BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Bond distances in polypeptide backbones depend on the local conformation

Roberto Improta,* Luigi Vitagliano* and Luciana Esposito*

Istituto di Biostrutture e Bioimmagini, CNR, via Mezzocannone 16, I-80134 Napoli, Italy. *Correspondence e-mail: robimp@unina.it, luigi.vitagliano@unina.it, luciana.esposito@cnr.it

Received 8 January 2015
Accepted 17 March 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: backbone geometry; conformational dependences; peptide-bond resonance; bond distances.

Supporting information: this article has supporting information at journals.iucr.org/d

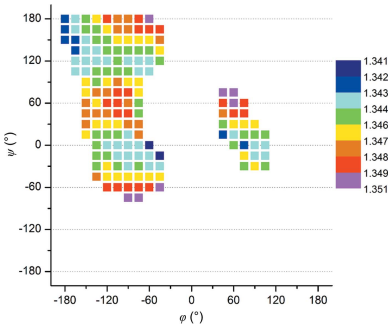
By combining quantum-mechanical analysis of small model peptides and statistical surveys of high-resolution protein structures, a systematic conformational dependence of bond lengths in polypeptide backbones has been unveiled which involves both the peptide bond (C—O and C—N) and those bonds centred on the C^{α} atom. All of these bond lengths indeed display a systematic variability in the ψ angle according to both calculations and surveys of protein structures. The overall agreement between the computed and the statistical data suggests that these trends are essentially driven by local effects. The dependence of C^{α} distances on ψ is governed by interactions between the σ system of the C^{α} moiety and the C—O π system of the peptide bond. Maximum and minimum values for each bond distance are found for conformations with the specific bond perpendicular and parallel to the adjacent CONH peptide plane, respectively. On the other hand, the variability of the C—O and C—N distances is related to the strength of the interactions between the lone pair of the N atom and the C—O π^* system, which is modulated by the ψ angle. The C—O and C—N distances are related but their trends are not strictly connected to peptide-bond planarity, although a correlation amongst all of these parameters is expected on the basis of the classical resonance model.

1. Introduction

How the amino-acid sequence determines the protein structure (folding code) is still a central open question in structural biology. Prediction of the three-dimensional arrangements of atoms in these macromolecules represents a formidable challenge owing to their peculiar architecture that often combines complexity with an intrinsic fragility, which is essential for functionality and turnover (degradation). The structure and function of proteins depend on the subtle balance of intraresidue and interresidue effects, including those very distant in the primary structure, which in many cases depend on weak noncovalent interactions (hydrogen bonds, van der Waals contacts) also involving the environment (solvent, other proteins *etc.*)

Full comprehension of the functions of proteins often requires the elucidation of very fine details of their structure. An illuminating example in this context, unveiled by sub-angstrom crystallographic studies, is represented by the mechanism of phosphate/arsenate ion discrimination exhibited by phosphate-binding proteins (Elias *et al.*, 2012). Indeed, it has been demonstrated that the 500-fold higher affinity of these proteins for phosphate compared with arsenate exclusively relies on a distortion of a unique low-barrier hydrogen bond observed caused by the 4% larger arsenate ion.

A complete understanding of the principles that dictate protein structures cannot avoid the quantification/definition of



all factors involved, starting from the discrimination between the effects of local and nonlocal interactions. From a very basic point of view, proteins are made of building blocks (the amino acids) that undergo a condensation reaction to form the peptide bond, thus generating long chains occasionally wrapped in folded states. Once the role of intraresidue effects have been fully elucidated, it is also easier to fully appreciate the consequences of interresidue and environmental interactions.

The elucidation of peptide-bond geometrical properties has proven to be crucial for the prediction of basic elements of protein structure (Edison, 2001; Pauling & Corey, 1951; Pauling *et al.*, 1951). Indeed, the seminal work by Pauling on the discovery of protein secondary structures would not have been possible without the assumption of the planarity of peptide bonds derived by resonance theory. However, over the years, the rigid and static vision of peptide geometry has progressively evolved toward a more plastic view. In the last two decades, careful inspections of fine structural details of protein structures have revealed a strong correlation between peptide-bond geometrical parameters (bond angles, dihedral angles and pyramidalization) and the local conformation (Van Alsenoy *et al.*, 1998; Karplus, 1996; Jiang *et al.*, 1997; Tronrud & Karplus, 2011; Jaskolski *et al.*, 2007; Esposito, Vitagliano, Sica *et al.*, 2000; Esposito, Vitagliano, Zagari *et al.*, 2000*a*; Esposito *et al.*, 2002, 2005, 2013; Improta *et al.*, 2011).

It is important to mention that the notion of peptide geometry variability has attracted the attention of the protein crystallography community, as it can be used both in protein structure refinement and validation (EU 3-D Validation Network, 1998; Kleywegt, 2009). In particular, the development of specific conformational-dependent libraries (CDL) by Karplus and coworkers (Berkholz *et al.*, 2009, 2010; Tronrud *et al.*, 2010; Tronrud & Karplus, 2011; Moriarty *et al.*, 2014) represents a promising novel approach in protein refinement.

An open question in this field is represented by the dependence of bond distances on the local conformation (Carugo, 2003; Berkholz *et al.*, 2009). Previous statistical analyses carried out on ultrahigh resolution structures have highlighted a significant correlation between peptide-bond C—O/C—N distances (Esposito, Vitagliano, Zagari *et al.*, 2000*b*; Howard *et al.*, 2004; Bönisch *et al.*, 2005). These findings have been interpreted in the framework of Pauling's resonance model, which indicates that the C—N bond elongates when the C=O bond contracts. Recent investigations suggest that bond distances in peptide bonds may also be related to the local conformation of the peptide (Berkholz *et al.*, 2009). In particular, the analysis of Berkholz and coworkers indicated that both C—O and C—N distances present a dependence on the ψ angle (see Fig. 5 in Berkholz *et al.*, 2009). However, the authors did not consider the implications of this finding, as the range of the detected variability of these distances was close to the standard deviations of their distributions (0.012–0.016 Å).

We have shown that calculations performed on small peptide-like model systems are able to correctly reproduce the variability of dihedral and bond angles found in protein structures (Improta *et al.*, 2011 and in preparation).

In this scenario, we have analyzed here peptide-bond distances in (φ, ψ) space using the same model systems and compared the results with those obtained from statistical surveys of protein structures. The close agreement that we found between the experimental and the theoretical trends has both methodological and chemical implications. Our investigations clearly indicate that the structural features of the peptide bond can be satisfactorily predicted by (ordinary) quantum-mechanics approaches with the use of small-sized models. Moreover, although long-range effects have an important role in protein structure and function, the present study corroborates and extends the notion that peptide-bond geometrical parameters are mainly dictated by local effects. Indeed, we show here that the distances involved in the peptide bond (C—O and C—N), as well as those centred on the C^{α} atom, strongly depend on the values of the adjacent ψ dihedral angle. Interestingly, C—O and C—N bond lengths are instead unrelated to the planarity of the peptide bond. This means that an essentially planar peptide bond is not necessarily associated with short C—N and long C—O backbone distances. This observation challenges too strict an application of the Pauling resonance model that predicts a connection between peptide-bond planarity and the bond order of the C—N and C—O bonds.

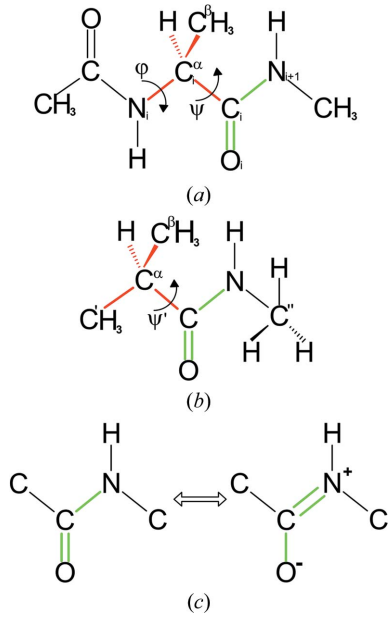


Figure 1
Schematic representation of the peptide models used in QM calculations. (*a*) *N*-Acetyl *N*′-methylalaninamide (Ala1). (*b*) Simplified model with the *N*-acetyl group replaced by a methyl group (Pep). (*c*) Classical representation of the major resonance forms of the peptide bond (lone pairs are not shown).

2. Materials and methods

2.1. Computations and models

Calculations have been performed by considering peptide models of different complexity as described previously (Fig. 1;

Improta *et al.*, 2011); most of them were conducted on a dipeptide analogue, *N*-acetyl *N*′-methylalaninamide (Ala1; Fig. 1*a*), which contains all of the relevant geometrical parameters to be analysed. Bulk-solvent effects have been taken into account by means of the polarizable continuum model (PCM; Tomasi *et al.*, 2005).

The PCM is not expected to give an account of possible explicit hydrogen bonds to solvent molecules or to other residues within a protein. On the other hand, with the possible exception of residues in ordered repeating structures, these effects are expected to be averaged out by our statistical analysis. On the other hand, the PCM can be important to reduce the weight of intraresidue electrostatic interaction, avoiding the artificial structural distortion that could occur in gas-phase calculations.

Calculations have been also performed in the gas phase on a simplified system (Pep; Fig. 1*b*) in order to obtain a simpler picture of the influence of the C^{α} substituents on the structural parameters of the amide moiety. Calculations on Pep provide a more direct picture of the orbital interactions, *i.e.* depending only on the interaction between the substituent at C^{α} and the peptide bond, and are not affected by the onset of electrostatic interactions between the two peptide groups present in Ala1. As a consequence, the analysis of different trends found for Pep and Ala1 gives significant insights into the different role played by 'sterical' and electrostatic interactions in determining the experimental results.

Geometry optimizations were carried out once the conformation of the molecule had been blocked at fixed values. In particular, the conformational space of Ala1 was sampled by using a grid of fixed (φ, ψ) values of (15 × 15°), whereas for Pep model ψ varies systematically in increments of 15° (Improta *et al.*, 2011). Except for the fixed (φ, ψ) values, all of the other geometrical parameters are free to vary.

The bulk of our computational analysis has been performed at the density functional theory (DFT) level using the PBE0 hybrid functional (Adamo *et al.*, 1999). PBE0 is a parameter-free hybrid functional commonly used to treat several classes of compounds. On the other hand, in common with many functionals without long-range correction, it suffers from limitations in the description of electron interaction between doubly occupied orbitals. As a consequence, some problems are encountered when studying the conformational behaviour of polypeptides (Improta & Barone, 2004), since PBE0 underestimates the relative stability of 'compact' conformations, such as α-helices, especially when large basis sets are used. This limitation is less important in the present study: the intraresidue structural effects that we are dealing with depend mainly on the interaction between filled and empty molecular orbitals and on the electrostatic interaction between the different peptide groups. These kinds of interaction are correctly treated by PBE0 (Improta & Barone, 2004), as shown by the very good performance of this functional in the study of the conformational behaviour of dipeptide analogues (Benzi *et al.*, 2002; Improta *et al.*, 2001; Improta & Barone, 2004; Langella *et al.*, 2002) also when using a relatively small basis set such as 6-31G(d) and, in fact, by the agreement with the experimental geometry trends that we discuss in the present paper.

All calculations were performed using the *Gaussian* 03 package (Frisch, 2004). The *MOLDEN* package was used for visualization purposes (Schaftenaar & Noordik, 2000).

2.2. Statistical analyses of the bond-length distributions in protein structure databases

Statistical analyses of peptide-bond geometrical parameters have been performed by considering a subset of nonredundant (pairwise sequence identity of <90%) highly accurate structures from the Protein Data Bank (PDB; July 2013 release) refined at resolutions in the range 0.5–0.9 Å (Berman *et al.*, 2007). The use of a generous sequence-identity threshold (90%) and a stringent resolution limit was made to obtain a significantly large protein ensemble of accurate structures. As confirmed by the results of the present paper, we expected that the protein-topology context has a limited impact on the distances. Obviously, for proteins sharing a sequence identity higher than 90% only the highest resolution model was considered. To avoid bias in the ensemble, for entries containing multiple copies of the protein monomer a single chain was considered in the analysis. Structures containing unconventional residues which could negatively affect the analysis were excluded. The survey resulted in the 74 structures listed in Supplementary Table S1.

Since even very well refined ultrahigh-resolution protein structures may contain regions that are locally disordered, peptide bonds were further selected based on their occupancy and *B*-factor values. In particular, either residues modelled with partial occupancy (<1.0) or residues at positions of sequence heterogeneity were omitted. After this selection, the average *B* factor of the entire protein was calculated. Residues containing either a main-chain or a C^{α} atom with a *B* factor value of higher than 1.20 times the average *B* factor of the protein were also excluded. To better compare the results with our calculations, we excluded Gly residues from the analysis. The applied filters selected 7109 residues from which bond distances were calculated.

Bond-distance variability was analysed by generating two-dimensional plots of bond distances as function of (φ, ψ). Indeed, we divided the experimental measurements into 15 × 15° bins centred on the same fixed (φ, ψ) grid points used in the calculations. We considered only bins where the population was larger than ten (81 bins in total). The mean values of each distance calculated in each bin were compared with the results of QM calculations by using linear regression analyses.

In addition, to assess the statistical significance of the observed trends, we grouped the bond distances into 40° bins of ψ angle centred on the expected minimum and maximum values for each distance. A mean value for the bond distance was calculated in each bin. Taking into account the different populations of ψ regions, we chose a 40° bin size in order to have at least 150 measures in each bin.

Simple statistical t-tests implemented in *Microsoft Excel* were used to evaluate the significance of the difference

between the means from two adjacent ψ bins (Supplementary Tables S2 and S3). In particular, we applied t-test differences between two unpaired means assuming either equal or unequal variances (depending on the results of an F-test to determine the equality of variance in the two groups of distances); a one-tailed test with 5% significance level was chosen in order to state that one distance is greater than the other in the adjacent ψ bin.

3. Results

3.1. Variability of bond distances in QM calculations

We identified two major points in the polypeptide chain for the analysis of the backbone bond distances: the distances directly involved in the peptide bond (C—O and C—N; highlighted in green in Fig. 1) and those centred on the C^{α} atom (C^{α}—C^{β}, N—C^{α}, C^{α}—H^{α} and C^{α}—C; highlighted in red in Fig. 1).

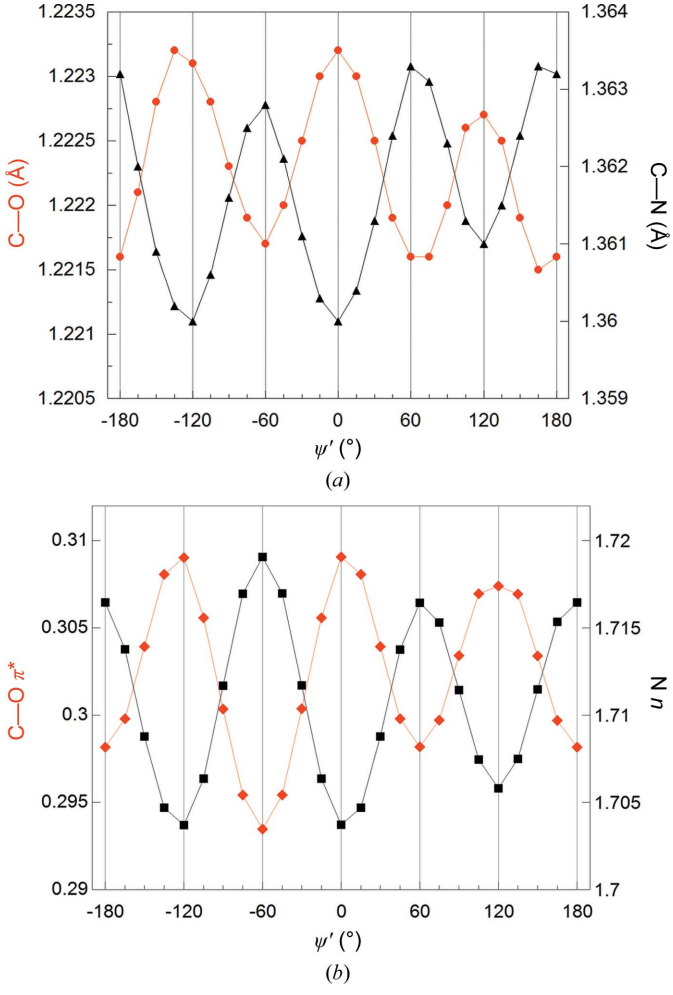


Figure 2
Dependence of the peptide bond on ψ' dihedral angles. (*a*) C—O/C—N distances *versus* ψ' in Pep. (*b*) Population of the C—O nonbonding π^* orbital (red diamonds) and the N lone-pair orbital *n* (black squares) as a function of ψ' in Pep (this figure is adapted from Supplementary Fig. S13 of Improta *et al.*, 2011).

In our calculations we used two peptide models: *N*-acetyl-*N*′-methylalaninamide (Ala1), which contains a central Ala residue, and a simplified model (Pep) which contains a methyl group replacing the *N*-acetyl group of Ala1. Therefore, in Pep the N atom of the Ala residue is missing and there is a CH_3(N-like) group. Given that Pep lacks a proper φ angle, this model essentially provides information about the dependence on the ψ angle (named ψ' in Pep). On the other hand, the dependences on the whole Ramachandran space (φ, ψ) are provided by the calculations on Ala1. In our analysis, the peptide bond formed by residues *i* and *i* + 1 is assigned to residue *i*.

3.1.1. C—O and C—N bond distances in Pep and Ala1 models. It is now well known that the analysis of experimental databases as well as of single high-resolution crystal structures of peptides and proteins provides evidence in support of a negative correlation existing between the C—O and C—N

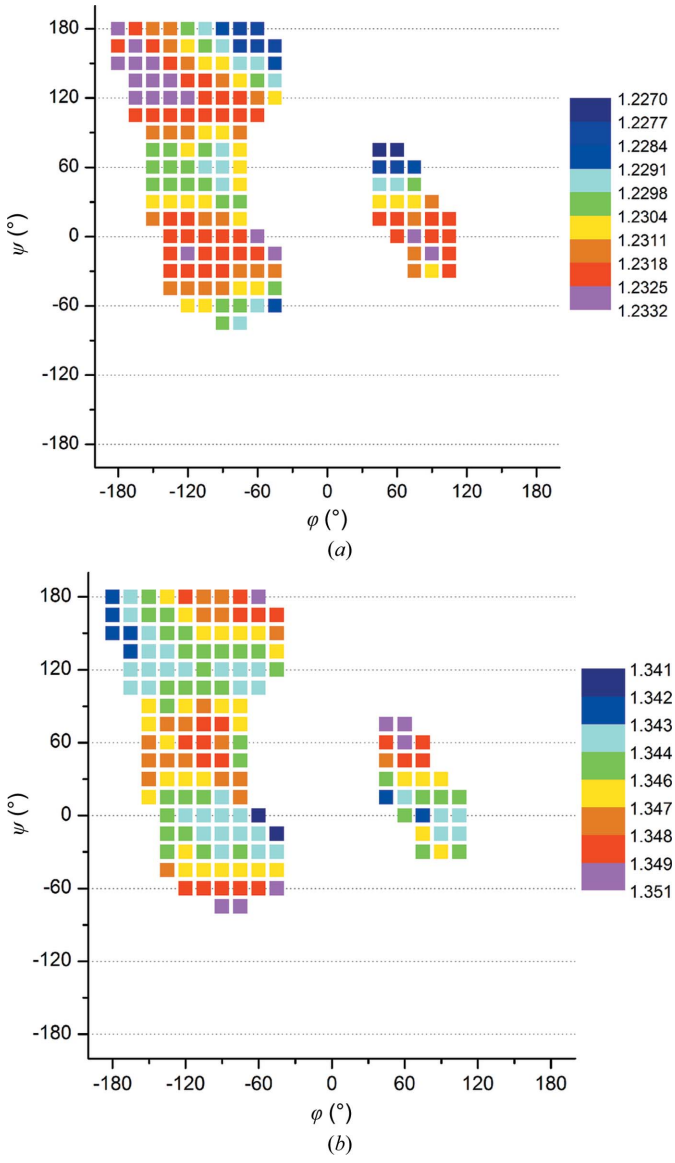


Figure 3
Dependence of distances on peptide conformation in Ala1. (*a*) C—O distances. (*b*) C—N distances.

bond lengths (Esposito *et al.*, 2005; Howard *et al.*, 2004; Bonisch *et al.*, 2005). We verified this feature in our calculations on the Pep model in the gas phase. The experimental trend is confirmed by the regression analysis, which shows a highly negative correlation coefficient [y_{C-N} (Å) = 3.37 − 1.94x_{C-O} (Å); R = −0.99; Supplementary Fig. S1]. Examination of the C—O/C—N bond distances separately reveals a marked sinusoidal variation with the ψ' angle (Fig. 2*a*). The pattern of variability shows a threefold periodicity (120°) along the ψ' angle.

The PCM/PBE0 calculations on Ala1 in water also indicate that the C—O/C—N bond distances are correlated. Regression analysis of the data shows a highly negative correlation [y_{C-N} (Å) = 2.97 − 1.32x_{C-O} (Å); R = −0.87] but lower than that observed in Pep, confirming that in the simplified model the correlations among the different geometrical parameters are overemphasized. In Ala1, the simple interdependence between C—O and C—N bonds is indeed attenuated by the existence of several intramolecular interactions (involving, for example, N—H and C—O moieties), which mimic some of the interactions existing in proteins (at least those involving adjacent residues). As a matter of fact, in a real protein structure determined at ultrahigh resolution (human aldose reductase refined at 0.66 Å), the correlation coefficient shows a comparable value (R = −0.69) to that determined by our calculations on Ala1 (Howard *et al.*, 2004). In agreement with the experimental results, our calculations thus indicate that shorter C—N bond lengths are associated with longer C—O bond lengths, which is in line with the predictions of the 'classical' resonance model (Fig. 1*c*).

We then analysed the variability of each bond in Ramachandran space for Ala1 (Fig. 3). Although a certain dependence on the φ angle is evident, the main variations occur with ψ (Fig. 3, Supplementary Fig. S2). The pattern of variability shows a threefold periodicity (120°) with the ψ angle. In particular, the C—O distance shows maximum values at $\psi \simeq 0$

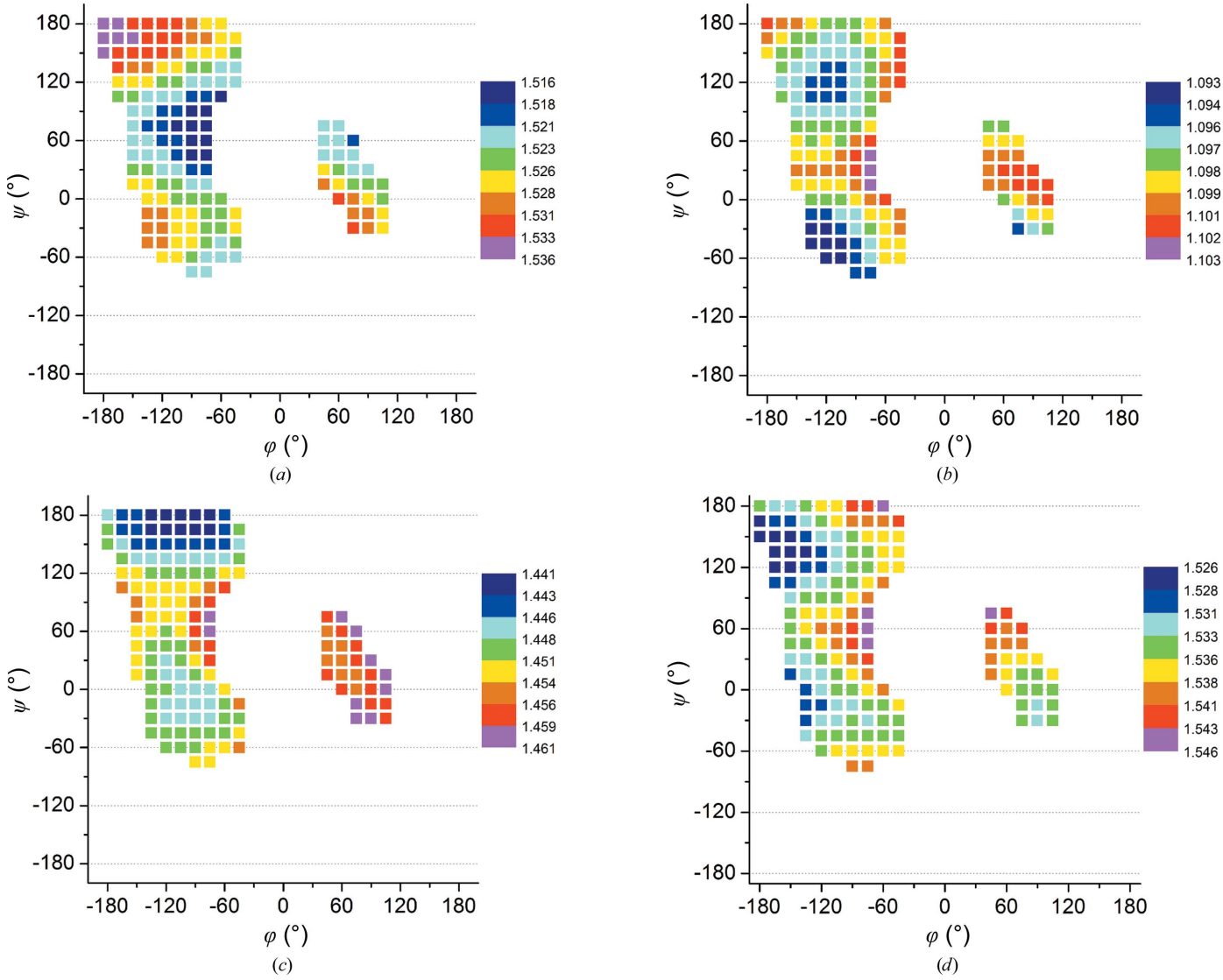


Figure 4
Dependence of distances at the C^{α} atom on peptide conformation in Ala1. (*a*) $C^{\alpha}-C^{\beta}$ distances. (*b*) $C^{\alpha}-H^{\alpha}$ distances. (*c*) $N-C^{\alpha}$ distances. (*d*) $C^{\alpha}-C$ distances.

and 120° and minima at $\psi \simeq -60$, 60 and 180° (Fig. 3*a*). A slight dependence on φ indicates that the bond length decreases with increasing φ values (less negative values) and is mainly evident in the region of extended conformation (ψ = 120–180°; Supplementary Fig. S3).

The opposite trend is displayed by the C—N distance, which shows minimum values at $\psi \simeq 0$ and 120° and maximum values at $\psi \simeq -60$, 60 and 180° (Fig. 3*b*). Minor variability with φ also occurs in the opposite direction with respect to the C—O distance.

3.1.2. Bond distances involving the C^{α} atom in the Pep and Ala1 models. As the next step in our analysis, we focused on the degrees of freedom involving the bond distances at the C^{α} atom ($C^{\alpha}-C^{\beta}$, $N-C^{\alpha}$, $C^{\alpha}-H^{\alpha}$ and $C^{\alpha}-C$). We analysed the $C^{\alpha}-X$ bond lengths obtained by calculations on Ala1 in solution by plotting them for each (φ, ψ) value of their respective residues (Fig. 4). $C^{\alpha}-C^{\beta}$, $N-C^{\alpha}$ and $C^{\alpha}-H^{\alpha}$ bonds exhibit a clear dependence on ψ, which dominates over the dependence on φ, which is present but is quantitatively less relevant. The ψ dependence is illustrated by a horizontal patterning of colour variations: horizontal regions of large values (magenta/red/orange/yellow) alternate with regions of

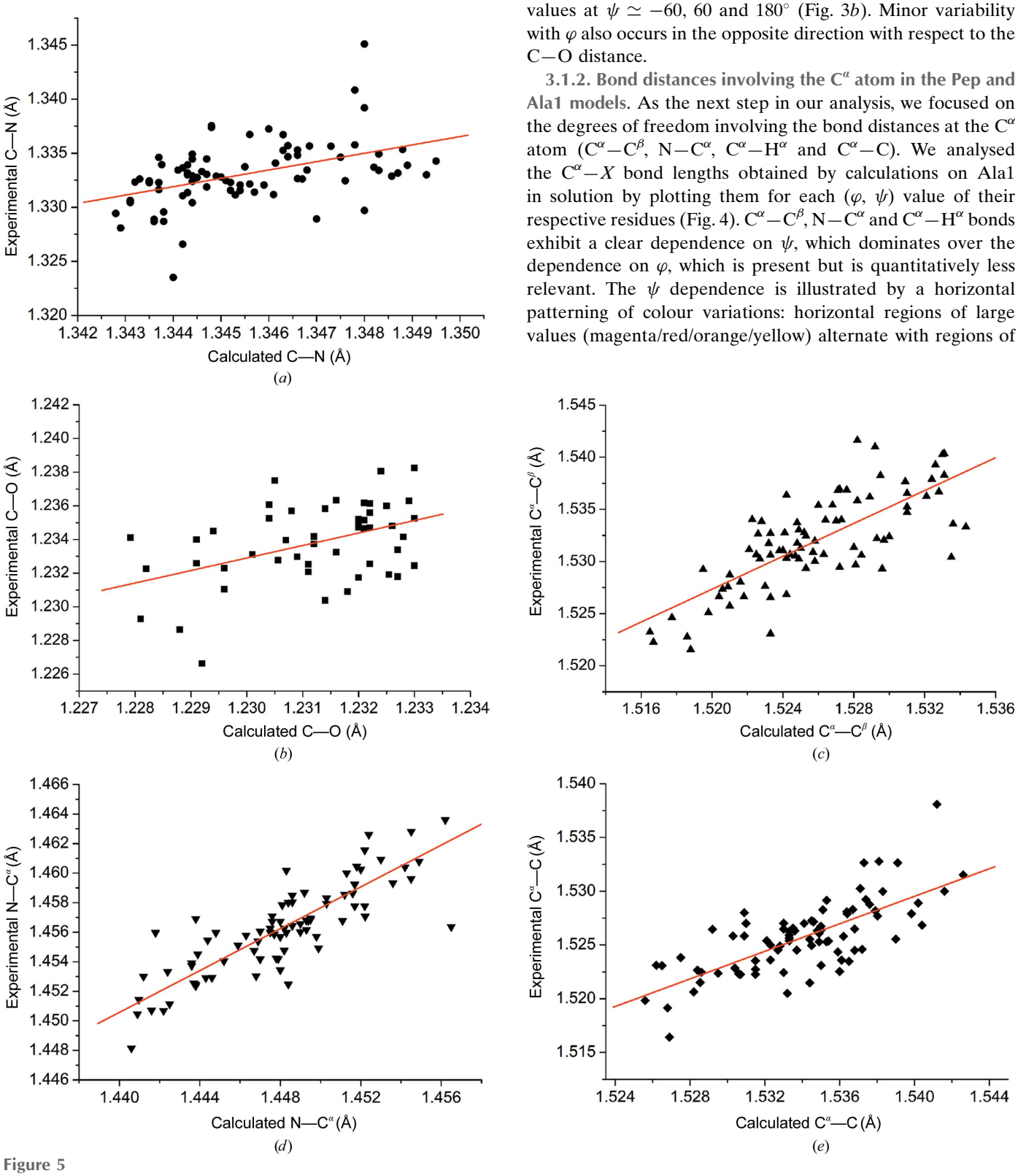


Figure 5
Comparison of calculated and experimental distances. The experimental values are derived by averaging the distances (all residue types except Gly) in (φ, ψ) bins of 15 × 15° in size centred on fixed grid points used for calculations. Only bins containing more than ten measures are considered (81 bins). In the case of C—O distances, only values from bins with $\psi > 70°$ are plotted (49 bins). (*a*) C—N. (*b*) C—O. (*c*) $C^{\alpha}-C^{\beta}$. (*d*) $N-C^{\alpha}$. (*e*) $C^{\alpha}-C$.

lower values (blu/cyan/green). In contrast to C—O and C—N bond lengths, which display a threefold periodicity (maxima or minima separated by 120°), these distances display an approximate twofold periodicity (maxima or minima separated by 180°). The dependence can be better visualized in a $C^{\alpha}-X$ *versus* ψ plot (Supplementary Fig. S4). As shown in Fig. 4(*a*), $C^{\alpha}-C^{\beta}$ presents maximum and minimum values for $\psi \simeq (-30, 150^{\circ})$ and 60°, respectively. A rather clear systematic variation with φ is also observed for negative values of φ: the distance decreases with increasing φ values. For $C^{\alpha}-H^{\alpha}$, maximum and minimum values are detected at $\psi \simeq 30^{\circ}$ and $\psi \simeq (-60, 120^{\circ})$ (Fig. 4*b*), respectively. Finally, $N-C^{\alpha}$ shows a less clear trend, with the maximum located at ψ approximately equal to 90°, while the minimum values are at $\psi \simeq (180, 0^{\circ})$ (Fig. 4*c*). In particular, the trend with ψ is not found when φ adopt positive values; in this region of the (φ, ψ) space large values for the distances are found as well as a diagonal patterning with bond lengths increasing towards higher values of φ.

The dependence of these $C^{\alpha}-X$ distances on ψ can be easily rationalized: maximum values are obtained when the $C^{\alpha}-X$ bond is perpendicular to the CONH amide plane, whereas minima are found for conformations with the $C^{\alpha}-X$ bond parallel to that plane (Supplementary Fig. S5).

The same trends in the equivalent $C^{\alpha}-X$ bond lengths are more clearly recognizable in the simpler Pep system [$X = C^{\beta}$, CH_3(N-like) or H^{α}; Supplementary Fig. S6].

The last geometrical parameter analysed is the $C^{\alpha}-C$ bond. The rotation around this bond defines the ψ torsion angle. In this case the trend is more complicated, with a significant dependence on both φ and ψ being evident (Fig. 4*d* and Supplementary Figs. S4*d* and S7). A maximum value for this distance is found around $\psi \simeq 60^{\circ}$, whereas minimum values are achieved around $\psi \simeq (-30, 150^{\circ})$ (Fig. 4*d*). The bond length becomes shorter as φ moves away from zero in either direction (positive or negative values; Supplementary Fig. S7).

3.2. Dependence of bond distances on peptide conformation: a survey of ultrahigh-resolution protein structures

In the previous section, we have thus shown that, according to the calculations on Ala1 and Pep, both C—O/C—N bond distances and distances involving the C^{α} atom exhibit clear conformational trends.

On the experimental side, a recent independent statistical survey of ultrahigh-resolution protein structures highlighted a systematic conformation-dependent variation of backbone bond angles, but also reported a preliminary analysis of C—O and C—N bond-distance variations (Berkholz *et al.*, 2009). A qualitative comparison of these latter results (see Fig. 5 in Berkholz *et al.*, 2009) with our QM calculation reports shows a similar threefold pattern of the C—O/C—N distance variability with the ψ angle (Fig. 3).

Here, we have re-evaluated the C—O/C—N variability in experimental polypeptide structures as function of φ and ψ (Supplemetary Fig. S8) by performing statistical surveys of protein structures at ultrahigh resolution (better than 0.9 Å; see §2). Moreover, we have extended this analysis to the distances involving the C^{α} atom ($N-C^{\alpha}$, $C^{\alpha}-C^{\beta}$ and $C^{\alpha}-C$), whose conformational trends have been highlighted in QM calculations. The distances were grouped into $15 \times 15^{\circ}$ (φ, ψ) bins and the mean values were calculated in each bin. The plot of average values shows that all of these distances are somewhat dependent on the local conformation. A qualitative comparison indicates that the calculated and statistical trends are similar (Supplementary Fig. S9). To obtain a quantitative comparison, we performed a linear regression analysis between calculated and experimentally averaged data for each distance (Fig. 5). For the C—N (Fig. 5*a*) and the $C^{\alpha}-C^{\beta}$ (Fig. 5*c*) distances, the correlation coefficients (R) between the calculated and experimentally derived values are 0.45 and 0.75, respectively. For the $N-C^{\alpha}$ (Fig. 5*d*) and the $C^{\alpha}-C$ (Fig. 5*e*) bonds the correlation coefficients are as high as 0.84 and 0.72, respectively. For all these distances the correlations were highly significant (P value < 0.0001; Fig. 5). In the case of the C—O distance, a linear correlation is hardly detected (R = 0.20, P = 0.07). Indeed, the comparison of the two trends (Supplementary Fig. S9*b*) indicates that major discrepancies arise within the helical region ($\varphi < 0^{\circ}$ and $\psi < 0^{\circ}$). Indeed, when the comparison is restricted to the highly populated regions with $\psi > 70^{\circ}$ the agreement greatly improves ($R = 0.45$, $P = 0.001$).

A rigorous evaluation of the reliability of statistical data would require knowledge of the standard uncertainties of the coordinates, which are not available. At very high resolution, estimated variances of the positional parameters may be obtained through inversion of the least-squares full matrix used in the refinement, but this was accomplished in a limited number of analyses, showing that the standard uncertainties in positions could be as small as 0.01 Å for atoms with low B-factor values (Cruickshank, 1999). The variations in the bond lengths (0.01–0.025 Å) analysed in the statistical survey reported here are close to the expected errors associated with the structural determinations. Nonetheless, the agreement between theoretical and statistical trends for different types of bond distances mutually supports the results of both approaches.

To further assess the internal consistency of statistical data and to highlight the dominant dependence of bond-distance variability on ψ dihedral angles, we calculated mean values by grouping the bond lengths into ψ bins (Supplementary Fig. S9, Supplementary Tables S2 and S3).

In particular, for each type of distance the values were grouped into 40° bins of ψ angle centred on the minimum and maximum values expected from the theoretical/experimental trends. The choice of the bin size was the result of a compromise aiming at considering a significant number of measures (see §2) and avoiding heterogeneity of the sample. The differences between two means in adjacent bins are lower than the standard deviations, hence the distance-distribution overlap. Nonetheless, we can investigate whether the differences between the means calculated in two adjacent bins are different at a statistically significant level. Indeed, we used statistical t-tests with a 5% significance level to evaluate the

significance of the differences in the bond lengths with the ψ angle (Supplementary Table S2). We used a one-tailed test to compare the means in adjacent bins and to test whether one value is greater than the other.

Altogether, these results show that, despite their mixed φ, ψ dependence, the variability of these distances as function of ψ follows clear trends.

As an example, in our QM calculations the trend of the C—N distance exhibits maximum values at $\psi \simeq -60$, 60 and 180° and minima at $\psi \simeq 0$ and 180° (Fig. 3*b*). We have therefore analysed the experimental distribution of C—N bond lengths in these populated ψ bins. We compared the mean C—N values observed in the two adjacent 40° bins centred for instance at $\psi = 180°$ (565 distances) and $\psi = 120°$ (1732 distances) (Supplementary Table S2). The mean values are 1.333 and 1.331 Å in the $\psi = 180°$ and $\psi = 120°$ bins, respectively. The t-test indicates that there is a statistically significant difference, at the 5% level, between the means of the two groups, with the mean value at $\psi = 180°$ being the larger (significance level = 5% level, P-value = 10^{-4}). A similar difference is found for mean values in the remaining pairs of bins: $\psi = 120/60°$, $\psi = -60/0°$ and $\psi = 0/60°$ (Supplementary Table S2). As expected, compared with C—N distances, the C—O bond distances show the reverse behaviour with respect to the maximum and minimum values in the above ψ bins. However, the mean values in adjacent bins are highly statistically different only for the $\psi = 120/180°$ bins (Supplementary Table S2). As regards the bond distances involving the C^{α} atom, they show maxima and minima with a 180° periodicity; hence, only three ψ bins are considered. For all $C^{\alpha}—C^{\beta}$, $N—C^{\alpha}$ and $C^{\alpha}—C$ distances the difference between means in adjacent bins is statistically significant (with P-values ranging from 10^{-7} to 10^{-24}; Supplementary Table S3).

3.3. Explaining the trends by orbital analysis

To gain insights into the stereoelectronic effects of the observed/calculated trend of the bond distances within the backbone of peptides/proteins, we performed natural bond orbital analyses (Foster & Weinhold, 1980; Reed & Weinhold, 1983; Glendening & Weinhold, 1998). These investigations were conducted on the simplified model Pep, in which the correlation between the conformation and the bond distances is more evident. By using this model, we can obtain more direct access to the interaction involving the molecular orbitals of the peptide group, without the perturbation owing to the hydrogen-bond-like interaction with the polar group that is present in Ala1. It is clear, on the other hand, that this choice overemphasizes the effects we are going to discuss.

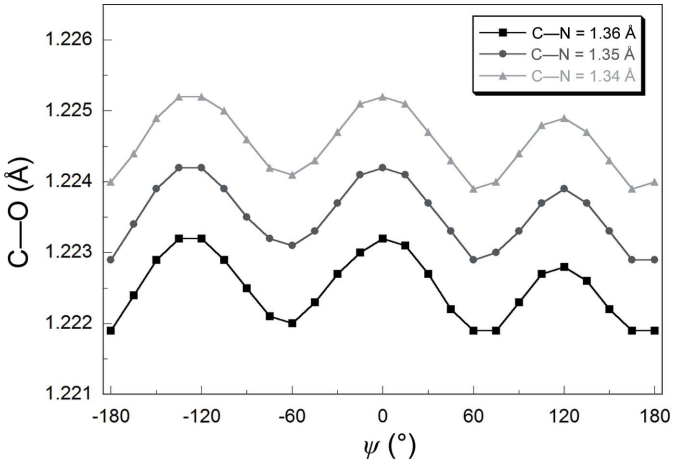


Figure 6
Pep model *in vacuo*: C—O bond length *versus* ψ for different fixed values of the C—N bond length.

3.3.1. C—O/C—N bond distances. In the Pep model we calculated the population of the N lone pair (n) and of the C—O π^* nonbonding orbitals depending on the ψ-like $C'—C^{\alpha}—C—N$ dihedral angle (ψ') which dictates the orientation of the C^{α} substituents with respect to the peptide plane (Fig. 1*b*). Fig. 2(*b*) clearly indicates that the orbital populations are strongly dependent on the ψ' angle. Indeed, the population of the N lone pair shows maximum values for $\psi' = 180$, 60 and −60°, which correspond to minimum values for the C—O π^* nonbonding orbital population (Fig. 2*b*). This result gives account of the experimental and calculated trends, which show short C—O and long C—N bond lengths for $\psi' = 180$, 60 and −60°. On the other hand, where a minimum of the nitrogen n population is found ($\psi' = 120$, 0 and −120°), short and long distances are observed for the C—N and C—O bonds, respectively, thus indicating a maximum contribution of the polar resonance amide structure (Fig. 1*c*, resonance form on the right). It is therefore clear that the orientation of the C^{α} substituents affects the strength of the N $n \rightarrow$ CO π^* interactions.

We also analysed the interactions between the σ system of the $-C^{\alpha}H(CH_3)_2$ moiety and the C—O π system ($\sigma \rightarrow \pi^*$ and $\pi \rightarrow \sigma^*$) which weaken the C—O π bond and could, in principle, increase its length. The NBO analysis shows, however, that the maximum of the $-C^{\alpha}H(CH_3)_2$/CO interactions is found for $\psi' = 180$, 60 and −60° (see Supplementary Fig. S11 of Improta *et al.*, 2011), for which the C—O bond length exhibits a minimum. These findings suggest that 'direct' interactions of the C^{α} σ skeleton with the C—O π bond play a minor role in affecting the C—O bond lengths. Therefore, the interaction between the C^{α} moiety and the nitrogen lone pair (Fig. 2*b*) is the most influential in determining the observed C—O bond distances.

In order to further assess the stereoelectronic effects underlying the dependence of the C—O bond length on ψ, we finally optimized Pep as a function of ψ' for different fixed values of the C—N bond length. As shown in Fig. 6, larger C—N distances correspond to smaller C—O distances, although substantial changes in the C—N bond length lead to much less significant shifts in the C—O bond length, which is in line with the results of previous studies, which highlight the interdependence between the σ and π systems in amides (Mujika *et al.*, 2006; Wiberg & Breneman, 1992; Milner-White, 1997). The sinusoidal dependence of the C—O distance on ψ' is conserved when the C—N distance is kept constant (Fig. 6). Interestingly, the variations of the C—O bond length owing to

changes in the ψ' value are of the same order of magnitude as those owing to a change of 0.01 Å in the C—N bond length (Fig. 6). The maxima of the C—O bond lengths are found for ψ' values for which, according to Fig. 2(*b*), the nitrogen lone-pair population exhibits a minimum (ψ' = 120, 0 and −120°). This result unambiguously shows that ψ' affects the C—O bond length mainly through an electronic interaction between the C^{α} moiety and the nitrogen lone pair, which is operative independently of the C—N bond length.

3.3.2. Bond lengths at the C^{α} atom. As discussed earlier, both calculations and statistical surveys indicate that the C^{α}—C^{β}, N—C^{α} and C^{α}—H^{α} distances exhibit the same twofold dependency on ψ: maximum and minimum values for each bond distance are detected for conformations with the specific bond perpendicular and parallel to the amide plane, respectively. As a representative example of the distances involving the C^{α} atom, we considered the C^{α}—C^{β} distance. NBO analysis of the interaction energies for the Pep model indicates that CO $\pi \rightarrow C^{\alpha}$—$C^{\beta}$ σ^* and C^{α}—C^{β} $\sigma \rightarrow$ CO π^* interaction energies are minimal when the C^{α}—C^{β} bond lies in the amide plane (ψ' = 60°), whereas they exhibit a maximum when this bond is close to being perpendicular to the amide plane ($\psi' \simeq -30$ and 150°; Fig. 7). Both of these interactions lead to lengthening of the C^{α}—C^{β} bond, thus explaining the experimental/computed trends. The same considerations also apply to the other N—C^{α} and C^{α}—H^{α} bond lengths.

A more complex conformational dependence is found for the C^{α}—C bond, *i.e.* the central bond around which the ψ dihedral angle rotation is defined. This bond length is significantly modulated by both the φ and ψ values. The ψ dependency of the C^{α}—C bond is anticorrelated with the C^{α}—C^{β} dependency, exhibiting minimum values when the C^{α}—C^{β} bond length is maximum, *i.e.* when this latter bond is perpendicular to the peptide plane, maximizing the interaction between the C^{α}—C^{β} σ bond and the C—O π system.

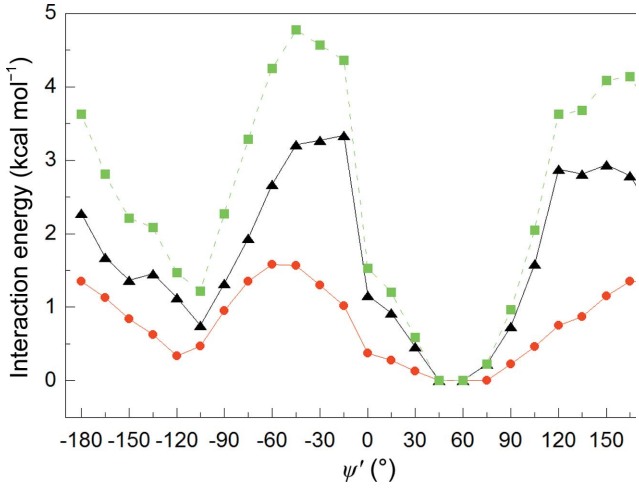


Figure 7
NBO analysis of Pep at the PBE0 level: the orbital interaction energy between the C^{α}—C^{β} σ system and the C—O π system as a function of ψ' is shown (black triangles, C^{α}—C^{β} $\sigma \rightarrow$ CO π^*; red circles, CO $\pi \rightarrow C^{\alpha}$—$C^{\beta}$ σ^*; green squares, sum of the above two contributions).

3.4. Effect of the peptide-bond planarity

As discussed in the previous sections, calculations and a statistical survey of protein structures agree in predicting that a clear interdependence exists between C—O and C—N bond lengths. Although we have shown that this relationship is strongly modulated by the ψ value, shorter C—N bonds correspond to longer C—O bonds and *vice versa*. This result is in line with the predictions of the 'resonance model': the resonance form with formal charges on the N and O atoms (Fig. 1*c*, right) shows a delocalization of charge from the nitrogen lone pair to the carbonyl O atom with an accompanying increase/decrease in the C—O/C—N bond lengths. The existence of this form, implying a double C═N bond, also gives an account of the planarity of the peptide bond (ω dihedral angle = 0 or 180°). According to a rigid application of the resonance model, we should thus expect that a perfectly planar peptide bond is associated with minimal distance of the C—N bond (and a maximal distance of the C—O bond) and that small deviations from planarity [measured by $\Delta\omega$, where $\Delta\omega = \omega - 180°$ (mod 360°)] are correlated with well defined and predictable variations in peptide distances. The present study instead shows that the significant ω distortions observed in proteins and peptides are not associated with specific expected trends of the C—O and C—N distances. Indeed, for instance, the conformations with ψ = −60, 60 and 180° are predicted to exhibit $\Delta\omega$ = 0° but, at the same time, long C—N and short C—O distances.

In order to further assess this issue, we thus analyzed C—O and C—N bond distances *versus* $|\Delta\omega|$: strict application of the resonance model would indeed provide a clear dependence of these distances on ω. Our calculations, both on Ala1 in solution (Figs. 8*a* and 8*b*) and on Pep in the gas phase (Supplementary Fig. S10), instead do not highlight any clear trend of C—O and C—N *versus* $|\Delta\omega|$, indicating that the anticorrelation of C—O and C—N distances cannot straightforwardly be related to the planarity of the peptide linkage. A statistical survey of proteins provides the same picture: C—O/C—N bond lengths are not correlated with the peptide-bond planarity (Figs. 8*c* and 8*d*).

4. Discussion

Recent QM calculations on model peptides have highlighted the central role of the dihedral angle ψ in affecting the planarity of the amide group ($\Delta\omega$, θ_C; Improta *et al.*, 2011). The present data extend the investigation to the conformational dependence of the length of several bonds, *e.g.* C—O, C—N and the bonds involving the C^{α} atom, by combining theoretical data and statistical analyses of very accurate protein structures. Both approaches offer a consistent view of the conformational dependence of bond distances. The central role of the ψ dihedral angle emerges from the systematic variability of bond distances in the Ramachandran plot. In particular, the pattern of variability of backbone C—O and C—N distances displays a threefold periodicity (120°) with the

ψ angle. On the other hand, the bond distances at the C^{α} atom display an approximate twofold periodicity (180°).

Besides unveiling the correlations between local conformation and peptide geometry, our study provides a unifying framework for their explanation. We have previously demonstrated that peptide-planarity deviations ($\Delta\omega$) are strongly dependent on the orientation of the C^{α} substituents (dictated by the ψ dihedral angle; Esposito *et al.*, 2005; Improta *et al.*, 2011). In particular, we found that peptide conformations with significant deviations from planarity ($\Delta\omega \neq 0°$) are predominant. $\Delta\omega$ exhibits a clear-cut sinusoidal dependence on ψ: positive and negative values of $\Delta\omega$ alternate every 60° in ψ. Indeed, planar peptides are detected only for specific ψ values (180, 120, 60, 0, −60 and −120°). The ψ angle is found to modulate the most influential orbital interaction between the N lone pair *n* and the C—O π^* orbitals ($n\rightarrow\pi^*$), thus determining the deviations from planarity. The present analysis shows that, even for the amide C—O and C—N bond lengths, the detected dependence on the ψ angle is owing to the same effect. Indeed, the ψ angle affects the population of the N *n* and the C—O π^* orbitals, thus directly modifying the bond lengths. These observations can be qualitatively interpreted by considering that for specific conformations the electronic repulsion between the C^{α} substituents and the electrons of the N lone pair may disturb the delocalization of the *n* lone pair in the π system.

On the other hand, as discussed in the previous paragraph, small deviations from peptide-bond planarity do not have any impact on the C—O and C—N bond lengths.

Some orientations of the C^{α} substituents with respect to the peptide plane (ψ = 180, −60, 60°) lead to a maximum of the population of the N *n* orbital as well as to a minimum of the population of the CO π^* orbital. This correspondingly yields maximal and minimal distances for the C—N and C—O bonds, respectively. The opposite situation is found for ψ angles (ψ = 0, −120 and 120°) which produce a minimum of the population of the N *n* orbital as well as a maximum of the population of the CO π^* orbital. Therefore, in agreement with the resonance model, the orbital analysis suggests that the N $n\rightarrow$CO π^* interaction is the most influential for determining both the

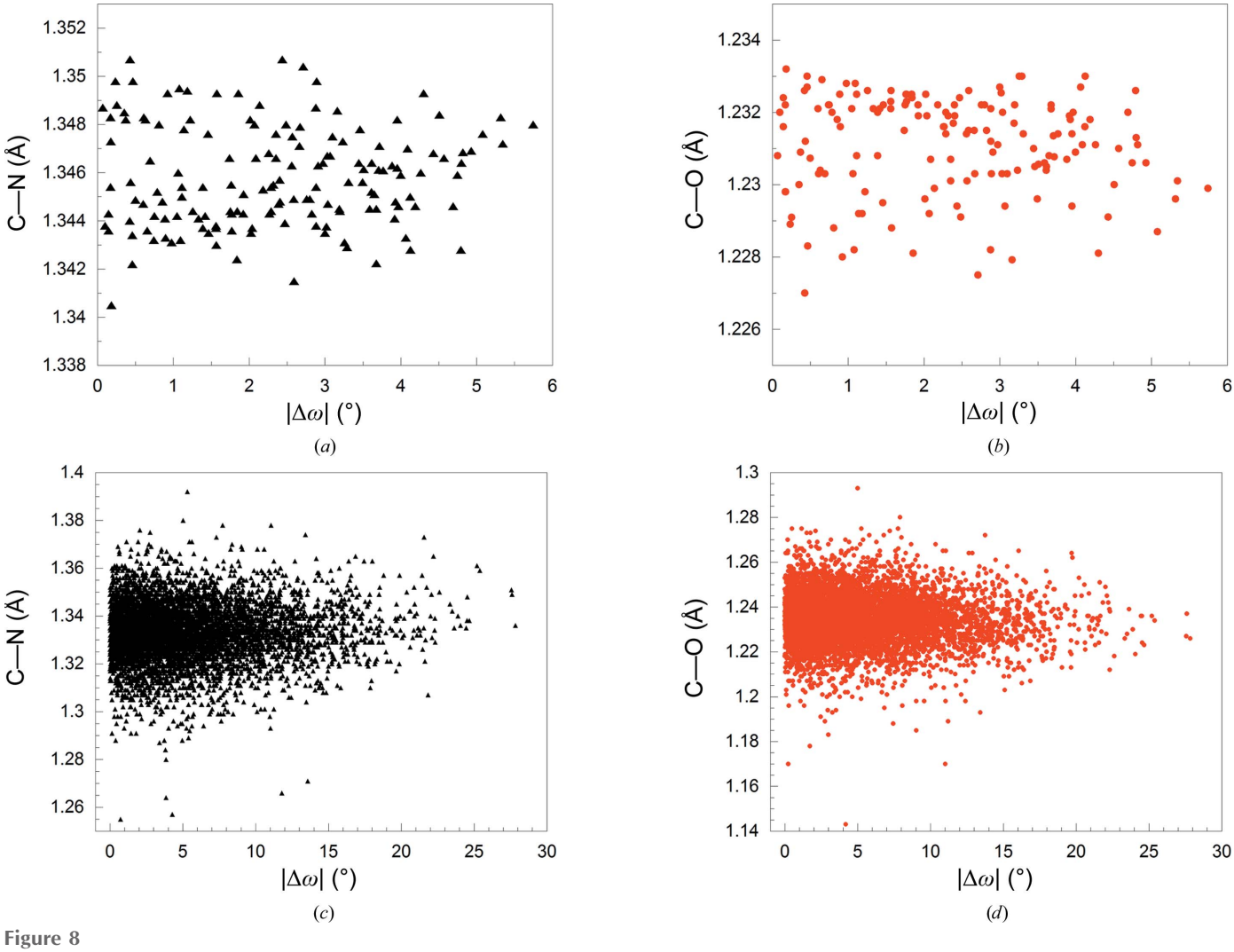


Figure 8
C—O/C—N distances *versus* $|\Delta\omega|$. (*a*) Ala1 model: C—N *versus* $|\Delta\omega|$. (*b*) Ala1 model: C—O *versus* $|\Delta\omega|$. (*c*) Proteins: C—N *versus* $|\Delta\omega|$. (*d*) Proteins: C—O *versus* $|\Delta\omega|$. It is worth noting that $\Delta\omega$ is defined as follows: $\Delta\omega = \omega - 180°$ (mod 360°).

C—N and the C—O bond lengths. However, the interaction of the C^{α} moiety with the π system has a noticeable effect on these distances by modulating the strength of N $n\rightarrow$CO π^* interactions.

The conformational variation of the bond distances of the C^{α} substituents can be explained by invoking a different interpretation in terms of orbital arguments. In these cases, the CO $\pi\rightarrow C^{\alpha}$ σ^* and C^{α} $\sigma\rightarrow$ CO π^* interactions are directly responsible for the dependence on ψ. Indeed, for each particular C^{α}—X bond these interactions are maximized when the X substituent is perpendicular to the amide plane (Supplementary Fig. S5). Since both CO $\pi\rightarrow C^{\alpha}$ σ^* and C^{α} $\sigma\rightarrow$CO π^* interactions lead to an increase in the C^{α}—X bond length, these bonds reach a maximum when the substituent is perpendicular to the amide plane.

In this paper, we report the results of a thorough analysis of the most significant stereoelectronic effects that modulate the local structural properties of proteins and peptides. We have integrated the results of quantum-mechanical calculations in the condensed phase of simple peptide model compounds with those of statistical analyses of high-resolution protein crystal structures. Both approaches show that peptide-bond lengths (C^{β}—C^{α}, N—C^{α}, C—N and C—O) are modulated by the arrangement of the C^{α} substituents. Interestingly, these conclusions hold for peptide systems of different sizes from small dipeptides to intricate protein architectures. This close correspondence emphasizes that basic local electronic effects play a role even in a complex folded protein structure. Long-range interresidue interactions, although obviously affecting the peptide geometry, are statistically averaged out, allowing the 'local' trends that we have highlighted here to emerge. Structured and oriented secondary structures, such as α-helices, can however induce trends that are not 'purely statistical', which may explain the only significant deviation that we observe between the experimental and the QM trends, *i.e.* the C—O bond variability in the region ($\varphi < 0^{\circ}$ and $\psi < 0^{\circ}$).

Simple interpretative models based on Lewis structures, such as those underlying the resonance structures shown in Fig. 1(*c*), are useful tools to explain, for example, the double-bond character of the peptide linkage and its essential planarity. However, they are not capable of correctly explaining the subtle structural details detected in real protein structures such as small deviations from planarity and/or the fine adjustment of bond distances with conformations. The energetic costs associated with small peptide-bond deformations are limited and are easily compensated by the optimization of other effects. On the other hand, in the present study we clearly show that these effects are intrinsic to the nature of the peptide bond (*i.e.* they are not only owing to interresidue interactions). In other words, when going from a simple amide to a peptide, with its asymmetric arrangement of substituents at C^{α}, we cannot expect a strict relationship between planarity and C—N/C—O bond lengths. Our QM calculations on small peptide models have instead disclosed the delicate interplay of several different geometrical parameters of the polypeptide backbone which also included bond distances.

In conclusion, even though unravelling tiny geometric changes could be regarded as a mere theoretical exercise, it is of paramount importance to understand key molecular mechanisms of biological macromolecules (Elias *et al.*, 2012). A broader awareness among scientists of bond-distance variations with local conformation together with an improved understanding of the electronic origins of these structural effects may spur new efforts to incorporate these subtle features into molecular modelling, protein-folding calculations and crystallographic refinement. In the latter field, recent attempts (Berkholz *et al.*, 2009, 2010; Tronrud *et al.*, 2010; Tronrud & Karplus, 2011; Moriarty *et al.*, 2014) to use a restraint library including conformational dependences seems to be very encouraging for future developments. Finally, the good performance of present QM calculations hints at the possibility of highlighting specific features by reiterating this procedure on other types of amino acids. Moreover, we are confident that in the near future the ongoing impressive growth of the database of ultrahigh-resolution PDB structures will allow a direct comparison between QM computational and experimental data on a residue-type base.

Acknowledgements

The authors thank L. De Luca for technical assistance. This work was supported by MIUR grant PRIN 2010ERFKXL. RI thanks Regione Campania, Legge 5/2007 for financial support.

References

Adamo, C., Scuseria, G. E. & Barone, V. (1999). *J. Chem. Phys.* **111**, 2889–2899.
Benzi, C., Improta, R., Scalmani, G. & Barone, V. (2002). *J. Comput. Chem.* **23**, 341–350.
Berkholz, D. S., Krenesky, P. B., Davidson, J. R. & Karplus, P. A. (2010). *Nucleic Acids Res.* **38**, 320–325.
Berkholz, D. S., Shapovalov, M. V., Dunbrack, R. L. Jr & Karplus, P. A. (2009). *Structure*, **17**, 1316–1325.
Berman, H., Henrick, K., Nakamura, H. & Markley, J. L. (2007). *Nucleic Acids Res.* **35**, D301–D303.
Bönisch, H., Schmidt, C. L., Bianco, P. & Ladenstein, R. (2005). *Acta Cryst.* D**61**, 990–1004.
Carugo, O. (2003). *Acta Chim. Slov.* **50**, 505–511.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Edison, A. S. (2001). *Nature Struct. Biol.* **8**, 201–202.
Elias, M., Wellner, A., Goldin-Azulay, K., Chabriere, E., Vorholt, J. A., Erb, T. J. & Tawfik, D. S. (2012). *Nature (London)*, **491**, 134–137.
Esposito, L., Balasco, N., De Simone, A., Berisio, R. & Vitagliano, L. (2013). *Biomed. Res. Int.* **2013**, 326914.
Esposito, L., De Simone, A., Zagari, A. & Vitagliano, L. (2005). *J. Mol. Biol.* **347**, 483–487.
Esposito, L., Vitagliano, L. & Mazzarella, L. (2002). *Protein Pept. Lett.* **9**, 95–105.
Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A. & Mazzarella, L. (2000). *J. Mol. Biol.* **297**, 713–732.
Esposito, L., Vitagliano, L., Zagari, A. & Mazzarella, L. (2000*a*). *Protein Sci.* **9**, 2038–2042.
Esposito, L., Vitagliano, L., Zagari, A. & Mazzarella, L. (2000*b*). *Protein Eng.* **13**, 825–828.
EU 3-D Validation Network (1998). *J. Mol. Biol.* **276**, 417–436.
Foster, J. P. & Weinhold, F. (1980). *J. Am. Chem. Soc.* **102**, 7211–7218.
Frisch, M. J. *et al.* (2004). *Gaussian* 03, version C.02. Gaussian Inc., Wallingford, Connecticut, USA.

Glendening, E. D. & Weinhold, F. (1998). *J. Comput. Chem.* **19**, 593–609.
Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). *Proteins*, **55**, 792–804.
Improta, R. & Barone, V. (2004). *J. Comput. Chem.* **25**, 1333–1341.
Improta, R., Benzi, C. & Barone, V. (2001). *J. Am. Chem. Soc.* **123**, 12568–12577.
Improta, R., Vitagliano, L. & Esposito, L. (2011). *PLoS One*, **6**, e24533.
Jaskolski, M., Gilski, M., Dauter, Z. & Wlodawer, A. (2007). *Acta Cryst.* D**63**, 611–620.
Jiang, X., Yu, C.-H., Cao, M., Newton, S. Q., Paulus, E. F. & Schäfer, L. (1997). *J. Mol. Struct.* **403**, 83–93.
Karplus, P. A. (1996). *Protein Sci.* **5**, 1406–1420.
Kleywegt, G. J. (2009). *Acta Cryst.* D**65**, 134–139.
Langella, E., Improta, R. & Barone, V. (2002). *J. Am. Chem. Soc.* **124**, 11531–11540.
Milner-White, E. J. (1997). *Protein Sci.* **6**, 2477–2482.
Moriarty, N. W., Tronrud, D. E., Adams, P. D. & Karplus, P. A. (2014). *FEBS J.* **281**, 4061–4071.
Mujika, J. I., Matxain, J. M., Eriksson, L. A. & Lopez, X. (2006). *Chem. Eur. J.* **12**, 7215–7224.
Pauling, L. & Corey, R. B. (1951). *Proc. Natl Acad. Sci. USA*, **37**, 251–256.
Pauling, L., Corey, R. B. & Branson, H. R. (1951). *Proc. Natl Acad. Sci. USA*, **37**, 205–211.
Reed, A. E. & Weinhold, F. (1983). *J. Chem. Phys.* **78**, 4066–4073.
Schaftenaar, G. & Noordik, J. H. (2000). *J. Comput. Aided Mol. Des.* **14**, 123–134.
Tomasi, J., Mennucci, B. & Cammi, R. (2005). *Chem. Rev.* **105**, 2999–3094.
Tronrud, D. E., Berkholz, D. S. & Karplus, P. A. (2010). *Acta Cryst.* D**66**, 834–842.
Tronrud, D. E. & Karplus, P. A. (2011). *Acta Cryst.* D**67**, 699–706.
Van Alsenoy, C., Yu, C.-H., Peeters, A., Martin, J. M. L. & Schäfer, L. (1998). *J. Phys. Chem. A*, **102**, 2246–2251.
Wiberg, K. B. & Breneman, C. M. (1992). *J. Am. Chem. Soc.* **114**, 831–840.