Syn vs Anti Carboxylic Acids in Hybrid Peptides: Experimental and Theoretical Charge Density and Chemical Bonding Analysis

Rumpa Pal, M. M Reddy, B. B Dinesh, Manjunath A Venkatesha, Simon Grabowsky, Christian Jelsch, Tayur N Guru Row

Article **DOI:** 10.1021/acs.jpca.7b10939

Pal, R., Reddy, M. M., Dinesh, B., Venkatesha, M. A., Grabowsky, S., Jelsch, C., & Guru Row, T. N. (2018). Syn vs Anti Carboxylic Acids in Hybrid Peptides: Experimental and Theoretical Charge Density and Chemical Bonding Analysis. *The Journal of Physical Chemistry A*, *122*(14), 3665-3679.

Syn vs. Anti Carboxylic Acids in Hybrid Peptides: Experimental and Theoretical Charge Density and Chemical Bonding Analysis

Rumpa Pal,[*,§,#] M. B. Madhusudana Reddy,[&,!] B. Dinesh,[&] Manjunath A. Venkatesha,[&] Simon Grabowsky,[#] Christian Jelsch,[%] Tayur N. Guru Row[*§]

[§]Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore, 560012, India. Email:ssctng@sscu.iisc.ernet.in

[&]Molecular Biophysics Unit, Indian Institute of Science, Bangalore, 560012, India

[#]Institute of Inorganic Chemistry and Crystallography, University of Bremen, 28359, Bremen Germany. Email: rpal@uni-bremen.de

[!]School of Chemistry, REVA University, Bangalore, India

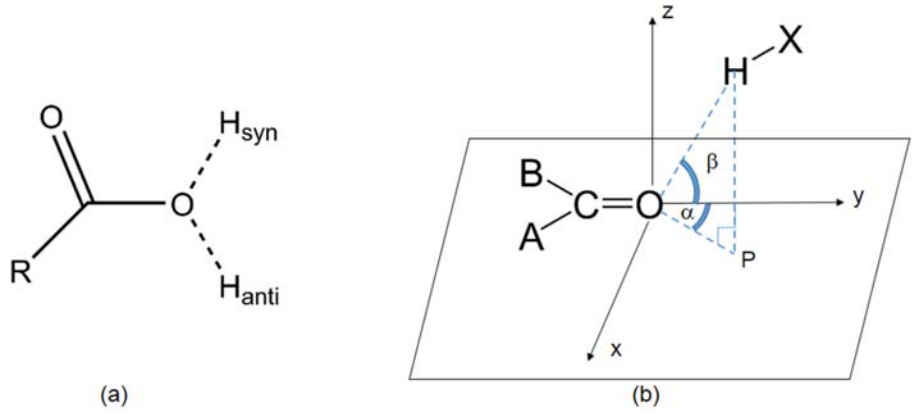
[%]CRM2, CNRS, Université de Lorraine, CNRS, CRM2 , Nancy, France

Abstract: A comparative study of *syn vs. anti* carboxylic acids in hybrid peptides based on experimental electron density studies and theoretical calculations shows that, in the *anti* form, all the three bond angles surrounding $C_{carboxyl}$ of the –COOH group are close to ~120°, as expected for a C-sp^2 atom, whereas in the *syn* form, the $\angle C_{\alpha}–C(O)–O_{hydroxyl}$ angle is significantly smaller by 5–10°. The oxygen atom in the carboxyl group is more electronegative in the *anti* form, so the polarity of the acidic O–H bond is higher in the *anti* form compared to the *syn* form,as observed within the limitations of H atom treatment in X-ray diffraction. Consequently, the investigated *anti* carboxylic acid forms the strongest O–H···O hydrogen bond among all model compounds. Furthermore, according to Natural Bond Orbital analysis, the oxygen lone pairs are clearly non-equivalent, as opposed to

the general notion of hybridization of equivalent sp^2 and sp^3 lone pairs on carbonyl or hydroxyl oxygen atoms. The hybridization of the lone pairs is directly related to the directionality and strength of hydrogen bonds.

Introduction

Carboxylic acids have a rich chemistry since they act as Brønsted and Lewis acids as well as Lewis bases. The carboxylate group serves as a general-base catalyst in various chemical and enzymatic reactions,[1,2] and proton transfer is the most common enzyme catalyzed reaction.[3] The stereoelectronic preferences of lone pairs in carboxylates for catalysis have been an active research topic.[4] In the 1980s, Gandour proposed that the *syn* lone pair of the carboxylic group (Scheme 1a) may be significantly more basic, by 10^3–10^4 in the K_a ratio, than the anti lone pair.[5] This was rationalized with the fact that *syn* carboxylic acids are more stable, hence weaker acids. Correspondingly, in the carboxylate anion, the conjugate base of carboxylic acids, the *syn* lone pairs are more basic (Scheme 1a; in this study, the hybrid peptides are referred to as *syn* and *anti* according to the H atom position.)



Scheme 1. a) *Syn* and *anti* H atom positions in carboxylic acids relative to the carbonyl oxygen atom. The hybrid peptides in this study are referred to as *syn* and *anti* according to the H atom position. Consequently, in *anti* carboxylic acid, the oxygen lone pair is *syn* relative to the carbonyl oxygen atom and vice versa. b) Definition of *α* and *β* angles for H···O=C interactions: P is the projection of H on the ABC=O plane. In the spirit of

spherical polar coordinate system, the angle between OP and C=O directions (*y* axis) in the ABC=O plane is α and the angle between OH and C=O directions (*y* axis) in the plane perpendicular to ABC=O is β.

Experimental and theoretical studies in gas phase and solution confirm that the *syn* lone pairs in carboxylates are more basic, although the difference in basicity of the two carboxylate lone pairs is not as large as originally proposed.[6,7]

The carboxylate group, like other hydrogen-bonding acceptors, can play a role for directing molecular organisation of catalytic components along with promoting catalysis. Although basicity and hydrogen bonding ability are two intrinsic properties of *syn* and *anti* lone pairs of oxygen atoms in carboxylates,[8-10] it has been addressed that stereoelectronic preferences for the two different roles that the carboxylate group plays - molecular organization by H-bond formation and catalytic activity governed by basicity- might not be similar. A Cambridge Structural Database (CSD) analysis showed no significant preference, statistical or geometric, for *syn* hydrogen bonds (involving the *syn* lone pairs of carboxylates) over *anti* hydroden bonds (involving the *anti* lone pairs of carboxylates), reflecting that both the lone pairs sites have equal hydrogen bond acceptor ability.[11] However, an in-depth knowledge of stereochemistry and relative strength of hydrogen bonds is a prerequisite in drug design and crystal engineering.[12,13]

In carboxylic acids, the *syn* form (Scheme 1a) is more stable, by ~6-8 kcal/mol, compared to the *anti* form in the gas phase.[14] In the 1970s, the general perception was that, in crystal structures, the –COOH group mostly adopts the *syn-planar* configuration and only in rare cases when it forms intramolecular O–H···O bonds, the *anti-planar* conformation can occur.[15] It has been shown that the preference for syn reduces to 1-2 kcal mol^{-1} upon formation of O–H···O hydrogen bonds.[10] Different packing motifs of the –COOH group as well as cyclic dimers *vs.* catameric forms have also been studied in detail.[16,17] It has been observed that the preference of *syn* over *anti* in presence of other potential hydrogen bond donors and acceptors is greatly influenced by the formation of hydrogen bonded dimers/catemers. Stabilization of anti –COOH···X (X=neutral or anionic O, Cl, N) has been explored in a recent CSD database survey.[18] A CSD database[19] search for this

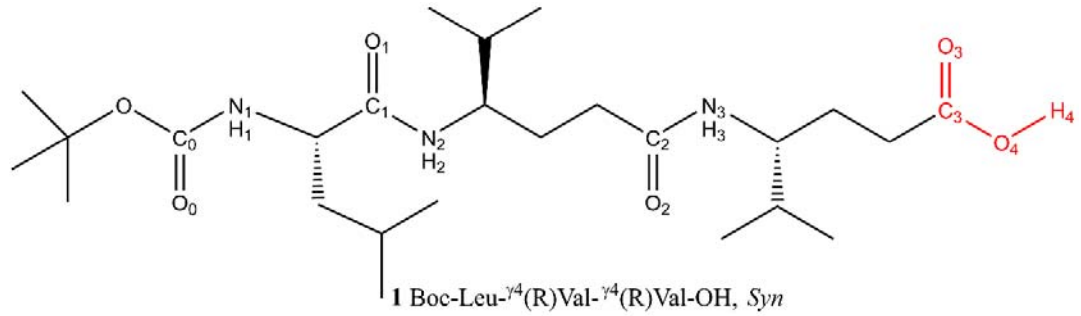
study [version 5.38, using ConQuest 1.19, May 2017] on carboxylic acids shows that *syn* conformations occur much more frequently than *anti* conformations, with 15482 *vs.* 1893 entries, respectively.

In this study, we have chosen hybrid peptides as model compounds to examine and evaluate the preference of *syn* over *anti* conformations of carboxyl functional groups in a variety of environments created through different hydrogen bonding patterns. Conformational analysis of peptides carried out so far in the literature generally focuses on classic H-bonds, N–H···O and O–H···O, with different carbonyl oxygen acceptors, namely (urethane, amide or carboxylic acid) and considerably less on the weaker C–H···O hydrogen bonds as the former dictate the overall conformation in peptide molecules. Hence, in the current study of *syn vs. anti* carboxylic acids, only N–H···O and O–H···O hydrogen bonds with oxygen as acceptor are considered. The insertion of additional atoms into the backbone of synthetic polypeptides greatly enhances the possibility of internally folded structures via hydrogen-bonding.[20,21] Although the common packing motifs associated with the carboxyl groups are dimers and catemers,[15] in hybrid peptides such motifs are absent due to the presence of other hydrogen bond donor and acceptor groups.

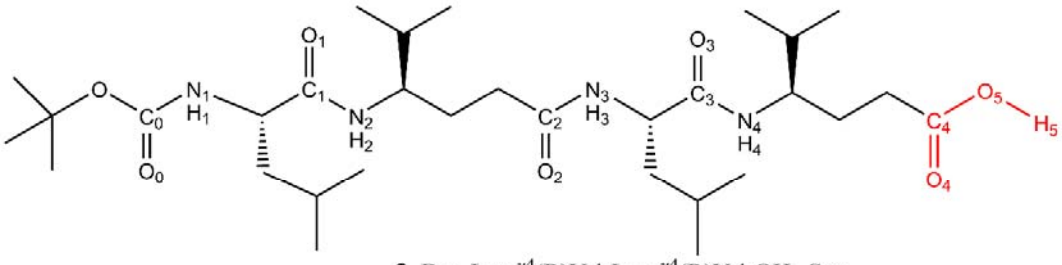
The specific nature of the interactions between donors and acceptors suggests that H-bonding is predominantly directional and there is a statistically significant tendency for hydrogen bonding to occur in the directions of the lone pairs on the acceptor atoms.[22,23] However, the directionality of the hydrogen bonds is still a matter of debate to date. A recent study indicates that the lone pair configuration as obtained from deformation electron density analysis shows striking differences for phenol, alcohol, carbonyl and other oxygen types acting as hydrogen bond acceptors.[24] Although there has been much interest in the geometry of the A–H···B systems (A, B = oxygen, nitrogen, *etc.*), comparatively fewer details are known on the angular distribution of proton donors around an acceptor atom, *eg.*, analysis of H···O=C angles. Also in the literature, the description of the H···O=C angle is not unique.[11,24,25] A CSD analysis study proposed that the variations of the H···O=C angle could be related to the hybridization, i.e. the extent of the sp^2 character on the concerned O atom,[11] but it is still an open question. In this study, we

have considered α, β angle definitions[24] to describe the H···O=C angle as has been implemented in the MoPro package[26,27] (Scheme 1b) and we have performed Natural Bond Orbital (NBO) analysis[28,29] to explore the hybridization of concerned atoms.

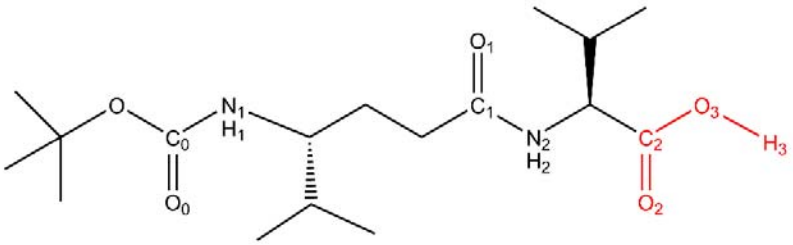
The acidic O–H bond in *syn vs. anti* carboxylic acids has not been investigated to the same extent as that of *syn vs. anti* lone pairs of carboxylates. In conventional X-ray crystallography, X–H distances are found to be shorter than from neutron diffraction and hence common practice is to normalize the X–H distances to values obtained from neutron diffraction experiments.[30,31] However, even, in standard tables that summarize results from neutron diffraction,[30] there are no separate entries for *syn* and *anti* O–H groups of the –COOH moiety. In fact, there is one single representative O–H neutron distance entry for all carboxylic acids.[30] It might be worth recalling that there is a linear correlation between bond length and pK_a values of the conjugate acid based on a CSD database study on ethers and esters.[32] All these analyses underpin the necessity for a systematic investigation of O–H bonds in syn and anti carboxylic acids and associated lone pairs.



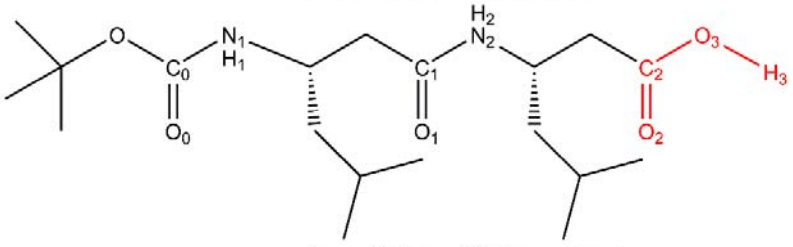
1 Boc-Leu-$^{\gamma 4}$(R)Val-$^{\gamma 4}$(R)Val-OH, *Syn*



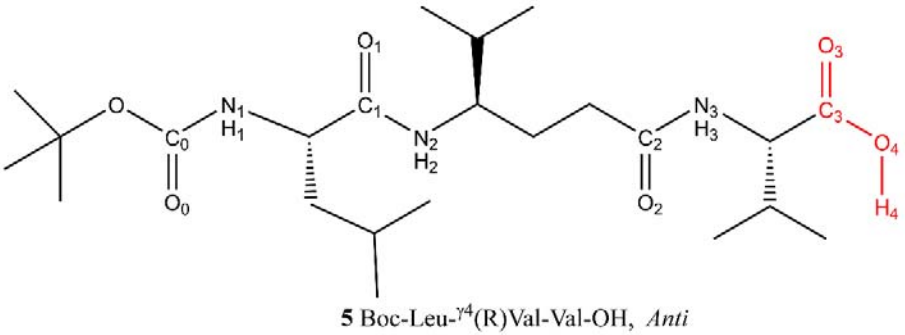
2 Boc-Leu-$^{\gamma 4}$(R)Val-Leu-$^{\gamma 4}$(R)Val-OH, *Syn*



3 Boc-$^{\gamma 4}$(R)Val-Val-OH, *Syn*



4 Boc-$^{\beta 3}$(S)Leu-$^{\beta 3}$(S)Leu-OH, *Syn*



5 Boc-Leu-$^{\gamma 4}$(R)Val-Val-OH, *Anti*

Scheme 2. Chemical structures of the model compounds; four *syn* and one *anti* hybrid peptides with atom labelling scheme.

In this report, we have determined molecular structures of four *syn* carboxylic acids, Boc–Leu–$^{\gamma}$Val–$^{\gamma}$Val–OH, **1**, Boc–Leu–$^{\gamma}$Val–Leu–$^{\gamma}$Val–OH , **2**, Boc–$^{\gamma}$Val–Val–OH, **3**, and Boc–$^{\beta}$Leu–$^{\beta}$Leu–OH, **4**, and one *anti* carboxylic acid Boc–Leu–$^{\gamma}$Val–Val–OH , **5**, (Scheme 2) from single-crystal X-ray diffraction. Among them molecules **1**, **2** and **3** are new compounds synthesized for this study. Because of the crystal quality, we managed to collect an experimental charge density dataset of *anti* **5**, only.

Experimental and Computational Procedures

Experimental Section.

Synthesis.

The three new hybrid peptides **1**, **2** and **3** were synthesized by the standard solution-phase method. The tert-butoxycarbonyl (Boc) group was used for N-terminus protection, and the C-terminus was protected as a methyl ester. Deprotections were performed using 98% formic acid for the Boc group and saponification for the methyl ester. The peptide couplings were performed using N,N′-dicyclohexylcarbodiimide (DCC)/1-hydroxybenzotriazole (HOBt) or 1-[Bis(dimethylamino)methylene]-1H- 1, 2, 3- triazolo [4,5-b] pyridinium 3-oxid hexafluorophosphate (HATU)/HOBt as the coupling reagent. The target peptides were purified by high-performance liquid chromatography (HPLC) on a reverse-phase C18 column (5–10 μm, 7.8 mm × 250 mm) using methanol–water gradients. All the final compounds were characterized by electrospray ionization mass spectrometry (ESI-MS) on a Broker Daltonics Esquire-3000 instrument and further details have been deposited in the supplementary information.

Single Crystal X-ray diffraction.

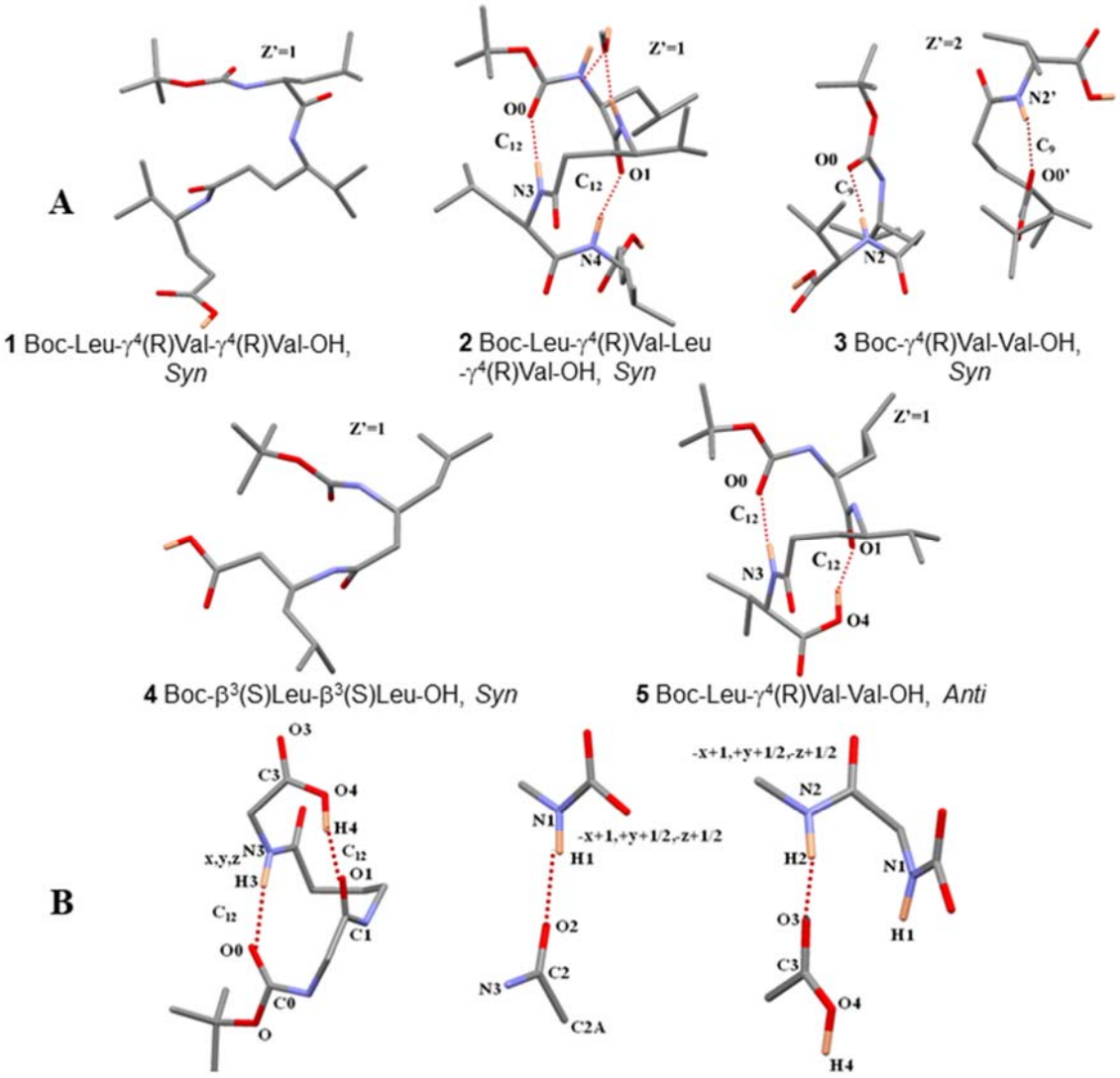
Good quality single crystals were chosen using a polarizing microscope and affixed to a Hampton Research Cryoloop using Paratone-N oil. The crystals were cooled to 100 K with a liquid nitrogen stream using an Oxford Instruments Cryojet-HT nitrogen gas-stream

cooling device. X-ray diffraction data were collected on an Oxford Xcalibur (Mova) diffractometer equipped with an EOS CCD detector using MoKα radiation (λ = 0.71073 Å). The crystal-to-detector distance was fixed at 45 mm and the scan width (Δω) was 1º per frame during the data collection. The data collection strategy was chosen in such a way to yield a high resolution X-ray data set (d = 0.45 Å) with high redundancy and completeness of 100% for *anti* **5**. For the other four *syn* compounds, i.e., **1**-**4**, routine 100(2) K data sets at 0.77 Å resolution were collected. Cell refinement, data integration and reduction were carried out using the program CrysAlisPro.[33] Face indexing was done for an accurate numerical absorption correction for the experimental charge density dataset of *anti* **5**. Sorting, scaling, and merging of the data sets were carried out using the program SORTAV.[34] The crystal structure was solved by direct methods using SHELXS97[35] and refined according to the spherical-atom approximation (based on F^2) using SHELXL97[35] included in the WinGX suite.[36] In most of the cases, the acidic hydrogen atoms were located whereas other hydrogen atoms were fixed stereochemically and the positions and isotropic thermal parameters were allowed to refine in the spherical atom model. The capped-stick representation[37] of the compounds, **1**-**5**, are shown in Figure 1.A and crystallographic details are summarized in Table 1.

Table 1. Crystallographic, measurement and refinement details.

	Syn **1**	*Syn* **2**	*Syn* **3**	*Syn* **4**	*Anti* **5**
Empirical formula	$C_{25}H_{47}N_3O_6$	$C_{32}H_{62}N_4O_8$	$C_{17}H_{32}N_2O_5$	$C_{19}H_{36}N_2O_5$	$C_{23}H_{43}N_3O_6$
CCDC number	1059075	1059077	1007207	1059078	1059076
Formula weight (g/mol)	485.65	630.85	344.45	372.50	457.6
Crystal size (mm)	0.06x0.21x0.38	0.58x0.11x0.04	0.46x0.30x0.24	0.07x0.09x0.55	0.17x0.25x0.56
Crystal system	Monoclinic	Monoclinic	Monoclinic	Orthorhombic	Orthorhombic
Space Group	$P2_1$	$P2_1$	$P2_1$	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$
a (Å)	10.224(1)	10.312(1)	9.868(1)	5.215(1)	9.4237(1)
b (Å)	12.325(1)	10.980(2)	17.500(2)	15.095(1)	15.9868(2)
c (Å)	12.340(1)	16.812(2)	12.292(1)	27.017(3)	17.3878(2)
β (°)	109.39(1)	106.62(1)	102.71(1)	90	90
Volume (Å^3),Z	1466.7(3),2	1824.1(4),2	2070.7(4),4	2126.8(4),4	2619.56(1),4
Calculated density (gcm^{-3})	1.10	1.15	1.11	1.16	1.16
F(0 0 0)	532	692	752	816	1000
λ (Å)	0.71073	0.71073	0.71073	0.71073	0.71073
μ (mm^{-1})	0.078	0.082	0.081	0.083	0.083

Temperature (K)	100(2)	120(2)	100(2)	100(2)	100(2)
θ range (°)	2.7-26.0	2.5-25.0	2.4-26.0	2.6-27.6	2.5-46.8
R_{int}	0.064	0.072	0.0(twin)	0.118	0.048
Measured reflections	11439	6938	12322	17022	95816
Unique reflections	5758	5201	12322	4902	23524
Completeness (%)	99.9	99.8	99.9	99.9	99.9
			Spherical Atom Refinement		
R_{all} / wR_2	0.106/0.144	0.183/0.235	0.046/0.092	0.151/0.188	0.074/0.116
R_{obs}	0.071/0.129	0.097/0.174	0.039/0.090	0.095/0.188	0.047
GoF	0.999	1.020	1.010	1.077	1.026
$\Delta\rho_{min}$, $\Delta\rho_{max}$ (eÅ^{-3})	-0.228,0.382	-0.304,0.289	-0.168, 0.189	-0.441,0.409	-0.251,0.447
Data to parameter ratio	5758/328=17.55:1	5201/410=12.69:1	12322/449=27.44:1	4902/278=17.63:1	23524/314=74.91:1
			Multipole refinement		
Refinement based on					F^2
No. of reflections					20830
Total number of parameters					1166
R(F)/$wR_2(F^2)$					0.036/0.057
GOF					0.967
$\Delta\rho_{min}$, $\Delta\rho_{max}$ (eÅ^{-3})					-0.35,0.43



1 Boc-Leu-γ^4(R)Val-γ^4(R)Val-OH, *Syn*

2 Boc-Leu-γ^4(R)Val-Leu -γ^4(R)Val-OH, *Syn*

3 Boc-γ^4(R)Val-Val-OH, *Syn*

4 Boc-β^3(S)Leu-β^3(S)Leu-OH, *Syn*

5 Boc-Leu-γ^4(R)Val-Val-OH, *Anti*

Figure 1. **A**. Capped stick representation[37] of molecular structures of the four *syn* **1**-**4** and one anti **5** compounds based on the experimental geometries obtained from single crystal X-ray diffraction, showing intramolecular hydrogen bonds. **B** Intra and intermolecular hydrogen bonds in *anti* **5** with different O acceptors.

Multipole Refinement

The charge density modeling and multipolar non-spherical atom refinements for *anti* **5**, were performed according to the Hansen and Coppens multipole formalism[38] with MoPro[39] and XD2006[40] packages. Initially in XD2006, the modelling was done with reflections $I > 2\sigma(I)$. The core and valence scattering factors of all atoms were derived

from Su, Coppens and Macchi wave functions.[41] In this model, no chemical symmetry constraints were applied to the structures, whereas separate κ and κ′ were used to define different atom types based on chemical environment.

In the second model obtained using MoPro, the least square minimization was carried out using all reflections. Chemical equivalence and local symmetry restraints were applied to the charge density (σ = 0.01). κ and κ' of H atoms were restrained to be similar (σ = 0.05). κ' of non H atoms were restrained to 1(σ =0.05). The core and valence scattering factors of all atoms were derived from Clementi & Roetti wave functions.[42] Initially, the scale factor was refined, then positional and thermal displacement parameters were refined against all reflections. The X−H bond lengths were constrained to the values determined by neutron diffraction experiments[43]. The isotropic displacement parameters for H atoms were riding on their carrier atom U_{iso} = 1.2 U_{eq} (CH_2, CH, NH) or 1.5 U_{eq} (CH_3, OH). For H atoms, bond directed dipole (d_z) components were allowed to refine. For non-hydrogen atoms, multipole populations P_{lm} were modelled up to octupole level (l = 3). The different charge density parameters P_{lm}, κ', P_{val} and κ were introduced in a stepwise manner in the refinement. The scale, positional, thermal anisotropic displacement parameters, P_{lm}, P_{val}, κ and κ' were refined successively, until convergence was reached. Near convergence, the model was used to calculate anisotropic displacement parameters of H atoms using the SHADE3 analysis[44,45]. Estimated ADPs for H atoms were kept fixed during the subsequent multipole refinements. The VMoPro tool[26,27] was used to generate residual, deformation and Laplacian plots and to analyze the electron lone pairs.

The second model with all reflections using chemical equivalence and local symmetry restraints produces a superior description of the charge density features and hence this model has been considered for all electron-density results shown in this paper. The fractal dimension plot for residual density corresponding to the second model is shown in Figure S1 in the supporting information.

Computational Methods

Theoretical charge density analysis was carried out using GAUSSIAN09[46] at B3LYP-D/6-311++g(2d,2p) level followed by the AIMAll package[47] for single molecules. NBO analysis

was performed with NBO6.0 as implemented in Gaussian09 at B3LYP-D/6-311++g(2d,2p) level for all the model compounds, **1**-**5**. As input geometry, for syn **1**-**4** compounds, the X-ray structure with H distances elongated to neutron values was considered, and for *anti* **5**, the geometry resulting from the experimental multipole model [as per $I > 2\sigma(I)$] was used.

For periodic calculations, two approaches were considered: (i) generating theoretical structure factors from CRYSTAL14[48] at the B3LYP/TZVP level[49,50] followed by multipole modeling in XD2006[40] (ii) obtaining topological features directly from wave function analysis using TOPOND[51,52] as implemented in CRYSTAL14.

(i) The input geometries of compounds for single point periodic quantum chemical calculations using CRYSTAL14[48] were taken from (a) the experimental multipole model on *anti* **5**, (b) from the routine data sets at 100 K for *syn* **1** and **2**. The single point calculations were performed at the B3LYP/TZVP level of theory. The shrinking factors (IS1, IS2, and IS3) along with the reciprocal lattice vectors were set to 4 (30 k-points in irreducible Brillouin zone). The bielectronic Coulomb and exchange series values for the truncation parameter were set as ITOL1−ITOL4 = 6 and ITOL5 = 14, respectively. The level shifter was set to 0.7 Hartree/cycle for better convergence. Upon convergence on energy ($\sim 10^{-7}$ Hartree), the periodic wave functions were obtained, and subsequently theoretical structure factors at the same resolution as observed from the experiments were calculated by a standard procedure implemented in CRYSTAL14. All theoretical structure factors were assigned unit weights during the refinements based on the methodology followed in the literature.[53,54] The anisotropic displacement parameters were set to zero to consider a static model, and multipolar refinements of the theoretical data were carried out up to the same levels as those used for the experimental charge density modeling.

(ii) TOPOND[55] is considerably different from other existing implementations of QTAIM[56] for crystalline systems[52] due to its interface with the CRYSTAL package. It thus becomes a powerful tool for applying QTAIM to molecules, polymers, surfaces, and crystals, exploiting the full symmetry of each of these systems. It calculates full topological features of $\rho(r)$ and $-\nabla^2\rho(r)$ scalar fields along with other QTAIM descriptors directly from wave function analysis. The level of theory used is B3LYP/ TZVP in our systems.

Results and Discussion

Geometry of the carboxyl group in *syn* and *anti* conformation

The geometric parameters of the carboxyl moiety, relevant bond distances, bond angles and torsion angles in compounds **1**-**5** are shown in Figure 2. There is a slight change in C=O and C–$O_{hydroxyl}$ bond lengths of ~0.01 Å in *syn vs. anti* compounds, but a rearrangement of bond angles surrounding the carbonyl C atom in these two conformations is more pronounced. In anti **5**, all the three angles, $\angle C_{\alpha}$–C(O)–$O_{carbonyl}$ (coloured in blue in Figure 1), $\angle C_{\alpha}$–C(O)–$O_{hydroxyl}$ (green), $\angle O_{carbonyl}$–C(O)–$O_{hydroxyl}$ (red) are symmetrically arranged, with values around ~120° each, as per sp^2 hybridization of carbonyl C atom. But in case of *syn*, two of the angles, mainly $\angle O_{carbonyl}$–C(O)–$O_{hydroxyl}$ (red) along with $\angle C_{\alpha}$–C(O)–$O_{carbonyl}$ (blue) are greater than 120° and as a result $\angle C_{\alpha}$–C(O)–$O_{hydroxyl}$(green) has shrunk by up to ~6-8° compared to that of *anti* form.

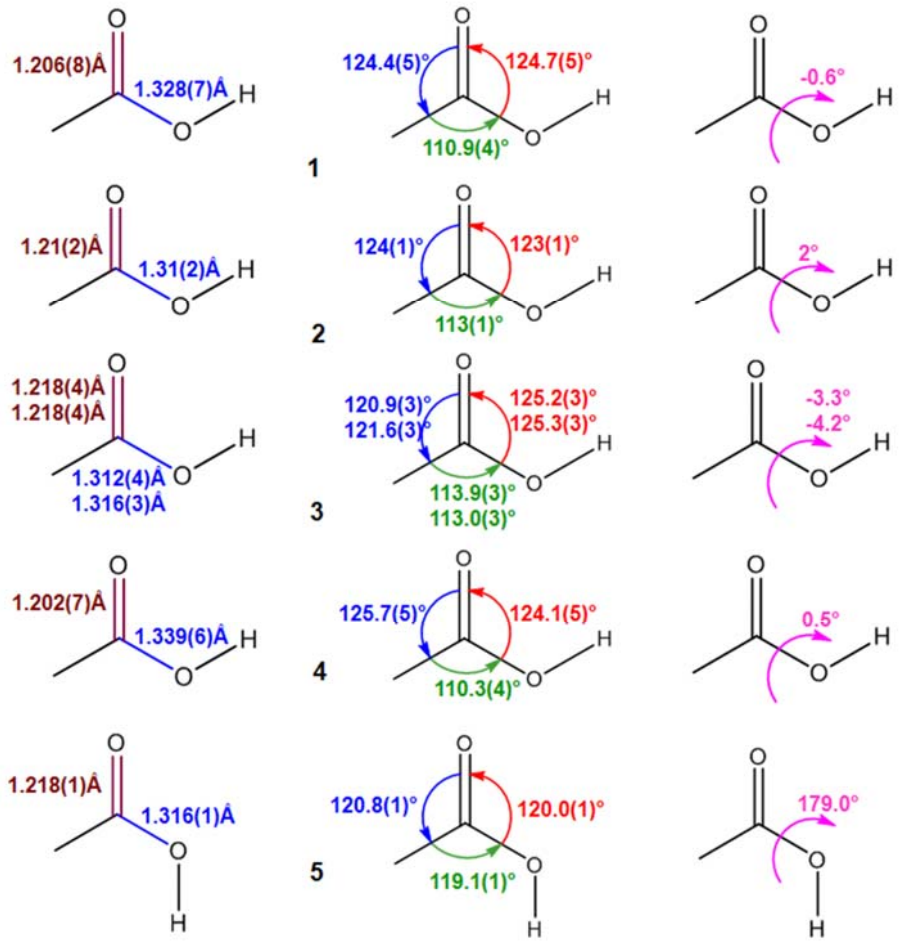


Figure 2. Comparison of bond lengths, bond angles and torsion angles of the carboxyl group in *syn* (compounds **1**-**4**) and *anti* (compound **5**) carboxylic acids. The angle marked in 'green' shows the largest deviations between *syn* and *anti* conformations.

A subsequent CSD database survey on relative populations of *syn vs. anti* carboxylic acids, restricted only to 'amino acids, peptides and complexes' class of compounds, has produced 2000 *vs.* 137 entries of individual –COOH moieties with *syn* and *anti* conformations, respectively. The distribution of torsion angles, bond angles and bond lengths of the –COOH moiety (Figure 3) supports the observation in our model systems as a general trend. With a mean value of 112.5° (2.0) for the angle $\angle C\alpha–C(O)–O_{hydroxyl}$ in *syn* compared to 117.6°(2.8) in *anti* compounds, there is a shrinkage of ~5.1° in *syn*. To balance, the increase in the remaining two angles, $\angle C_\alpha–C(O)–O_{carbonyl}$ and $\angle C_\alpha–C(O)–$

$O_{hydroxyl}$ in *syn* are ~3.6° and ~1.5° larger, respectively, compared to those in the *anti* form. The changes in the bond lengths remain small (~0.01 Å) (Figure 3).

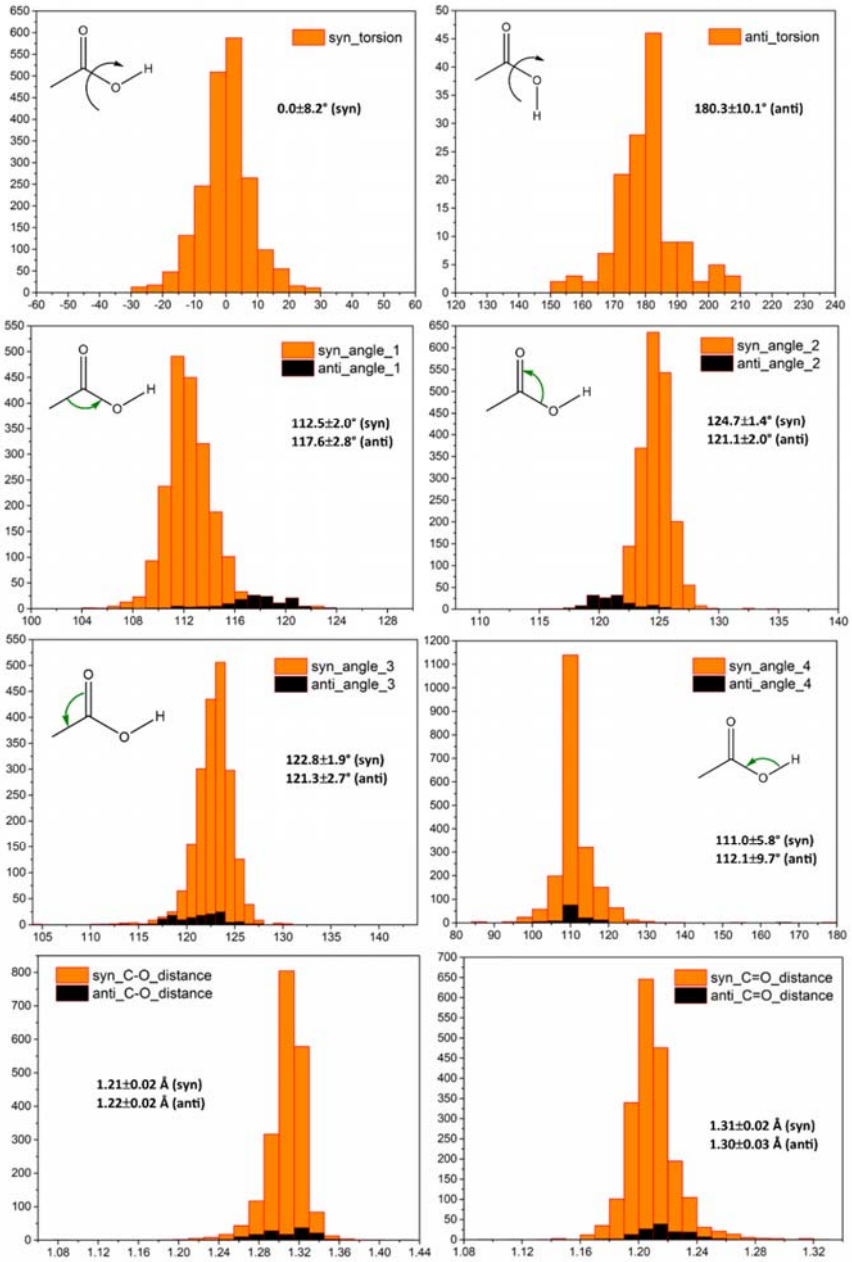


Figure 3. Histograms comparing bond lengths, bond angles and torsion angles in *syn* and *anti* carboxylic acids in peptides from a CSD analysis.

H-bonding pattern in the crystal structures

The molecular conformations of compounds **1**-**5** indicate that the insertion of additional atoms in the backbone of the polypeptide enhances the possibility of hydrogen-bonded folded structures, as observed in compound **2**, **3** and **5** (Figure 1A). The backbone and side chain torsion angles for all the compounds are listed in supplementary Table S1.

Compound **2**, Boc–Leu–γ^4(R)Val–Leu–γ^4(R)Val–OH, adopts a folded conformation stabilized by two intramolecular C_{12} hydrogen bonds, a nomenclature often used in proteins to distinguish between different local folding patterns of polypeptide backbones[57]: [Leu(1)] C=O···H–N [γ^4(R)Val (4)] and [Boc] C=O···H–N [Leu (3)]. It crystallizes as a solvate[58] with methanol and forms another intramolecular hydrogen bond, [methanol] O···H–N [γ^4(R)Val (2)]. In compound **3**, an intramolecular C_9 hydrogen bond is observed, [Boc] C=O···H–N [Val(2)]. In compound **5**,[59] two intramolecular C_{12} hydrogen bonds are formed: [Boc] C=O···H–N [Val (3)] and [Leu(1)] C=O···H–O [COOH]. The hydrogen bonding patterns in compounds **1**-**5** have been analyzed in terms of acceptor carbonyl groups in amide, acid and urethane regions in each structure, Figure 1.B depicts the same for *anti* 5. The geometrical details of the hydrogen bonds in **1**-**5** are given in Table S2. The hydrogen bonds are mostly not far from linearity, the ∠X–H···O (X=O/N) angle being generally > 150°. The angular distribution of proton donors around the acceptor, i.e., the ∠C=O···H angle reveals more about the lone pair directions since a nearly linear hydrogen bond is formed between a proton and a directed lone pair of electrons. Hence, the α and β angles of all $O_{carbonyl}$ acceptors in O–H···O or N–H···O bonds are listed in Table S2 as well. In the *syn* compounds, the acidic O–H bond is not involved in intramolecular hydrogen bonds, only intermolecular ones; however, in anti **5**, the acidic O–H bond forms an intramolecular H-bond which plays a crucial role in the resulting folded structure. As discussed below in more detail, the [COOH] O–H···O=C [Leu(1)] hydrogen bond is the strongest among all hydrogen bonds formed by all **5** compounds.

It is known that the $O_{carbonyl,amide}$ atom is a significantly stronger proton acceptor than the $O_{carbonyl,COOH}$ atom.[60] Hence, in crystal structures, sometimes doubly H-bonded systems O-H…O (amide) and N–H···O(amide) are observed, but O–H…O(carboxyl) and N–H…O(carboxyl) systems are generally less bifurcated. In the crystal structures **1**-**5**,

following the above trend, a few $O_{carbonyl,amide}$ atoms form bifurcated intermolecular H-bonds, but none of the $O_{carboxyl}$ atoms form bifurcated H-bonds. Also, the NPA charges clearly support the trend that O(carboxyl) has less negative charge compared to O(amide) (Table S3).

Lewis picture and NBO analysis

Natural bond orbital (NBO) methods, generally regarded as a 'chemist's basis set', have a strong connection to the traditional concept of resonance, hybridization and bonding in chemistry.[28,29] NBO analysis suggests higher polarity of the acidic O–H bond in *anti* carboxylic acids compared to the O–H bond in *syn* carboxylic acids through various descriptors. For example, natural population analysis (NPA) charges (Table 2) [and Bader atomic charges (Table S1)] on acidic H in *anti* **5** show slightly higher positive values compared to *syn* compounds and alcoholic H in methanol. The Wiberg bond index which roughly resembles the covalent contributions is lower by 0.1 for acidic O–H bond, in *anti* **5**, compared to the *syn* compounds. Similarly the Ionicity Parameter i_{OH} (Table 2) for the O–H bond is higher in *anti* **5**, compared to the *syn* acids. The solvent methanol O–H bond in **5** has a higher Wiberg bond index and a lower i_{OH} value, relative to all carboxylic acids indicating the less polar nature of alcoholic OH compared to acidic OH groups.

Table 2. NBO based descriptors for the model compounds.

	Syn **1**	*Syn* **2***	*Syn* **3**	*Syn* **4**	*Anti* **5**
		NPA charges			
$C_{carbonyl}$	0.82	0.82	0.83	0.83	0.82
$O_{carbonyl}$	-0.61	-0.62	-0.62	-0.61	-0.62
$O_{hydroxyl}$	-0.70	-0.69	-0.69	-0.70	-0.71
$H_{hydroxyl}$	0.49	0.49	0.50	0.49	0.51
		Wiberg bond indices			
C=O	1.74	1.72	1.73	1.75	1.72
C–O(H)	1.06	1.06	1.07	1.04	1.10
O–H	0.73	0.73	0.72	0.73	0.63
		Ionicity			
C=O, π-bond	0.41	0.41	0.41	0.41	0.40
C=O,α-bond	0.30	0.30	0.30	0.31	0.30
C–O(H)	0.37	0.37	0.38	0.37	0.36
O–H	0.51	0.50	0.51	0.50	0.56

* *Syn* **2** crystallizes as a solvate with methanol present in the asymmetric unit. For comparison with the acidic OH groups, the corresponding values the methanol OH group are: the NPA charges on C, O and $H_{hydroxyl}$ are -0.19, -0.75 and 0.46. The Wiberg bond indices of C–O and O–H are 0.92 and 0.88. The ionicity of C–O and O–H are 0.36 and 0.47.

Each $O_{hydroxyl}$ atom has two lone pairs according to the NBO analysis: one is unhybridized π-type, n_O^{π} (pure p_z), the other one is hybridized σ-type, n_O^{σ} . Considering the second order perturbative estimate of the donor-acceptor stabilization (E2 energies) as obtained from NBO analysis (Figure 4), the –COOH group in both *syn* and *anti* conformations is mostly stabilized by the π-type non-bonding orbital at the hydroxyl oxygen atom to the anti-bonding π-orbital of the C=O bond [$n_O(\pi, O_{hydroxyl}) \rightarrow \pi^*$C=O delocalization]. However, in the *syn* form, $n_O(\sigma, O_{hydroxyl}) \rightarrow \sigma^*$C=O delocalization is also feasible. In the *anti* form, such delocalization is almost negligible, but $\sigma_{OH} \rightarrow \sigma^*$C=O delocalization compensates hyperconjugation involving the *s*-rich lone pair to a large extent. However, E2 energies alone cannot explain the greater stability of *anti* forms over *syn* forms; dipolar repulsion between the lone pairs of $O_{hydroxyl}$ and C=O in the *anti* form contributes significantly to the destabilization of the *anti* form, as previously reported.[61]

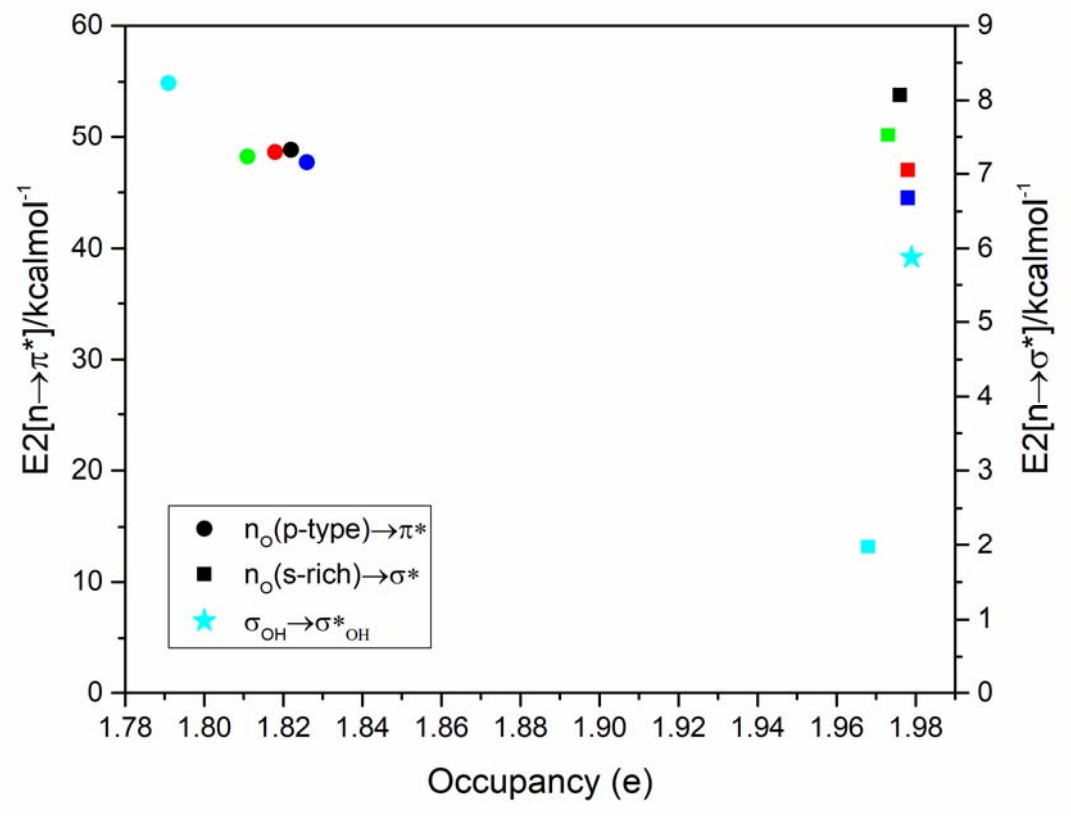


Figure 4. Second order perturbative estimate of the donor-acceptor stabilization, E2 energies of –COOH group in *syn* and *anti* conformation.

The localized Natural Resonance Theory (NRT) analysis,[62,63] of the –COOH moiety in *syn* and *anti* orientations in compounds **1**, **3** and **5** shows that mostly three Resonance Structures (RS) contribute with RS1 ~71%, RS2 ~25% and RS3 ~1% (Figure 5). Generally, carboxylic acids are less stable in *anti* conformation compared to *syn* because of the greater dipolar repulsion between the lone pairs of $O_{hydroxyl}$ and C=O in the *anti* form.[61] The RS2 resembles reduced dipolar repulsion due to the partial positive charge on $O_{hydroxyl}$. The higher relative percentage of RS2 in *anti* ~27%, compared to ~25% in *syn*, adds to the stability of the *anti* form to some extent. In RS3, since the carbonyl C atom is *sp* hybridized towards C=O, it uses more p character to form the C(=O)–O bond. The higher relative contribution of RS3 in *syn* compared to *anti* supports the shrinkage of the $\angle C_{\alpha}$–C(O)–$O_{hydroxyl}$ angle in *syn*, from a standard value of 120° corresponding to sp^2 hybridization. Likewise, the E2 stabilization energy corresponding to the orbital interaction

of the parent Lewis structure that gives rise to the secondary form RS3, namely, [$n_O(\pi, O_{carbonyl}) \rightarrow \sigma^* C–O_{hydroxyl}$ delocalization], is ~3.5 kcal mol^{-1} higher in *syn* (Table S4). This is in agreement with the findings summarized in Figure 3. Although the reason behind the redistribution of angles in the *syn* form is not yet fully understood in this study, localized natural resonance theory analyses provide some explanation for the observed deviations. However, since these hybrid peptides are comparatively large molecules, for computational purposes, we have not investigated the full hyperconjugative delocalization effect on the geometry. We are currently investigating this effect on small organic molecules.

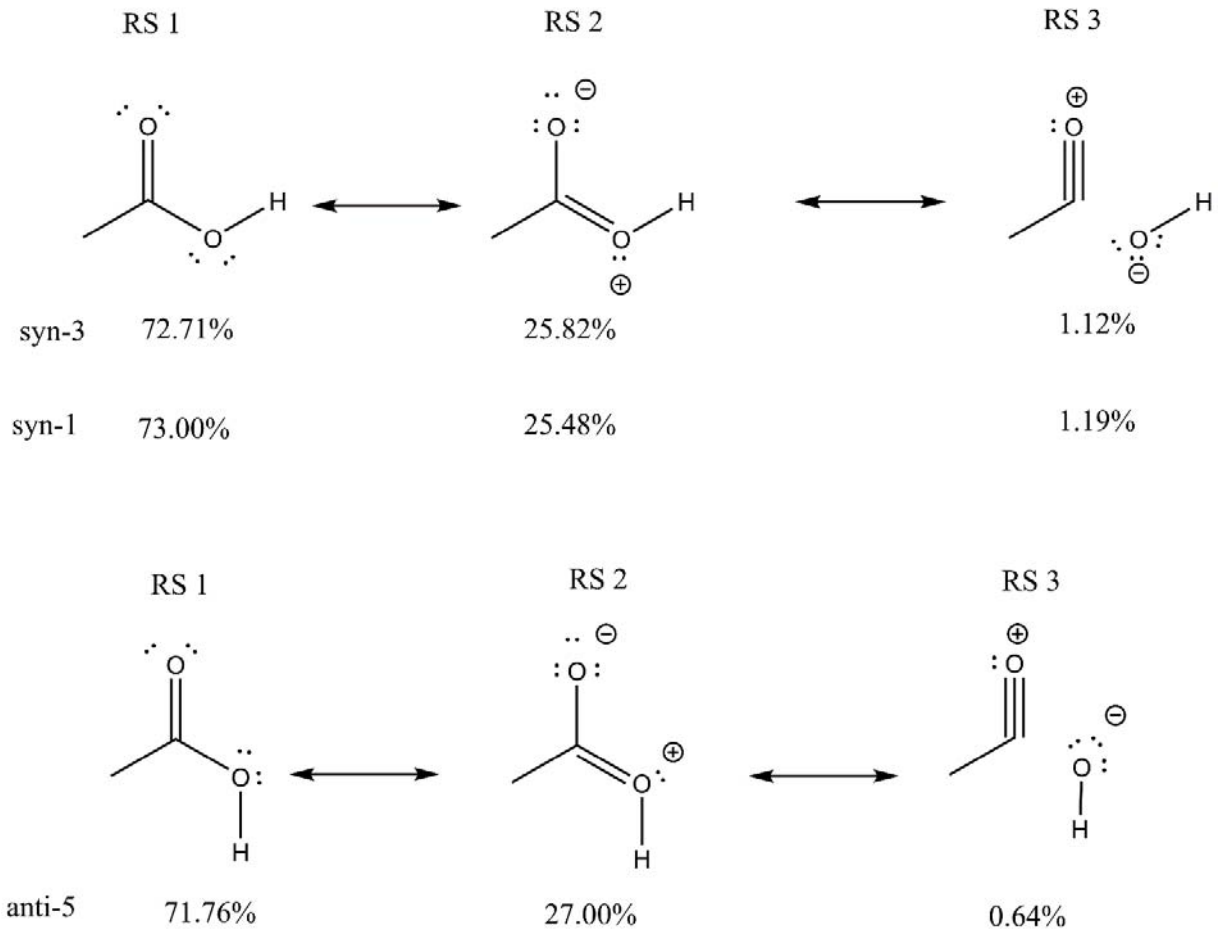


Figure 5. Localised NRT analysis on –COOH group in *syn* and *anti* conformations in compounds syn **1** and syn **3** and anti 5.

Oxygen Lone pairs and H-bond directionality

It is generally assumed that the electron lone pairs on the carbonyl O atom are sp^2 hybridized and arranged in a trigonal geometry. In an extensive CSD database analysis on short hydrogen bonds involving $O_{carbonyl}$ as acceptor, α and β angles were found to be spread around 60° and 0°, respectively.[24] In our model systems the distributions of α and β angles are quite widespread, 3.3°< α <59.8° with average value 32(16)° and 3.16°< β < 59.3°, with average value 26(16)° (Table S2). For the intermolecular O–H···O bond in syn **1**, α = 51.2° and β = 8.0°; but in case of the intramolecular O–H···O bond in anti **5**, α= 36.7° and β=31.3°. The d(O···H) distances in the two hydrogen bonds are 1.721 Å and 1.634 Å, respectively.

Natural bond orbital (NBO) analysis calculates the number of lone pairs, the percentage contribution of s and p orbitals in a particular lone pair and which lone pairs participate to the stabilization of a particular donor– acceptor interaction. In our model systems, NBO analysis shows that the hybridization of O lone pairs is no longer idealized sp^2 for carbonyl and sp^3 for hydroxyl groups. In fact, for both functional groups, the two lone pairs are not equivalent. Similar to the oxygen lone pairs in water molecule,[64] one is unhybridized π-type (mostly pure p_z), $n_O^{(\pi)}$ and the other one is hybridized σ-type, $n_O^{(\sigma)}$ (s-rich) (Figure 6).

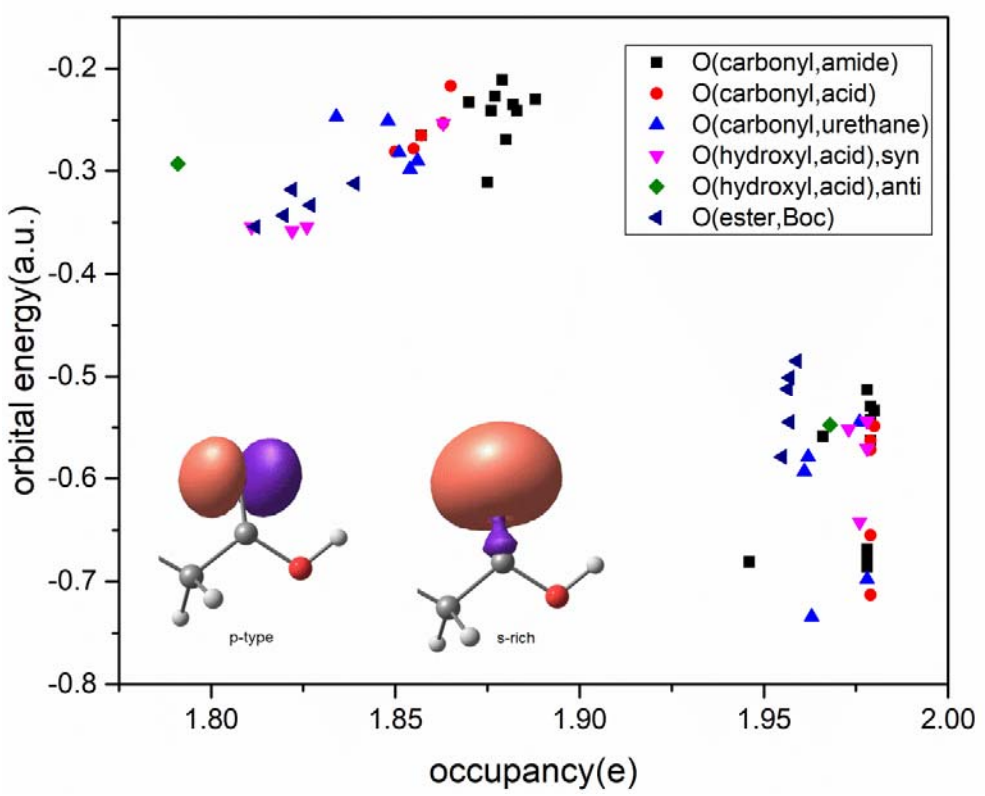


Figure 6. Scatter plot of occupancies and orbital energies of O lone pairs in the five compounds for $O_{carbonyl}$, $O_{hydroxyl}$, O_{ester} atoms. The shape of $n_O^{(\pi)}$ (p-type) and $n_O^{(\sigma)}$ s-rich lone pair lobes are shown at the bottom left corner in the graph. The $n_O^{(\sigma)}$ lone pairs have lower occupancy and weaker energy in magnitude (less negative, clustered at left upper corner in the figure) compared to $n_O^{(\pi)}$ lone pairs (right bottom corner).

The hybridizations of O atom lone pairs are tabulated in Table S2. In fact, a plot of occupancies vs. orbital energies (Figure 6) clearly indicates how different the two types of lone pairs are. The $n_O^{(\pi)}$ lone pairs appear at the right bottom corner in Figure 6, with higher occupancy and stronger energy in magnitude (more negative values), i.e. they are more stabilized, compared to the $n_O^{(\sigma)}$ lone pairs which are clustered at the left upper corner in Figure 6. The $n_O^{(\sigma)}$ lone pairs have a larger span with respect to the x axis, i.e., they have a varying occupancy compared to p-type lone pairs. Moreover, the $n_O^{(\sigma)}$ lone pairs associated to the various $O_{carbonyl}$ types appear at the higher occupancy site compared to $O_{hydroxyl,acid}$ and $O_{ester,Boc}$. Apparently, the $n_O^{(\sigma)}$ lone pair of the $O_{hydroxyl,\ acid}$ in

anti **5** has the lowest occupancy and weakest energy in magnitude (least negative value) compared to all the other lone pairs (green diamond in Figure 6). However, there is no separation for $n_O^{(\pi)}$ lone pairs between $O_{carbonyl}$,$O_{hydroxyl}$ or O_{ester}. In a strong H-bond, the proton tends to be oriented towards the oxygen-lone pair. Hence, the distribution of α, β angles with a wide range, not coinciding with idealized sp^2 direction, can be explained with the different extent of s-contributions to the lone pairs.

A closer look at NBO results indicates one striking difference between the intra- and intermolecular O–H···O bonds in *anti* **5**, and compared to *syn* **1**. In the intramolecular O4–H4···O1 hydrogen bond in *anti* **5**, both the lone pairs, with $sp^{0.76}$ [$n_O^{(\sigma)}$] and $sp^{25.8}$ [$n_O^{(\pi)}$] hybridizations, on O1 (amide) participate to an almost equal extent as proton acceptor. The E2 energy values for $n_O \rightarrow \sigma^*_{OH}$ are 14.74 and 12.64 kcal mol^{-1}, respectively (Table 3), with even slightly greater stabilization involving more directional $n_O^{(\sigma)}$ lone pairs. Whereas in *syn* **1**, the calculation on a dimeric motif in which one molecule is linked via the intermolecular O4–H4···O2 hydrogen bond reveals that between the two lone pairs on acceptor O2 (amide), the $n_O^{(\pi)}$ lone pair ($sp^{23.43}$ hybridization) contributes 16.65 kcal mol^{-1} and the $n_O^{(\sigma)}$ ($sp^{0.74}$ hybridization) contributes only 5.87 kcal mol^{-1}, towards $n_O \rightarrow \sigma^*_{OH}$. Also it is noteworthy that in the dimeric unit, the lone pair hybridizations on O2 (amide) in the two molecules are not identical. In the free molecule where O2 is not interacting, the hybridization for $n_O^{(\sigma)}$ is $sp^{0.63}$ and pure p_z for $n_O^{(\pi)}$, while in the other molecule, which acts as proton acceptor in the intermolecular O4–H4···O2 interaction, the hybridizations are $sp^{0.74}$ and $sp^{23.43}$, respectively for $n_O^{(\sigma)}$ and $n_O^{(\pi)}$ lone pairs (Table 3). For the same dimer of *syn* 1, the lone pair hybridization of the non-interacting O3 atom is $sp^{0.69}$ and pure p_z for $n_O^{(\sigma)}$ and $n_O^{(\pi)}$ lone pairs, whereas in the other molecule upon N2–H2···O3 hydrogen bond formation it changes to $sp^{0.71}$ and pure p_z respectively(Table 2) with corresponding E2 values for $n_O \rightarrow \sigma^*_{NH}$ of 4.17 and 2.21 kcal mol^{-1}. These small but noticeable differences in s/p mixing for the lone pairs which are directly involved as proton acceptors suggest that the p character gets redistributed to increase the directionality while participating in hydrogen bond formation. The analysis so far strongly suggests that there must be a relationship between H-bond strength, directionality and s/p hybridization of the lone pairs.

Table 3. The acceptor oxygen lone pair hybridization and E2 energies [$n_O \rightarrow \sigma^*_{OH}$] involved in intra vs. intermolecular O–H···O / intramolecular O–H···O vs. N–H···O bonds, according to the NBO perspective.

compound	H bond	Lone pair type (n_o)	E_2 (kcal mol^{-1})
Syn **1**	Intermolecular	$sp^{0.74}$	5.87
	O4–H4···O2	$sp^{23.43}$	16.65
Anti **5**	Intramolecular	$sp^{0.76}$	14.74
	O4–H4···O1	$sp^{25.8}$	12.64
	Intramolecular	$sp^{0.63}$	7.86
	N3–H3···O0	p	1.75
Syn **2**	intramolecular	$sp^{0.61}$	8.79
	N3–H3···O0	p	2.17
	intramolecular	$sp^{0.61}$	5.29
	N4–H4···O1	p	2.61
	intramolecular	$sp^{1.97}$	5.15
	N2–H2···O6s*	$sp^{5.49}$	5.43

*In N2–H2···O6s in *syn* **2**, the acceptor O is methanol O atom, whereas in all other H-bonds the acceptor O is $O_{carbonyl}$.

Different approaches to address one chemical question complement each other and provide view points from various perspectives. The theory of Atoms in Molecules (AIM) pioneered by Bader and coworkers is based on the scalar field of molecular electron density, to extract chemical insight for bonding, molecular structure and the concept of an atom in a molecule.[56] The experimental multipole model of anti **5** ('ExpMul'), has been compared with a theoretical multipole model (henceforth termed 'TheoMul') of the same compound. Among the four *syn* compounds, we have derived TheoMul models for two compounds, **1** and **2**, and mostly the results from compound **1** have been compared with the ExpMul and TheoMul models of the *anti* form. Static deformation density and two-dimensional Laplacian maps of the carboxylic acid region are shown in Figure 7. At this scale, the maps appear quite similar. A major difference between the theoretical and experimental models can be found at the carbonyl oxygen lone pairs. It is not clear if this is an experimental feature or an artefact of the model.

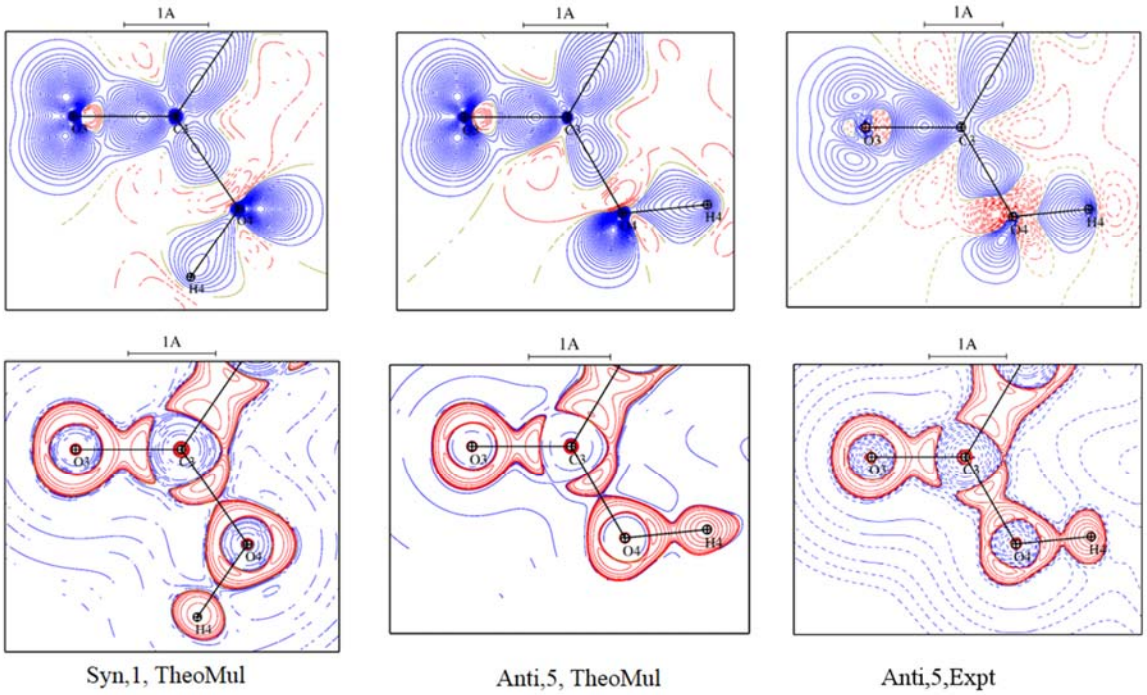


Figure 7. Deformation density and Laplacian plots of –COOH are shown in first and second rows, respectively, for TheoMul model for syn 1 and both theoretical and experimental multipole models for anti 5. Contours are drawn at the intervals of ±0.05 eÅ^{-3} for deformation maps. Laplacian is drawn in logarithmic contours. The colour code is: for deformation maps, blue solid lines indicate positive, red dashed line negative and green dashed lines zero contour; for Laplacian maps, red solid lines indicate negative and blue dashed lines indicate positive contour.

Integrated Bader charges after multipole modelling (Table S3) do not provide a clear trend either for O atoms in different chemical environments. For example, it does not show less negative charges on $O_{carboxyl}$ atoms compared to O_{amide} atoms as speculated from the H-bonding pattern and supported by NPA charges Table S3. The topological parameters of the covalent bonds corresponding to the –COOH moiety in compounds **1**, **2** (TheoMul) and **5** (ExpMul and TheoMul) are listed in Table 4.

Table 4. Topological properties at the bond critical point (BCP) of carboxylic acid groups in compounds **1**, **2** and **5**. d_1 and d_2 are the distances from the BCP to the first atom (eg, O3 in O3–C3) and second atom (eg, C3 in O3–C3), respectively. The interaction length, $R_{ij}=d_1+d_2$.

	Bond	ρ(r) (eÅ^{-3})	$\nabla^2\rho$(r) (eÅ^{-5})	R_{ij} (Å)	d_1 (Å)	d_2 (Å)	ε
Syn 1							
	O3–C3	3.04	-37.7	1.2058	0.7574	0.4484	0.14
	O4–C3	2.23	-19.5	1.3260	0.7810	0.5450	0.13
	O4–H4	1.97	-16.9	1.0180	0.7608	0.2572	0.02
Syn 2							
	O4–C4	2.98	-37.8	1.2100	0.7560	0.4540	0.14
	O5–C4	2.31	-20.9	1.3180	0.7710	0.5470	0.13
	O5–H5	1.95	-16.2	1.0180	0.7652	0.2528	0.02
	O6S–H6S	2.22	-21.3	0.9700	0.7288	0.2412	0.03
Anti 5							
Expt	O3–C3	3.04(2)	-26.6(3)	1.2176	0.8047	0.4130	0.14
Theory		*2.90*	*-35.7*	*1.2196*	*0.7517*	*0.4678*	*0.18*
Expt	O4–C3	2.30(1)	-25.3(2)	1.3160	0.8637	0.4524	0.07
Theory		*2.22*	*-20.2*	*1.3169*	*0.7854*	*0.5315*	*0.16*
Expt	O4–H4	2.21(6)	-24.4(5)	0.9791	0.7419	0.2372	0.04
Theory		2.09	-27.9	*1.0180*	*0.7651*	*0.2529*	*0.02*

The electron density [ρ(r)] at the bond critical point (bcp) of the acidic O–H bond in the –COOH group is higher in the *anti* form compared to *syn* form, by ~0.1 eÅ^{-3} and the Laplacian values [$\nabla^2\rho$(r)] is more negative by ~11 eÅ^{-5}. These differences are outside the experimental standard uncertainties and outside the limit of statistical errors from multiple experiments and reproducibility indices as reported in literature.[65,66] Therefore, these differences are statistically significant. The solvent of crystallization, methanol, present in the crystal structure of *syn* **2** allows us to compare the alcoholic O–H *vs.* acidic O–H bond. The $\rho(r_{bcp})$ value of the acidic O–H bond in the *anti* form appears to be lower than that of the alcoholic O–H bond in methanol by ~ 0.1 eÅ^{-3}; but the Laplacian value is more negative by 6.6 eÅ^{-5}.

The topological parameters of intra and intermolecular hydrogen bonds in compounds **1**, **2** and **5** are listed in Table S5, including hydrogen bonding energies roughly estimated using the Espinosa−Molins−Lecomte (EML) method, solely based on topological

parameters at bond critical points.[67-69] It establishes that the intramolecular hydrogen bond [COOH] O–H···O=C [Leu(1)] in anti **5** is the strongest among all hydrogen bonds, contributing ~-34.6 kcal mol^{-1} towards the stability of the crystal structure of **5**, compared to -12.3 and -14.4 kcal mol^{-1} for the intermolecular O–H···O bond in compounds **1** and **2**, respectively. It agrees with the earlier reports that the preference of the *syn* form over the *anti* form decreases significantly in the solid state due to strong H-bond formation in the crystal.

The deformation density maps of lone pairs on $O_{carbonyl}$ in urethane, amide and carboxylic acid moieties as obtained from the TheoMul models are shown in Figure 8 for *syn* **1** and *anti* **5**. Similar plots for $O_{hydroxyl}$ in –COOH in different orientations are shown in Figure 9 and O_{ester} in Boc is shown in Figure S2. In all different chemical environments $O_{carbonyl}$ appears as two separate lone pairs maxima, whereas $O_{hydroxyl}$ appears as merged, similarly to the results from previous studies.[70]

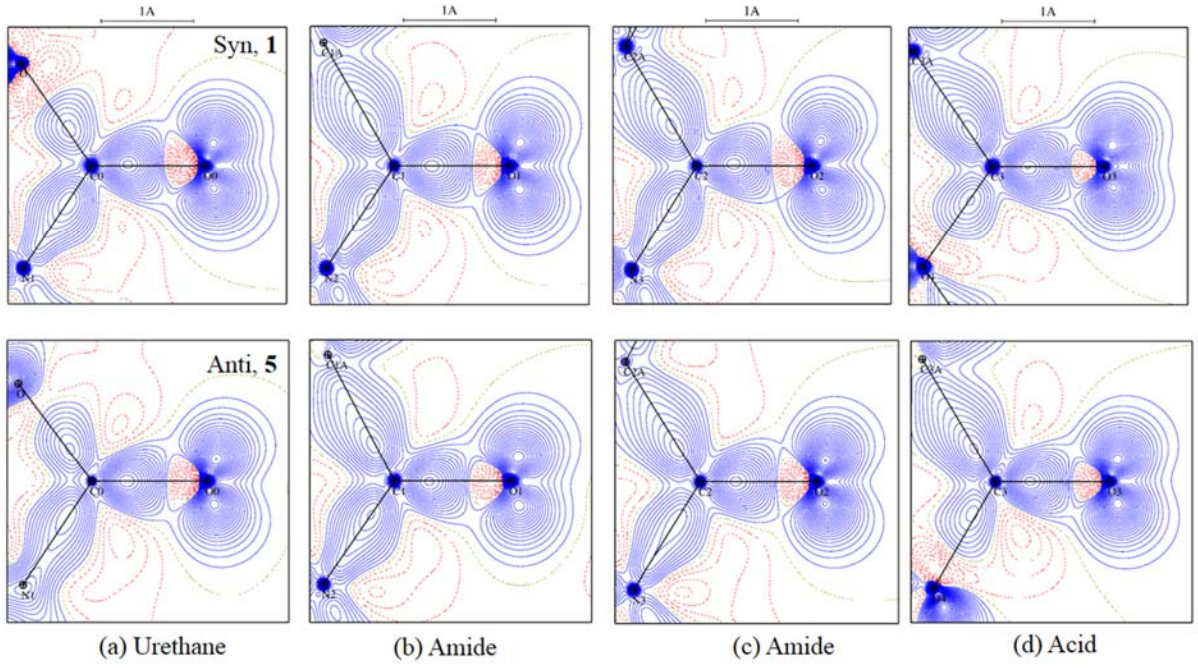


Figure 8. Lone pair lobes of $O_{carbonyl}$ in various chemical environments as visualized from deformation density maps for *syn* **1** and *anti* **5**, corresponding to the respective TheoMul models. Contours are drawn at the intervals of ±0.05 eÅ^{-3} for deformation maps. The colour code is: blue solid lines indicate positive, red dashed line negative and green dashed lines zero contour.

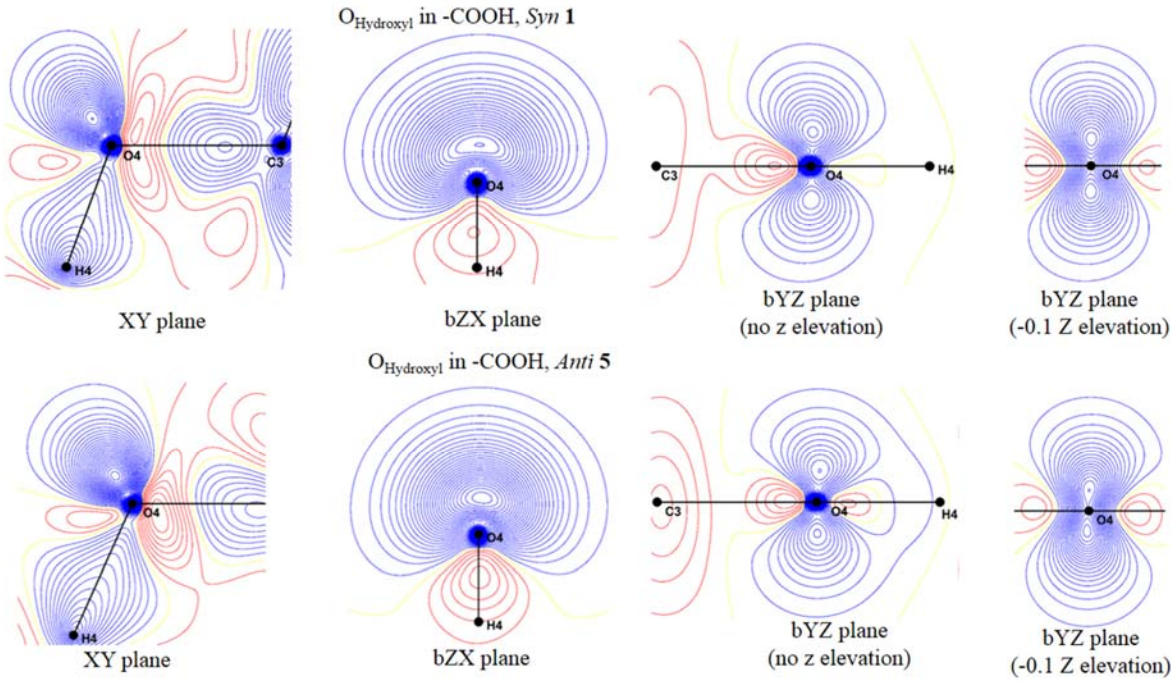


Figure 9. Lone pair lobes on $O_{hydroxyl}$ in the –COOH group as visualized from deformation density maps in different reference planes for *syn* **1** (upper row) and *anti* **5** (lower row) compounds, as obtained from TheoMul models. Definition of the planes: XY contains the C-O-C moiety; bZX contains the two LPs; bYZ: perpendicular to the O–LP1 and O–LP2 inner bisecting direction. Contours are drawn at the intervals of ±0.05 eÅ^{-3} for deformation maps. The colour code is: blue indicate positive, red negative and yellow zero contour.

The lone pairs sites were located by means of (3,-3) critical points of the Laplacians. Both gas phase single-molecule and crystalline-state calculations show two widely separated lone pairs, much wider than 120° of a trigonal geometry for the $O_{carbonyl}$ belonging to urethane, amide and acids. The LP–O–LP' angle and other relevant angles corresponding to the orientation of the lone pairs have been compiled in supplementary Table S6-S8 and Figure S3 and S4 for compounds *syn* **1** and *anti* **5**, respectively. In general, the lone pair lobes of the carbonyl oxygen atoms are closer to each other by ~5-10° in the crystal phase compared to the isolated molecule. Two distinct lone pair lobes on O_{ester} and $O_{hydroxyl}$ oxygen atoms, could not be identified in all cases in the electron density deformation maps. But whenever two separate lone pair lobes could be identified,

the LP–O–LP' angle was close to 109°, similar to a tetrahedral geometry. Although the angular orientation of the lone pairs of $O_{hydroxyl}$ with respect to the $C_{carbonyl}$–$O_{hydroxyl}$ bond is similar in both *syn* and *anti* forms, the value of electron density and the Laplacian at the lone pair position is larger in *anti* compared to *syn*, the difference being $\Delta\rho(r)$ = 0.2 eÅ^{-3}, $\Delta[\nabla^2\rho(r)]$ = 8.9 eÅ^{-5} (Table S8). Similarly for different carbonyl groups, although the orientation of the lone pairs are quite similar in all the compounds, the associated Laplacian values (Table S8) show noticeable differences and reflect the asymmetry in the hydrogen bond features around the carbonyl oxygen atoms.

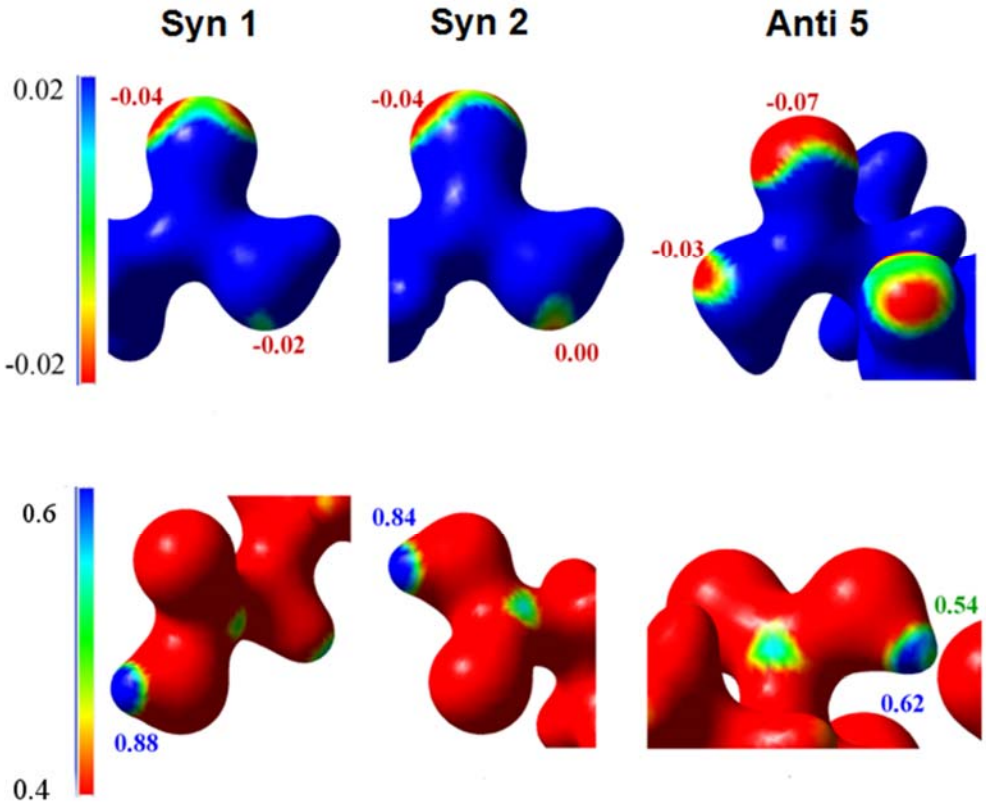


Figure 10. Electrostatic potential mapped on 0.074 a.u. isodensity surfaces of the –COOH group for compounds *syn* **1**, syn **2** and *anti* **5**, according to TheoMul models. The second row highlights the acidic H by imposing a different colour scale and choosing a different orientation. The blue, red and green regions represent electropositive, electronegative and neutral regions.

The electrostatic potential mapped on isodensity surfaces (0.074 a.u.) of the carboxyl group has been analyzed (Figure 10). Electropositive, electronegative and neutral regions are indicated by blue, red and green surfaces, respectively. The $O_{carbonyl}$ and $O_{hydroxyl}$ atoms of the carboxyl group appear to be more electronegative in *anti* 5, compared to the *syn* 1 (Figure 10, first row). The same electrostatic potential maps with a different scale to highlight the acidic hydrogen atom (second row in Figure 10) indicate that the acidic hydrogen in *anti* **5** is more polarized compared to *syn* **1**. This is in agreement with the NBO analyses as well.

Conclusions

A comparative study on *syn* and *anti* carboxylic acids has been carried out with four *syn* and one *anti* hybrid peptides. The molecular conformations in the crystal structures obtained in this study as well as a thorough CSD analysis establish the difference in geometry of the carboxyl groups in *syn* and *anti* forms, especially the rearrangement of the bond angles surrounding the carbonyl C atom in the *syn* –COOH group. In the *anti* conformation, all the angles, namely, $\angle C_{\alpha}$–C(O)–$O_{carbonyl}$, $\angle C_{\alpha}$–C(O)–$O_{hydroxyl}$, $\angle O_{carbonyl}$–C(O)–$O_{hydroxyl}$ are ~120°, as per sp^2 hybridization of $C_{carbonyl}$, whereas in the *syn* form $\angle O_{carbonyl}$–C(O)–$O_{hydroxyl}$ is significantly smaller by 5–10° and as a result the remaining angles are greater than 120°.

The hybrid peptides **1**-**5** show various hydrogen bonding patterns in the crystalline state. These interactions have been characterized quantitatively in terms of experimental and theoretical charge density analysis. The topological parameters indicate that the intramolecular [acid] O–H···O=C hydrogen bond in *anti* **5** is the strongest among all the other intra and intermolecular hydrogen bonds present in the crystal structures. Thus it supports that the possibility of H-bond formation stabilizes the *anti* conformation in the crystalline state and thus the higher preference of *syn* over *anti* carboxylic acids decreases in the crystal state compared to the gas phase.

The deformation density maps from charge density multipole models and NBO analyses reveal complementary viewpoints on oxygen lone pairs and O–H bonds in *syn* and *anti* –

COOH. In deformation density maps $O_{carbonyl}$ in various chemical environments, urethane, amide, carboxylic acid, bears two distinct equivalent lone pairs, whereas O_{ester} or $O_{hydroxyl}$ atoms mostly show merged lone pair lobes. In contrast, the NBO analysis finds always two lone pairs on all types of O atoms, $O_{carbonyl}$, O_{ester}, $O_{hydroxyl}$, but they are not equivalent. One of them appears as unhybridized π-type, $n_O^{(\pi)}$, mostly of pure p_z-type, and the other one resembles hybridized σ-type, $n_O^{(\sigma)}$,s-rich, (similar to the oxygen lone pairs in water molecule), as opposed to the general notion of two equivalent sp^2 lone pairs for $O_{carbonyl}$ and two sp^3 lone pairs for O_{ester} or $O_{hydroxyl}$.

Although the relative configurations of the two lone pairs of the $O_{hydroxyl}$ oxygen atom are quite similar in *syn* and *anti* forms, the electron density and Laplacian values at the lone pairs positions indicate higher accumulation of electron density in the *anti* compared to the *syn* form. Also, a plot of occupancy *vs.* orbital energy indicates separate regions of σ-type [$n_O^{(\sigma)}$] lone pairs for $O_{carbonyl}$ compared to $O_{hydroxyl}$ and O_{ester}. The higher the occupancy of the lone pair orbital, the more stable it is.

A closer look at the –COOH group in electrostatic potential maps indicates a greater polarization of the acidic O–H bond in *anti* carboxylic acids compared to *syn*. NPA charges, Wiberg bond indices and ionicity parameters from NBO analysis also establish the higher polarity of the O–H bond in *anti* carboxylic acids. However, in this study, the O–H distances in both *syn* and *anti* carboxylic acid crystal structures have been fixed to the same averaged distance as obtained from neutron-diffraction experiments reported in the literature. The literature lacks separate entries for O–H bond lengths in *syn* and *anti* carboxylic acids. The usage of identical bond lengths certainly has a bias on any comparative studies on *syn* and *anti* carboxylic acids. Hence, in order to obtain an accurate comparison, separate neutron data on both *syn* and *anti* systems would be preferred. Therefore, we plan neutron-diffraction studies of simple and small model systems bearing *syn* and *anti* carboxylic acid groups in the near future.

Supporting information

Table S1-S8 and Figures S1-S5; Table S1 lists backbone and side chain torsion angles of the five peptides. Table S2 lists lone pair properties of different oxygen acceptors.

Table S3 contains atomic charge and volume, Table S4 shows E2 stabilization energies from NBO calculation, Table S5 topological parameters of intra and intermolecular H-bonds. Table S6-S8 lists lone pair orientation of different oxygen atoms in *Syn* 1, *Syn* 2 and *Anti* 5, respectively. Table S9 contains E2 energies related to –COOH group in the five peptides. Figure S1 depicts fractal dimension plots of residual density from experiment multipole model (second model, with all reflections) of compound Anti 5. Figure S2 shows lone pair lobes on O(ether) in the Boc group from deformation density maps in Syn 1 and Anti 5. Figure S3-S4 contain lone pair orientations of different oxygen atoms in Syn 1 and Anti 5 and Figure S5 shows mass spectra of the five peptides.

Acknowledgements

RP thanks P. Balaram for suggesting the problem, for providing the peptide samples and discussions in the early stages of this project. S. Grabowsky thanks the German Research Foundation (DFG) for funding within the Emmy Noether grant GR4451/1-1.

References

(1) A. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. Macmillan: 1999.
(2) K.D. Cramer; S.C. Zimmerman, Kinetic effect of a syn-oriented carboxylate on a proximate imidazole in catalysis: a model for the histidine-aspartate couple in enzymes. *J. Am. Chem. Soc.* **1990,** *112*, 3680-3682.
(3) P. Ädelroth, Special issue on proton transfer in biological systems. *BBA - Bioenergetics* **2006,** *1757*, 867-870.
(4) S.C. Zimmerman; K.D. Cramer, Stereoelectronic effects at carboxylate: A syn-oriented model for the histidine-aspartate couple in enzymes. *J. Am. Chem. Soc.* **1988,** *110*, 5906-5908.
(5) R.D. Gandour, On the importance of orientation in general base catalysis by carboxylate. *Bioorg. Chem.* **1981,** *10*, 169-176.
(6) T.A. Montzka; S. Swaminathan; R.A. Firestone, Reversal of Syn-Anti Preference for Carboxylic Acids along the Reaction Coordinate for Proton Transfer. Implications for Intramolecular Catalysis. *J. Phys. Chem.* **1994,** *98*, 13171-13176.
(7) J. Pranata, Relative basicities of carboxylate lone pairs in aqueous solution. *J. Comput. Chem.* **1993,** *14*, 685-690.

(8) J. Rebek, Molecular recognition: model studies with convergent functional groups. *J. Mol. Recognit.* **1988,** *1*, 1-8.
(9) J.B. Huff; B. Askew; R.J. Duff; J. Rebek, Stereoelectronic effects and the active site of the serine proteases. *J. Am. Chem. Soc.* **1988,** *110*, 5908-5909.
(10) Y. Li; K.N. Houk, Theoretical assessments of the basicity and nucleophilicity of carboxylate syn and anti lone pairs. *J. Am. Chem. Soc.* **1989,** *111*, 4505-4507.
(11) C.H. Gorbitz; M.C. Etter, Hydrogen bonds to carboxylate groups. Syn/anti distributions and steric effects. *J. Am. Chem. Soc.* **1992,** *114*, 627-631.
(12) J.E.J. Mills; P.M. Dean, Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput. Aided Mol. Des.* **1996,** *10*, 607-622.
(13) G.R. Desiraju; J.J. Vittal; A. Ramanan, *Crystal engineering: a textbook*. World Scientific: 2011.
(14) K.B. Wiberg; K.E. Laidig, Barriers to rotation adjacent to double bonds. 3. The carbon-oxygen barrier in formic acid, methyl formate, acetic acid, and methyl acetate. The origin of ester and amide resonance. *J. Am. Chem. Soc.* **1987,** *109*, 5935-5943.
(15) L. Leiserowitz, Molecular packing modes. Carboxylic acids. *Acta Cryst. B* **1976,** *32*, 775-802.
(16) D. Das; G.R. Desiraju, Packing Modes in Some Mono- and Disubstituted Phenylpropiolic Acids: Repeated Occurrence of the Rare syn,anti Catemer. *Chem. Asian J.* **2006,** *1*, 231-244.
(17) L. D'Ascenzo; P. Auffinger, A comprehensive classification and nomenclature of carboxyl-carboxyl(ate) supramolecular motifs and related catemers: implications for biomolecular systems. *Acta Cryst. B* **2015,** *71*, 164-175.
(18) M.G. Medvedev; I.S. Bushmarinov; K.A. Lyssenko, Z-effect reversal in carboxylic acid associates. *Chem. Commun.* **2016,** *52*, 6593-6596.
(19) F.H. Allen, The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Cryst. B* **2002,** *58*, 380-388.
(20) P.G. Vasudev; K. Ananda; S. Chatterjee; S. Aravinda; N. Shamala; P. Balaram, Hybrid peptide design. Hydrogen bonded conformations in peptides containing the stereochemically constrained γ-amino acid residue, gabapentin. *J. Am. Chem. Soc.* **2007,** *129*, 4039-4048.
(21) K. Ananda; P.G. Vasudev; A. Sengupta; K.M. Poopathi Raja; N. Shamala; P. Balaram, Polypeptide helices in hybrid peptide sequences. *J. Am. Chem. Soc.* **2005,** *127*, 16668-16674.
(22) R. Taylor; O. Kennard, Hydrogen-bond geometry in organic crystals. *Acc. Chem. Res.* **1984,** *17*, 320-326.
(23) P. Murray-Rust; J.P. Glusker, Directional hydrogen bonding to sp2- and sp3-hybridized oxygen atoms and its relevance to ligand-macromolecule interactions. *J. Am. Chem. Soc.* **1984,** *106*, 1018-1025.
(24) M. Ahmed; C. Jelsch; B. Guillot; C. Lecomte; S. Domagała, Relationship between Stereochemistry and Charge Density in Hydrogen Bonds with Oxygen Acceptors. *Cryst. Growth Des.* **2013,** *13*, 315-325.
(25) R. Taylor; O. Kennard; W. Versichel, Geometry of the imino-carbonyl (N-H...O:C) hydrogen bond. 1. Lone-pair directionality. *J. Am. Chem. Soc.* **1983,** *105*, 5761-5766.
(26) C. Jelsch; B. Guillot; A. Lagoutte; C. Lecomte, Advances in protein and small-molecule charge-density refinement methods using MoPro. *J. Appl. Crystallogr.* **2005,** *38*, 38-54.
(27) B. Guillot In *MoProViewer: A molecule viewer for the MoPro charge-density analysis program*, 27th European Crystallographic Meeting, ECM 27, Bergen, *Acta Cryst. A,* Bergen, 2012, s204.
(28) A.E. Reed; L.A. Curtiss; F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **1988,** *88*, 899-926.
(29) A.E. Reed; R.B. Weinstock; F. Weinhold, Natural population analysis. *J. Chem. Phys.* **1985,** *83*, 735-746.

(30) F.H. Allen; I.J. Bruno, Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data. *Acta Cryst. B* **2010,** *66*, 380-386.
(31) A.A. Hoser; P.M. Dominiak; K. Woźniak, Towards the best model for H atoms in experimental charge-density refinement. *Acta Cryst. A* **2009,** *65*, 300-311.
(32) F.H. Allen; A.J. Kirby, Stereoelectronic effects at carboxylate oxygen: similar basicity of the E and Z lone pairs in solution. *J. Am. Chem. Soc.* **1991,** *113*, 8829-8831.
(33) O.D. CrysAlisPRO, Agilent Technologies UK Ltd: Yarnton. England: 2012.
(34) R. Blessing, Outlier Treatment in Data Merging. *J. Appl. Crystallogr.* **1997,** *30*, 421-426.
(35) G. Sheldrick, A short history of SHELX. *Acta Cryst. A* **2008,** *64*, 112-122.
(36) L. Farrugia, WinGX and ORTEP for Windows: an update. *J. Appl. Crystallogr.* **2012,** *45*, 849-854.
(37) J.L. Atwood; L.J. Barbour, Molecular graphics: from science to art. *Cryst. Growth Des.* **2003,** *3*, 3-8.
(38) N.K. Hansen; P. Coppens, Testing aspherical atom refinements on small-molecule data sets. *Acta Cryst. A* **1978,** *34*, 909-921.
(39) B. Guillot; L. Viry; R. Guillot; C. Lecomte; C. Jelsch, Refinement of proteins at subatomic resolution with MOPRO. *J. Appl. Crystallogr.* **2001,** *34*, 214-223.
(40) A. Volkov; P. Macchi; L. Farrugia; C. Gatti; P. Mallinson; T. Richter; T. Koritsanszky *XD2006–a computer program package for multipole refinement, topological analysis of charge densities and evaluation of intermolecular energies from experimental and theoretical structure factors*, 2006.
(41) Z. Su; P. Coppens, Nonlinear Least-Squares Fitting of Numerical Relativistic Atomic Wave Functions by a Linear Combination of Slater-Type Functions for Atoms with Z = 1-36. *Acta Cryst. A* **1998,** *54*, 646-652.
(42) E. Clementi; C. Roetti, Roothaan-Hartree-Fock atomic wavefunctions: Basis functions and their coefficients for ground and certain excited states of neutral and ionized atoms, Z≤ 54. *At. Data Nucl. Data Tables* **1974,** *14*, 177-478.
(43) F. Allen, A systematic pairwise comparison of geometric parameters obtained by X-ray and neutron diffraction. *Acta Cryst. B* **1986,** *42*, 515-522.
(44) A. Madsen, SHADE web server for estimation of hydrogen anisotropic displacement parameters. *J. Appl. Crystallogr.* **2006,** *39*, 757-758.
(45) P. Munshi; A.O. Madsen; M.A. Spackman; S. Larsen; R. Destro, Estimated H-atom anisotropic displacement parameters: a comparison between different methods and with neutron diffraction results. *Acta Cryst. A* **2008,** *64*, 465-475.
(46) M. Frisch; G. Trucks; H. Schlegel; G. Scuseria; M. Robb; J. Cheeseman; G. Scalmani; V. Barone; B. Mennucci; G. Petersson *01; Gaussian, Inc*, 2009.
(47) T.A. Keith, AIMAll (Version 13.05. 06). *TK Gristmill Software, Overland Park KS, USA* **2013**.
(48) R. Dovesi; V. Saunders; C. Roetti; R. Orlando; C. Zicovich-Wilson; F. Pascale; B. Civalleri; K. Doll; N. Harrison; I. Bush, *CRYSTAL14 user's manual*. 2014.
(49) C. Lee; W. Yang; R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988,** *37*, 785-789.
(50) A.D. Becke, Density‐functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.
(51) C. Gatti; V.R. Saunders; C. Roetti, Crystal field effects on the topological properties of the electron density in molecular crystals: The case of urea. *J. Chem. Phys.* **1994,** *101*, 10686-10696.
(52) P. Rabiller; M. Souhassou; C. Katan; C. Gatti; C. Lecomte, Accuracy of topological analysis of gridded electron densities. *J. Phys. Chem. Solids* **2004,** *65*, 1951-1955.
(53) A. Volkov; Y. Abramov; P. Coppens; C. Gatti, On the origin of topological differences between experimental and theoretical crystal charge densities. *Acta Cryst. A* **2000,** *56*, 332-339.

(54) P. Coppens; A. Volkov, The interplay between experiment and theory in charge-density analysis. *Acta Cryst. A* **2004,** *60*, 357-364.
(55) L. Bertini; F. Cargnoni; C. Gatti, Chemical insight into electron density and wave functions: software developments and applications to crystals, molecular complexes and materials science. *Theor. Chem. Acc.* **2007,** *117*, 847-884.
(56) R.F. Bader, *Atoms in molecules*. Wiley Online Library: 1990.
(57) P.G. Vasudev; S. Chatterjee; N. Shamala; P. Balaram, Structural chemistry of peptides containing backbone expanded amino acid residues: conformational features of β, γ, and hybrid peptides. *Chem. Rev.* **2010,** *111*, 657-687.
(58) E. Grothe; H. Meekes; E. Vlieg; J. Ter Horst; R. De Gelder, Solvates, salts, and cocrystals: a proposal for a feasible classification system. *Cryst. Growth Des.* **2016,** *16*, 3237-3243.
(59) K. Basuroy; B. Dinesh; N. Shamala; P. Balaram, Promotion of Folding in Hybrid Peptides through Unconstrained γ Residues: Structural Characterization of Helices in (αγγ)n and (αγα)n Sequences. *Angew. Chem.* **2013,** *125*, 3218-3221.
(60) Z. Berkovitch-Yellin; S. Ariel; L. Leiserowitz, The comparative roles of the proton-acceptor properties of amide and carboxyl groups in influencing crystal packing patterns: doubly vs. singly hydrogen-bonded systems in N-acylamino acids and in other amide-acid crystals. *J. Am. Chem. Soc.* **1983,** *105*, 765-767.
(61) X. Wang; K. Houk, Theoretical elucidation of the origin of the anomalously high acidity of Meldrum's acid. *J. Am. Chem. Soc.* **1988,** *110*, 1870-1872.
(62) E.D. Glendening; F. Weinhold, Natural resonance theory: I. General formalism. *J. Comput. Chem.* **1998,** *19*, 593-609.
(63) E.D. Glendening; J. Badenhoop; F. Weinhold, Natural resonance theory: III. Chemical applications. *J. Comput. Chem.* **1998,** *19*, 628-646.
(64) A.D. Clauss; S.F. Nelsen; M. Ayoub; J.W. Moore; C.R. Landis; F. Weinhold, Rabbit-ears hybrids, VSEPR sterics, and other orbital anachronisms. *Chem. Educ. Res. Pract.* **2014,** *15*, 417-434.
(65) S. Grabowsky; R. Kalinowski; M. Weber; D. Förster; C. Paulmann; P. Luger, Transferability and reproducibility in electron-density studies–bond-topological and atomic properties of tripeptides of the type l-alanyl-Xl-alanine. *Acta Cryst. B* **2009,** *65*, 488-501.
(66) R. Kamiński; S. Domagała; K.N. Jarzembska; A.A. Hoser; W.F. Sanjuan-Szklarz; M.J. Gutmann; A. Makal; M. Malińska; J. Bąk; K. Wozńiak, Statistical analysis of multipole-model-derived structural parameters and charge-density properties from high-resolution X-ray diffraction experiments. *Acta Cryst. A* **2014,** *70*, 72-91.
(67) E. Espinosa; E. Molins; C. Lecomte, Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chem. Phys. Lett.* **1998,** *285*, 170-173.
(68) P. Coppens; Y. Abramov; M. Carducci; B. Korjov; I. Novozhilova; C. Alhambra; M.R. Pressprich, Experimental Charge Densities and Intermolecular Interactions: Electrostatic and Topological Analysis of dl-Histidine. *J. Am. Chem. Soc.* **1999,** *121*, 2585-2593.
(69) M.A. Spackman, How Reliable Are Intermolecular Interaction Energies Estimated from Topological Analysis of Experimental Electron Densities? *Cryst. Growth Des.* **2015,** *15*, 5624-5628.
(70) N. Dadda; A. Nassour; B. Guillot; N. Benali-Cherif; C. Jelsch, Charge-density analysis and electrostatic properties of 2-carboxy-4-methylanilinium chloride monohydrate obtained using a multipolar and a spherical-charges model. *Acta Cryst. A* **2012,** *68*, 452-463.