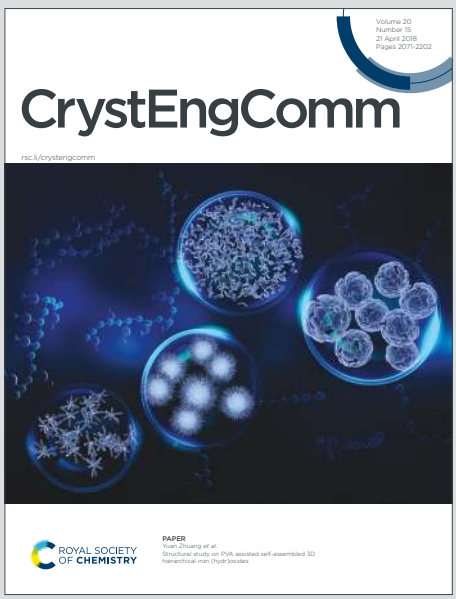
CrystEngComm

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: S. K. K. Mandal, B. Guillot and P. Munshi, *CrystEngComm*, 2020, DOI: 10.1039/D0CE00577K.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/crystengcomm

View Article Online
DOI: 10.1039/D0CE00577K

Electron Density Based Analysis of N-H···O=C Hydrogen Bonds and Electrostatic Interaction Energies in High-resolution Secondary Protein Structures: Insights from Quantum Crystallographic Approaches

Suman K. Mandal[a], *Benoit Guillot*[b] *and Parthapratim Munshi*[*a]

[a] Chemical and Biological Crystallography Laboratory, Department of Chemistry, School of Natural Sciences, Shiv Nadar University, Dadri, Uttar Pradesh – 201314, India

[b] Laboratoire de Cristallographie, Institut Jean Barriol, Université de Lorraine, 34 Cours Léopold, 54000 Nancy, France

*E-mail: parthapratim.munshi@snu.edu.in.

View Article Online
DOI: 10.1039/D0CE00577K

Keywords: High-resolution Protein Structure, Electron Density Analysis, Quantum Crystallography, Hydrogen Bonds and Electrostatic Interaction Energy

ABSTRACTS

In proteins, the main-chain N-H•••O=C hydrogen bonds (HBs) play crucial role in the formation of the α-helices and β-sheets. Accurate analysis of such hydrogen bonds and their electrostatic interaction energies are essential for studying binding interactions and for better understanding of the energetics involved in protein folding. Here, we studied 22 high-resolution (0.87 Å to 0.48 Å) secondary protein structures (4.7 kDa to 54.5 kDa) from RCSB PDB and performed topological analyses of 1443 N-H•••O=C HBs (750 in α-helices and 693 in β-sheets) using the multipole analysis based experimental electron densities as transferred from the *ELMAM2* database. This is the first study of its kind involving by far the largest number of high-resolution protein structures and HBs from both α-helices and β-sheets. Further, based on the accurate estimation of the electrostatic interaction energies the excellent correlations with various topological parameters have

View Article Online
DOI: 10.1039/D0CE00577K

been demonstrated. The excellent correlations have also been observed between the topological parameters. Thereby, we identified the limiting values of the topological parameters and the electrostatic interaction energies to establish the presence of the true N-H•••O=C HBs in protein main-chain *via* quantitative and qualitative analyses of electron densities using quantum crystallographic approaches – Quantum Theory of Atoms in Molecules (QTAIM) and Noncovalent Interaction (NCI) index.

INTRODUCTION

Folding of the polypeptide chain to adopt a three-dimensional protein structure upon formation of α-helices and β-sheets is essentially govern by the interatomic interactions among the amino acids. The two main stabilizing forces which contribute to the protein folding and its stability are the hydrophobic effects and the hydrogen bonds (HBs)[1]. During protein folding the nonpolar side chains entomb due to hydrophobic effects and the main-chains form HBs *via* N-H donors and C=O acceptors. However, the debate on which of these two energetic factors predominates is still active. While some earlier studies[2,3] direct towards the hydrophobic effects the other recent experimental[4–6] and

View Article Online
DOI: 10.1039/D0CE00577K

theoretical[7] studies point to the HBs. The crucial role of HBs for the formation of the α-helices and β-sheets has been elucidated long ago by Mirsky and Pauling[8]. The key building blocks of both α-helix and β-sheet structures are N-H•••O=C HBs and they contribute about 5 – 10 kcal mol^{-1} to the stability of protein, as estimated by Pauling's group[9,10]. The electrostatic interaction energy, which dictates the structure-function correlation for proteins[11], is believed to be the major contributor to the HB energy[12]. Accurate analysis of HBs and knowledge of the local electrostatics in proteins are essential for their proper thermodynamic modelling, estimation of energies for the folding and binding interactions.

However, an accurate study of HBs demands precise atomic locations, especially for the H-atoms, which can be located[13] using high-resolution (sub-atomic) single-crystal X-ray diffraction (SCXRD) and complimentary neutron diffraction experiments[14]. Although, achieving high-resolution (better than ~0.40 Å) SCXRD data on small molecule crystals has now become trivial[15–17] but it remains an uphill task for protein crystals. Because, growing good quality protein crystals, its poor stability, high atomic thermal motions, high solvent contents and large number of H-atoms etc. are some of the common factors,

View Article Online
DOI: 10.1039/D0CE00577K

which limit the resolution of the diffraction data. However, the number of protein structures determined at high-resolutions is increasing steadily. Currently, only 82 unique structures (excluding peptides, hormones and oligonucleotides) with resolutions 0.87 Å or better have been deposited in the Protein Data Bank (PDB: http://www.rcsb.org/pdb/home/home.do). A few of the HB studies using the protein structures from PDB have provided wealth of information. However, these studies were based on either the less number of data points or only the distance-angle criteria or low structural resolutions (1.8 Å to 1.4 Å and 0.95 Å to 0.87 Å)[18,19]. Moreover, the electrostatic interaction energies (E_{elec}) derived purely based on geometrical parameters provide information on dipole-dipole interactions only and ignore significant contributions from the higher order multipoles. Further, the molecular modeling studies based on point-charge potentials have some inherent drawbacks, such as modeling of lone pairs and aromatic rings[20]. Alternatively, the multipole modeling based charge density analysis[21] allow extracting precise structural information in terms of its detailed electron density distributions. The electron densities, $\rho(\mathbf{r})$ can be utilized to calculate electrostatic potential, $\phi(\mathbf{r})$. The derived $\rho(\mathbf{r})$ and $\phi(\mathbf{r})$ can be treated together to estimate E_{elec}[22]. Topological

View Article Online
DOI: 10.1039/D0CE00577K

analysis of $\rho(\mathbf{r})$ can be performed using Bader's Quantum Theory of Atoms in Molecules (QTAIM)[23], based on which one can identify the bond critical point (BCP), where the gradient of $\rho(\mathbf{r})$ vanishes, *i.e.* $\nabla\rho(\mathbf{r}) = 0$. The bond path (BP), the length of which is represented as R_{ij}, not necessarily be the interatomic distance and can be traced along the electron density gradient. Essentially, BP represents a path of maximum electron density between two bonded nuclei[24,25]. The second derivative of $\rho(\mathbf{r})$, *i.e.* $\nabla^2\rho(\mathbf{r})$, the Laplacian, which can be decomposed into the contributions from the three principle axes, represented by the three eigenvalues (λ_1, λ_2 and λ_3) of the electron density Hessian matrix[26]. The signs of the eigenvalues and their sum ($\lambda_1 + \lambda_2 + \lambda_3$) can be used for the characterization of the chemical bonds.

The characterizations of HBs *via* topological analysis of $\rho(\mathbf{r})$ using QTAIM have been commonly employed in small molecule crystals[17,27–29] or in model polypeptides[30] and are becoming popular in the cases of protein systems[31–38]. Although the experimental change density analysis (ECDA) using high-resolution SCXRD data is a well-established methodology in the case of small molecules but it has remained challenging for the biological macromolecules. Consequently, so far, only six of the high-resolution X-ray

View Article Online
DOI: 10.1039/D0CE00577K

protein structures were subjected to ECDA (**Table 1**)[31–36,39]. However, for all of these, the initial charge density models were constructed based on the experimental $\rho(\mathbf{r})$ transferred from *ELMAM*/*ELMAM2* databases[40,41] using *MoPro*[42,43]. For modeling the deformation of the atomic densities, due to the interatomic interactions in a crystal, the aspherical description of $\rho(\mathbf{r})$ is essential and for macromolecules, this can be conveniently achieved using electron density database transfer approach. Liebschner *et al.* have utilized the *ELMAM* database for the topological analysis of HBs and weak interactions in six selected proteins, out of which five structures with resolutions ranging from 0.99 Å to 0.89 Å were retrieved from PDB[37]. Their study also included protein human aldose reductase (hAR) at the resolution of 0.66 Å, however, focused only on the helices. Recently, some of us performed topological analysis of main-chain N-H•••O=C HBs in both α-helices and β-sheets in hAR based on the electron densities transferred from *ELMAM2* database for the estimation of the E_{elec}[39]. The energies were found to be comparable to those measured experimentally[9]. That insightful study motivated us to perform quantitative and qualitative analysis of the main-chain N-H•••O=C HBs in several high-resolution secondary protein structures for accurate characterizations and estimation of their E_{elec}.

View Article Online
DOI: 10.1039/D0CE00577K

Noncovalent interaction (NCI) descriptor, which can be derived from $\rho(\mathbf{r})$ and its derivative and it is based on the maps of the reduced density gradient (RDG), has been proved to be an extremely useful approach for qualitative analysis of HBs[44,45]. The NCI approach complements to the QTAIM approach. Further, the combination of these two approaches for reliable analysis of the weaker interactions appeared to be an efficient strategy in quantum crystallography[46–48]. The program *NCIPLOT* can be utilized for plotting the reduced electron density gradient isosurfaces in the interaction regions - the BCPs and its surroundings[49]. The concept of sudden change in RDG, as implemented in NCI, helps visualizing intermolecular interactions.

In this study, from RCSB PDB, we retrieved 22 protein structures ranging from 4.7 kDa to 54.5 kDa and with resolutions 0.87 Å to 0.48 Å for studying main-chain N-H•••O=C HBs in both α-helices and β-sheets (**Table 1**). Accordingly, we performed topological analyses of 1443 N-H•••O=C HBs (750 in α-helices and 693 in β-sheets) using multipole-based electron densities as transferred from the *ELMAM2* database. Subsequently, we estimated E_{elec} and examined its relationships with the various topological parameters. The relationships between the topological parameters have also been discussed.

View Article Online
DOI: 10.1039/D0CE00577K

Thereby, we identified the limiting values of the various topological parameters and E_{elec} of the protein main-chain N-H•••O=C HBs *via* quantitative and qualitative analyses of electron densities using quantum crystallographic approaches - QTAIM and NCI index.

Table 1. List of selected high-resolution secondary protein structures from RCSB PDB

No.	Protein systems	PDB ID	Resolution (Å)	Molecular Weight (kDa)	Main-chain Average *B*-factor (Å^2)	Year of Publication
*1	high-potential iron-sulfur protein (HIPIP)	5D8V	0.48	8.8	3.43	2016
*2	Crambin	1EJG	0.54	4.7	2.65	2000
*3	Hen Egg White Lysozyme (HEWL)	2VB1	0.65	14.3	4.77	2007
*4	Human Aldose Reductase (hAR)	1US0	0.66	35.9	5.58	2004
5	*Pyrococcuss abyssi* Rubredoxin	1YK4	0.69	5.8	4.95	2005
*6	Cholesterol Oxidase	4REK	0.74	54.5	8.72	2015
7	Serine Protease	1GCI	0.78	26.7	7.33	1998

View Article Online
DOI: 10.1039/D0CE00577K

*8	Cytochrome b5 reductase	5GV8	0.78	30.8	8.68	2017
9	Trypsin	1PQ7	0.80	22.2	5.81	2003
10	Human Parvulin 14	3UI4	0.80	11.2	4.19	2011
11	α-Lytic protease	2H5C	0.82	19.9	4.78	2006
12	Triose-Phosphate Isomerase	2VXN	0.82	27.2	6.27	2010
13	Proteinase K	2PWA	0.83	29.0	7.39	2007
14	Ponsin	2O9S	0.83	7.7	5.12	2007
15	Cytochrome c6	4EIC	0.84	9.4	7.58	2013
16	Acutohaemonlysin	1MC2	0.85	14.1	7.93	2003
17	Diisopropyl Fluoro Phosphatase (DFP)	1PJX	0.85	35.1	8.05	2003
18	1,2-alpha mannosidase	4AYO	0.85	50.5	5.15	2012
19	Cyclophilin D	4O8H	0.85	17.7	5.11	2014
20	Glutaredoxin NrdH	4HS1	0.87	9.3	7.86	2013
21	Cu-containing nitrite reductase	5AKR	0.87	40.8	8.01	2015
22	Fatty acid-binding protein	4TJZ	0.87	14.9	8.8	2015

*ECDA has been performed

View Article Online
DOI: 10.1039/D0CE00577K

MATERIALS AND METHODS

Structure selection and preparation

Only monomeric (not multimeric) structures, which were deposited in the RCSB PDB as a complete functional protein and not a subdomain of a larger protein, published in scientific journals with resolutions of 0.87 Å or better and with main-chain average *B*-factor of lower than 9 $Å^2$ were selected for this study. In case of the structures with multiple entries, the one with highest resolution and satisfying the above criteria was considered in this analysis. Thus, 22 protein structures were identified as listed in **Table 1**. As mentioned earlier that the previous report on topological properties of HBs using electron densities from *ELMAM* database included five out of six structures with resolutions ranging from 0.99 Å to 0.89 Å.[37] Further, our PDB search with the aforementioned selection criteria, did not result any structures between the resolutions of 0.89 Å to 0.87 Å. Therefore, in this study, we have chosen 0.87 Å as the lower limit.

For the convenience of transferring of the multipole modelling based electron densities from *ELMAM2* database, for the structures with multiple conformers, only the major

View Article Online
DOI: 10.1039/D0CE00577K

conformers were selected using the module '*pdbset*' as implemented in the *CCP4* suite[50] (**Scheme S1**). The module *pdbset* was also used to remove the solvents including the water molecules. Moreover, most of these high-resolution structures were lacking the information on H-atom positions. This could be because of the use of inconsistent methods of refinement using diffraction data from various sources. Therefore, in our analysis, for consistency, all the H-atoms were removed with the tool *pdbset*. Subsequently, in the cleaned structures (**Scheme S1**), the H-atoms were added to the standard neutron distances (e.g. N-H = 1.02 Å)[51] after maintaining the appropriate geometries using *MoPro*. In the absence of neutron diffraction data, this approach is generally adopted while dealing with the multipole modelling based electron densities of the H-atoms as the experimentally observed electron density peak does not truly correspond to the position of the H-atom nucleus. Since the non H-atom coordinates were determined using very high resolutions (0.87 Å to 0.48 Å) X-ray diffraction data, a further structural refinement using a common refinement scheme was felt not necessary.[52] The modified structures with the new H-atom positions and with the non H-atom coordinates as deposited in the PDB were then considered for transferring the multipole modelling

View Article Online
DOI: 10.1039/D0CE00577K

based electron densities using *MoPro*. Some of these structures contained metals in ionic or complexed form in addition to some heavy elements as a part of the ligands/substrates (**Table S1**). However, due to the unavailability of the multipole information of such ions and atoms in the *ELMAM2* database, the electron densities were transferred only on the protein atoms (e.g. chain A, i.e. monomer) and not on the metal ions/atoms and on the ligands. The level of multipoles up to which the electron densities transferred were; quadrupole for H-atoms, octupole for C-, N- and O-atoms and hexadecapole for S-atoms. In this context, it is noteworthy that our present study focuses on the main-chain N-H•••O=C HBs only.

In the case of α-helices, we consider the standard *i → i+4* type of N-H•••O=C HBs only. However, no such N-H•••O=C HBs are identified in the α-helices of proteins ponsin and DFP. Similarly, no N-H•••O=C HBs are identified in the β-sheet of protein cytochrome C6 (**Table 2**). The main-chain N-H•••O=C HBs in this study were identified based on the following criteria; H•••O distances to ≤ 2.65 Å, the N-H•••O=C and C=O•••H-N angles to 120-180°, similar criteria to those considered in some earlier studies[19,37,53]. Subsequently,

View Article Online
DOI: 10.1039/D0CE00577K

based on the QTAIM approach, the BCPs were located for all such N-H•••O=C HBs present in these 22 protein structures.

Electrostatic Interaction Energy

For the N-H•••O=C HBs, the E_{elec} based on the $\rho(\mathbf{r})$ at the BCPs and the corresponding $\phi(\mathbf{r})$ were calculated following the same procedure as discussed earlier and using the formula as given below[37,39]. The C=O group and the N-H group of a HB were treated as two different entities (A and B) for the derivations of the corresponding $\rho(\mathbf{r})$ and $\phi(\mathbf{r})$.

$$E_{elec} = \int \phi_B(\boldsymbol{r})\rho_A(\boldsymbol{r})\mathrm{d}\mathbf{r} = \int \phi_A(\boldsymbol{r})\rho_B(\boldsymbol{r})\mathrm{d}\mathbf{r} \qquad (1)$$

The integrations of the product of $\rho(\mathbf{r})$ and $\phi(\mathbf{r})$ over the whole space with ρ as non-zero were perfomed using program *VMoPro*.

Noncovalent Interactions Analysis

The methods utilized and the protocol followed for the generation of the NCI isosurfaces are given in the form of a flowchart in **Scheme S1**. For NCI analysis, the 3D electron density grids generated around the main-chain N-H•••O=C HBs using *ELMAM2* and

View Article Online
DOI: 10.1039/D0CE00577K

VMoPro were fed in to the program *NCImilano*[54] to calculate the RDG of the electron densities and Abramov's energy densities. Subsequently, the module *MolIso* as implemented in *MolecoolQT*[55] was used for the visualization of the NCI isosurfaces. Additionally, the electron densities obtained upon quantum-mechanical calculations on those N-H•••O=C HBs using *Gaussian09*[56] at the MP2/6-311G(*d,p*) level of theory and experimental geometry were used for the generation of NCI isosurfaces using *NCIPLOT*[49]. The NCI isosurfaces were also generated using the program *NCIPLOT* based on the promolecular densities - the sum of the spherically averaged atomic charge densities. The electron density isosurfaces thus obtained from the latter approaches were visualized using *VMD*[57].

RESULTS AND DISCUSSION

Based on our HB criteria, from the 22 high-resolution secondary protein structures, we identify 1443 N-H•••O=C HBs, out of which 750 are from the α-helices and 693 are from the β-sheets (166 parallel and 527 antiparallel). We list the total number of N-H•••O=C HBs in the α-helices and the β-sheets for each protein system in **Table 2**.

View Article Online
DOI: 10.1039/D0CE00577K

Table 2. Average electrostatic energies (E_{elec}) of N-H•••O=C HBs in α-helices and β-sheets

Sr. No.	Proteins	α-helices HBs		β-sheets HBs	
		No.	Average E_{elec} (kCal/mol)	No.	Average E_{elec} (kCal/mol)
1	high-potential iron-sulfur protein (HIPIP)	5	-8.60	6	-10.05
2	Crambin	10	-8.42	2	-10.75
3	Hen Egg White Lysozyme (HEWL)	31	-8.90	4	-9.88
4	Human Aldose Reductase (hAR)	77	-9.06	20	-9.37
5	Pyrococcuss abyssi Rubredoxin	1	-3.03	5	-10.04
6	Cholesterol Oxidase	91	-8.95	57	-10.23
7	Serine Protease	45	-8.91	31	-9.76
8	Cytochrome b5 reductase	36	-8.44	60	-10.56

View Article Online
DOI: 10.1039/D0CE00577K

9	Trypsin	12	-8.99	50	-9.65
10	Human Parvulin 14	22	-8.10	15	-10.80
11	α-Lytic protease	4	-7.85	61	-10.04
12	Triose-Phosphate Isomerase	73	-9.07	22	-10.21
13	Proteinase K	50	-9.42	40	-9.41
14	Ponsin	-	-	14	-9.83
15	Cytochrome c6	35	-9.37	-	-
16	Acutohaemonlysin	38	-9.03	3	-9.95
17	Diisopropyl Fluoro Phosphatase (DFP)	-	-	100	-10.00
18	1,2-alpha mannosidase	159	-9.27	23	-9.20
19	Cyclophilin D	16	-8.89	33	-9.31
20	Glutaredoxin NrdH	22	-8.98	12	-11.60
21	Cu-containing nitrite reductase	10	-10.10	61	-10.38
22	Fatty acid-binding protein	13	-9.27	74	-10.82
Total number of HBs and average energies		750	-9.04	693	-10.08

The static deformation of electron density and the Laplacian maps as plotted for the representative strong and weak N-H···O=C HBs in α-helices and β-sheets demonstrate the accuracy of electron density features around the N-H and C=O groups of the protein main-chain (**Figure S1 and S2)**. The directionality of the lone pair of electrons of the O-

View Article Online
DOI: 10.1039/D0CE00577K

atoms and for the H-atoms, the different degrees of polarizations of the electron densities due to the formation of HBs are clearly noticeable in these maps.

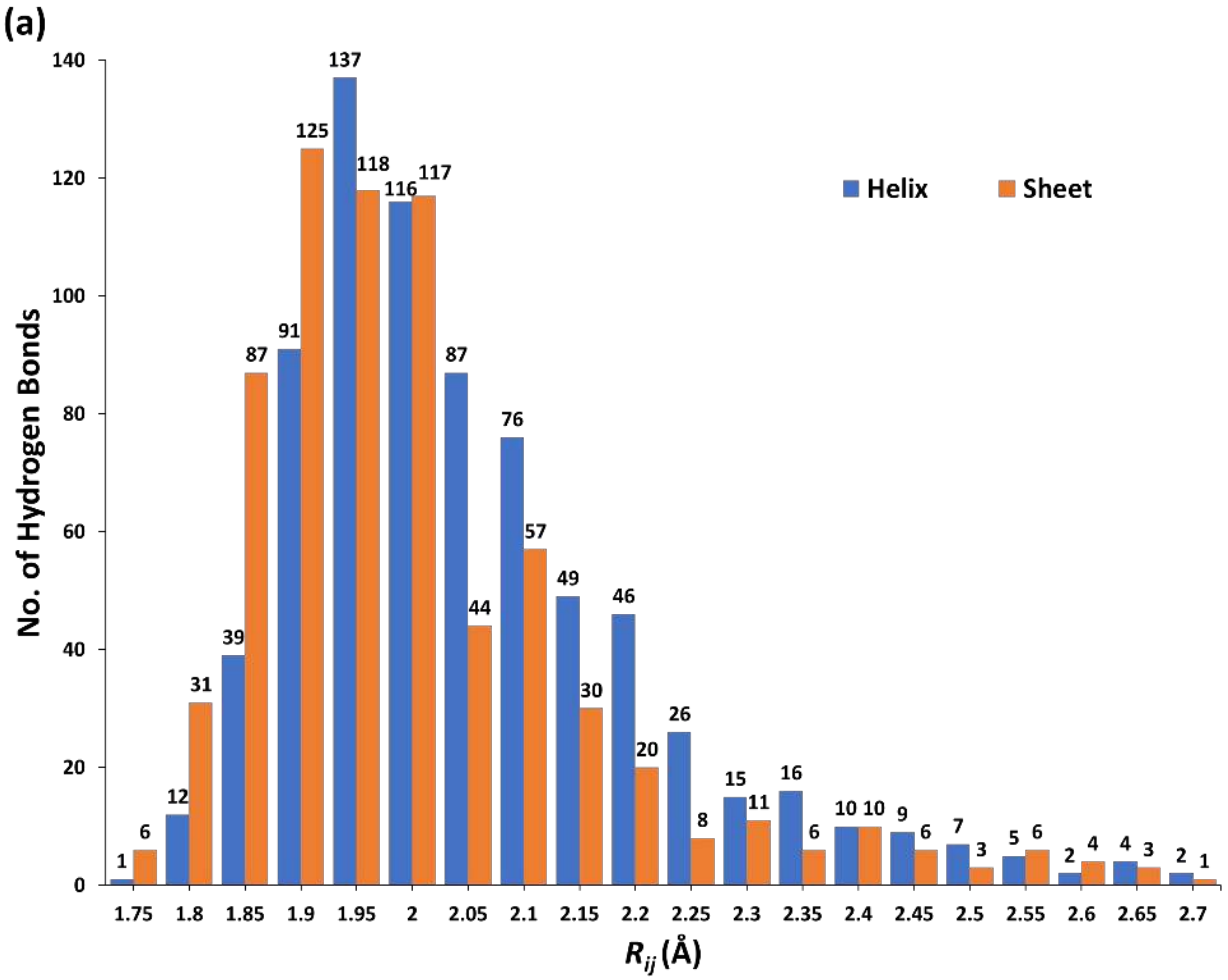
The bond path along the H•••O contacts and the corresponding ρ_{BCP} are found to exist for all the 1443 N-H•••O=C HBs. The succession of N-H•••O=C HBs in terms of the bond path and the corresponding BCPs for an α-helix and a β-sheet of protein hAR is shown in **Figure S3.** Details of the topological properties and the ∠C=O···H and ∠N-H···O of the i→i+4 α-helix and β-sheets hydrogen bonds are listed in **Table S2** and **Table S3**, respectively. Further, the distributions of the ∠N-H···O and the ∠C=O···H across the ranges of R_{ij} and E_{elec} for the α-helices and the β-sheets hydrogen bonds are shown in **Figures S4 – S7**. The ranges of these ∠N-H···O and ∠C=O···H are listed in **Table S4 and S5**, respectively. The overall ranges (considering the extremities of the R_{ij} and E_{elec} based ranges) of ∠N-H···O and ∠C=O···H for the α-helices, β-sheets and both (α-helices and β-sheets together) HBs are also listed in **Table S6**.

The population of N-H•••O=C HBs and their distributions in α-helices and β-sheets in the different shells of R_{ij} and ρ_{BCP} with shell width of 0.05 Å and 0.01 eÅ^{-3}, respectively, are shown in **Figure 1a** & **1b**. There are 50 short HBs with 1.72 Å < R_{ij} ≤ 1.80 Å and with

View Article Online
DOI: 10.1039/D0CE00577K

~0.30 eÅ^{-3} > ρ_{BCP} ≥ 0.25 eÅ^{-3} and the majority (89.3%) of the HBs are found to populate between 1.8 Å < R_{ij} ≤ 2.2 Å which corresponds to the ρ_{BCP}in the range of 0.25 eÅ^{-3} > ρ_{BCP} ≥ ~0.1 eÅ^{-3}. However, the shell of 1.9 Å < R_{ij} ≤ 1.95 Å registered the maximum number of HBs (17.7%). Whereas the peak of number of HB *vs* the ρ_{BCP} distribution is noticed between ~0.20 eÅ^{-3} to 0.17 eÅ^{-3}. There are 117 HBs populated between the R_{ij} 2.2 Å < R_{ij} ≤ 2.45 Å and the corresponding ρ_{BCP} values are in the range of ~0.1 eÅ^{-3} to 0.06 eÅ^{-3}. Thereafter and up to R_{ij} of 2.7 Å, the population diminished to ~10 HBs or less per shell (< 1%), which have ρ_{BCP}< 0.06 eÅ^{-3}. The aforementioned population analysis based on the large number of data points (total 1443) clearly convey the trend of the N-H•••O=C HBs in protein systems.

View Article Online
DOI: 10.1039/D0CE00577K



View Article Online
DOI: 10.1039/D0CE00577K

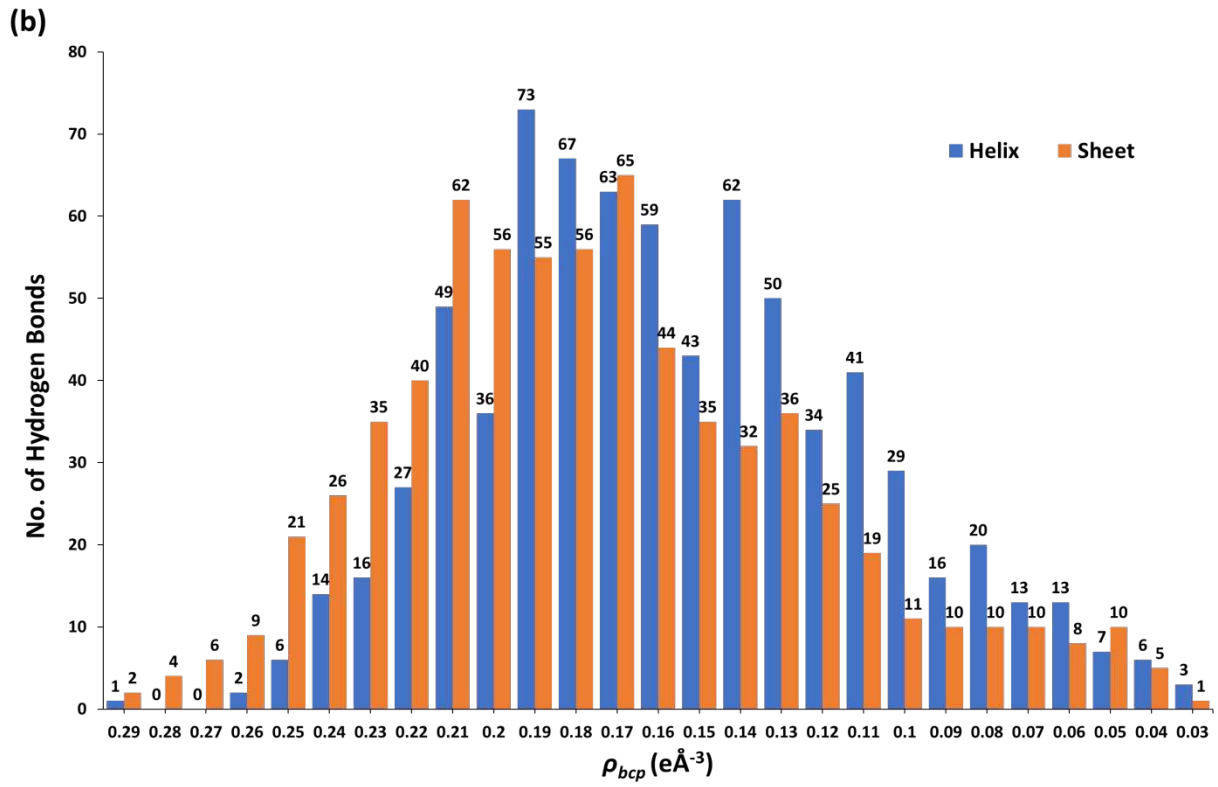


Figure 1. Distributions of (a) R_{ij} (Å) and (b) ρ_{BCP} (eÅ^{-3}) of N-H•••O=C HBs in α-helices and β-sheets

The R_{ij} and the ρ_{BCP} values in both α-helices and β-sheets are found to follow an exponential relationship with regression coefficient (R^2) better than 99% (**Figure 2**). A similar exponential relationship was also observed in the case of small molecules[17,28,58] and NADH-cytochrome *b*5 reductase[36]. While the values of R_{ij} and ρ_{BCP} for the 750 N-H···O=C HBs, in the α-helices, are in the range of 1.720 Å to 2.670 Å and 0.297 eÅ^{-3} to 0.038 eÅ^{-3}, respectively, those for the 693 N-H•••O=C HBs, in the β-sheets, are varying

View Article Online
DOI: 10.1039/D0CE00577K

from 1.715 Å to 2.650 Å and 0.297 eÅ^{-3} to 0.039 eÅ^{-3}, respectively. For the α-helices, the ranges of R_{ij} and ρ_{BCP} are found to be in good agreement with those reported earlier[37]. In our study, covering both α-helices and β-sheets, the R_{ij} *vs* ρ_{BCP} plot (**Figure 2**) not only highlighting the variation of the HB population across the R_{ij} values but also demonstrating the demarcation of N-H•••O=C HBs. However, the ρ_{BCP} for the α-helices and the β-sheets appeared to follow the same trend as can be noticed from their indistinguishable exponential fitting curves. While the shorter N-H•••O=C HBs exist with high values of ρ_{BCP} (~0.30 eÅ^{-3} to 0.25 eÅ^{-3}) the majority of the N-H•••O=C HBs have values of ρ_{BCP} between 0.25 eÅ^{-3} to ~0.10 eÅ^{-3}. Whereas the longer N-H•••O=C HBs have ρ_{BCP} between ~0.10 eÅ^{-3} to ~0.06 eÅ^{-3} and the longest ones have ρ_{BCP} < ~0.06 eÅ^{-3}. These demarcations are in accordance with the observations made from the distribution plots as shown in the **Figures 1a & 1b**. Further, as expected, all ~~of~~ the 1443 N-H•••O=C HBs have positive values of Laplacian (**Table S2 & S3**), negative values of λ_1 and λ_2 and positive values of λ_3 (**Table 3)**.

View Article Online
DOI: 10.1039/D0CE00577K

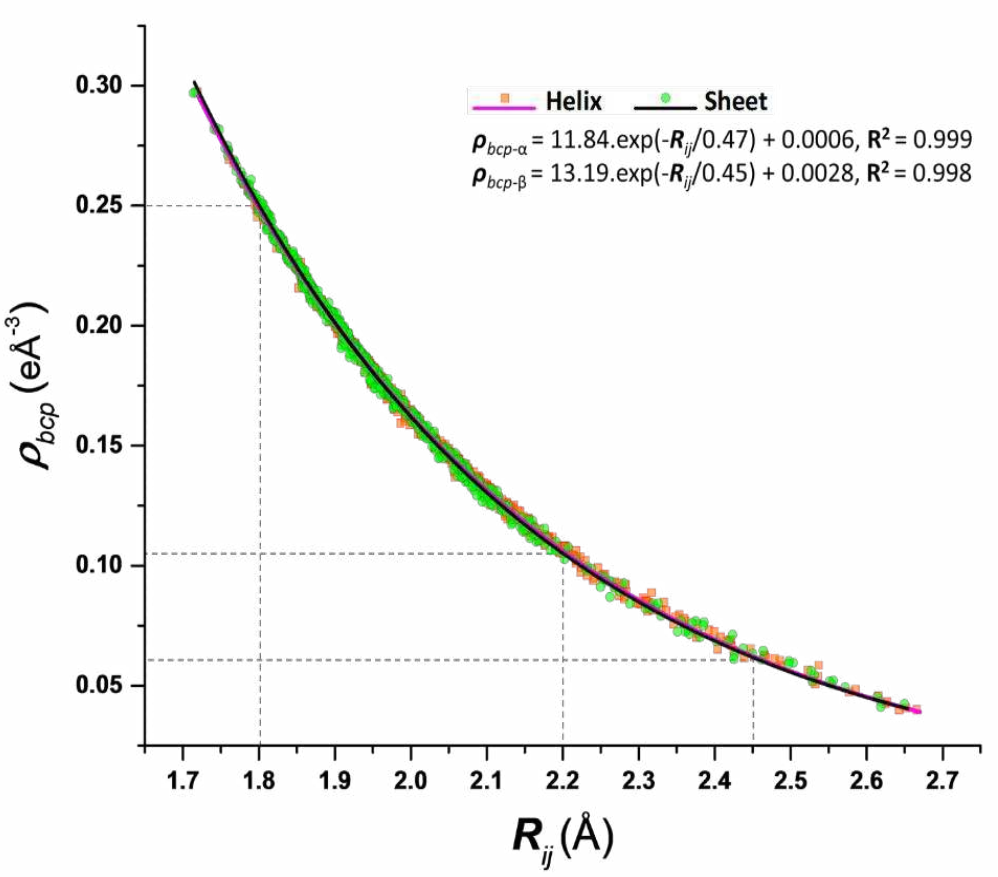


Figure 2. Exponential relationship between the R_{ij} (Å) and ρ_{BCP} (eÅ^{-3}) values of 1443 N-H•••O=C HBs in α-helices (750) and β-sheets (693). The dashed lines denote the demarcations of shortest, most populated, longer and longest N-H•••O=C HBs in 22 high-resolution proteins.

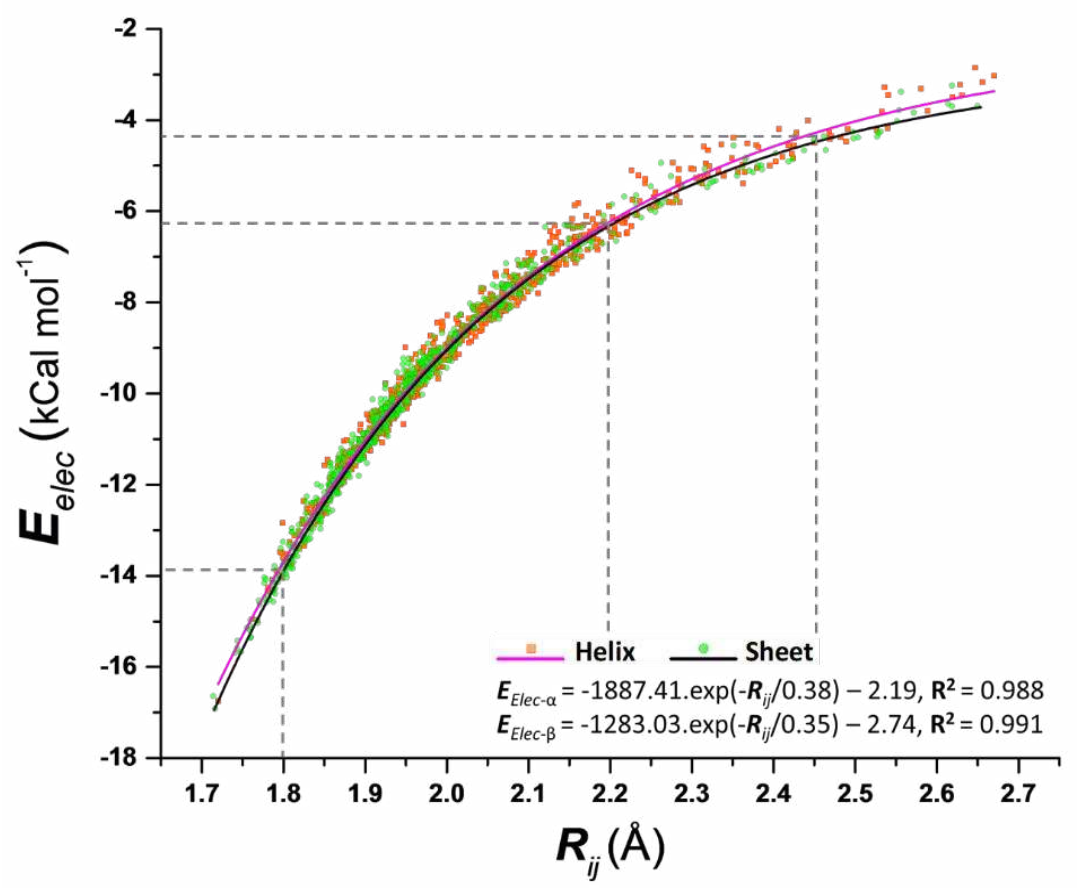
The excellent correlation between the R_{ij} and the ρ_{BCP} and the estimated error of 0.05 eÅ^{-3} (see ESI) for the ρ_{BCP} clearly suggest the accuracy of electron density modelling in protein systems *via* experimental database transfer approach, at least for these 22 high-

View Article Online
DOI: 10.1039/D0CE00577K

resolution proteins. Given these superior results, we have estimated the electrostatic interaction energy, E_{elec} for all of 1443 HBs using equation **(1)**. The values of average E_{elec} for the N-H•••O=C HBs in α-helices and β-sheets for each protein system are listed in **Table 2**. The estimated values of E_{elec} vary from -16.75 kCal mol^{-1} to -2.85 kCal mol^{-1} for the α-helices and -16.99 kCal mol^{-1} to -3.26 kCal mol^{-1} for the β-sheets (**Table S2 & S3**). The values of average E_{elec} of -9.04 kCal mol^{-1} for the 750 HBs in the α-helices and -10.08 kCal mol^{-1} for the 693 HBs in the β-sheets (**Table 2**) are slightly overestimated but lies almost within the range of the experimental values (-5 kCal mol^{-1} to -10 kCal mol^{-1}) ~~as~~ of the overall energy of the N-H•••O=C HBs.[9] For the 750 HBs in α-helices and the 693 HBs in β-sheets, the E_{elec} (**Figure 3**) display an exponential relationship with their corresponding values of R_{ij} and once again, find exceptional correlations. From E_{elec} vs R_{ij} plot, it can be inferred that the majority of the N-H•••O=C HBs with R_{ij} values between 1.8 Å to 2.2 Å have E_{elec} values in the range of about -6 kCal mol^{-1} to -14 kCal mol^{-1} with an estimated error of 1.9 kCal mol^{-1}, which are almost within the range of the experimental values.[9] Moreover, in this range, the E_{elec} values for the two secondary structures are indistinguishable. The stronger HBs have E_{elec} values approximately between -14 kCal

View Article Online
DOI: 10.1039/D0CE00577K

mol^{-1} to -17 kCal mol^{-1} and those of the weaker and weakest HBs vary approximately between –6 kCal mol^{-1} to –4.5 kCal mol^{-1} and up to approximately -3 kCal mol^{-1}, respectively. In these three regions, the E_{elec} values for the α-helices are found to be systematically slightly lower than those of the β-sheets. The error on E_{elec} (over the entire range) is estimated to be of 2.52 kCal mol^{-1}.



View Article Online
DOI: 10.1039/D0CE00577K

Figure 3. Exponential relationship between the R_{ij} (Å) and E_{elec} (kCal mol^{-1}). The dashed lines denote the demarcations of the strongest, most populated, weaker and weakest N-H•••O=C HBs.

Although both the E_{elec} and the ρ_{BCP} of the HBs in the α-helices and the β-sheets vary exponentially with R_{ij} but they themselves found to follow a quadratic relationship (**Figure 4**) with slightly better R^2 than a linear relationship (**Figure S8a**). However, the E_{elec} and the ρ_{BCP} follow a linear relationship for the shorter HBs with ρ_{BCP} > ~0.06 eÅ^{-3} (**Figure S8b & S8c)**. The relationships between the various topological parameters and their values are found to agree well with those reported earlier for small molecules[17,28,58] and protein systems[36,37]. No significant differences in topological parameters of the N-H•••O=C HBs were noticed between the parallel and anti-parallel β-sheets (**Table S3)**.

View Article Online
DOI: 10.1039/D0CE00577K

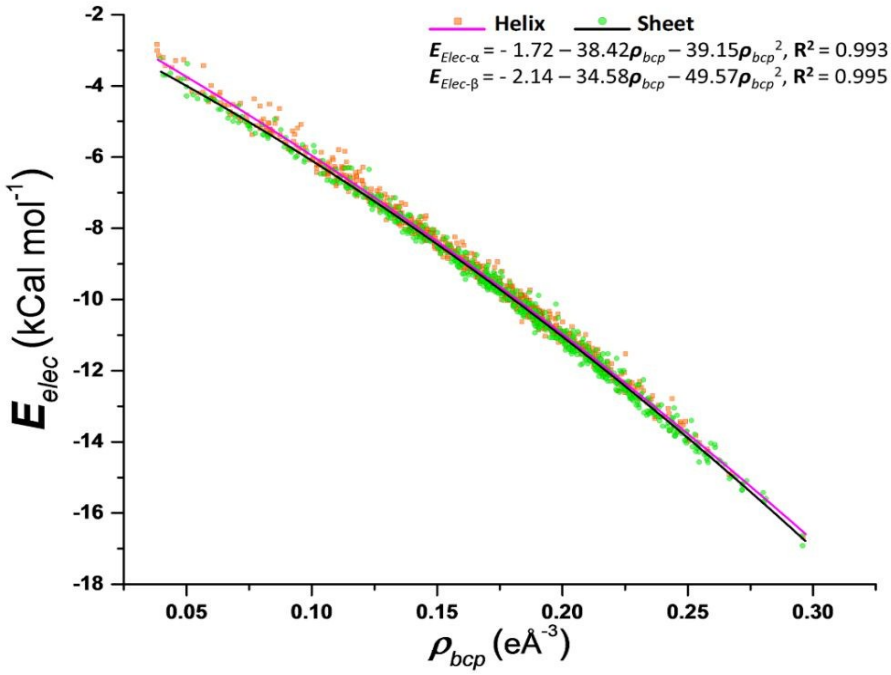


Figure 4. Quadratic relationship between the ρ_{BCP} (eÅ^{-3}) and E_{elec} (kCal mol^{-1}).

Further, to investigate if the N-H•••O=C HBs follow a different trend along the α-helices, we studied the terminal HBs and focused on the population of the weaker and the weakest HBs. Upon inspection of the results, as summarized in **Table S7**, we notice that out of the total of 20 weakest HBs in α-helices (**Figure 1a**), 12 (60%) HBs belong to the terminal residues. Whereas, in the weaker HB region, only ~32% HBs are found to be originated from the terminal residues. There are almost equal numbers of HBs identified from the N-terminal (96) and C-terminal (95) residues. However, the E_{elec} vs R_{ij} plot (**Figure S9**) for these entire terminal HBs (total 191) display a similar exponential relationship (R^2 =

View Article Online
DOI: 10.1039/D0CE00577K

98.4%) to that noticed for the entire HBs in the α-helices (**Figure 2**, R^2 = 99.9%), which also includes the HBs from the central part of the α-helices. In this context, it is noteworthy that in this study the HBs have been identified purely based on the QTAIM approach and not based on the distance-angle criteria, which often employed for studying the HBs in proteins.

Moreover, to verify if the populations of the HBs are affected by the resolutions of the structures, we analyzed percentage distribution of the population of HBs in the different regions (strongest, most populated, weaker and weakest) of the HBs, in terms of both R_{ij} and E_{elec} for two systems with resolution difference of 0.22 Å and having almost same number (~35) of N-H•••O=C HBs; HEWL (PDB ID: 2VB1) with resolution of 0.65 Å and glutaredoxin NrdH (PDB ID: 4HS1) with resolution of 0.87 Å. Interestingly, the distributions display similar trend of HB populations for both the structures (**Table S8)**.

View Article Online
DOI: 10.1039/D0CE00577K

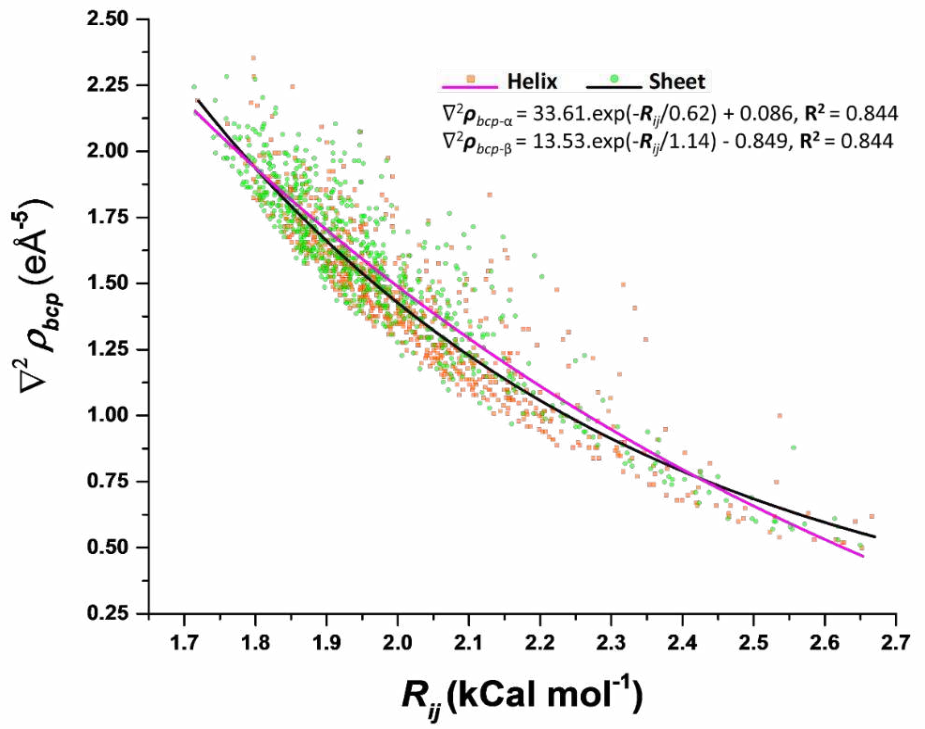


Figure 5. Exponential relationship between the R_{ij} (Å) and $\nabla^2\rho_{BCP}$ (eÅ-5)

The Laplacian of electron density, $\nabla^2\rho_{BCP}$, provides a measure of the local charge depletion (positive value) or concentration (negative value) in the interatomic region, i.e. for the noncovalent interactions (**Figure S10**) or covalent bonds, respectively. However, the demarcation (strong, weak etc.) of HBs from the plot of $\nabla^2\rho_{BCP}$ *vs* R_{ij} (**Figure 5**), which follow an exponential relationship, appeared to be not as obvious as that noticed from the plot of ρ_{BCP} vs R_{ij}. Therefore, we have employed the complementary NCI approach for the qualitative analysis of the N-H•••O=C HBs, especially focusing on the weaker HBs. To monitor the variation in NCI isosurfaces across the regions of strongest to weakest N-

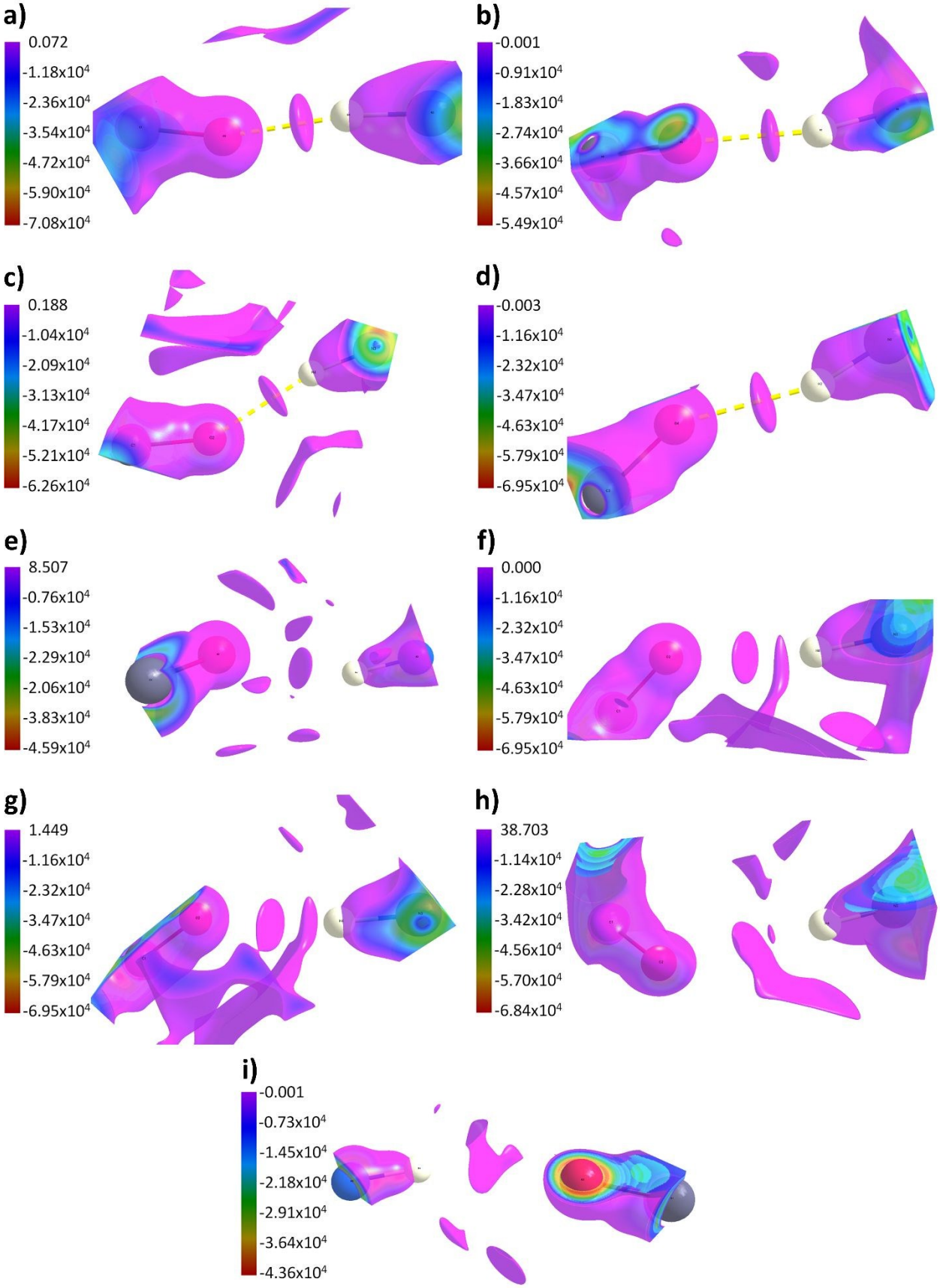
View Article Online
DOI: 10.1039/D0CE00577K

H•••O=C HBs, we have chosen some representative N-H•••O=C HBs from both α-helices and β-sheets (**Table 3**). NCI isosurfaces are usually plotted using *NCIPLOT* based on either the promolecular densities[49] or the densities based on fully quantum-mechanical calculations[56]. However, recently, extremely localized molecular orbital (*ELMO*) based NCI isosurfaces are used to detect NCI in biosystems[59]. Here, to the best of our knowledge, for the first time, we report the NCI isosurfaces in protein systems using *ELMAM2* based experimental electron densities. The comparison of NCI isosurfaces plotted using the aforementioned methodologies, except the *ELMO* based electron densities, are shown in **Figure S11**. The *NCI-ELMAM2* isosurfaces are found to compare well with those based on the molecular densities. The QTAIM based topological parameters along with the E_{elec} of these representative HBs are listed in **Table 3**. The NCI isosurfaces for the N-H•••O=C HBs in the shortest or strongest ($R_{ij} \leq 1.8$ Å), most populated (1.8 Å < $R_{ij} \leq 2.2$ Å), longer or weaker (2.2 Å < $R_{ij} \leq 2.45$ Å) and longest or weakest ($R_{ij} \geq 2.45$ Å) regions are shown in **Figures 6a – i**. The NCI isosurfaces with well-defined shapes are clearly visible for the strongest and up to the weaker N-H•••O=C HBs (**Figures 6a – f**). Subsequently, at ~2.46 Å, the NCI isosurface starts diffusing and merging

View Article Online
DOI: 10.1039/D0CE00577K

with the other isosurfaces in the vicinity (**Figure 6g**). Afterwards, the isosurfaces disappear for the weakest N-H•••O=C HBs (**Figures 6h & i**). These observations also suggest that the limit of R_{ij} for the protein main-chain N-H•••O=C HB is ~2.45 Å. Further, these NCI isosurface analyses suggest that the limits of $\nabla^2\rho_{BCP}$ and λ_3 (curvature along the bond) for the N-H•••O=C HB are ~0.62 eÅ^{-5} and ~1.1 eÅ^{-5}, respectively (**Table 3**). The linear dependence of $\nabla^2\rho_{BCP}$ (eÅ^{-5}) with λ_3 (eÅ^{-5}) is shown in **Figure S12**.

View Article Online
DOI: 10.1039/D0CE00577K



View Article Online
DOI: 10.1039/D0CE00577K

Figure 6. NCI isosurfaces of N-H•••O=C HBs **(a)** in the strongest region (~1.72 Å) in one of the α-helices of proteinase K (2pwa), **(b)** in the strongest region (~1.74 Å) in one of the β-sheets of human aldose reductase (1us0), **(c)** from the most populated region (~1.95 Å) in one of the α-helices of 1us0, **(d)** from the most populated region (~1.95 Å) in one of the β-sheets of DFP-ase (1pjx), **(e)** from the weaker region (~2.44 Å) in one of the β-sheets of 1us0, **(f)** at the interface of weaker and weakest regions (~2.45 Å) in one of the α-helices of 1us0, **(g)** at the interface of weaker and weakest regions (~2.46 Å) in one of the α-helices of 1us0, **(h)** from the weakest region (~2.55 Å) in one of the β-sheets of 2pwa and **(i)** from the weakest (~2.6 Å) regions in one of the α-helices of 1us0. The color scale for sign(λ_2).ρ *au* is given on the left.

Table 3. Topological parameters of N-H···O=C HBs across the strongest to weakest regions as shown in the Laplaican maps (**Figures S10**) and NCI isosurfaces (**Figure 6a–i**).

Figure 6	PDB ID	C=O	H-N	ρ_{BCP}	R_{ij}	E_{elec}	$\nabla^2\rho_{BCP}$	λ_1	λ_2	λ_3
		of Residue		eÅ^{-3}	Å	kCal/mol	eÅ^{-5}			

View Article Online
DOI: 10.1039/D0CE00577K

(a)	2pwa	phe_113	asp_117	0.297	1.719	-16.75	2.188	-1.92	-1.88	5.99
(b)	1us0	val_258	ala_208	0.282	1.739	-15.66	2.190	-1.76	-1.74	5.69
(c)	1us0	val_27	val_31	0.179	1.949	-9.89	1.450	-0.98	-0.95	3.38
(d)	1pjx	tyr_203	leu_190	0.179	1.949	-10.00	1.520	-0.98	-0.94	3.44
(e)	1us0	ile_74	leu_106	0.060	2.441	-4.38	0.670	-0.23	-0.23	1.13
(f)	1us0	arg_232	ile_236	0.061	2.453	-4.28	0.620	-0.25	-0.24	1.11
(g)	1us0	glu_53	ala_57	0.061	2.457	-4.41	0.610	-0.25	-0.24	1.10
(h)	2pwa	gly_92	lys_57	0.048	2.545	-3.85	0.579	-0.16	-0.13	0.87
(i)	1us0	ala_143	leu_147	0.044	2.585	-3.50	0.520	-0.14	-0.14	0.81

CONCLUSION

View Article Online
DOI: 10.1039/D0CE00577K

The quantitative and qualitative analyses of the topological parameters and the electrostatic interaction energies have been performed on 1443 main-chain N-H•••O=C hydrogen bonds (750 in α-helices and 693 in β-sheets) in 22 high-resolution (0.87 Å to 0.48 Å) secondary protein structures ranging from 4.7 kDa to 54.5 kDa, as retrieved from RCSB PDB. This study stands out to be the first of its kind involving by far the largest number of high-resolution protein structures and HBs both from the α-helices and β-sheets. The topological analyses have been performed based on the experimental electron densities as transferred from the *ELMAM2* database. Further, the estimated values of the E_{elec} compare well with the experimental values as reported in the literature and demonstrate excellent correlations with the topological parameters. Moreover, the relationships and the excellent correlations between the topological parameters are found to be in good agreement with those observed in the cases of small molecules and some protein systems. The limiting values of the topological parameters ($R_{ij} \leq$ 2.45 Å; $\rho(\mathbf{r})$ = ~0.06 eÅ^{-3}; $\nabla^2\rho(\mathbf{r})$ = ~0.62 eÅ^{-5}; λ_3 = ~1.1 eÅ^{-5}) and the corresponding E_{elec} (~ –4.3 kCal mol^{-1}) to establish the presence of true N-H•••O=C HBs in protein main-chain are thus identified upon employing the QTAIM in conjunction with NCI index – the quantum

View Article Online
DOI: 10.1039/D0CE00577K

crystallographic approaches. To the best of our knowledge, the *NCI-ELMAM2* approach using *NCImilano* has been employed for the first time for the qualitative analysis of HBs. Further, similar studies for the quantitative and qualitative analysis of other types of HBs (N-H•••N, C-H•••O etc.) present in these (and possibly some more) protein systems are planned. Lastly, this study based on a large set of high-resolution protein structures demonstrates that the QTAIM approach can be employed not only for the accurate estimation of the electrostatic interaction energies but also to establish the limiting values of the topological parameters of the true HBs in proteins, when studied together with the NCI approach.

CONFLICTS OF INTEREST

There are no conflicts to declare.

ACKNOWLEDGMENT

SKM thanks UGC and SNU for research fellowships. We are grateful to SNU for infrastructure and high-performance computer cluster facility. PM thanks SERB, India

View Article Online
DOI: 10.1039/D0CE00577K

Govt. for research grant (EMR/2014/000491). We thank Dr. Gabriel Saleh and Dr. Christian Jelsch for their help on the use of *NCIMilano* and the *MoPro* utility programs.

NOTES

The following information are provided as supplementary materials:. Details of NCI isosurface generation methods. List of metal ions/atoms and other heavy atoms present in the protein structures. 3D deformation density maps and 2D Laplacian maps of strong and weak HBs. The bond path along with the bond critical points of the HBs in α-helices and β-sheets for a specific protein;. Topological parameters, E_{elec} and ∠C=O···H and ∠N-H···O of i → i+4 α-helices HBs and parallel and antiparallel β-sheets HBs. Details of error estimation, The quadratic and linear relationships of ρ_{BCP} and E_{elec}, Details of the terminal N-H…O=C HBs in the α-helices. R_{ij} vs E_{elec} plot for the terminal HBs. Percentage distributions of the HB populations in HEWL and glutaredoxin NrdH. Laplacian maps of the representative N-H···O=C HBs and the NCI isosurfaces of some selected α-helices and β-sheets from electron densities as derived using the various approaches. Plot of λ_3 (eÅ^{-5}) and $\nabla^2\rho_{BCP}$ (eÅ^{-5}). The file name is "CEC-ESI-20_05_20-revised"

View Article Online
DOI: 10.1039/D0CE00577K

REFERENCES

1 C. N. Pace, *Nat. Struct. Mol. Biol.*, 2009, **16**, 681–682.

2 J. D. Bernal, *Nature*, 1939, **143**, 663–667.

3 W. Kauzmann, *Adv. Protein Chem.*, 1959, **14**, 1–63.

4 C. N. Pace, S. Trevino, E. Prabhakaran and J. M. Scholtz, *Philos. Trans. R. Soc. London B Biol. Sci.*, 2004, **359**, 1225–1235.

5 P. J. Fleming and G. D. Rose, *Protein Sci.*, 2005, **14**, 1911–1917.

6 G. D. Rose, P. J. Fleming, J. R. Banavar and A. Maritan, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, 103, 16623–16633.

7 J. Kyte, *Structure in protein chemistry*, Garland Science, 2006.

8 A. E. Mirsky and L. Pauling, *Proc. Natl. Acad. Sci. U. S. A.*, 1936, **22**, 439–447.

9 L. Pauling, R. B. Corey and H. R. Branson, *Proc. Natl. Acad. Sci.*, 1951, **37**, 205–211.

View Article Online
DOI: 10.1039/D0CE00577K

10 L. Pauling and R. B. Corey, *Proc. Natl. Acad. Sci. U. S. A.*, 1951, **37**, 729–740.

11 A. Warshel, P. K. Sharma, M. Kato and W. W. Parson, *Biochim. Biophys. Acta - Proteins Proteomics*, 2006, **1764**, 1647–1676.

12 S. Scheiner, in *Reviews in Computational Chemistry*, eds. D. Boyd and K. Lipkowitz, VCH Publishers, New York, 1991, pp. 165–218.

13 E. I. Howard, R. Sanishvili, R. E. Cachau, A. Mitschler, B. Chevrier, P. Barth, V. Lamour, M. Van Zandt, E. Sibley, C. Bon, D. Moras, T. R. Schneider, A. Joachimiak and A. Podjarny, *Proteins Struct. Funct. Bioinforma.*, 2004, **55**, 792–804.

14 P. Munshi, S.-L. Chung, M. P. Blakeley, K. L. Weiss, D. A. A. Myles and F. Meilleur, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2012, **68**, 35–41.

15 P. Coppens, *X-ray charge densities and chemical bonding*, International Union of Crystallography, 1997, vol. 4.

16 R. Herbst-Irmer and D. Stalke, *Acta Crystallogr. Sect. B*, 2017, **73**, 531–543.

17 P. Munshi and T. N. Guru Row, *CrystEngComm*, 2005, **7**, 608.

View Article Online
DOI: 10.1039/D0CE00577K

18 E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol.*, 1984, **44**, 97–179.

19 R. S. Lipsitz, Y. Sharma, B. R. Brooks and N. Tjandra, 2002, **124**, 10621–10626.

20 D. J. Willock, S. L. Price, M. Leslie and C. R. A. Catlow, *J. Comput. Chem.*, 1995, **16**, 628–647.

21 N. K. Hansen and P. Coppens, *Acta Crystallogr. Sect. A*, 1978, **34**, 909–921.

22 B. Fournier, E.-E. Bendeif, B. Guillot, A. Podjarny, C. Lecomte and C. Jelsch, *J. Am. Chem. Soc.*, 2009, **131**, 10929–10941.

23 R. F. W. Bader, *Atoms in molecules*, Wiley Online Library, 1990.

24 R. F. W. Bader, *J. Phys. Chem. A*, 1998, **102**, 7314–7323.

25 G. R. Runtz, R. F. W. Bader and R. R. Messer, *Can. J. Chem.*, 1977, **55**, 3040–3045.

26 R. F. W. Bader and H. Essén, *J. Chem. Phys.*, 1984, **80**, 1943–1960.

27 P. Munshi and T. N. Guru Row, *J. Phys. Chem. A*, 2005, **109**, 659–672.

View Article Online
DOI: 10.1039/D0CE00577K

28 P. R. Mallinson, G. T. Smith, C. C. Wilson, E. Grech and K. Wozniak, *J. Am. Chem. Soc.*, 2003, **125**, 4259–4270.

29 C. Gatti, E. May, R. Destro and F. Cargnoni, *J. Phys. Chem. A*, 2002, **106**, 2707–2720.

30 R. Parthasarathi, S. S. Raman, V. Subramanian and T. Ramasami, *J. Phys. Chem. A*, 2007, **111**, 7141–7148.

31 C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing and C. Lecomte, *Proc. Natl. Acad. Sci.*, 2000, **97**, 3171–3176.

32 Y. Hirano, K. Takeda and K. Miki, *Nature*, 2016, **534**, 281–284.

33 J. Held and S. van Smaalen, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2014, **70**, 1136–1146.

34 B. Guillot, C. Jelsch, A. Podjarny and C. Lecomte, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2008, **64**, 567–588.

35 B. Zarychta, A. Lyubimov, M. Ahmed, P. Munshi, B. Guillot, A. Vrielink and C.

View Article Online
DOI: 10.1039/D0CE00577K

Jelsch, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2015, **71**, 954–968.

36 K. Takaba, K. Takeda, M. Kosugi, T. Tamada and K. Miki, *Sci. Rep.*, 2017, **7**, 43162.

37 D. Liebschner, C. Jelsch, E. Espinosa, C. Lecomte, E. Chabrière and B. Guillot, *J. Phys. Chem. A*, 2011, **115**, 12895–12904.

38 M. Holcomb, R. Adhikary, J. Zimmermann and F. E. Romesberg, *J. Phys. Chem. A*, 2018, **122**, 446–450.

39 S. K. Mandal and P. Munshi, in *Understanding Intermolecular Interactions in the Solid State: Approaches and Techniques*, ed. D. Chopra, The Royal Society of Chemistry, 2018, pp. 189–210.

40 B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte and C. Jelsch, *Acta Crystallogr. Sect. A Found. Crystallogr.*, 2007, **63**, 108–125.

41 S. Domagała, B. Fournier, D. Liebschner, B. Guillot and C. Jelsch, *Acta Crystallogr. Sect. A Found. Crystallogr.*, 2012, **68**, 337–351.

View Article Online
DOI: 10.1039/D0CE00577K

42 B. Guillot, L. Viry, R. Guillot, C. Lecomte and C. Jelsch, *J. Appl. Crystallogr.*, 2001, **34**, 214–223.

43 C. Jelsch, B. Guillot, A. Lagoutte and C. Lecomte, *J. Appl. Crystallogr.*, 2005, **38**, 38–54.

44 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498–6506.

45 G. Saleh, C. Gatti, L. Lo Presti and J. Contreras-García, *Chem. - A Eur. J.*, 2012, **18**, 15523–15536.

46 L. Massa and C. F. Matta, *J. Comput. Chem.*, 2018, **39**, 1021–1028.

47 A. Genoni, L. Bučinský, N. Claiser, J. Contreras-García, B. Dittrich, P. M. Dominiak, E. Espinosa, C. Gatti, P. Giannozzi, J. Gillet, D. Jayatilaka, P. Macchi, A. Ø. Madsen, L. Massa, C. F. Matta, K. M. Merz, P. N. H. Nakashima, H. Ott, U. Ryde, K. Schwarz, M. Sierka and S. Grabowsky, *Chem. - A Eur. J.*, 2018, **24**, 10881–10905.

View Article Online
DOI: 10.1039/D0CE00577K

48 S. Grabowsky, A. Genoni and H.-B. Bürgi, *Chem. Sci.*, 2017, **8**, 4159–4176.

49 J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625–632.

50 M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 235–242.

51 F. H. Allen, *Acta Crystallogr. Sect. B Struct. Sci.*, 1986, **42**, 515–522.

52 K. A. Majorek, M. D. Zimmerman, M. Grabowski, I. G. Shabalin, H. Zheng and W. Minor, in *Structural Biology in Drug Discovery*, Wiley, 2020, pp. 253–275.

53 U. Koch and P. L. A. Popelier, *J. Phys. Chem.*, 1995, **99**, 9747–9754.

54 G. Saleh, L. Lo Presti, C. Gatti and D. Ceresoli, *J. Appl. Crystallogr.*, 2013, **46**, 1513–1517.

55 C. B. Hübschle and B. Dittrich, *J. Appl. Crystallogr.*, 2011, **44**, 238–240.

View Article Online
DOI: 10.1039/D0CE00577K

56 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V Ortiz, J. Cioslowski and D. J. Fox, Gaussian09 Revision D.01.

57 W. Humphrey, A. Dalke and S. Klaus, *J. Mol. Graph.*, 1996, **14**, 33–38.

58 E. Espinosa, M. Souhassou, H. Lachekar and C. Lecomte, *Acta Crystallogr. Sect.*

View Article Online
DOI: 10.1039/D0CE00577K

B, 1999, **55**, 563–572.

59 D. Arias-Olivares, E. K. Wieduwilt, J. Contreras-García and A. Genoni, *J. Chem. Theory Comput.*, 2019, **15**, 6456–6470.

View Article Online
DOI: 10.1039/D0CE00577K

Table of Content

The limiting values of the topological parameters and the electrostatic interaction energies to establish the presence of the true N-H···O=C HBs in protein main-chain have been identified *via* quantitative and qualitative analyses of electron densities using quantum crystallographic approaches – Quantum Theory of Atoms in Molecules (QTAIM) and Noncovalent Interaction (NCI) index.