BMC Structural Biology

BioMed Central

Research article

Open Access

The Ramachandran plots of glycine and pre-proline

Bosco K Ho*[1] and Robert Brasseur[2]

Address: [1]Department of Pharmaceutical Chemistry, University of California San Francisco, 600 16th St, San Francisco, CA 94107, USA and [2]Centre de Biophysique Moléculaire Numérique, 2 Passage des déportés, B-5030 Gembloux, Belgium

Email: Bosco K Ho* - bosco@maxwell.ucsf.edu; Robert Brasseur - brasseur.r@fsagx.ac.be

* Corresponding author

Published: 16 August 2005

BMC Structural Biology 2005, **5**:14 doi:10.1186/1472-6807-5-14

Received: 15 March 2005

Accepted: 16 August 2005

This article is available from: http://www.biomedcentral.com/1472-6807/5/14

© 2005 Ho and Brasseur; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The Ramachandran plot is a fundamental tool in the analysis of protein structures. Of the 4 basic types of Ramachandran plots, the interactions that determine the generic and proline Ramachandran plots are well understood. The interactions of the glycine and pre-proline Ramachandran plots are not.

Results: In glycine, the ψ angle is typically clustered at $\psi = 180°$ and $\psi = 0°$. We show that these clusters correspond to conformations where either the N_{i+1} or O atom is sandwiched between the two H^{α} atoms of glycine. We show that the shape of the 5 distinct regions of density (the α, α_L, β_S, β_P and β_{PR} regions) can be reproduced with electrostatic dipole-dipole interactions. In pre-proline, we analyse the origin of the ζ region of the Ramachandran plot, a region unique to pre-proline. We show that it is stabilized by a $CO_{i-1}\cdots C^{\delta}H^{\delta}_{i+1}$ weak hydrogen bond. This is analogous to the $CO_{i-1}\cdots NH_{i+1}$ hydrogen bond that stabilizes the γ region in the generic Ramachandran plot.

Conclusion: We have identified the specific interactions that affect the backbone of glycine and pre-proline. Knowledge of these interactions will improve current force-fields, and help understand structural motifs containing these residues.

Background

The Ramachandran plot [1] is the 2d plot of the ϕ-ψ torsion angles of the protein backbone. It provides a simple view of the conformation of a protein. The ϕ-ψ angles cluster into distinct regions in the Ramachandran plot where each region corresponds to a particular secondary structure. There are four basic types of Ramachandran plots, depending on the stereo-chemistry of the amino acid: generic (which refers to the 18 non-glycine non-proline amino acids), glycine, proline, and pre-proline (which refers to residues preceding a proline [2]). The generic and proline Ramachandran plots are now well understood [3] but the glycine and pre-proline Ramachandran plots are not.

The generic Ramachandran plot was first explained by Ramachandran and co-workers in terms of steric clashes [1]. This has become the standard explanation for the observed regions in the Ramachandran plot [4,5]. However, recent studies found significant discrepancies between the classic steric map and the Ramachandran plot of high-resolution protein structures [6-9]. These discrepancies have now been resolved. The first discrepancy is that the $N\cdots H_{i+1}$ and $O_{i-1}\cdots C$ steric clashes in the classic steric map have no effect in the observed Ramachandran plot [3]. By removing these steric clashes, a better steric map can be constructed. The second discrepancy is that the Ramachandran plot cluster into distinct regions within the sterically-allowed regions of the

Ramachandran plot [8,10]. These clusters have now been explained in terms of backbone dipole-dipole interactions [3,11,12].

The proline Ramachandran plot has been reproduced in a calculation [13]. The proline Ramachandran plot is severely restricted by the pyrrolidine ring, where the flexibility in the pyrrolidine ring couples to the backbone [14].

The observed glycine Ramachandran plot has a distinctive distribution (Figure 1A) quite different to the generic Ramachandran plot. An early attempt to explain the observed Ramachandran plot in terms of a steric map of glycine [15] (Figure 2A) fails to account for the observed distribution. It does not explain the observed clustering at $\psi = 180°$ and $\psi = 0°$, nor the clustering into 5 distinct regions [8]. Using a molecular-dynamics simulation of Ace-Gly-Nme [16], Hu and co-workers found that the glycine Ramachandran plot generated by standard force-fields reproduced the original steric map but not the observed Ramachandran plot. They calculated a somewhat better result with a quantum-mechanics/molecular-mechanics model, which reproduced the observed clustering along ψ, but not the partitioning into the 5 clusters. In this study, we identify the specific interactions that define the observed glycine Ramachandran plot by studying the conformations of glycine in the structural database. We test these interactions with a simple model based on electrostatics and Lennard-Jones potentials.

Although the overall shape of the pre-proline Ramachandran plot (Figure 1B) is well understood, there exists a region unique to pre-proline that remains unexplained. The basic shape of pre-proline was predicted by Flory using steric interactions [17]. This was later confirmed in a statistical analysis of the protein database [2]. However, the statistical analysis also revealed the existence of a little leg of density poking out below the β-region (Figure 1B; purple in Figure 2C), which Karplus called the ζ region [10]. More recent calculations using standard molecular mechanics force-fields reproduced the energy surface of the original Flory calculation [13,18] but not the ζ region. In this study, we focus on the physical origin of the ζ region.

Results

A non-redundant PDB data-set

To extract the statistical distributions of the glycine and pre-proline Ramachandran plots, we chose a high-resolution subset of the PDB [19] provided by the Richardson lab [9] of 500 non-homologous proteins. These proteins have a resolution of better than 1.8 Å where all hydrogen atoms have been projected from the backbone and optimized in terms of packing. Following the Richardsons, we only consider atoms that have a B-factor of less than 30.

Regions in the glycine Ramachandran plot

Glycine is fundamentally different to the other amino acids in that it lacks a sidechain. In particular, glycine does not have the C^{β} atom, which induces many steric clashes in the generic Ramachandran plot. We call the hydrogen atom that is shared with the other amino acids, the $H^{\alpha 1}$ atom. We call the hydrogen atom that replaces the C^{β} atom, the $H^{\alpha 2}$ atom. The absence of the C^{β} atom allows the glycine Ramachandran plot to run over the borders at -180° and 180° (Figure 1A).

The observed glycine map has 5 regions of density [8]. In order to display the observed density in one continuous region, we shift the coordinates from ϕ-ψ to ϕ'-ψ' where ϕ': $0° < \phi' < 360°$, and ψ': $-90° < \psi' < 270°$. With the shifted glycine Ramachandran plot (Figure 3A), we can clearly identify the different regions. Along the horizontal strip $\psi' \sim 180°$, there are three separate regions. One of these is an elongated version of the β_P region of the generic Ramachandran plot. The β_P region corresponds to the polyproline II structure, which forms an extended left-handed helix along the protein chain [20]. The β_{PR} region is a reflection of the β_P region where a sequence of glycine residues in the β_{PR} conformation will form a right-handed helix. Finally, there is a region that corresponds to the β_S region of the generic Ramachandran plot. This region corresponds to the extended conformation of residues in β-sheets. However, the glycine β_S region, centered on $(\phi', \psi') = (180°, 180°)$, is slightly displaced from the β_S region of the generic Ramachandran plot. There is also the diagonal α and α_L regions (Figure 3A), which are associated with helices and turns [3]. Unlike the generic Ramachandran plot, the glycine α region is symmetric to the α_L region [8,21]. In the generic Ramachandran plot, there is also a γ region corresponding to the hydrogen bonded γ-turn [12]. The glycine Ramachandran plot does not have any density in the γ region.

Steric interactions in glycine

The original steric map of glycine (Figure 2A) [15] fails to explain large parts of the observed glycine Ramachandran plot (Figure 1A). In the observed glycine Ramachandran (Figure 3A), there are two large excluded horizontal strips at $50° < \psi' < 120°$ and $-120° < \psi' < -50°$, which are not excluded in the glycine steric map (Figure 2A). Conversely, the glycine steric map excludes a horizontal strip at $-30° < \psi' < 30°$ (Figure 2A), but this region is populated in the observed plot (Figure 1A). There are also diagonal steric boundaries in the observed glycine Ramachandran plot (Figure 1A), whereas the steric map predicts vertical boundaries (Figure 2A).

We carried out a re-evaluation of the steric map of glycine (Figure 2B) by following the methodology of Ho and co-workers [3]. For each interaction in the glycine backbone,

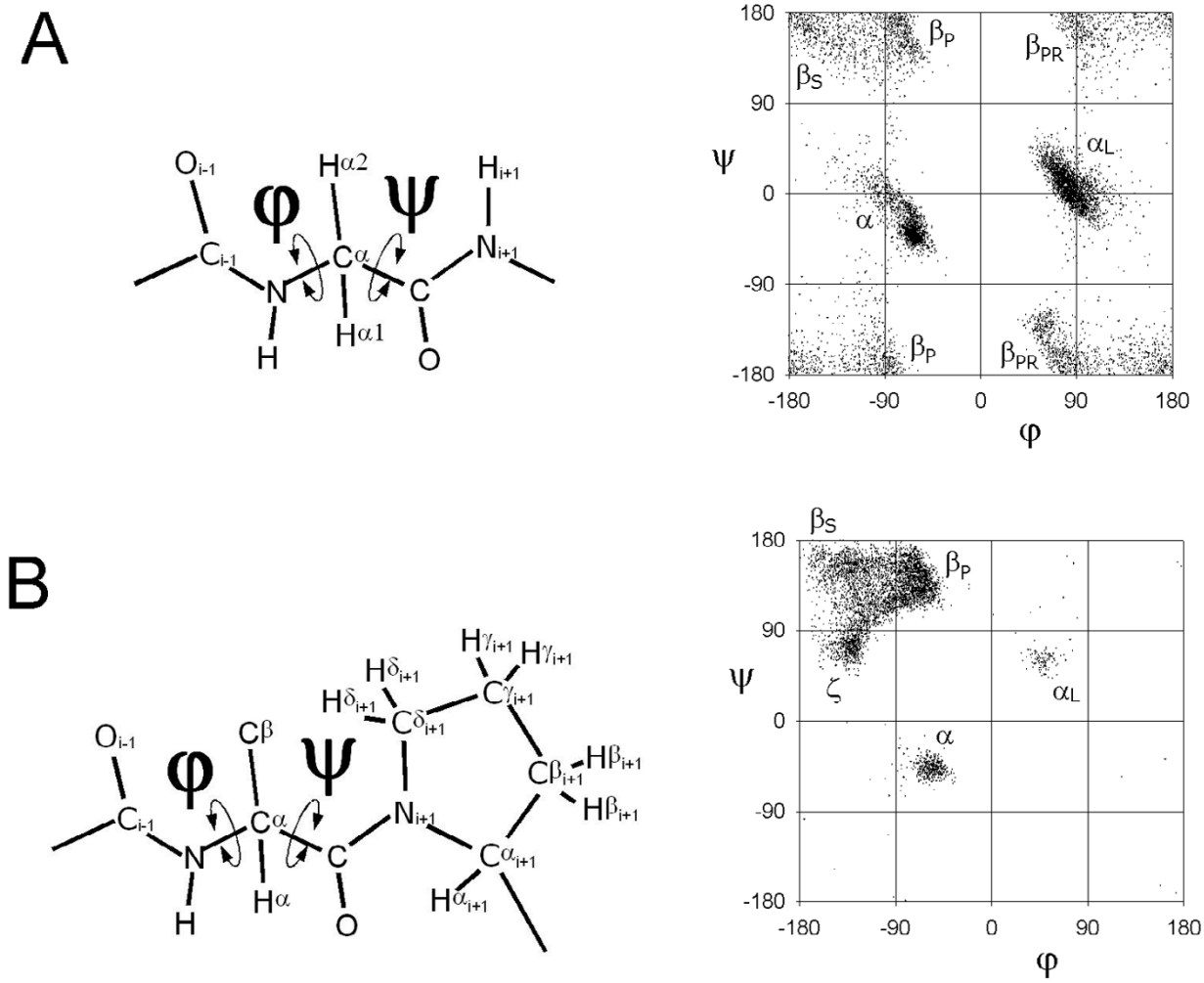

Figure 1
Backbone conformations of glycine and pre-proline. Backbone schematic (left) and observed Ramachandran plot (right) of (A) glycine and (B) pre-proline. Taken from the data-set of Lovell et al. (2003). The clustered regions are labeled on the Ramachandran plots.

we consider the variation of the inter-atomic distance with respect to the ϕ'-ψ' angles. We compare the observed variation to the variation generated from a model that uses canonical backbone geometry. We divide these interactions into 3 categories: the ϕ' dependent, ψ' dependent and ϕ'-ψ' co-dependent distances.

For some of the interactions, the results for glycine are identical to that of the generic Ramachandran plot [3]. For brevity, we omit the analysis of these interactions and summarize the results. The excluded horizontal strip -30° < ψ' < 30°, due to the $N \cdots H_{i+1}$ steric interaction in the glycine steric map (Figure 2A), does not exist in the observed distribution (Figure 1A). Similarly, the $O_{i-1} \cdots C$ steric clash in the original glycine steric map, which excludes a vertical strip centered on ϕ' = 0° (Figure 2A), does not exist in the observed distribution (Figure 1A). We ignore the effect of the $N \cdots H_{i+1}$ and $O_{i-1} \cdots C$ steric clashes. The diagonal boundaries of the observed distribution are defined by the ϕ'-ψ' co-dependent steric interactions $O_{i-1} \cdots O$ and $O_{i-1} \cdots N_{i+1}$. In Figure 3A, we show the fit of these steric interactions to the data.

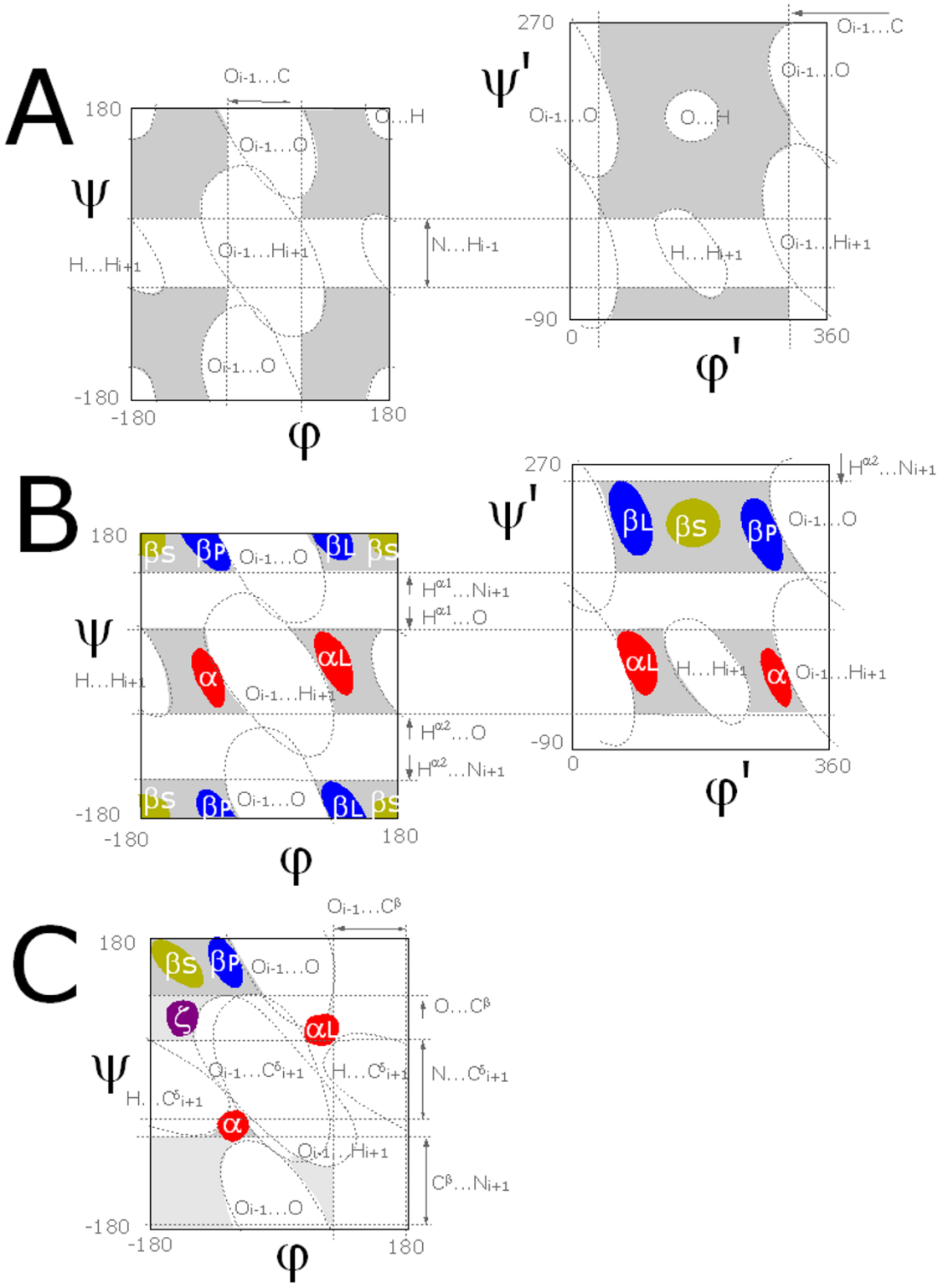

Figure 2

Schematic of the Ramachandran plot. (A) original steric map of glycine, in standard (left) and shifted (right) coordinates; (B) revised schematic of glycine, in standard (left) and shifted (right) coordinates; (C) pre-proline. The clustered regions are: grey – sterically allowed; red – α and α_L; yellow – β_S; blue – β_P and β_{PR}; purple – ζ. See text for explanation of the regions.

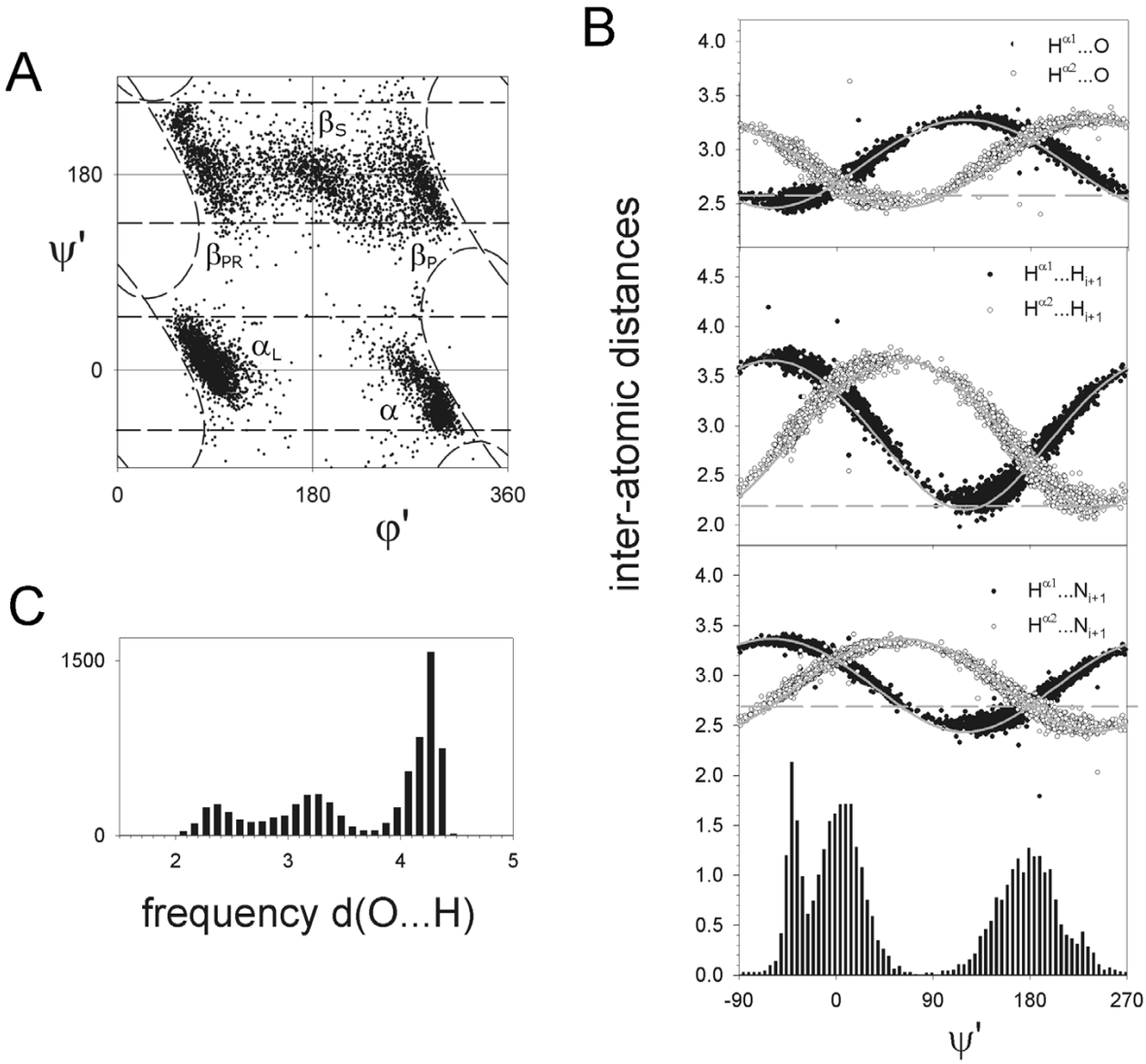

Figure 3

Glycine parameters. (A) The Ramachandran plot in shifted coordinates ϕ'-ψ'. The dashed lines show the steric clashes that define the boundaries of the observed densities (Figure 2B describes the specific interactions). (B) The distributions of various inter-atomic interactions as a function of ψ'. The dashed line show the limit of the VDW diameters. The grey line gives the model curve calculated with ideal geometry. At the bottom is the frequency distribution of the ψ' angle. (C) Frequency distribution of the inter-atomic distance d(O···H). There are 3 peaks, of which, the smallest at d(O···H) = 2.4 Å, which corresponds to the β_S region.

Here, we analyze the most distinctive feature of the glycine Ramachandran plot – the tendency for ψ' to cluster near 180° and 0°. We focus on the ψ'-dependent interactions. For each interaction, we first calculate the model curve of the corresponding inter-atomic distance as a function of ψ' (see Methods). We then compare the

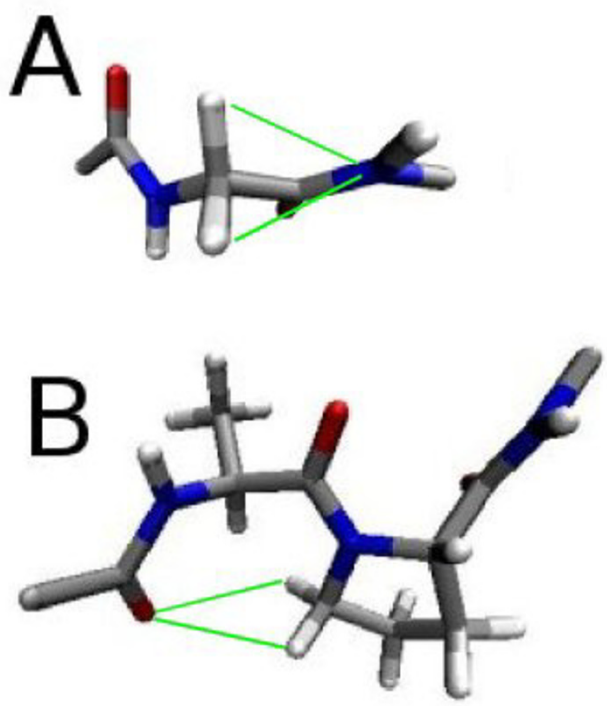

Figure 4
Stick figure representation of glycine and pre-proline. (A) glycine in the ψ ~ 180° conformation where the N_{i+1} atom is sandwiched between the two H^{α} atoms, and (B) pre-proline in the ζ conformation where the O_{i-1} atom interacts with the H^{δ} atoms of the succeeding proline.

observed ψ' distribution (bottom of Figure 3B) to the curve. If a hard-sphere repulsion restricts ψ', then, in regions of ψ' where the model curve is below the van der Waals (VDW) diameter (horizontal dashed line in Figure 3B), the ψ' frequency distribution should drop correspondingly.

In the region (60° < ψ' < 100°), we find that the drop-off in the ψ frequency distribution (bottom of Figure 3B) corresponds to values of $H^{\alpha 1} \cdots N_{i+1}$ (bottom of Figure 3B) and $H^{\alpha 2} \cdots O$ (top of Figure 3B) that are smaller than their VDW diameters. In the region (-90° < ψ' < -60; 210° < ψ' < 270°), the drop-off in the ψ frequency distribution corresponds to regions where $H^{\alpha 2} \cdots N_{i+1}$ and $H^{\alpha 1} \cdots O$ are found below their VDW radii. In contrast, the values of $H^{\alpha 1} \cdots H_{i+1}$ and $H^{\alpha 2} \cdots H_{i+1}$ are never found significantly below their VDW diameter (middle of Figure 3B).

The observed ψ' dependence in glycine is due to the $H^{\alpha 1} \cdots O$, $H^{\alpha 2} \cdots O$, $H^{\alpha 1} \cdots N_{i+1}$ and $H^{\alpha 2} \cdots N_{i+1}$ steric clashes. A simple interpretation is that the ψ' dependence in glycine arise from conformations that place either the N_{i+1} or O atom between the two H^{α} atoms (Figure 4A). The observed limits in the distributions have been drawn in Figure 3A as horizontal lines.

We thus obtain a revised steric map of glycine, consisting of the steric clashes $O_{i-1} \cdots O$, $O_{i-1} \cdots N_{i+1}$, $H^{\alpha 1} \cdots O$, $H^{\alpha 2} \cdots O$, $H^{\alpha 1} \cdots N_{i+1}$ and $H^{\alpha 2} \cdots N_{i+1}$. Using parameters from CHARMM22 [22], we calculate the Lennard-Jones 12-6 potential due to the revised steric clashes (Figure 5A). The minimum-energy region accounts for much of the shape of the observed distribution (Figure 3A).

Dipole-dipole interactions in glycine

The revised glycine steric map does not explain the diagonal shape of the α, α_L, β_P, β_{PR} and β_S regions. In the generic Ramachandran plot, it was found that the diagonal shape of regions could be reproduced using electrostatic dipole-dipole interactions [3] but only when the dipole-dipole interactions were considered individually. The overall electrostatic interaction does not reproduce the observed Ramachandran plot [23]. Here, we use the same approach of treating individual electrostatic dipole-dipole interactions along the glycine backbone.

We calculate the energy map of ϕ-ψ for the 4 dipole-dipole interactions in the glycine backbone interaction: $CO_{i-1} \cdots CO$, $NH \cdots NH_{i+1}$, $CO \cdots NH$ and $CO_{i-1} \cdots NH_{i+1}$ (Figure 5C-F). The electrostatic interactions are calculated with the Lennard-Jones potentials of the steric clashes identified in the section above. We find that the shapes of the different regions of the glycine Ramachandran plot (Figure 3A) are reproduced (Figure 5). The $CO \cdots NH$ interaction produces the diagonal α_L, α and β_S region (Figure 5E). The $NH \cdots NH_{i+1}$ interaction also produces a diagonal α_L and α region (Figure 5D). The α region is symmetric to the α_L region. The $CO_{i-1} \cdots CO$ interaction produces minima corresponding to the β_P and β_{PR} regions (Figure 5C).

In the original glycine steric map (Figure 2A), the region near (ϕ, ψ) = (-180°, 180°) is forbidden due to a steric clash between O and H. Yet glycine has density in this region in the observed Ramachandran plot (Figure 3A). This can also be seen in the frequency distribution of $d(O \cdots H)$ (Figure 3C), where there is a peak at $d(O \cdots H) \sim 2.4$ Å. At this peak, the O and H atoms are in contact, as the VDW diameter is 2.5 Å. Thus, in the β_S region of glycine, the favorable $CO \cdots HN$ dipole-dipole interaction overcomes the steric repulsion of the O and H atoms (Figure 5E).

The pre-proline Ramachandran plot

Schimmel and Flory argued in 1968 that pre-proline – amino acids preceding proline – has a particularly

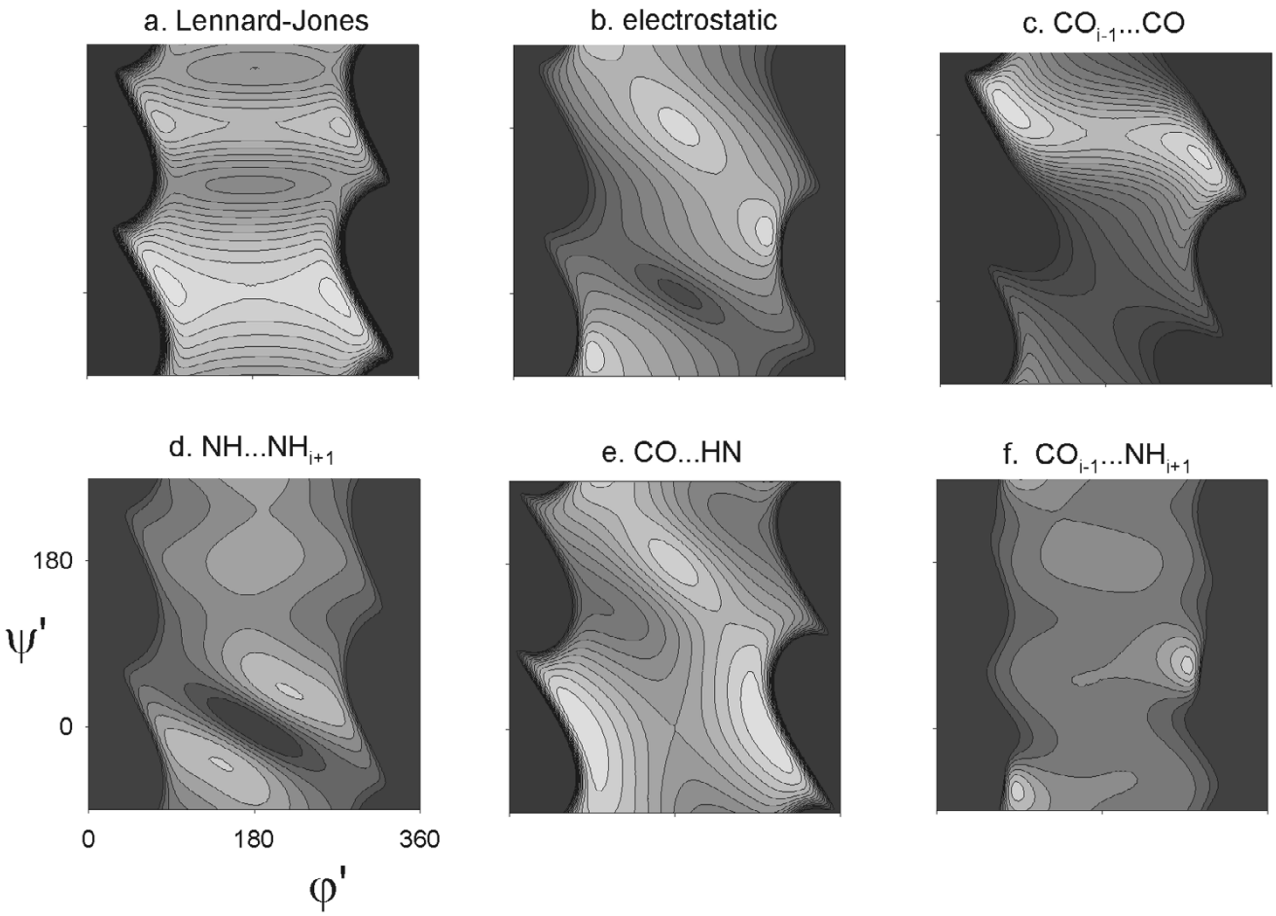

Figure 5
Dipole-dipole interactions in glycine. Axes are shown in the shifted φ'-ψ' angles [°]. Energy plots [kcal/mol] of (a) the Lennard-Jones 12-6 potentials of the revised set of steric clashes; (b) all electrostatic interactions; (c)-(f) the individual dipole-dipole interactions of the glycine backbone (see Figure 1A for backbone schematic of the dipoles). Energy parameters were taken from CHARMM22. The light areas show regions of minimum energy.

restricted Ramchandran plot, compared to the generic Ramachandran plot [17]. This was finally observed in the protein database by MacArthur and Thornton (Figure 1B) [2].

There are three main differences between the pre-proline Ramachandran plot and the generic Ramachandran plot. In the pre-proline Ramachandran plot, there is a large excluded horizontal strip at $-40° < \psi < 50°$, which restricts α_L and α regions. The α_L region is shifted up higher. These two features were reproduced in the Schimmel-Flory calculation [17] and subsequent calculations [13,18]. The third feature is a little leg of density poking out below the β-region (Figure 1B; purple in Figure 2C). Karplus called this the ζ region [10], which is unique to pre-proline.

Previous calculations [2,17,18] did not focus on the individual interactions, and did not account for the ζ region. Here, we identify the exact steric clashes that determine the pre-proline Ramachandran plot. We will then analyse the interactions responsible for the ζ region.

Steric interactions in the pre-proline backbone

In pre-proline, instead of an interaction with the ^{N}H atom in the succeeding generic amino acid, the pre-proline interacts with a CH_2 group of the succeeding proline (Figure 1B). The CH_2 group exerts a much larger steric effect on the pre-proline Ramachandran plot. MacArthur and Thornton [2] suggested that the dominant effect is due to the $N \cdots C^{\delta}_{i+1}$ and $C^{\beta} \cdots C^{\delta}_{i+1}$ steric clashes. Here we can

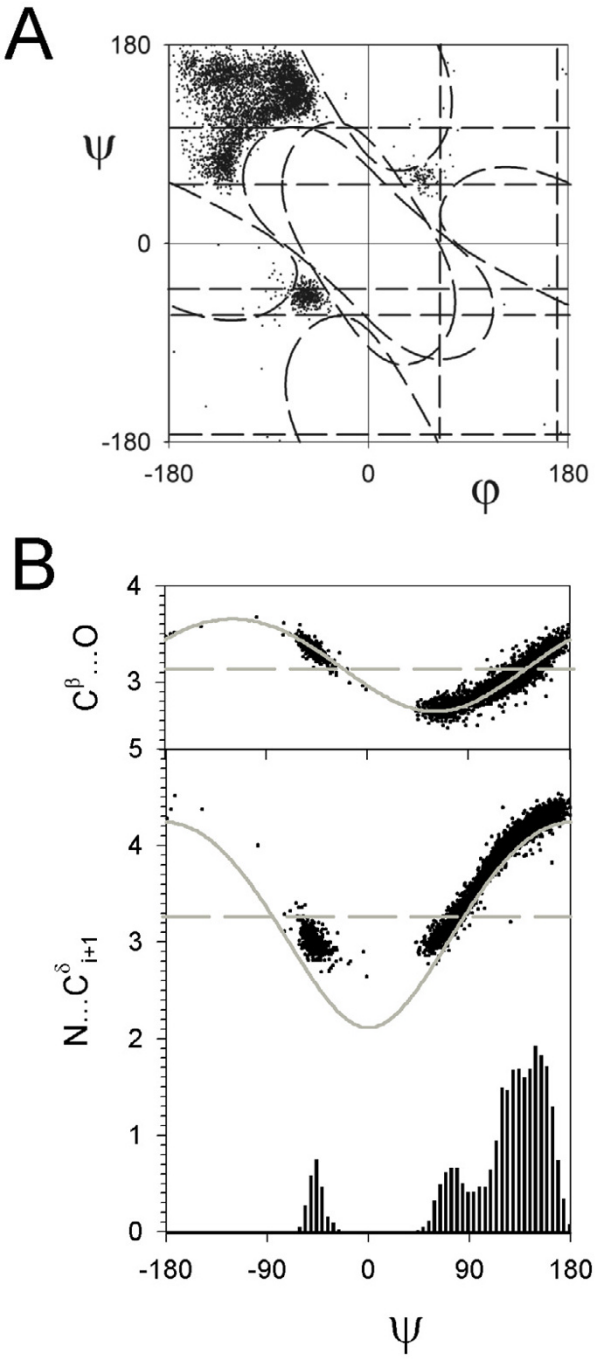

Figure 6

Pre-proline parameters. (A) The Ramachandran plot. The dashed lines show the steric clashes that define some of the boundaries of the observed densities (see Figure 2C). (B) The distributions of various inter-atomic interactions as a function of ψ. The dashed lines show the limit of the VDW diameters. The solid grey line gives the model curve calculated with ideal geometry. At the bottom is the frequency distribution of the ψ angle.

analyse the efficacy of each clash by analysing the statistical distributions directly.

We consider the ϕ-ψ co-dependent interactions that involve the C^{δ}, $H^{\delta 1}$ and $H^{\delta 2}$ atoms of the succeeding proline (Figure 1B). For each interaction, we generate the contour plot in ϕ-ψ of the VDW diameter distance. By comparing the contour plot to the observed density in the pre-proline Ramachandran plot, we identify the interactions that induce the best match in the boundaries (Figure 6A, the interactions are identified in Figure 2C). We found that the chunk taken out of the bottom-left β-region of the observed density is due to the $O_{i-1} \cdots C^{\delta}_{i+1}$ steric clash. Another restriction on the α_L and α regions is due to the $H \cdots C^{\delta}_{i+1}$ steric clash.

We next consider the ψ dependent interactions. In the pre-proline ψ frequency distribution, we found three distinct peaks (bottom Figure 6B). The left-most peak at ψ ~ -50° corresponds to the α region of pre-proline. We focus on the two peaks in the β-region 50° < ψ < 180° The larger peak centred on ψ ~ 150° corresponds to the β_S region of the generic Ramachandran plot. In the generic Ramachandran plot, this β_S region is bounded by the $C^{\beta} \cdots O$ and $C^{\beta} \cdots N_{i+1}$ steric clashes. In pre-proline, the smaller peak centred on ψ ~ 70° corresponds to the ζ region and occurs in a region that would be excluded by the $C^{\beta} \cdots O$ steric clash. Instead the smaller peak is bounded from below by the $N \cdots C^{\delta}_{i+1}$ steric clash. This can be seen by comparing the ψ distribution to the model curve of $N \cdots C^{\delta}_{i+1}$ vs. ψ (middle of Figure 6B).

Using parameters from CHARMM22, we calculate the Lennard-Jones 12-6 potential due to the revised steric clashes (Figure 7A). Lennard-Jones potentials cannot account for the ζ region.

Interactions that stabilize the pre-proline ζ region

As the ζ region (purple in Figure 2B) brings the $C^{\beta} \cdots O$ interaction into steric conflict, there must be a compensating interaction that stabilizes the ζ region. What is this interaction? To understand this interaction, we consider an analogy with the γ region in the generic Ramachandran plot. In the γ region, a distorted $CO_{i-1} \cdots HN_{i+1}$ hydrogen bond is formed, which brings the H_{i+1} atom into contact with the O_{i-1} atom. Similarly, in the ζ region of pre-proline, the O_{i-1} atom of pre-proline is in contact with the $H^{\delta 1}$ and $H^{\delta 2}$ atoms (see Figure 4B; Table 1), suggesting that the CO_{i-1} group interacts with the $C^{\delta}H^{\delta}_{i+1}$ group of the succeeding proline.

Can the $C^{\delta} H^{\delta}_{i+1}$ group interact with CO_{i-1}? Such an interaction would fall under the class of the $CH \cdots O$ weak hydrogen bond, a well-documented interaction in proteins [24]. Studies of the $CH \cdots O$ weak hydrogen bond

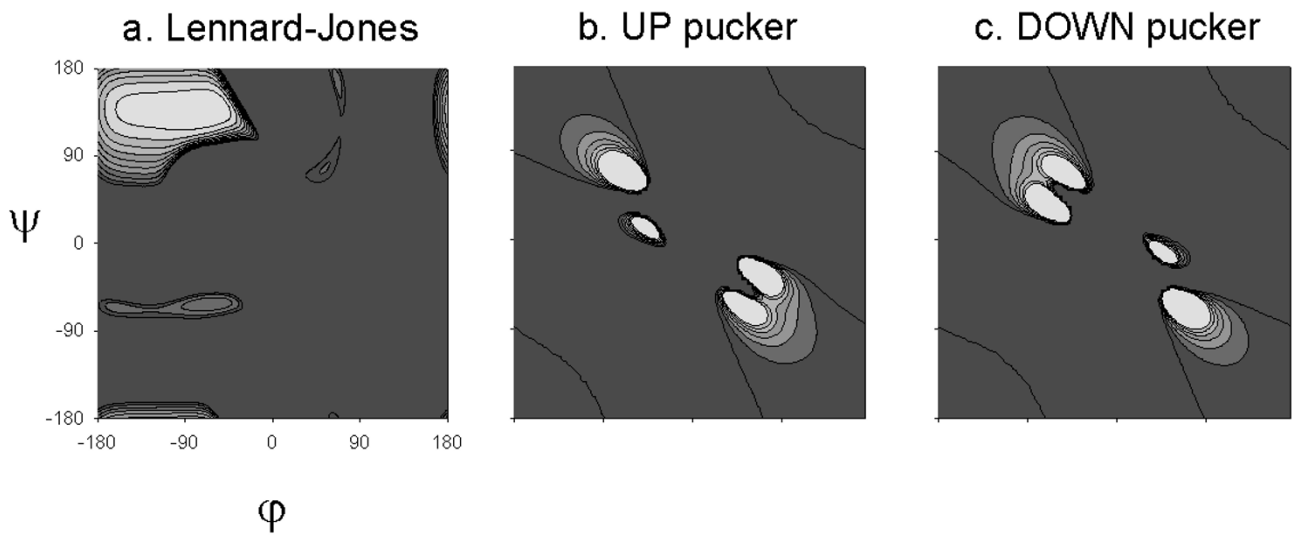

Figure 7
Energy plots in pre-proline as a function of ϕ-ψ. Energy plots [kcal/mol] of (a) the Lennard-Jones 12-6 potentials of the revised set of steric clashes; the $CO_{i-1}\cdots C^{\delta}H^{\delta}_{i+1}$ dipole-dipole interactions when the succeeding proline ring is in (b) the UP pucker and (c) the DOWN pucker. The light areas show regions of low energy.

Table 1: Parameters of the CO···HX hydrogen bond

	n	ϕ	ψ	O···H	∠COH	∠OHX	$\angle_{dih}$COHX
	γ region of the generic amino acid						
CO···HN	518	-85(3)°	81(11)°	2.39(0.24) Å	107(5)°	123(10)°	178(11)°
	ζ region of pre-proline: UP PUCKER						
$CO\cdots H^{\delta 1}C^{\delta}$	105	-129(6)°	80(6)°	3.64(0.27) Å	79(8)°	75(10)°	25(9)°
$CO\cdots H^{\delta 2}C^{\delta}$				2.59(0.22) Å	99(12)°	144(14)°	-108(18)°
	ζ region of pre-proline: DOWN PUCKER						
$CO\cdots H^{\delta 1}C^{\delta}$	406	-129(6)°	74(9)°	3.16(0.38) Å	95(32)°	101(20)°	22(11)°
$CO\cdots H^{\delta 2}C^{\delta}$				2.98(0.36) Å	71(10)°	113(22)°	-118(14)°

use a distance criteria of $d(H\cdots O) < 2.8$ Å [25-27]. There is little angular dependence found in the $CH\cdots O$ bond around the H atom where an angle criteria of $\angle OHX > 90°$ is often used. This is much more permissive than the geometry of the canonical hydrogen bond. In Table 1, we list the hydrogen bond parameters of the $CO_{i-1}\cdots C^{\delta}H^{\delta}_{i+1}$ interaction in the ζ region. As proline can take on two different major conformations, the UP and DOWN pucker, measurements of the geometry of the $CO_{i-1}\cdots C^{\delta}H^{\delta}_{i+1}$ interaction must also be divided in terms of the UP and DOWN pucker. The observed geometry of the $CO_{i-1}\cdots C^{\delta}H^{\delta}_{i+1}$ geometry satisfies the geometric criteria of the weak hydrogen bond (Table 1).

As the $CO_{i-1}\cdots C^{\delta}H^{\delta}_{i+1}$ weak hydrogen bond is a close contact, we need to model the interaction in order to understand its dependence on the ϕ-ψ angles. For the modelling, we consider strategies that have been used for the analogous $CO_{i-1}\cdots HN_{i+1}$ hydrogen bond. The $CO_{i-1}\cdots HN_{i+1}$ hydrogen bond has been modelled in quantum-mechanical studies where the γ region was found to be the minimum energy conformation in vacuum [12]. A

simpler approach, which modelled the hydrogen bond with electrostatic dipole-dipole interactions, also find a minimum in the γ region [23].

Here, we model the $CO_{i-1} \cdots C^{\delta}H^{\delta}_{i+1}$ weak hydrogen bond as an electrostatic dipole-dipole interaction (see Methods). How do we model the $C^{\delta}H^{\delta}_{i+1}$ group as an electrostatic dipole? Bhattacharyya and Chakrabarti [28] found that, of the CH groups in proline, the $C^{\delta}H^{\delta}$ group forms the most $CH \cdots O$ hydrogen bonds. The C^{δ} atom sits next to the electron-withdrawing N atom and thus, is more acidic than the other C atoms. Consequently, we place a small negative partial charge on the C^{δ} atom. In our model, we find an energy minimum in the ζ region for both the UP pucker (Figure 7B) and the DOWN pucker (Figure 7C). We conclude that the $CO_{i-1} \cdots C^{\delta}_{i+1}H^{\delta 1}_{i+1}$ weak hydrogen bond stabilizes the ζ region in pre-proline.

Conclusion

We have identified the interactions that determine the high-resolution Ramachandran plots of glycine and pre-proline.

For glycine, the Ramachandran plot of the glycine backbone modeled by standard force-fields fails to reproduce the observed Ramachandran plot [16]. Instead the modeled Ramachandran plot resembles the original steric map of glycine [1]. The failure of these calculations arises from the inadequate treatment of the H^{α} atoms. We have identified a revised set of steric interactions that can reproduce the observed glycine Ramachandran plot. These are $O_{i-1} \cdots O$, $O_{i-1} \cdots N_{i+1}$, $H^{\alpha 1} \cdots O$, $H^{\alpha 2} \cdots O$, $H^{\alpha 1} \cdots N_{i+1}$ and $H^{\alpha 2} \cdots N_{i+1}$ (Figure 2B). These steric interactions constrain either the N_{i+1} or O atom to be sandwiched between the two H^{α} atoms, which clusters glycine to $\psi = 180°$ and $\psi = 0°$. The five clustered regions can be traced to electrostatic dipole-dipole interactions: the $CO \cdots NH$ interaction induces diagonal α_L, α and β_S regions; and the $CO_{i-1} \cdots CO$ interaction induces the diagonal β_P and β_{PR} regions.

Previous calculations of the pre-proline Ramachandran reproduced most of the observed pre-proline Ramachandran plot with the notable exception of the ζ region. Previous studies did not identify the specific steric interactions involved in defining the pre-proline Ramachandran plot. Here, we have identified them: $N \cdots C^{\delta}_{i+1}$, $O_{i-1} \cdots C^{\delta}_{i+1}$ and $H \cdots C^{\delta}_{i+1}$ (Figure 2C). We have also identified the physical mechanism that stabilizes the ζ region (purple in Figure 2C). It is the $CO_{i-1} \cdots C^{\delta}H^{\delta}_{i+1}$ weak hydrogen bond, which is directly analogous to the $CO_{i-1} \cdots NH_{i+1}$ hydrogen bond that stabilizes γ-turns in the generic amino acid.

Combined with the analysis of the generic Ramachandran plot [3] and the proline Ramachandran plot [13,14], we have identified the interactions that define the high-resolution Ramachandran plots of all 20 amino acids. Although our analysis uses simple modeling techniques, the interactions identified here suggest concrete ways to resolve the inadequacies in current force-fields.

Methods

VDW radii

In the steric clash analysis, we used the VDW radii given by the Richardson lab [29]: H^{α} = 1.17Å, H = 1.00Å, C = 1.65Å, $C^{\alpha} = C^{\beta}$ = 1.75Å, O = 1.40Å and N = 1.55Å. From the database, we extracted 7277 glycine and 4336 pre-proline residues.

Local conformations of the φ-ψ map

To calculate the model curves of the inter-atomic distances as a function of the φ-ψ angles, we modeled the glycine and pre-proline protein fragments shown in Figure 1. Covalent bond lengths and angles were fixed to CHARMM22 values [22]. Only the φ-ψ angles vary. The φ-ψ angles of the central residue were incremented in 5° steps and the corresponding distance parameters and energies of the inter-atomic interactions were calculated. We used 2 types of interactions, partial charge electrostatics, $E_{elec} = 331 \cdot (q_1 \cdot q_2)$ kcal · mol^{-1}, and Lennard-Jones 12-6 potentials, $E_{LJ} = \varepsilon\ (\sigma/d)^{12} - 2\ (\sigma/d)^{6})$ kcal · mol^{-1}, where the parameters were taken from CHARMM22 [22]. There are no parameters in CHARMM22 for the H^{δ} and C^{δ} atoms. As such, we have assigned a partial charge of -0.20 to C^{δ} and 0.10 to $H^{\delta 1}$ and $H^{\delta 2}$. These are not based on any detailed arguments but are merely used to estimate the effect that such charges would have.

Authors' contributions

Both authors conceived the study. BKH carried out the data analysis and modelling, and drafted the manuscript. RB provided guidance and mentorship.

Acknowledgements

BKH was supported by a post-doctoral grant from the Fonds National de la Recherche Scientifique (FNRS), Belgium. RB is director of research of FNRS.

References

1. Ramachandran GN, Ramakrishnan C, Sasisekharan V: **Stereochemistry of polypeptide chain configurations.** *J Mol Biol* 1963, **7:**95-99.
2. MacArthur MW, Thornton JM: **Influence of proline residues on protein conformation.** *Journal of Molecular Biology* 1991, **218:**397-412.
3. Ho BK, Thomas A, Brasseur R: **Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix.** *Protein Sci* 2003, **12:**2508-2522.
4. Mandel N, Mandel G, Trus BL, Rosenberg J, Carlson G, Dickerson RE: **Tuna cytochrome c at 2.0 A resolution. III. Coordinate optimization and comparison of structures.** *J Biol Chem* 1977, **252:**4619-4636.

5. Richardson JS: **The anatomy and taxonomy of protein structure.** *Adv Protein Chem* 1981, **34:**167-339.
6. Kleywegt GJ, Jones TA: **Phi/psi-chology: Ramachandran revisited.** *Structure* 1996, **4:**1395-1400.
7. Chakrabarti P, Pal D: **The interrelationships of side-chain and main-chain conformations in proteins.** *Prog Biophys Mol Biol* 2001, **76:**1-102.
8. Hovmöller S, Zhou T, Ohlson T: **Conformations of amino acids in proteins.** *Acta Crystallogr D Biol Crystallogr* 2002, **58:**768-776.
9. Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC: **Structure validation by Calpha geometry: phi,psi and Cbeta deviation.** *Proteins* 2003, **50:**437-450.
10. Karplus PA: **Experimentally observed conformation-dependent geometry and hidden strain in proteins.** *Protein Sci* 1996, **5:**1406-1420.
11. Maccallum PH, Poet R, Milner-White EJ: **Coulombic interactions between partially charged main-chain atoms not hydrogen-bonded to each other influence the conformations of alpha-helices and antiparallel beta-sheet. A new method for analysing the forces between hydrogen bonding groups in proteins includes all the Coulombic interactions.** *J Mol Biol* 1995, **248:**361-373.
12. Milner-White EJ: **Situations of gamma-turns in proteins. Their relation to alpha-helices, beta-sheets and ligand binding sites.** *Journal of Molecular Biology* 1990, **216:**386-397.
13. Summers LN, M. K: **Modeling of globular proteins. A distance-based data search procedure for the construction of insertion/deletion regions and Pro-non-Pro mutations.** *Journal of Molecular Biology* 1990, **216:**991-1016.
14. Ho BK, Coutsias EA, Seok C, Dill KA: **The flexibility in the proline ring couples to the protein backbone.** *Protein Sci* 2005, **14:**1011-1018.
15. Ramakrishnan C, Ramachandran GN: **Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units.** *Biophys J* 1965, **5:**909-933.
16. Hu H, Elstner M, Hermans J: **Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution.** *Proteins* 2003, **50:**451-463.
17. Schimmel PR, Flory PJ: **Conformational energies and configurational statistics of copolypeptides containing L-proline.** *J Mol Biol* 1968, **34:**105-120.
18. Hurley JH, Mason DA, Matthews BW: **Flexible-geometry conformational energy maps for the amino acid residue preceding a proline.** *Biopolymers* 1992, **32:**1443-1446.
19. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Research* 2000, **28:**235-242.
20. Stapley BJ, Creamer TP: **A survey of left-handed polyproline II helices.** *Protein Sci* 1999, **8:**587-595.
21. Eswar N, Nagarajaram HA, Ramakrishnan C, Srinivasan N: **Influence of solvent molecules on the stereochemical code of glycyl residues in proteins.** *Proteins* 2002, **49:**326-334.
22. Mackerell ADJ, Bashford D, Bellott M, Dunbrack Jr. RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir LKK, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WEIII, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M: **All-atom empirical potential for molecular modeling and dynamics Studies of proteins.** *Journal of Physical Chemistry* 1998, **B102:**3586-3616.
23. Roterman IK, Lambert MH, Gibson KD, Scheraga HA: **A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. II. Phi-psi maps for N-acetyl alanine N'-methyl amide: comparisons, contrasts and simple experimental tests.** *J Biomol Struct Dyn* 1989, **7:**421-453.
24. Derewenda ZS, Lee L, Derewenda U: **The occurrence of C-H...O hydrogen bonds in proteins.** *J Mol Biol* 1995, **252:**248-262.
25. Taylor R, Kennard O: **Crystallographic evidence for the existence of C-H...O, C-H...N, and C-H...Cl hydrogen bonds.** *J Am Chem Soc* 1982, **104:**5063-5070.
26. Steiner T, Saenger W: **Role of C-H...O hydrogen bonds in the coordination of water molecules. Analysis of Neutron Diffraction Data.** *J Am Chem Soc* 1993, **115:**4540-4547.
27. Vargas R, Garza J, Dixon DA, Hay BP: **How strong is the C{alpha}-H...O=C hydrogen bond?** *J Am Chem Soc* 2000, **122:**4750-4755.
28. Bhattacharyya R, Chakrabarti P: **Stereospecific interactions of proline residues in protein structures and complexes.** *J Mole Biol* 2003, **331(4):**925-940.
29. Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC: **Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms.** *J Mol Biol* 1999, **285:**1711-1733.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral