Noncovalent Carbon-Bonding Interaction in Proteins

V. Rao Mundlapati[1,2], Dipak Kumar Sahoo[1,2], Suman Bhaumik[1,2], Subhrakant Jena[1,2], Amol Chandrakar[1,2], and Himansu S. Biswal[1,2]*

[1]*School of Chemical Sciences, National Institute of Science Education and Research (NISER), PO- Bhimpur-Padanpur, Via-Jatni, District- Khurda, PIN - 752050, Bhubaneswar, India*

[2]*Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai 400094, India*

* Corresponding Author's E-mail: himansu@niser.ac.in, Phone No: - +91-674-2494 185/186

Abstract: The importance of conventional non-covalent interactions such as hydrogen bond (H-bond) and halogen bond (X-bond) in structure and function of biological molecules are well established while carbon bond (C-bond), yet another non-covalent interaction is to be recognized. In proteins, the occurrences and importance of carbon bonds (C-bonds), the highly directional non-covalent interactions between carbonyl-oxygen acceptors and the sp^3–hybridized–carbon σ–hole donors through $n \rightarrow \sigma^*$ electron delocalization are still unknown. For the first time, we discovered ubiquitous existences of C–bonds in proteins with the help of careful protein structure analysis, quantum calculations, nuclear magnetic resonance and infrared spectroscopy and determined C-bond energies precisely. We demonstrated the importance of C-bonds in explaining photochemistry of oxygen-storage protein myoglobin and protein-DNA interaction. We anticipate inclusion of C-bonds in computational force fields would unravel many more implications of C-bonds in structure, function and dynamics of proteins and protein-ligand/drug complexes.

Significance: The importance of conventional non-covalent interactions such as hydrogen bond (H-bond) and halogen bond (X-bond) in structure and function of biological molecules are well established while carbon bond (C-bond), yet another non-covalent interaction is to be recognized. Like X-bond, C-bond is a σ-hole interaction, discovered in the year 2003 in small organic molecules. However, the occurrence, strength and significance of C-bonds are still unknown. Here by using solution NMR spectroscopy, detail PDB analysis, and quantum calculations, we investigated the C-bonds in proteins and revealed that a significant number of C-bonds are present in proteins, contribute enthalpically to the overall hydrophobic interaction and play a significant role in photo dissociation mechanism of myoglobin and binding of nucleobase to the protein.

The cumulative contributions from diverse type of several non-covalent interactions govern protein folding and establish their structures.[1] Exploring novel non-covalent interactions in biomolecules is fascinating; determining their strength leading to the discovery of many new non-covalent interactions in proteins has remained even compelling and challenging in the last decade[1h, 2]. One such non-covalent interaction is "Tetrel Bond"[3] in general and "Carbon Bond" (C-bond)[4] to be specific. Like halogen bond (X-bond), C-bond is a σ-hole interaction, discovered in the year 2003 by Arunan and coworkers.[4] For the first time they reported the existences of the C-bonds in small organic molecules and the estimated C-bond energy at the CCSD(T) level can be as much as ~-15 kJ/mol. Later on through the X-ray charge density analysis Guru Row and coworkers[5] provided the experimental evidences of C-bonds in natural products and hormones. The C-bond concept was further extended to protein-ligand systems[6] and smaller intermolecular complexes[6-7] by Frontera and coworkers. In a very recent article they showed that C-bonds can stabilize cis conformation in acylhydrazones.[8] Hence the concept of C-bonds is slowly getting recognized by the research communities and yet to be explored in detail in proteins. The fact that the C=O groups in proteins are known to be engaged in varieties of non-covalent interactions, viz. X-H•••O=C (H-bond), X•••O=C (X-bond), and C=O•••C=O interactions through n→π* electron delocalization.[2] However, in proteins the possibility of Z-C•••O=C (Z=C, N, O & S) interaction through n→σ* electron delocalization has never been reckoned and explored. The O atom of C=O can interact with the sp^3-hybridized carbon (C-sp^3) through n→σ* electron delocalization, thereby forming a Z-R_3C•••O=C (R = H & C) interaction or C-bond as depicted in Figure-1. Owing to the ubiquitous presence of C=O and aliphatic side chains in proteins it is expected that the

C^{sp3}•••O=C C-bonds be present in large numbers that would lead to better understanding of the hydrophobic interactions in proteins.

The objective of this work is to scrutinize the protein structures thoroughly to concur C^{sp3}•••O=C C-bonds in proteins, to estimate their enthalpic contributions to the stability of protein structures and to address the role of C-bonds in protein structure and function. Herein, we for the first time provide evidences for C-bonds in proteins with the help of a detail PDB analysis and various quantum chemical calculations. We noticed that there were no software or codes available to search C-bonds in proteins that encouraged us to write our own codes and python scripts to look for C-bonds in proteins.

Carbon Bonds in Proteins through PDB Structure Analysis: The protein structure coordinates were retrieved from the RCSB website[9] which satisfied the following criteria: structure by X-ray crystallography was at less than 2.0 Å resolution (resolution range is 0.5 Å to 2 Å) and with less than 30% sequence identity among the proteins. The distance and angle criteria used to search and identify the Z-R_3C•••O=C are (a) 2.5 Å ≤ $d_{C\cdots O}$ ≤ 3.6 Å, (b) 160º ≤ θ_1 (∠C•••O=C) ≤ 180º, (c) 160º ≤ θ_2 (∠O•••C−Z) ≤ 180º, (d) 1.18 Å ≤ $d_{C=O}$ ≤ 1.35 Å, (e) 1.41 Å ≤ d_{C-Z} ≤ 1.56 Å . The C−bonds are highly directional, hence restrictions on θ_1 and θ_2 is crucial. The lower and upper limit for θ_1 and θ_2 were chosen 160º and 180º, respectively by carefully analyzing the geometrical parameters of several C-bond complexes reported in the literature.[3-8, 10] We found 8358 interactions satisfying these structural criteria and forming C−bonds. Out of the 8358 interactions 109, 1986 and 6263 interactions were found in the ranges of ≤1 Å, 1.01−1.5 Å and 1.51−2.0 Å,

respectively. A histogram of C•••O=C C-bond distance distribution was produced using 0.1 Å bar width (Figure 2A). The shorter distances between O and C in proteins could be attributed to the interpenetration of donor-acceptor orbitals. The donor-acceptor orbital interactions that lead to C-bonds were investigated using natural bond orbital (NBO) analysis, vide infra. Figure 2B shows the 3D heat plots of C•••O=C distance vs. the ∠O•••C-Z and .∠C•••O=C while Figure 2C displays the Ramchandran Plot for the donor residues. Similar Ramchandran plot for the acceptor residues was also obtained, suggesting that the C-bonds can be found in all types of secondary structures. We also observed that almost all amino acids can contribute to the formations of C-bonds in proteins. Among them ALA, LEU, ASP, and GLU are the potential C-bond acceptors summing up to 40% of total interactions whereas LYS is engaged most frequently as C-bond donors. Figure 2D presents a representative example of C•••O=C interactions in a RAD51 protein (PDB ID: 3NTU). The C=O of ALA^{269} interacts with the C_α of LYS^{261}. The distance between O and C is 2.872 Å and the angles (θ_1 and θ_2) are very close to 170º, suggesting the formation of a strong C-bond of binding energy of -21.7 kJ/mol (*vide infra*). The donor- acceptor pair analysis suggests that LYS-ALA, LYS-ASP, and ASN-ALA pairs are the prominent ones (Figure S1). Figure 2E displays different types of C-bond acceptors and donors involved in C-bond formation in proteins. In all most all the cases, the C-bond acceptors are the backbone or side chain carbonyls whereas in 80.4 % cases C-C bonds are involved as C-bond donor followed by C-N and C-O bonds. Several examples of Z-R_3C•••O=C (R = H & C; Z = C, N, & O) interactions are shown in Supplementary Figure S2. To explore the possible location of C-bonds in proteins, we made a secondary structure analysis in detail. A histogram of secondary structure assignment and distribution is shown in Figure 2E. We found C-bond donors are evenly distributed in coils, turns, α-helix and β-strand, whereas

~50% of C-bond acceptors are located in α-helix. The analysis also indicates that the C-bonds are mostly observed in α-helix, coils, turns, summing up to ~70% of C-bonds found in proteins. To ascertain that the C-bonds in proteins were the consequences of $n \rightarrow \sigma^{\star}$ electron delocalization, we carried out natural bond orbital (NBO) analysis of 29 amino acid pairs covering complete distance (2.75 Å ≤ $d_{C\cdots O}$ ≤ 3.6Å) and angle (160º ≤ θ_i ≤ 180º). The donor–acceptor interaction energy (E_{DA}) or the $n_O \rightarrow \sigma^{\star}_{C\text{-}Z}$ electron delocalization energy were calculated using second order perturbation theory and are provided in Table S1. The E_{DA} values were plotted against the distance between interacting O and C atoms and the graph is provided in Figure 3A. An exponential relation between E_{DA} and $d_{C\cdots O}$ was deduced as shown in equation 3.

$$E_{DA} = 3.5 \times 10^{7} \times e^{\frac{-d_{C\cdots O}}{0.184}} \quad \text{for } 2.75\ \text{Å} \leq d_{C\cdots O} \leq 3.6\text{Å} \ \&\ (160^{\circ} \leq \theta_i \leq 180^{\circ}) \qquad (1)$$

There is substantial donor-acceptor orbital overlap i.e. the overlap between the oxygen lone pair orbital (n) with the antibonding sigma orbital of C-Z (σ*). The $n_O \rightarrow \sigma^{\star}_{C\text{-}Z}$ interaction energies are very similar to those of two $n \rightarrow \pi^{\star}$ interaction energies observed through reciprocal C=O•••C=O interaction.[2d] Hence, we would like to draw attention to the fact that one $n_O \rightarrow \sigma^{\star}_{C\text{-}Z}$ interaction is equivalent to two $n \rightarrow \pi^{\star}$ interactions, confirming that C-bonds are strong and viable in proteins. Hence their effects on protein structures and functions should not be ignored. It is noteworthy to mention that carbon bonds have already been proven in small molecule crystal structures through the X-ray charge density analysis by T. N. Guru Row and coworkers.[5] Hence we did not perform the CCDC analysis and will not be discussed henceforth as the purpose of this article is to investigate C-bonds in proteins.

Strength of Carbon Bonds in Proteins: The determination of accurate strength of C-bonds in protein is an insurmountable task owing to the interferences of other non-covalent interactions. The alternative strategy is to choose some model molecular systems that mimic C-bonds in proteins. We chose N,N-dimethylacetamide (NNDMA) as the C-bond acceptor mimicking the amide-C=O group in proteins and several C-bond donors bearing different functional groups. As shown in Figure 2A the C-bond donors include acetonitrile (ACN), nitromethane (NM), alanine (ALA), acetic acid (AA), acetylchloride (AcetylCl), and trifluoroacetic acid (TFACE). We would like to clearly mention here that in ALA, AA, AcetylCl and TFACE the strongest sigma hole site is not on the C atom of the CH_3 group but on the carbon atom in the C=O and CF_3 units. However the whole exercise was done to find out how different functional groups control the Z-$H_3C\oplus\oplus\oplus O{=}C$ C-bond strength and consequently the C-bond distance. The optimized structures (space-filling) of C-bond complexes are shown in Figure 1D. We did not observe any steric hindrance from other groups for the oxygen to approach carbon of CH_3 group and form C-bond. We performed several quantum chemical analyses to theoretically establish the C-bond formation in the model donor-acceptor pairs. Figure 1B displays the interacting natural bond orbitals. The NBO[11] analysis reveals that there is substantial donor-acceptor orbital overlap i.e. the in phase orbital overlap between oxygen lone pair (n_O) of NNDMA and the C-C antibonding sigma orbital ($\sigma^*_{C\text{-}C}$) of ACN leading to ~3 kJ/mol of $n\rightarrow\sigma^*$ interaction energy. The C-C•••O=C interaction was further supported by Bader's atoms in molecules (AIM) electron density analysis. The molecules presented in Figure 2D show a bond critical point (BCP) along the C•••O bond vector, inferring accumulation of electron density in between non-covalently bonded O and C atoms forming the C-bond. We also performed Non-Covalent Interaction (NCI) index analysis and molecular electrostatic potential (MESP) calculations to establish that the

C•••O=C interactions are attractive and σ-hole interactions. Figure 2E depicts reduced density gradient isosurfaces in real 3D space for C•••O=C C-bond, the blue isosurface indicates attractive noncovalent interaction between NNDMA and ACN. Figure 2F displays MESP contours of NNDMA-ACN complex attesting the σ-hole interaction. We carried out all the aforementioned analyses for the rest of the donor-acceptor pairs and confirmed the existences of C-bonds of variable strengths in these complexes (Supplementary Figure S3-S8). We noticed same exponential relation between E_{DA} and $d_{C\cdots O}$ for these model compounds (point 1-6 in Figure 3A).

Now it is computationally affordable to calculate the C-bond energies (D_0: dimerisation energy with zeropoint vibrational energy correction) of the model complexes with a very accurate method like coupled-cluster with single and double and perturbative triple i.e. at the CCSD(T) level. We found an exponential correlation between D_0 and E_{DA} as shown in Figure 3B D_0 and E_{DA} are related as

$$D_0 = 22.2 \times e^{\frac{-E_{DA}}{1.6}} - 22.2 \qquad (2)$$

Since the calculation of E_{DA} is not as computationally expensive as CCSD(T), the above expression will be useful in predicting CCSD(T) binding energy with the prior knowledge of n→σ* interaction energy. By combining equation (1) and (2), we derived another expression as shown below

$$D_0 \approx e^{(lnln\,(22.2)\,-22.2\times 10^6\times (e^{\frac{-d_{C\cdots O}}{0.184}}))} - 22.2 \qquad (3)$$

The equation 3 can be used to estimate Z-C•••O=C C-bond energy very precisely just by knowing the C•••O distances. The estimated D_0 of NNDM-ACN C-bond complex by using the

above formula is -16.4 kJ/mol that is comparable with that of D_0 obtained at CCSD(t) level. The same is true for other model complexes. We are now in a convincing position to use this empirical equation to estimate the C-bond energy in proteins from the C•••O distances. The estimated C-bond energies in proteins are in the range of -2 to -22 kJ/mol. As shown in Figure S9, we devised a C-bond energy protractor that can be used to measure C-bond strengths in proteins.

Possible Role of C-bonds in Structure and Function of Proteins: Since the C-bonds are discovered few years back, we did not find a single report on the role of C-bonds in protein functions. Here we provide two examples where the concept of C-bonds can be evoked to explain many biophysical phenomena. In a recent report[12] entitled "Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation", the authors attribute the ultrafast structural changes in the carbonmonoxy myoglobin complex upon photolysis of the Fe-CO bond to the coupling of the ultrafast heme doming mode to the large scale modes of the proteins through the low frequency vibrations. The low frequency modes of the protein are mostly assigned to the H-bonded residues, e.g. Lys^{98} O–Lys^{42} N_ξ distance shows oscillations with a period of 500±150 fs through an N-H•••O=C H-bond. However these H-bonded residues are remotely located from the heme center and the oscillations are transmitted from residues that are directly bonded to the heme center through multiple H-bonds. A careful structure analysis of the CD corner of CO myoglobin revealed that Lys^{42}O is in direct contact with the heme through a C•••O=C C-bond (Figure 3D). Using our aforementioned empirical relation we estimated the C-bond energy from the C•••O distance and it is found to be ~15.7 kJ/mol.

We performed relaxed potential energy scan as a function $d_{C\cdots O}$ (Figure 3C) and fitted to Morse potential to deduce intermolecular stretching frequency ($\sigma_{C\cdots O}$) for NNDMA-ACN that has

similar $d_{C\cdots O}$ distance as in myoglobin. The $\sigma_{C\cdots O}$ is ~75 cm^{-1} (equivalent to 2.25 THz). The corresponding time period for this C-bond is ~450 fs which is highly in synergy with the ultrafast heme doming mode (417-430 fs)[12]. Hence the heme doming oscillation can easily be transmitted directly through Lys^{42} to other parts of the protein via a C-bond. However it is not explored in the aforementioned article as the C-bond is yet to be recognized as another important non-covalent interaction that can influence the structures and functions of proteins. One more example as shown in Figure S10 can be cited where 6-methyleadenine (6MA) can be held in the active center of the ribosome-inactivating proteins with a C-bond. In their article Yu Wang and coworkers[13] reported that the ligand (6MA) and protein interaction is due to "hydrophobic forces", aromatic stacking and H-bonds. We observed a C-bond between Glu^{160}-C=O and CH_3-6MA. The estimated C-bond energy (using equation 3) in this case is -21.7 kJ/mol. This C-bond energy is strong enough to be considered as a prominent non-covalent interaction to bind 6MA in the active center of the protein. Hence, the concepts of C-bonds can unequivocally be considered to explain a debatable term frequently used in biomolecules called "hydrophobic forces or hydrophobic interactions". These two are just representative examples. We expect the importance of C-bonds would be explored in many more systems in near future.

Conclusion:

Our data after careful examination of various protein structures suggest that C-bonds are ubiquitous in proteins. We proposed a functional to estimate the C-bond energy in proteins from the C•••O distances. We further established that the C-bonds can contribute to the functions and structures of proteins, e.g. the transmission of the heme doming oscillation in monocarboxy myoglobin can be possible directly through Lys^{42} to other parts of the protein via a C-bond. The C-bond was also found to be strong enough to be considered as a prominent non-covalent

interaction to bind 6MA in the active centre of ribosome inactivating proteins. We hope that this work will encourage further research on the implications of C-bonding in biomolecules. The C-bonds will definitely compliment other non-conventional noncovalent and they deserve to be included in the new generation force field for bio-molecular structure and function simulation.

Figure Legends

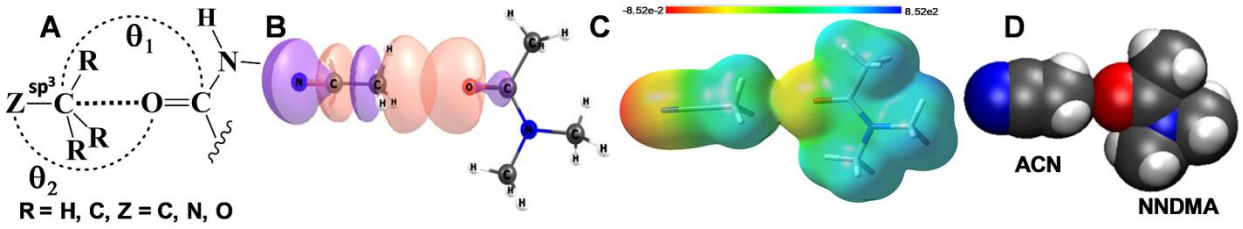

Figure 1. ***(A)*** *Schematic representation of carbon bond.* ***(B)*** *The C-C•••O=C C-bond in NNDMA-ACN characterized by overlap of p-type carbonyl-oxygen lone pair and C-C ƒ* orbital i.e.* $n_O \rightarrow f^*_{C\text{-}C}$ *electron delocalization.* ***(C)*** *Molecular electrostatic potential showing a ƒ-hole on C atom of C-donor. The molecular surface is defined by the isovalue of the electron density (0.004 au). Negative values are shown in red and positive values in blue.* ***(D)*** *Optimized structures of C-bond complexes obtained at B97-D/aug--cc-pVDZ level.*

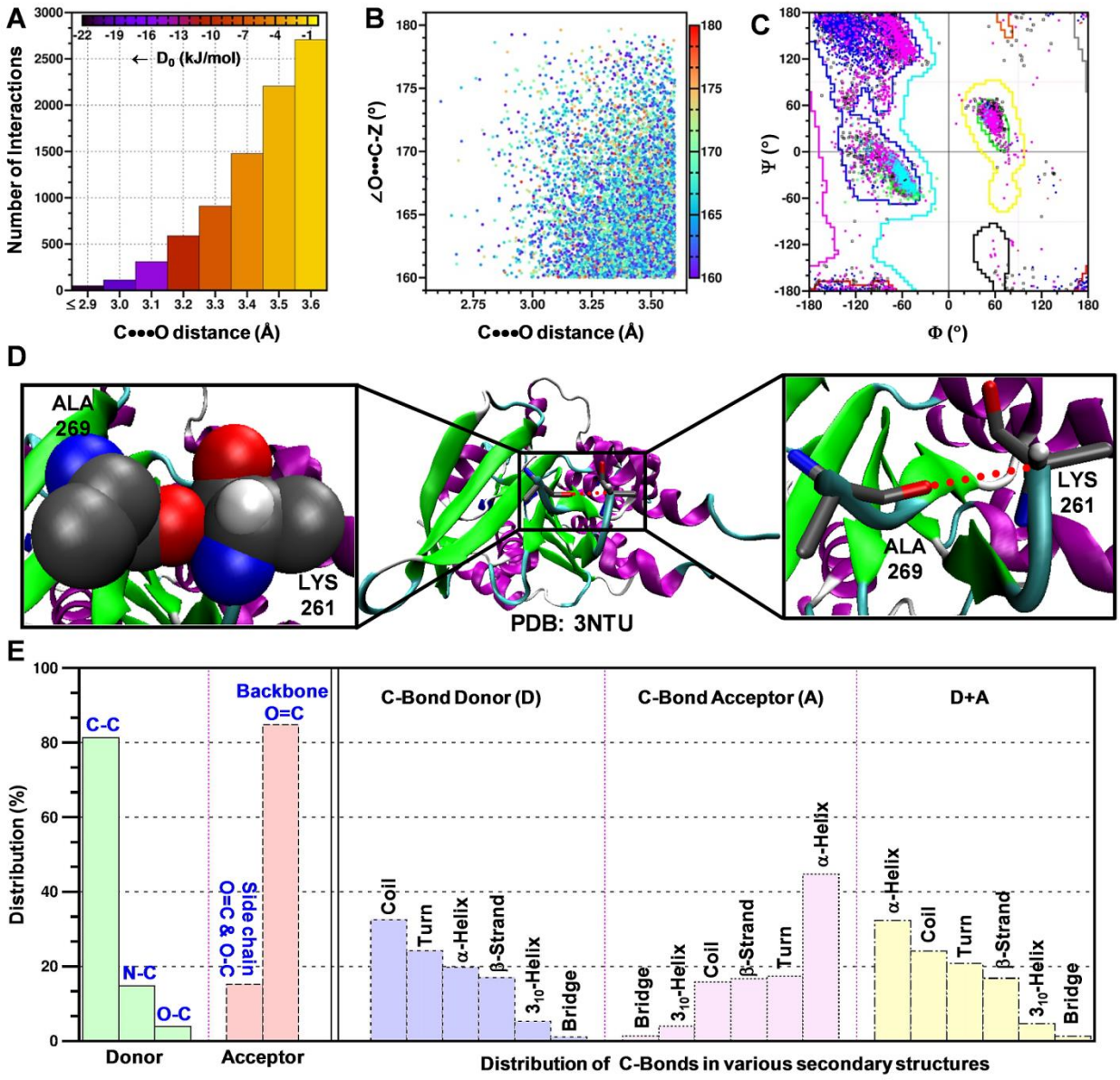

Figure 2. (A) *Frequencies of C•••O carbon bond donor–acceptor distances ($d_{C\cdots O}$) in proteins. Values of the C•••O distances are from protein crystal structures. The histogram represents the distribution of C-bond length within the +.0.4 Å of the sum of the van der Waal's radii of O and C (r_O = 1.52 Å, r_C = 1.70 Å).* ***(B)*** *Scatter plot showing the distribution of proteins in C⊕⊕⊕O distance vs Z-C⊕⊕⊕O angle and color code represents the C⊕⊕⊕O=C Plot showing the percentage distribution of amino acids involved in C-bonds.* ***(C)*** *Ramachandran plot generated by plotting torsion angles (φ, ψ) of all residues involved as C-bond donor.* ***(D)*** *A representative example of Z-C•••O=C C-bond observed in RADA recombinase D302K mutant in complex with AMP-PNP (PDB: 3NTU). The C-bond acceptor residue is alanine and the Ala^{269} and the donor*

*residue is Lys[261]. **(E) Left:** Percentage distribution plot for different type of C-bond acceptors and donors in proteins. **Right:** Histogram plot showing percentage location of C-bond acceptor and donor residues in various secondary structures of the proteins.*

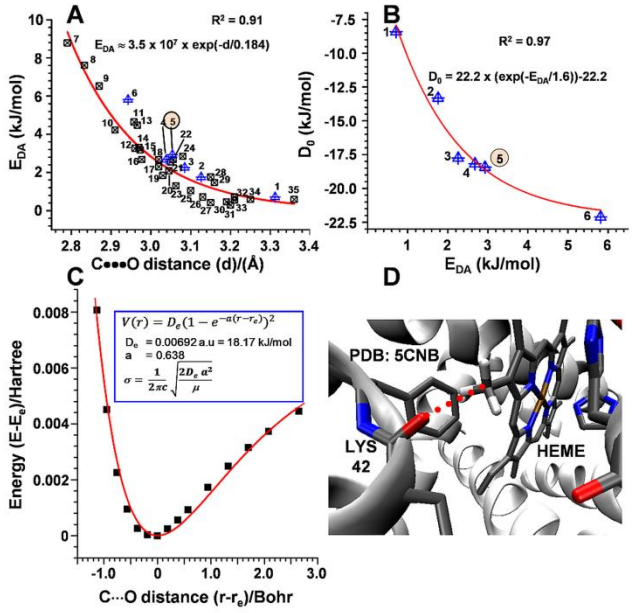

***Figure 3. (A)** The exponential correlation plot between the donor-acceptor energies (E_{DA}) of model C-bond complexes and proteins and C•••O bond distances (d), refer the supplementary information for detail of numbering and C-bond donor-acceptor pairs. The NNDMA-ACN complex is highlighted in circle **(B)** The exponential correlation plot between carbon bond energies (D_0) obtained at CCSD-T level and donor-acceptor interaction energies (E_{DA}) of model C-bond complexes. for which NMR experiments were performed to determine the C-bond energy. (C) Potential energy scan of energy differences ($E-E_e$) over the change in C⊕⊕⊕O C-bond distances($r-r_e$). E is the energy for C⊕⊕⊕O C-bond distance(r) and E_e is the minimum energy for C⊕⊕⊕O equilibrium C-bond distance (r_e) for **C**-bond interaction**.** The solid red line represents Morse potential fitted curve. **(D)** Representative example of implication of C-bonds in protein structure and function Z-C•••O=C, i.e. C-bonds observed in C-C•••O=C C-bond in myoglobin (PDB: 5CNB) formed by Lys[42]-C=O and Heme-CH_3, responsible for the transmission*

of the heme doming oscillation directly through Lys^{42} during CO photolysis in myoglobin to other parts of the protein.

References:

[1] a) R. H. Crabtree, *Science* **1998**, *282*, 2000-2001; b) P. Auffinger, F. A. Hays, E. Westhof, P. S. Ho, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16789-16794; c) N. H. Joh, A. Min, S. Faham, J. P. Whitelegge, D. Yang, V. L. Woods, J. U. Bowie, *Nature* **2008**, *453*, 1266-U1273; d) H. S. Biswal, E. Gloaguen, Y. Loquais, B. Tardivel, M. Mons, *J. Phys. Chem. Lett.* **2012**, *3*, 755-759; e) P. Goymer, *Nat. Chem.* **2012**, *4*, 863-864; f) V. Rao Mundlapati, S. Ghosh, A. Bhattacherjee, P. Tiwari, H. S. Biswal, *J. Phys. Chem. Lett.* **2015**, *6*, 1385-1389; g) V. R. Mundlapati, S. Gautam, D. K. Sahoo, A. Ghosh, H. S. Biswal, *J. Phys. Chem. Lett.* **2017**, *8*, 4573-4579; h) V. R. Mundlapati, D. K. Sahoo, S. Ghosh, U. K. Purame, S. Pandey, R. Acharya, N. Pal, P. Tiwari, H. S. Biswal, *J. Phys. Chem. Lett.* **2017**, *8*, 794-800.

[2] a) A. R. Voth, P. Khuu, K. Oishi, P. S. Ho, *Nat. Chem.* **2009**, *1*, 74-79; b) M. J. Plevin, D. L. Bryce, J. Boisbouvier, *Nat. Chem.* **2010**, *2*, 466-471; c) R. W. Newberry, R. T. Raines, *Nat. Chem. Biol.* **2016**, *12*, 1084-1088; d) A. Rahim, P. Saha, K. K. Jha, N. Sukumar, B. K. Sarma, *Nat Commun* **2017**, *8*, 78.

[3] A. Bauzá, T. J. Mooibroek, A. Frontera, *Angew. Chem. Int. Ed.* **2013**, *52*, 12317-12321.

[4] D. Mani, E. Arunan, *Phys. Chem. Chem. Phys.* **2013**, *15*, 14377-14383.

[5] S. P. Thomas, M. S. Pavan, T. N. Guru Row, *Chem. Commun.* **2014**, *50*, 49-51.

[6] A. Bauza, A. Frontera, *Crystals* **2016**, *6*, 26/21-26/13.

[7] a) A. Bauzá, T. J. Mooibroek, A. Frontera, *Chem. Commun.* **2014**, *50*, 12626-12629; b) A. Bauzá, A. Frontera, T. J. Mooibroek, *Phys Chem Chem Phys* **2016**, *18*, 1693-1698; c) A. Bauza, T. J. Mooibroek, A. Frontera, *Chem. Rec.* **2016**, *16*, 473-487; d) S. Scheiner, *J. Phys. Chem. A* **2018**, *122*, 2550-2562.

[8] M. M. Naseer, M. Hussain, A. Bauzá, K. M. Lo, A. Frontera, *ChemPlusChem* **2018**, *83*, 881-885.

[9] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235-242.

[10] a) A. Bauza, T. J. Mooibroek, A. Frontera, *Phys. Chem. Chem. Phys.* **2014**, *16*, 19192-19197; b) D. Mani, E. Arunan, *J. Phys. Chem. A* **2014**, *118*, 10081-10089; c) A. Bauzá, T. J. Mooibroek, A. Frontera, *ChemPhysChem* **2015**, *16*, 2496-2517; d) E. C. Escudero-Adán, A. Bauzá, A. Frontera, P. Ballester, *ChemPhysChem* **2015**, *16*, 2530-2533; e) P. Pal, S. Konar, P. Lama, K. Das, A. Bauzá, A. Frontera, S. Mukhopadhyay, *J. Phys. Chem. B* **2016**, *120*, 6803-6811.

[11] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899-926.

[12] T. R. M. Barends, L. Foucar, A. Ardevol, K. Nass, A. Aquila, S. Botha, R. B. Doak, K. Falahati, E. Hartmann, M. Hilpert, M. Heinz, M. C. Hoffmann, J. Koefinger, J. E. Koglin, G. Kovacsova, M. Liang, D. Milathianaki, H. T. Lemke, J. Reinstein, C. M. Roome, R. L. Shoeman, G. J. Williams, I. Burghardt, G. Hummer, S. Boutet, I. Schlichting, *Science* **2015**, *350*, 445-450.

[13] Q. Huang, S. Liu, Y. Tang, S. Jin, Y. Wang, *Biochem. J.* **1995**, *309* 285-298.