein structures to rebuild their precise electron density.

Atomic charges and multipolar parameters (up to octupoles for C, O, and N atoms and up to dipoles for H atoms) were transferred from the ELMAM database to the six protein models listed above. Afterward, the covalent X−H bonds were elongated up to average distances observed in small molecule neutron crystallography.[34] This procedure accounts for the fact that the electron density of a hydrogen atom is shared within a covalent bond. Hence, the experimentally observed electron density peak in the Fourier synthesis does not truly correspond to the position of the hydrogen nucleus. Modifying in this way the hydrogen atom position (N−H distance elongated up to 1.009 Å) is compatible with the transfer procedure, which corrects the electron density description with a dipole oriented along the X−H covalent bond.

The topological analysis of the electron density has been carried out with the VMoPro software.[35] Water molecules and ligand molecules have not been taken into account and the search was limited to interactions between the main chain N−H and C=O groups. The positions of the CPs are calculated by minimizing the electron density gradient norm. To quicken the search process, a distance cutoff of 2.9 Å has been applied between the main chain H and O atoms. Additionally, the following criteria, based on a theoretical study of hydrogen bonds,[36] were used to consider a critical point as relevant for describing such an interaction: (a) The critical points must satisfy a correct topology, that is, the CP lies between the two interacting atoms and a bond path exists; (b) The electron density (ρ_{cp}) as well as its Laplacian ($\nabla^2\rho_{cp}$) at the critical point are larger than 0.002 a.u. (0.013 e/Å^3) and 0.024 a.u. (0.57 e/Å^5), respectively.

Hydrogen Bond Parameters. The calculation resulted in finding the critical points of all interactions between amide N−H and carbonyl C=O groups. These interactions were then classified according to secondary structure elements (helix, antiparallel sheet, parallel sheet). Secondary structure assignment was carried out with program STRIDE,[37] allowing a +1 residue tolerance to take into account that the results of various assignment programs are often slightly different,[38] especially at the extremities of the structural elements.

For helices, the interactions are categorized according to their type: $i \rightarrow i+3$ or $i \rightarrow i+4$. For each critical point, topological and geometrical parameters were calculated, some of them being different from those usually reported for HB geometries:

- H···O and N···O distances, N−H···O, N···O=C, and H···O=C angles (Figure 1a).
- Values of the electron density (ρ_{cp}), the Laplacian ($\nabla^2\rho_{cp}$), and the eigenvalues of the Hessian matrix (λ_1, λ_2, λ_3).
- H···LP, namely, the distance between the hydrogen atom and the nearest oxygen atom lone pair.
- H···plane, namely, the distance between the hydrogen atom and the peptide bond plane, which includes the oxygen atom lone pairs (Figure 1b). To estimate the positions of the carbonyl oxygen atoms lone pairs, the maxima of ($\nabla^2\rho(r)$) near the O atom were located. The O···LP distance and the C=O···LP angle are 0.34 Å and 109.5°, respectively, as estimated from the parameters of the ELMAM charge density database.
- The atoms located at the extremities of each bond path. Indeed, as shown later on, it is important to test which atoms are actually connected by a bond path, as the N···O interaction in helices, reported by LaPointe et al.,[19] can be misinterpreted as a hydrogen bond.

Most of the geometrical parameters take explicitly into account the position of the hydrogen atoms, contrarily to several surveys in the literature, which use only heavy atom positions (N, C, O) for the identification and analysis of HBs.[2,8,39] As the orientation of the N−H group in relation to the C=O group is crucial for the interpretation of a hydrogen bond, it is preferable to consider the position of the H atoms.

It should be stressed that new parameters, related to the positions of the oxygen electron lone pairs, have been introduced. Indeed, surveys of small molecule structures in the Cambridge Structural Database[40] showed a tendency for H-bonding to occur in the direction of the oxygen atom lone pairs.[41,42]

Energy Computations. The electrostatic interaction energy between two charge distributions A and B is given by

$$E_{elec} = \int V_B(\vec{r})\rho_A(\vec{r})\,\mathrm{d}\,\vec{r} \qquad (1)$$

where V_B is the electrostatic potential created by the charge distribution B and ρ_A is the charge distribution A. The integration is on total space, that is, where ρ_A is nonzero. For HBs and N···O contacts, the C=O group was chosen as fragment A, and the interacting N−H group as fragment B. Each fragment was therefore composed of two atoms. The integration was carried out with VMoPro software,[35] using the same transferred electron density parameters as for the topological analysis.

In order to estimate the total interaction energy of a HB (E_{HB}), the phenomenological equation of Espinosa et al.[14] has been used, which relates E_{HB} to the potential energy density V_{cp} at the

Table 2. Distribution of the Closed-Shell Bond Critical Points Among Secondary Structure Elements of the Six Proteins Studied

interaction	helix	AP sheet[a]	P sheet[b]	other[c]
all	672[d]	144	296	215
$i \rightarrow i+4$	364			
$i \rightarrow i+3$	141			
N···O	162	1	3	15

[a] AP = antiparallel β-sheet. [b] P = parallel β-sheet. [c] The CPs in the category "other" do not belong to helices or sheets. [d] The five missing interactions in helices correspond to $i \rightarrow i+5$ hydrogen bonds.

bond critical point:

$$E_{HB} = \frac{1}{2}V_{cp} \quad (2)$$

The proportionality factor is in volume atomic units and V_{cp} is in energy per volume atomic units. Knowing ρ_{cp} and $\nabla^2\rho_{cp}$, V_{cp} can be computed using the local form of the virial theorem[11] and the Abramov expression for the electron kinetic energy density at CPs in closed-shell interactions.[43]

The figures displaying the gradient lines and bond paths are computed with VMoPro software and drawn with MoProViewer.

RESULTS AND DISCUSSION

Overall Results of the CP Analysis. A total of 1327 critical points have been found between carbonyl and amide groups in the six proteins used for this study. Table 2 shows the distribution of the CPs among secondary structural elements. A total of 51% of the interactions are in helices, 11% and 22% occur in antiparallel (AP) and parallel (P) β-sheets, respectively, and 16% involve residues that do not adopt a specific structural motif.

By checking the nature of the atoms linked by a bond path, a significant number of N···O contacts, 181 in total, was revealed. As far as the critical point is located on the bond path connecting the amide N and carbonyl O atoms of the peptide units, this type of contact can be considered as a bonding interaction. With 162 CPs, the majority of these N···O contacts occurs in helices. Among them, 138 adopt a $N_{i+3}\cdots O_i$ interaction pattern, whereas the remaining 24 contacts are found between residues separated by 4, 2, 5, and 6 residues (16, 5, 2, and 1 interactions, respectively). If not stated otherwise, the following considerations (geometry, topology, energy) and plots involve $N_{i+3}\cdots O_i$ contacts only, the number of the other interactions being too small to carry out a statistical analysis. The N···O contacts are absent in β-sheets and very few are observed between residues of coiled structure. As the interaction has also been detected in a theoretical study of helical segments,[19] the existence of the N···O interactions characterized by our data does not seem to be an artifact due to the method used here, which is based on transferred multipolar electron densities.

In helices, the large majority of CPs belongs to $i \rightarrow i+4$ hydrogen bonds (54%), whereas only 21% of the interactions correspond to $i \rightarrow i+3$ HBs. The number of N···O contacts constitutes 24% and is therefore not negligible. Only five hydrogen bonds of type $i \rightarrow i+5$ have been identified, as this quantity is small, they will be disregarded in this study. Comparing the number of CPs versus the number of oxygen atoms, it turns out that one O atom is involved in 1.3 interactions (HBs or N···O contacts) on average. Therefore, in addition to the classical approach stating that protein helices are exclusively composed of either $i \rightarrow i+4$ or $i \rightarrow i+3$ hydrogen bonds, our results strongly suggest that helices are in fact maintained by a rich network of interactions (hydrogen bonds and N···O contacts). Figure 2 shows an example of the three interactions in a helix of CTS-M-9-β-lactamase.

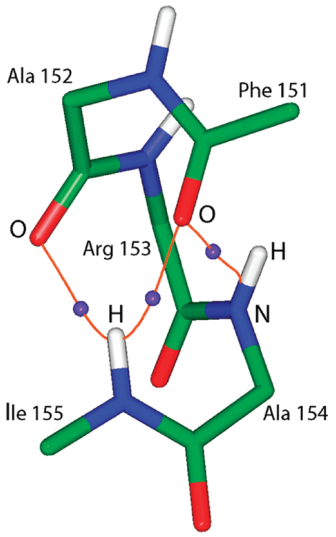


Figure 2. Succession of N···O contact ($O_{151}\cdots N_{154}$), $i \rightarrow i+4$ ($O_{151}\cdots H_{155}$), and $i \rightarrow i+3$ ($O_{152}\cdots H_{155}$) hydrogen bonds in the protein CTS-M-9-β-lactamase. The orange lines and violet spheres represent the bond paths and bond CPs, respectively. Only the main chain of the residues is represented.

Table 3. Average Geometrical[a] and Topological Parameters[b] of the 668 Hydrogen Bonds and N···O Contacts in Helices[c]

interaction	$i \rightarrow i+4$	$i \rightarrow i+3$	N···O
distance [Å]			
N···O	2.98(13)	3.09(13)	3.17(8)
H···O	2.03(16)	2.22(19)	2.66(11)
H···LP	1.86(14)	2.16(22)	2.65(12)
H···plane	0.79(30)	2.01(33)	2.63(13)
angle [°]			
N···O=C	155(7)	118(9)	109(5)
N-H···O	158(10)	147(15)	111(5)
H···O=C	149(8)	110(11)	93(6)
topology			
ρ_{cp} [e/Å^3]	0.10(4)	0.07(3)	0.05(1)
$\nabla^2\rho_{cp}$ [e/Å^5]	2.5(8)	1.6(7)	0.7(1)

[a] See text and Figure 1 for definition of H···LP and H···plane distances. [b] Standard deviations are indicated between parentheses. [c] For N···O interactions, the parameters involving the hydrogen atom, highlighted in gray, are indicated for the sake of comparison.

Geometry of the Interactions. The average values of geometrical HB parameters are summarized in Table 3. The classical parameters are in agreement with the results of other surveys, such as those of Thomas et al.[10] and Koch et al.[8] Accordingly, the geometry of $i \rightarrow i+3$ hydrogen bonds is not as directional as for the $i \rightarrow i+4$ bonds. In the former, the average H···O distance is

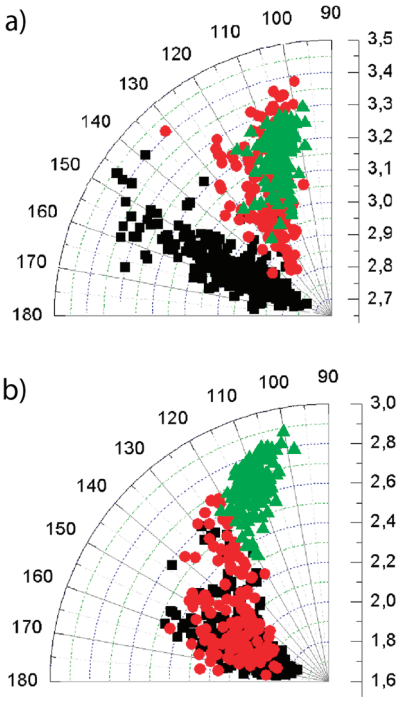


Figure 3. (a) Polar representation of the C═O···N angle. (b) Polar representation of the N−H···O angle. The radial component corresponds to the distances and the angular component to the angles; $i \rightarrow i+3$, $i \rightarrow i+4$ hydrogen bonds, and N···O contacts are depicted as red circles, black squares, and green triangles, respectively.

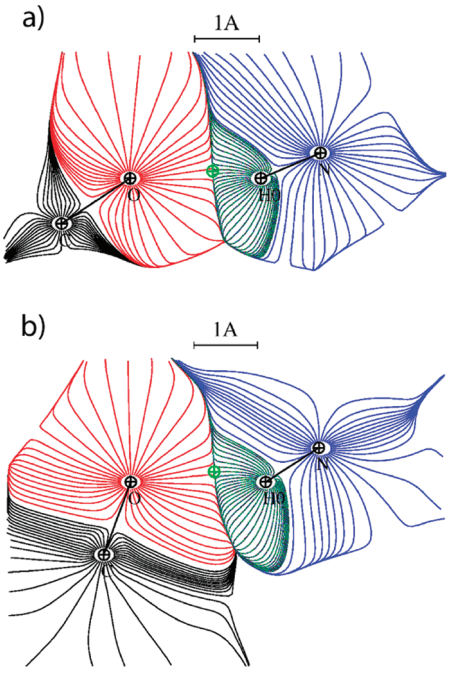


Figure 4. (a) Gradient lines of a $i \rightarrow i+4$ hydrogen bond in the protein xylanase (between the residues Ser 217 and Ala 221). (b) Gradient lines of a $i \rightarrow i+3$ hydrogen bond in the protein aldose reductase (between the residues Ala 45 and Tyr 48). Both figures are computed in the C═O···H plane. The position of the hydrogen bond CP is marked by a green cross.

0.19 Å longer and the average N···O═C and H···O═C angles are approximately 40° smaller than in the latter. In contrast, the N−H···O angles are comparable for both types of HBs (147(15)° and 158(10)° for $i \rightarrow i+3$ and $i \rightarrow i+4$ HBs, respectively). It can be further noted that the distance between the hydrogen atom and the oxygen electron lone pairs plane (H···plane) is significantly shorter in $i \rightarrow i+4$ HBs (0.8(3) Å) than in $i \rightarrow i+3$ HBs (2.0(3) Å). Therefore, in $i \rightarrow i+4$ hydrogen bonds, the H atom interacts with the oxygen atom nearly in the carbonyl plane. This is also reflected by the distances between the hydrogen atom and the nearest oxygen lone pair (H···LP), which are 1.9(1) and 2.2(2) Å for $i \rightarrow i+4$ and $i \rightarrow i+3$ hydrogen bonds, respectively.

For N···O interactions, the N−H and C═O groups of the peptide units are less aligned. The characterized angles are smaller and the distances (N···O and H···O) are longer than in hydrogen bonds. The significant decrease of the average N−H···O angle, dropping from approximately 150° for HBs to 111(5)° for N···O contacts, can be noted in particular.

Figure 3 shows, for all three interactions occurring in helices, a polar representation of the C═O···N angle as a function of the O···N distance (Figure 3a) as well as the N−H···O angle as a function of the H···O distance (Figure 3b). The distributions of the C═O···N angles belonging to $i \rightarrow i+3$ hydrogen bonds and N···O contacts overlap. The points corresponding to $i \rightarrow i+4$ HBs occupy a region of higher angles and smaller distances. In contrast, the distribution of the N−H···O angles is different for hydrogen bonds and N···O interactions. The former adopt geometries with larger N−H···O angles and shorter H···O distances. The latter are in a region of donor−acceptor angles ranging from 90 to 123°. Interestingly, there is almost no overlap between the points representing HBs and N···O interactions. The maximum N−H···O angle of *all* N···O contacts (including those with $i \rightarrow i+2$, $i \rightarrow i+4$, $i \rightarrow i+5$, and $i \rightarrow i+6$ interaction pattern) is 125°, whereas the lowest angle for the HBs is 120°. Providing the position of the H-atom is accurate enough, this parameter can hence serve as a geometrical criterion allowing for the discrimination between hydrogen bonds and N···O contacts in helices. The N−H···O angle of 125° is therefore proposed as discriminating value for HBs, while contacts in the narrow region 120−125° should be analyzed in more detail. For angles lower than 120°, only N···O interactions should be considered. Consequently, using the N−H···O angle as criterion, hydrogen bonds can be differentiated from N···O interactions without performing a topological analysis, which proves time-consuming in routine inspections of protein structures. This is of interest for future surveys, as most analyses use geometrical criteria for the identification of hydrogen bonds.

The Ramachandran plot of residues belonging to helices has been drawn for the protein aldose reductase (Figure S1, see Supporting Information). Residues, whose N and/or O atoms are involved in N···O contacts are clustered in the region $\varphi = -65°$, $\psi = -45°$, which corresponds to the most favorable dihedral angles for α helix. Hence, N···O contacts occur in helices if the residues adopt this particular geometry.

Example Schemes of the Interactions. According to our results, helices are stabilized by three types of interactions between N−H and C═O groups: (1) $i \rightarrow i+4$ hydrogen bonds; (2) $i \rightarrow i+3$ hydrogen bonds; N···O contacts.

All three interactions are illustrated with example schemes (Figures 4 and 5). The geometrical parameters of the chosen examples (Table 4) are close to the average values obtained for all interactions (Table 3). Figure 4a shows the gradient lines of a $i \rightarrow$

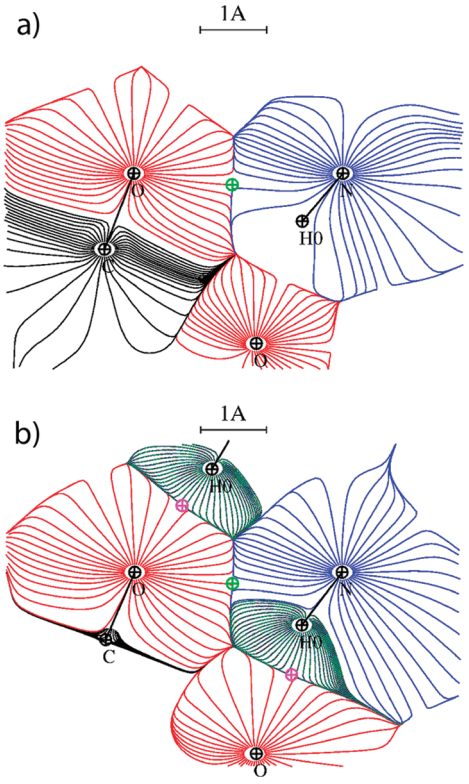


Figure 5. Gradient lines of a N···O contact in the protein Aldose Reductase (between the residues Thr 244 and Val 247), calculated in the C═O···N plane (a) and in the O···H−N plane (b). The positions of the bond CPs are marked by green and magenta crosses for the O···N contact and HBs, respectively.

Table 4. Geometrical and Topological Parameters[a] of the Interactions Used for the Example Schemes

interaction	$i \rightarrow i+4$	$i \rightarrow i+3$	N···O
figure	4a	4b	5
distance [Å]			
N···O	2.98	3.02	3.15
H···O	2.03	2.13	2.66
H···LP	1.77	2.11	2.64
H···plane	0.12	2.00	2.64
angle [°]			
N···O=C	154	121	111
N-H···O	156	146	110
H···O=C	146	110	95
topology			
ρ_{cp} [e/Å^3]	0.09	0.08	0.05
$\nabla^2\rho_{cp}$ [e/Å^5]	2.4	1.8	0.7

[a] Nomenclature as in Table 2.

$i + 4$ hydrogen bond, plotted in the C═O···H plane. It can be noted that the amide and carbonyl groups of the peptide unit are almost aligned. Furthermore, the interatomic surface of the H and O atomic basins is large and the CP is nearly located on a straight line connecting the hydrogen and oxygen atoms. The gradient lines of a $i \rightarrow i + 3$ hydrogen bond are depicted Figure 4b

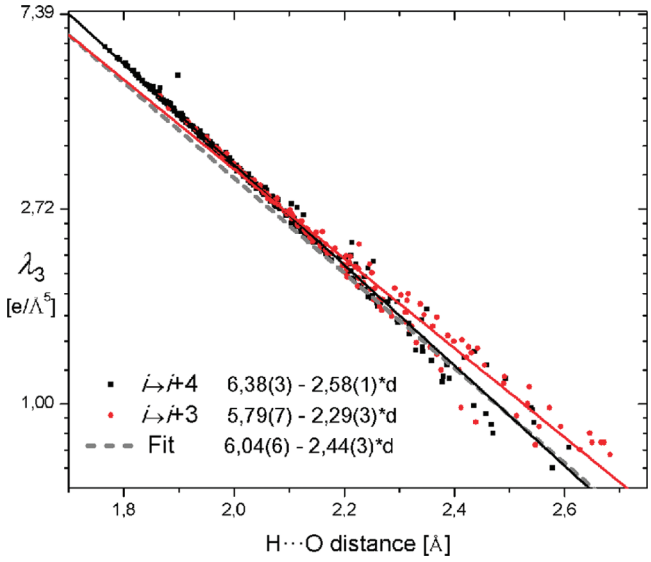


Figure 6. Log−linear plot of the positive electron density curvature λ_3 at the hydrogen bond CP as a function of the H···O distance. Points corresponding to $i \rightarrow i + 4$ and $i \rightarrow i + 3$ hydrogen bonds are depicted as black squares and red spheres, respectively. Dashed gray line: linear regression derived from Mata et al.[16] The black and red lines correspond to the linear regressions of the $i \rightarrow i + 4$ and $i \rightarrow i + 3$ hydrogen bonds, respectively; the values of the fitted parameters are indicated on the plot, the parameter d is the H···O distance.

in the C═O···H plane. The interface of the atomic basins is still large, but located differently, as the N−H and C═O groups are less well aligned, the C═O···H angle being smaller than for $i \rightarrow i + 4$ HBs. Figure 5a shows the N···O contact, plotted in the C═O···N plane. The CP is located on a bond path linking the nitrogen and oxygen atoms. It can be noted that the hydrogen atom is out of the plane; its gradient lines are therefore absent in this figure. In order to ascertain the difference between HBs and N···O contacts, the gradient lines were also drawn in the O···H−N plane (Figure 5b), which contains the hydrogen atom. Here, the orientation of the amide group with respect to the C═O group is such that the H atom is almost placed alongside the oxygen atom. Indeed, the average C═O···H angle of the N···O interactions is 93(6)° (Table 3). Like in Figure 4a,b, it can be noted that the atomic basin of the H atom is compressed by the surrounding atoms and is thus smaller than those of N and O atoms. It is noticeable that the shared interatomic surface between the noninteracting H and O atoms is significantly reduced compared to the one observed for H···O HBs. This trend appears as a consequence of both (i) the $i \rightarrow i + 4$ interaction, which involves the nitrogen of the N···O contact and an oxygen atom in the neighboring region (Figure 5b) and (ii) the geometric configuration of the donor and acceptor moieties in N···O contacts, both most probably related to each other. Accordingly, the nitrogen and the oxygen atoms are linked by the N···O bond path and, concomitantly, the CP exhibits in the internuclear region.

Hydrogen Bonds: Topological Behavior. The average values of the electron density at the bond CPs are 0.10(4), 0.07(3), and 0.05(1) e/Å^3 for $i \rightarrow i + 4$ and $i \rightarrow i + 3$ HBs and N···O contacts, respectively. The Laplacian $\nabla^2\rho_{cp}$ is positive for all bonds, as expected for closed shell interactions, classifying in the same order the interactions (average values: 2.5(8), 1.6(7) and 0.7(1) e/Å^5,

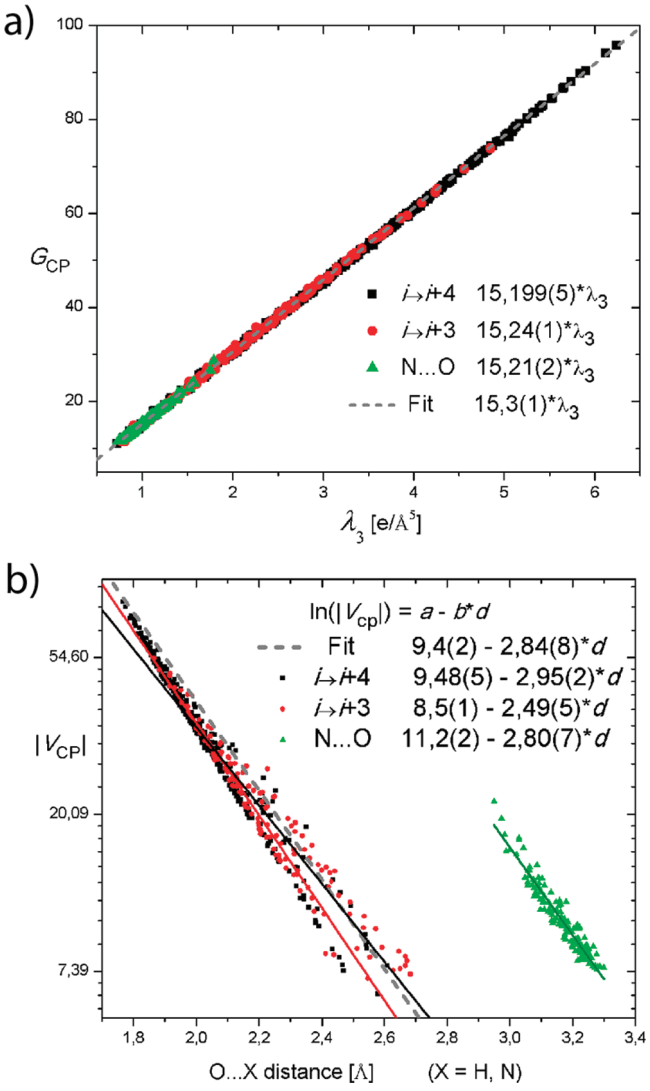


Figure 7. (a) Kinetic energy density G_{cp} (kJ/mol per atomic unit volume) as a function of the positive curvature λ_3 at the critical point. The values of the fitted parameters are indicated in the plot. (b) Log−linear plot of the potential energy density modulus $|V_{cp}|$ (kJ/mol per atomic unit volume) as a function of the interaction distance O···X (X = H, N). The dashed gray line corresponds to the linear regression from ref 15 for a) and ref 16 for b). The black and red solid lines represent the regression lines for $i \rightarrow i + 4$ and $i \rightarrow i + 3$ HBs, respectively. The parameter d corresponds to the O···X distance.

respectively). Both trends point toward a decreasing interaction intensity along the series $i \rightarrow i + 4 > i \rightarrow i + 3 >$ N···O.[13,16]

Figure 6 shows the positive curvature λ_3 of the electron density at the CP as a function of the distance between the H and O atoms. The H···O distance of the $i \rightarrow i + 4$ hydrogen bonds is distributed between 1.7 and 2.6 Å, whereas the $i \rightarrow i + 3$ HBs show up for distances between 1.9 and 2.7 Å. The depicted data show a linear distribution in the log−linear representation and superimpose well for both types of HBs. Only at larger H···O distances, approximately greater than 2.2 Å, data are more dispersed. The function used for the linear regression is

$$\ln(\lambda_3) = a - b \cdot d \qquad (3)$$

where d is the H···O distance and $a > 0$ and $b > 0$ are adjustable parameters. The fitting parameters a and b, which are indicated in Figure 6, are different for $i \rightarrow i + 3$ and $i \rightarrow i + 4$ hydrogen bonds. It can be noted that at large d values, $i \rightarrow i + 3$ HBs are predominant and exhibit more scattered points, affecting thereby the parameters of the regression line.

The exponential fit obtained from HBs in small molecules[13,16] (represented as dashed gray line in Figure 6) reproduces well the behavior of the present data. The latter deviate systematically from the dashed gray regression line only at short H···O distances (H···O < 2.0 Å), the maximum difference 0.5 e/Å^5 being found at H···O = 1.8 Å. Here, it should be pointed out that (i) the hydrogen bonds are constrained to yield the spatial arrangement of the intramolecular helical structure and (ii) the donor and acceptor moieties involved in the interaction (C═O···H−N) are the same for all HBs. On the other hand, the HBs analyzed by Espinosa et al.[13] are more diverse and occur in intermolecular regions between small molecules. Although the maximum difference is quite small, the particular geometry of α-helices could account for the systematic deviation of the data from the small molecules exponential fitting curve.

Figure 7a shows the electron kinetic energy density G as a function of the positive curvature λ_3 along the bond path direction, both at CPs. The proportional relationship between G and λ_3 was first shown for hydrogen bond interactions.[15] The present data, including N···O contacts, show the same phenomenological behavior. Data were fitted by a simple unweighted linear function $G_{cp} = a \cdot \lambda_3$ (the fitting parameter a is indicated in the plot for each type of interaction) and superimpose well to the linear regression line obtained from small molecule HBs (dashed gray line, Figure 7a).[15] While smaller G_{cp} values are similar for HBs and N···O contacts, the larger ones differ significantly (30, 70, and 100 kJ/mol per atomic unit volume for N···O contacts, $i \rightarrow i + 3$ and $i \rightarrow i + 4$ HBs, respectively). Thus, N···O contacts exhibit the same phenomenological behavior as weak hydrogen bonds, which display a small overlap between the electron clouds in the interaction region and are for the most part electrostatic in nature.

Figure 7b shows the log−linear representation of the potential energy density modulus $|V_{cp}|$ as a function of the interaction distance O···X (X = H, N). The data were fitted by using the function $\ln(|V_{cp}|) = a - b \cdot d$(O···X). For hydrogen bonds, the data superimpose well and can be reproduced by the linear regression line obtained from HBs observed in small molecules (dashed gray line).[16] For N···O contacts, $|V_{cp}|$ is plotted against the N···O distance. The corresponding data do not superimpose to those of hydrogen bonds, however, the slope of the linear regression −2.84(8) is found between the values obtained for $i \rightarrow i + 3$ and $i \rightarrow i + 4$ HBs (2.49(5) and 2.95(2), respectively).

Summarizing, the intramolecular hydrogen bonds in helices clearly show the same topological behavior as the intermolecular HBs characterized with small molecules, while N···O contacts exhibit the same topological characteristics as weak HBs due to their weak closed-shell interaction nature.

Characteristics of N···O Contacts. Our findings heretofore suggest that helices are stabilized not only by hydrogen bonds, but also by N···O contacts between carbonyl oxygen and amide nitrogen atoms (mostly separated by three residues). While these two atoms are both partially negatively charged (δ is −0.31e for both O and N atoms, according to the ELMAM database), the presence of a CP between them implies the existence of a bonding contact. To characterize further this kind of interaction

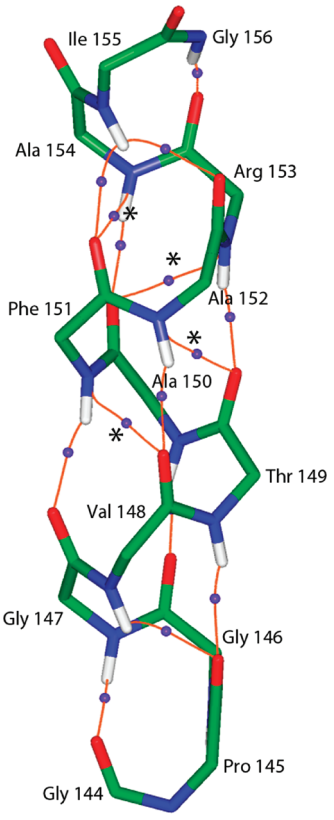


Figure 8. Network of interactions in a helix of the protein CTS-M-9-β-lactamase (residues 144−156). The orange lines and violet spheres represent the bond paths and bond CPs, respectively. N···O contacts are marked with an asterisk. For a clearer view, only the main chain of the helix is represented. It should be noted that $i \rightarrow i + 3$ H bonds are only present at the extremities of the helix.

and to ascertain how the N···O contact differs from hydrogen bonds, the electrostatic interaction energies between N−H and C═O groups in a helix of the protein CTS-M-9-β-lactamase (residues 144−156) were calculated. All the interactions taking place in the helix, which is composed of 13 amino acids, are depicted in Figure 8. According to the usual definition of α-helices, there should be nine (i.e., 13 minus 4) $i \rightarrow i + 4$ hydrogen bonds. However, the topological analysis revealed seven $i \rightarrow i + 4$ hydrogen bonds, four $i \rightarrow i + 3$ H bonds, and four N···O contacts, adding to a total of 15 interactions (1.5 per oxygen atom). The $i \rightarrow i + 3$ HBs are located at both extremities of the helix, whereas the $i \rightarrow i + 4$ hydrogen bonds and the N···O contacts involve residues inside.

The electrostatic interaction energy and the estimation of the total HB interaction energy of these interactions (E_{elec} and E_{HB}) are summarized in Table 5. The aim of this calculation is to undertake a qualitative analysis that allows the comparison between the interactions, taking into account that total interaction energies are described by a larger number of components (electrostatic, charge transfer, polarization, dispersion, and exchange repulsion). For $i \rightarrow i + 4$ and $i \rightarrow i + 3$ hydrogen bonds, the average electrostatic component (E_{elec}) of the interaction energy is −8.5 and −6.1 kcal/mol, respectively. They are thus larger than the average E_{elec} value corresponding to N···O contacts, which reaches only −1.5 kcal/mol. The energy balance of the four atoms (C═O···N−H) is stabilizing, in spite of the

Table 5. Electrostatic Energy Component, Along with the Estimated Interaction Energy ($E_{HB} = 1/2V_{cp}$), for All HB Interactions and N···O Contacts in a Helix of the Protein CTS-M-9-β-lactamase[a]

C═O group	N−H group	type of interaction	E_{elec}	E_{HB}
144	147	$i \rightarrow i + 3$	−7.3	−5.1
145	148		−5.2	−2.8
152	155		−4.6	−1.9
153	156		−7.3	−4.1
145	149	$i \rightarrow i + 4$	−8.5	−2.1
146	150		−9.3	−5.8
147	151		−8.0	−1.9
148	152		−9.0	−2.9
149	153		−9.2	−3.7
150	154		−8.6	−3.5
151	155		−7.0	−2.1
148	151	N···O	−1.8	−1.1
149	152		−1.1	−1.5
150	153		−1.2	−0.9
151	154		−1.9	−2.1
average values		$i \rightarrow i + 3$	−6.1	−3.5
		$i \rightarrow i + 4$	−8.5	−3.1
		N···O	−1.5	−1.4

[a] Energetic quantities are in kcal/mol.

Table 6. Average, Minimum, and Maximum Interaction Energy Magnitudes ($E_{HB} = 1/2V_{cp}$, in kcal/mol) of the Bonding Contacts in Helices

type of interaction	average[a]	min	max
$i \rightarrow i+4$	-4.3(1.7)	-0.8	-9.3
$i \rightarrow i+3$	-2.7(1.3)	-0.8	-6.5
N···O	-1.2(3)	-0.9	-2.6

[a] Standard deviations are between parentheses.

partial negative net charges of N and O atoms (−0.31e). Indeed, the N−H and C═O groups are dipolar and these dipoles are arranged parallel to each other in the helix. The stabilizing electrostatic energy balance is therefore likely to stem from the dipole−dipole interaction between the N−H and C═O groups.

The total HB interaction energy (E_{HB}) has been estimated by using the observed proportionality between the dissociation energy of a hydrogen bond (D_e) and the potential energy density at the critical point ($E_{HB} = -D_e = 1/2V_{cp}$; Table 6).[14] As N···O contacts show the same topological behavior as hydrogen bonding interactions (Figures 6 and 7), their interaction energy has been also estimated by this method. Using the entire sample, the average values are −4.3 and −2.7 kcal/mol for $i \rightarrow i + 4$ and $i \rightarrow i + 3$ HBs, respectively. The minimum energies are similar for both types of hydrogen bonds, but $i \rightarrow i + 4$ HBs reach higher maximum energies (−9.3 kcal/mol) than $i \rightarrow i + 3$ HBs (−6.5 kcal/mol). These values are in agreement with theoretical calculations, as those carried out by Deshmuk et al. who found 4−6 kcal/mol for helical fragments in $i \rightarrow i + 3$ geometry.[44] In comparison to $i \rightarrow i + 4$ and $i \rightarrow i + 3$ HBs, N···O contacts show the lowest average and minimum interaction energy magnitudes

(−1.2(3) and −2.6 kcal/mol, respectively), indicating again their weaker interaction intensity. We estimated the accuracy of E_{HB} by the error propagation method applied to both Abramov's functional and the local form of the virial theorem. Thus, using 0.05 e/Å^3 and 0.1 e/Å^5 as mean estimated errors for the electron density and its Laplacian at the N···O bond critical points, the mean calculated error on E_{HB} is 0.9 kcal/mol. This value is indeed large compared to the one of E_{HB} itself. However, the systematic observation of small values (compared to standard HBs) permits us to conclude about the very weak intensity of this closed shell interaction.

In summary, the N···O contact is caused by the specific interplay of the N−H and C═O groups. It has been shown that their respective orientation is electrostatically favorable. As a consequence, they take part in the rich interaction network in helices and contribute to their stabilization.

CONCLUSION

In this study, a topological analysis of the transferred electron density in six high-resolution protein structures has been carried out, focusing on helical motifs of the polypeptide chain. Geometrical parameters (involving the position of the oxygen lone pairs) have been defined to better characterize the hydrogen bonds. This investigation enabled us to identify a N···O bonding contact between the carbonyl oxygen and amide nitrogen atoms in helices, which has been often misleadingly interpreted as a classical hydrogen bond. A detailed analysis accounting for the geometry and the topology of the electron density distribution in hydrogen bonds and N···O contacts taking place in helices is provided. A geometrical criterion, based on the donor−acceptor N−H···O angle, has been elaborated, allowing for the discrimination between hydrogen bonds and N···O contacts. Thus, while for angles N−H···O < 120° the interaction corresponds to a N···O contact, a hydrogen bond takes place for N−H···O > 125°. In the intermediate range 120−125°, a careful analysis should be done to identify the exact nature of the interaction.

It is pointed out that hydrogen bonds in helices show the same electron density topological features as those observed in intermolecular regions when involving small molecules. In addition, N···O contacts exhibit the same topological characteristics as weak HBs due to their weak closed-shell interaction nature. N···O contacts have been further characterized by calculating the electrostatic interaction energy component, and by estimating the total interaction energy. Accordingly, they have been identified as electrostatically slightly favorable interactions between the N−H and C═O groups, which benefit from their specific orientation in helices. From geometric, topological and energetic considerations, the interactions intensities observed within helices are classified along the series $i \rightarrow i+4 > i \rightarrow i+3 >$ N···O.

ASSOCIATED CONTENT

Supporting Information. Ramachandran plot of residues in helical geometry of the protein aldose reductase. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +33 383 68 48 99. Fax: +33 383 68 43 00. E-mail: benoit.guillot@crm2.uhp-nancy.fr.

ACKNOWLEDGMENT

This work was supported by Nancy University and CNRS. D.L. was a doctoral fellow supported by the French Ministry of Research. The authors thank Prof. R. F. W. Bader for discussions during his stay in Nancy in October 2009.

REFERENCES

(1) Branden, C.; Tooze, J. *Introduction to Protein Structure*; De Boeck: France, 1996.
(2) Fain, A. V.; Ukrainskii, D. L.; Dobkin, S. A.; Galkin, A. V.; Esipova, N. G. *Biophysics* **2008**, *53*, 125–133.
(3) Hol, W. G. J.; Halie, L. M.; Sander, C. *Nature* **1981**, *294*, 532–536.
(4) Hol, W. G. J. *Prog. Biophys. Mol. Biol.* **1985**, *45*, 149–195.
(5) Aqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2026–2030.
(6) Fodje, M. N.; Al-Karadaghi, S. *Protein Eng.* **2002**, *15*, 353–358.
(7) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
(8) Koch, O.; Bocola, M.; Klebe, G. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 310–317.
(9) Steiner, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 48–76.
(10) Thomas, A.; Benhabiles, N.; Meurisse, R.; Ngwabije, R.; Brasseur, R. *Proteins: Struct., Funct., Bioinf.* **2001**, *43*, 37–44.
(11) Bader, R. W. F. *Atoms in Molecules. A Quantum Theory*; Oxford University Press: New York, 1990.
(12) Souhassou, M.; Blessing, R. H. *J. Appl. Crystallogr.* **1999**, *32*, 210–217.
(13) Espinosa, E.; Souhassou, M.; Lachekar, H.; Lecomte, C. *Acta Crystallogr., Sect. B* **1999**, *55*, 563–572.
(14) Espinosa, E.; Molins, E.; Lecomte, C. *Chem. Phys. Lett.* **1998**, *285*, 170–173.
(15) Espinosa, E.; Lecomte, C.; Molins, E. *Chem. Phys. Lett.* **1999**, *300*, 745–748.
(16) Mata, I.; Alkorta, I.; Molins, E.; Espinosa, E. *Chem.—Eur. J.* **2010**, *16*, 2442–2452.
(17) Popelier, P. L. A.; Bader, R. F. W. *J. Phys. Chem.* **1994**, *98*, 4473–4481.
(18) Parthasarathi, R.; Raman, S. S.; Subramanian, V.; Ramasami, T. *J. Phys. Chem. A* **2007**, *111*, 7141–7148.
(19) LaPointe, S. M.; Farrag, S.; Bohórquez, H. J.; Boyd, R. J. *J. Phys. Chem. B* **2009**, *113*, 10957–10964.
(20) Vener, M.; Egorova, A.; Fomin, D.; Tsirel'son, V. *Russ. J. Phys. Chem. B* **2009**, *3*, 541–547.
(21) Zarychta, B.; Pichon-Pesme, V.; Guillot, B.; Lecomte, C.; Jelsch, C. *Acta Crystallogr., Sect. A* **2007**, *63*, 108–125.
(22) Pichon-Pesme, V.; Lecomte, C.; Lachekar, H. *J. Phys. Chem.* **1995**, *99*, 6242–6250.
(23) Berman, H.; Henrick, K.; Nakamura, H. *Nat. Struct. Biol.* **2003**, *10*, 980.
(24) Guillot, B.; Jelsch, C.; Podjarny, A.; Lecomte, C. *Acta Crystallogr., Sect. D* **2008**, *64*, 567–588.
(25) Lokanath, N.K.; I. S., Matsunaga, E.; Tanaka, T.; Kunishima, N. *Structure of β-Glucosidase at Atomic Resolution and Complex with Glucose*; The First Pacific-Rim International Conference on Protein Science, Pacifico Yokohama, Japan, 2004.
(26) Chen, Y.; Delmas, J.; Sirot, J.; Shoichet, B.; Bonnet, R. *J. Mol. Biol.* **2005**, *348*, 349–362.
(27) Natesh, R.; Manikandan, K.; Bhanumoorthy, P.; Viswamitra, M. A.; Ramakumar, S. *Acta Crystallogr., Sect. D* **2003**, *59*, 105–117.
(28) Shimizu, T.; Nakatsu, T.; Miyairi, K.; Okuno, T.; Kato, H. *Biochemistry* **2002**, *41*, 6651–6659.
(29) Deacon, A.; Gleichmann, T.; J. Kalb., A.; Price, H.; Raftery, J.; Bradbrook, G.; Yariv, J.; R. Helliwell, J. *J. Chem. Soc., Faraday Trans* **1997**, *93*, 4305–4312.
(30) Collaborative Computational Project Number 4. *Acta Crystallogr., Sect. D* **1994**, *50*, 760−763.

(31) Liebschner, D.; Elias, M.; Moniot, S. b.; Fournier, B.; Scott, K.; Jelsch, C.; Guillot, B.; Lecomte, C.; Chabrière, E. *J. Am. Chem. Soc.* **2009**, *131*, 7879–7886.
(32) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., III; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res.* **2007**, *35*, W375–383.
(33) Hansen, N. K.; Coppens, P. *Acta Crystallogr., Sect. A* **1978**, *34*, 909–921.
(34) Allen, F. H. *Acta Crystallogr., Sect. B* **1986**, *42*, 515–522.
(35) Jelsch, C.; Guillot, B.; Lagoutte, A.; Lecomte, C. *J. Appl. Crystallogr.* **2005**, *38*, 38–54.
(36) Koch, U.; Popelier, P. L. A. *J. Phys. Chem.* **1995**, *99*, 9747–9754.
(37) Heinig, M.; Frishman, D. *Nucleic Acids Res.* **2004**, *32*, W500–502.
(38) Martin, J.; Letellier, G.; Marin, A.; Taly, J.-F.; de Brevern, A.; Gibrat, J.-F. *BMC Struct. Biol.* **2005**, *5*, 1–17.
(39) Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. *J. Mol. Biol.* **1992**, *226*, 1143–1159.
(40) Allen, F. H. *Acta Crystallogr., Sect. B* **2002**, *58*, 380–388.
(41) Taylor, R.; Kennard, O.; Versichel, W. *J. Am. Chem. Soc.* **1983**, *105*, 5761–5766.
(42) Murray-Rust, P.; Glusker, J. P. *J. Am. Chem. Soc.* **1984**, *106*, 1018–1025.
(43) Abramov, Y. *Acta Crystallogr., Sect. A* **1997**, *53*, 264–272.
(44) Deshmukh, M. M.; Gadre, S. R. *J. Phys. Chem. A* **2009**, *113*, 7927–7932.