Determination of amino acids that favour the α_L region using Ramachandran propensity plots. Implications for α-sheet as the possible amyloid intermediate

Steven Hayward[1*] and E. James Milner-White[2*]

[1] Computational Biology Laboratory, School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

[2] College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK

***Contact:** Dr Steven Hayward, School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK. Tel: +44-1603-593542, e-mail: Steven.Hayward@uea.ac.uk

***Contact:** Prof E. James Milner-White, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK. Tel: +44-1413305283, e-mail: James.Milner-White@glasgow.ac.uk

Abstract

In amyloid diseases an insoluble amyloid fibril forms via a soluble oligomeric intermediate. It is this intermediate that mediates toxicity and it has been suggested, somewhat controversially, that it has the α-sheet structure. Nests and α-strands are similar peptide motifs in that alternate residues lie in the α_R and γ_L regions of the Ramachandran plot for nests, or α_R and α_L regions for α-strands. In nests a concavity is formed by the main chain NH atoms whereas in α-strands the main chain is almost straight. Using "Ramachandran propensity plots" to focus on the α_L/γ_L region, it is shown that glycine favours γ_L (82% of amino acids are glycine), but disfavours α_L (3% are glycine). Most charged and polar amino acids favour α_L with asparagine having by far the highest propensity. Thus, glycine favours nests but, contrary to common expectation, should not favour α-sheet. By contrast most charged or polar amino acids should favour α-sheet by their propensity for the α_L conformation, which is more discriminating amongst amino acids than the α_R conformation. Thus, these results suggest the composition of sequences that favour α-sheet formation and point towards effective prediction of α-sheet from sequence.

Keywords: Ramachandran propensity plot; nest; repeating dipeptide conformation; peptide plane flip, amyloid

Introduction

The aggregation of proteins into a fibrillar conformation composed largely of β-sheet termed amyloid is the cause of many different diseases including Alzheimer's Disease, Parkinson's Disease, type II diabetes and Huntington's Disease (Chiti and Dobson, 2017; Eisenberg and Jucker, 2012; Erskine et al., 2018; Lee et al., 2017). It is thought that self-association into the amyloid structure is an inherent property that all polypeptide chains share (Fandrich et al., 2001). Because many amyloid fibres do not exist as multiple structurally identical molecules and are often in the form of a non-crystallizable sticky mass, they are far more difficult to study by current biophysical techniques than folded proteins. Nevertheless, a number of β-sheet containing molecular structures of diverse amyloid proteins have been determined in atomic detail. We point out that these structures are those of mature amyloid, which are distinct from those of amyloid precursors or intermediates.

Amyloids of different systems have been analysed showing that their structure is less dependent on side chain composition than the folding of proteins to their native state. The toxic structure in amyloid diseases is not the mature amyloid fibril but a soluble intermediate. This was shown via the generation of antibodies that bind only the soluble intermediates, not the mature amyloid fibre or the soluble precursor proteins (Kayed et al., 2003). The antibodies bind all intermediate polypeptides regardless of sequence, and their toxicity is thereby greatly reduced. It has been suggested this is due to the antibodies recognizing the backbone conformation of α-sheet (Arai et al., 2012; Daggett, 2006). The inter-mainchain hydrogen bonding of α-sheet exhibits similarity, as shown in Fig 1, to β-sheet, but α-sheet has different electrostatic properties because the two edges of the sheet have a pronounced polarity (free NHs along one side, free COs on the other), which is absent in β-sheet.

Armen *et al* (Armen et al., 2004b) first proposed α-sheet as the toxic intermediate, formed by the amyloidogenic regions of proteins. The likely existence of α-sheet (polar-pleated sheet) as a protein conformation was originally suggested by Pauling and Corey (Pauling and Corey, 1951), along with α-helix and β-sheet. As it is rare in natural proteins it has tended to be overlooked. Ideas about amyloid, however, rekindled interest in it (Daggett, 2006). In recent years accelerating numbers of papers have been providing further evidence in support of α-sheet as the toxic amyloid precursor(Xu, 2007; Grillo-Bosch et al., 2009; Babin et al., 2011; Hopping et al., 2014; Kellock et al., 2016; Hilaire et al., 2018; Maris et al., 2018; Bi and Daggett, 2018; Shea et al., 2019; Meng et al., 2019; Childers and Daggett, 2019; Balupuri et al., 2019; Childers and Daggett, 2020; Balupuri et al., 2020; Wang et al., 2020). Certain extracellular proteins, named functional amyloids (Deshmukh et al., 2018; Erskine et al., 2018), are also thought to occur preferentially, rather than being misfolded, as amyloid; evidence indicates that that they too pass through an α-sheet intermediate stage (Bleem et al., 2017; Paranjapye and Daggett, 2018).

Nests (Afzal et al., 2014; Watson and Milner-White, 2002a; Watson and Milner-White, 2002b) are common 3-8 residue protein motifs; 8% of residues in proteins belong to one. Nests and α-strands (some within α-sheet) resemble each other in that both consist of amino acids with alternating right-handed (negative ϕ) and left-handed (positive ϕ) main chain conformations; they differ in that the main chain atoms of nests are curved to various degrees while those of α-strands are straight or nearly so. In previous work peptides with $\alpha_R\alpha_L$ or $\alpha_L\alpha_R$ conformations have been included as nests, but in this work, they are referred to as α-strands. α-strands are less common in most proteins than nests but occur for example in potassium channels and aquaporins (Milner-White et al., 2006; Watson and Milner-White, 2002a) where they play a key role. Both nests and α-sheet orientate adjacent main-chain carbonyl groups onto one side of the backbone, and NH groups onto the other side, creating significant polarity. Nests often utilise this charge separation to bind negative or δ-atoms. The curvature of the motif allows it to 'cup' the group or atom it binds to, termed the 'egg' (Kim et al., 2018; Milner-White et al., 2006). With α-sheet the polarity assists the self-association of the nearly linear (Hayward and Milner-White, 2008) adjacent strands of each sheet (Armen et al., 2004b), as seen in Fig 1.

Analysis of steric hindrance in the Ramachandran plot (Ramachandran et al., 1963; Richardson and Richardson, 1989; Hovmoller et al., 2002; Pal and Chakrabarti, 2002; Ho et al., 2003; Beck et al., 2008; Berkholz et al., 2009; Hollingsworth and Karplus, 2010; Porter and Rose, 2011; Zhou et al., 2011; Carugo and Carugo, 2013; Richardson et al., 2013; Laskowski et al., 2013; Hayward et al., 2014; Balasco et al., 2019; Ravikumar et al., 2019) indicates two major commonly allowed regions: α_R, and β. Towards the bottom of the α_R region a discrete area is for residues in the α-helical conformation. A third main region is commonly occupied by L amino acids in proteins, not as often as the other two but frequently none the less; it is often collectively referred to as α_L; towards the top of the region is the area for residues in the left-handed α-helical conformation.

In studies of β-turn occupancy of the α_L region Wilmot and Thornton (Wilmot and Thornton, 1990) and Efimov (Efimov, 1993) distinguished two distinct subregions centred on (60°,30°) and (90°,0°); we denote these α_L and γ_L respectively. Nests have two successive residues in $\alpha_R\gamma_L$ (RL) or $\gamma_L\alpha_R$ (LR) conformations; RL nests make up 80% of such motifs in proteins (Watson and Milner-White, 2002b), while 20% are LR. Strands of α-sheet (α-strands), on the other hand, consist of alternating $\alpha_{R/}\alpha_L$ residues, that are, approximately, adjacent right-handed/left-handed α-helical conformations.

The α_L at residue i+1 conformation need not be exactly the main-chain enantiomer of the α_R conformation at residue i, such that ϕ_{i+1}=-ϕ_i and ψ_{i+1}=-ψ_i. An alternative is the “mirror” condition, where in the Ramachandran plot, the point for residue i+1 is a reflection in the diagonal line ψ =-ϕ of the point for residue i; the mirror relationship is: ϕ_{i+1}=-ψ_i and ψ_{i+1}=-ϕ_i. It has been shown (Hayward and Milner-White, 2008) that mirror peptides (where this mirror relationship repeats along the chain), are helices with large radii and no twist, as expected for α-sheet. They are also the structures arrived at following peptide plane flipping in β-sheet as described below. We shall name the α_L conformation with a mirror relationship to α_R as α_L^m and the enantiomeric α_L conformation as α_L^e.

Interconversion of β-strands to α-strands has been shown to occur relatively easily via peptide plane flipping (Davis et al., 2006; Hayward, 2001; Keedy et al., 2015;

Milner-White et al., 2006; Yang et al., 2006), as suggested from molecular dynamics (MD) studies under denaturing conditions (Armen et al., 2004a; Daggett, 2006). In the straight-strand α-sheet model of Pauling and Corey (Pauling and Corey, 1951) (see Fig 1(A)) neighbouring residues do obey the enantiomer condition, $\phi_{i+1}=-\phi_i$ and $\psi_{i+1}=-\psi_i$. However, the average alternating ϕ,ψ angles derived from multiple MD simulations, $(-87°,-49°)_i$, $(45°,92°)_{i+1}$ reported by Daggett (Daggett, 2006) for α-sheet formed via peptide plane flipping from β-sheet, suggest the mirror condition. In fact, it was shown that the mirror condition is a natural consequence of a repeating dipeptide conformation arising from peptide plane rotations in β-strands with $\psi_i=-\phi_i$ (Hayward and Milner-White, 2008), and α-sheet is an example of this whereby alternate peptide planes rotate 180˚ (flip) (Milner-White et al., 2006). Peptide plane flipping is favourable in folded proteins where both ends of the main-chain of the peptide are reasonably well anchored, as it requires only minor adjustments in the adjacent main chain and side chain conformations. The mirror condition implies that for α-sheet successive residues have the $\alpha_R\alpha_L^m$ conformation, not $\alpha_R\gamma_L$. Here we focus on the amino acid compositions of the α_L and γ_L regions.

Readers may suppose that the distinction between the α_L and γ_L regions has been investigated fully previously, but two interacting factors indicate why this is not so. Firstly, the number of examples is small relative to those in the α_R or β regions so a large protein structure database, only available in recent years, is needed. Secondly, in terms of distribution on the Ramachandran plot, they overlap to give the appearance of one region, such that many authors, for the purpose of analysing amino acid content, have grouped them together.

We were influenced by the seminal work of Wilmot and Thornton (Wilmot and Thornton, 1990) and Hovmuller *et al.* (Hovmoller et al., 2002). Both groups drew attention to glycines and L-amino acids favouring distinct regions corresponding to the α_L and γ_L conformations in proteins. Having made the interesting observation, they did not pursue the topic further by analysing the amino acid compositions of the two regions. The reason would seem to be that, when the work was carried out, the number of protein three-dimensional structures available was insufficient to provide convincing statistics for comparing the two overlapping regions. A further point is that the $\alpha_{L/}\gamma_L$ conformation was regarded as a "turn" region then; nests were not thought of, and few considered α-sheet to be of any importance.

The aim of our paper is to fully investigate differences between types of amino acids occurring at α_L and γ_L conformations in the Protein Data Bank (PDB) and consider the implications for the likelihood of α-sheet formation.

Methods

X-ray crystallographic structure files of proteins from the entire PDB were selected at 30% sequence identity filtering for polypeptide chains and with a resolution of at most 2 Å. The selection was achieved using the advanced search tool at the Research Collaboratory for Structural Bioinformatics PDB (https://www.rcsb.org/) on 15 October 2019. For each resulting PDB file we selected only the first protein chain in order to eliminate repeating chains irrespective of whether the PDB file contained homooligomers or not. Chains with 10 or fewer amino acids were also removed. This selection process resulted in 14,008 chains. The PDB codes of the files used are listed in Supplementary data file (see Data S1).

Results are presented as frequencies, probabilities or propensities as described below. We constructed a 4°×4° grid on the Ramachandran plot and for all amino acids in the data set we evaluated their ϕ, ψ angles in order to count the frequency of occurrence, $N(X \cap G)$, of each amino acid, X, in each 4°×4° grid cell, G. The following conditional probabilities were evaluated:

$$p(X|G) = N(X \cap G)/\,N(G) \qquad (1a)$$

$$p(G|X) = N(X \cap G)/\,N(X) \qquad (1b)$$

where $N(G) = \sum_X N(X \cap G)$ and $N(X) = \sum_G N(X \cap G)$.

$p(X|G)$ is the probability of amino acid X, given the set ϕ,ψ angles in grid cell G, evaluated as the ratio of the number of occurrences of amino acids of type X in grid cell G to the number of all amino acids of any type in G. $p(G\,|\,X)$ is the probability of the set ϕ,ψ angles in grid cell G, given a particular amino acid X, evaluated as the ratio of the number of amino acids of type X in G to the total number of amino acids of type X in all grid cells. We also evaluated:

$$p(X) = N(X)/\sum_X N(X) \qquad (2a)$$

$$p(G) = N(G)/\sum_G N(G) \qquad (2b)$$

which are, respectively, the overall probability of amino acid X over the entire Ramachandran plot, and the overall probability of the set ϕ,ψ angles in grid cell G irrespective of amino acid type.

The “propensity” was first used by Shortle (Shortle, 2002) in this context, but their definition was the ratio of two probabilities whereas we take the log (base 10) of this ratio to give a propensity that is a log likelihood ratio:

$$Prop(X,G) = log(p(G|X)/p(G)) = log(p(X|G)/\ p(X)) \qquad (3)$$

Shortle describes this ratio as “*a relative measure of preference, and thus is always normalized to an average or mean residue type. A probability, on the other hand, is a measure of the absolute likelihood that an amino acid will adopt one out of a specified set of structures.*” For a particular region, propensity quantifies the probability of occurrence of the amino acid in that region, relative to the probability of occurrence of the amino acid overall. $Prop(X,G) = 0$ means amino acid X behaves like the “average amino acid” for cell G (mathematically: $Prop(X,G) = 0 \Rightarrow p(G|X) = p(G) = \sum_{X'} p(G|X')p(X')$; the right-hand-side is the weighted *average* of the probabilities of each amino acid for cell G, the weighting factor being the overall probability of occurrence of the amino acid); $Prop(X,G) > 0$ means that, compared to the average amino acid, X favours cell G; $Prop(X,G) < 0$ means that compared to the average amino acid, X disfavours cell G. A $Prop(X,G) = 1$ means that it is 10 times more frequent in the region G than the average amino acid and $Prop(X,G) = -1$ that it is 10 times less frequent in the region G than the average amino acid. Shortle also describes this measure (the log of their propensity) as the free energy cost of replacing the average amino acid with the specific one, X.

Ramachandran plots are presented colour-coded according to the numerical value of the probability or propensity. For the Ramachandran propensity plots we accumulate the data over the 4°×4° grid cells to calculate the propensity in 36°×36° windows centred on each grid cell. The propensity calculated in each window is colour-coded according to value and the grid cell at the centre of the window is filled with this colour. As this is performed at each grid-cell the overlapping of the windows smooths the results.

The web facility Motivated Proteins ((Leader and Milner-White, 2009) and the desktop application Structure Motivator (Leader and Milner-White, 2012) were employed to analyse small protein motifs in the PDB, notably the ϕ,ψ angles. The dihedral angle data derives originally from DSSP files (Kabsch and Sander, 1983); the hydrogen bonds defining motifs are from HBplus (McDonald and Thornton, 1994). The images of the structural models in Figs 1 and 6 were produced using PyMol (www.pymol.org).

Results & Discussion

A filtered set of high-resolution structures from the entire PDB was used to calculate the ϕ,ψ angles of each amino acid type. The results for alanine and glycine are plotted as frequency distributions in the Ramachandran plots in Figs 2(A) and 2(B).

Plots like this were produced by Hovmuller *et al* (Hovmoller et al., 2002) for all 20 amino acids.

Fig 3 shows Ramachandran propensity plots for individual amino acids. Propensity, defined in the Methods Section, is a measure of the amino acid’s relative, rather than absolute, frequency, compared to that of all 20 amino acids. The logarithm of this ratio is used. Results for four key amino acids including glycine are presented; those for all 20 amino acids are in Fig S1 and interested readers should find them illuminating. Although a related measure was used by Shortle (Shortle, 2002), we believe this is the first time that a propensity plot for each amino acid has been presented in this way.

A propensity of zero means it neither disfavours or favours the region and behaves like the “average amino acid”; a positive propensity means that, compared to the average amino acid, it is common in the region; a negative propensity means that, compared to the average amino acid, it is rare in the region (see Methods section for fuller interpretations of this quantity). Colour coding is employed to illustrate this. Note the black colouring for the extensive areas in Fig 3(B) where glycine is strongly preferred. When viewing such plots, readers should bear in mind that, as seen in Fig 2, some areas of the plot are much more densely populated than others and propensity does not show this.

A major aim of our work is to compare the α_L and γ_L regions of the plot. Though not the highest areas of population density, these areas are none the less well populated. The distributions in the Ramachandran plot for alanine and glycine in Fig 2 reveal a somewhat linear shape, diagonally orientated, for the α_L and γ_L regions. To analyse this region, we defined points α_L^m, α_L^e and γ_L, at ϕ,ψ positions (42˚,62˚), (62˚,42˚) and (78˚,2˚), shown as red, green and cyan spots in Fig 2(A), respectively. They are approximately co-linear and lie within the diagonal populated area. That for α_L^m is at the “mirror” position in relation to α_R at (-62˚,-42˚), the average angles for α-helix (Hovmoller et al., 2002), while α_L^e is the main-chain enantiomer of the α-helix. The point for $\gamma_{L.}$ although slightly shifted from the (90˚,0˚) given by Wilmot and Thornton (Wilmot and Thornton, 1990), lies on the same diagonal linear area as α_L and coincides with “additionally allowed” PROCHECK (Laskowski et al., 1993) regions of that part of the Ramachandran plot.

For the analysis of amino acid occurrences, a sliding 36°×36° window was moved along the ”sampling line” defined by α_L^m and γ_L and passing close to α_L^e, as shown in Fig 2(A). Fig 4(A) gives the frequency of all amino acids regardless of type at each window position and the equivalent frequencies for glycine and non-glycine, showing that the overall single peak can be decomposed into two distributions each comprising a single peak, one for glycine and the other for non-glycine. Fig 4(B) gives the probability of individual amino acids at each window position indicating glycine and non-glycine. The γ_L and α_L^e points lie essentially within the broad peak of Fig 4(A), while α_L^m is well to the side of the peak. At γ_L the probability of glycine is high (82%), and the probability of non-glycine amino acids is low (18%). At α_L^m, non-glycine amino acids predominate (97%) and the proportion of glycines is very low (3%), whereas at α_L^e the non-glycine/glycine amounts are 64%/36%.

The distributions of glycine and non-glycine in Fig 4 might help define the α_L and γ_L regions, although the degree of correspondence is not known. Fig 4 suggests a natural boundary between the two distributions at ϕ=65° where the probabilities for glycine and non-glycine are both 0.5. The α_L^m point is clearly in the α_L region and our selected point for γ_L is clearly within the γ_L region. The α_L^e point with ϕ=62°, whilst within the α_L region so defined, is close to the boundary where there is considerable overlap of the two distributions.

Propensities along this line are shown in Fig 5(A). The collective propensity for non-glycines is slightly above zero in the α_L region indicating they favour this region, but not strongly, whereas the propensity for glycine is negative indicating it disfavours this region. By contrast, glycine strongly favours the γ_L region, whereas non-glycines strongly disfavour it. The dashed blue lines in Fig 5(A) indicate charged and polar amino acids. All apart from threonine (see below), tyrosine, and serine have propensities above zero at α_L^m and in a broad region around it. Cysteine has a small peak centred on (14°, 65°), also reflected in the colouring of Fig 3(C), but it should be borne in mind this is a sparsely populated area. As seen in Fig 5(A), the β-branched amino acids, Thr, Ile and Val, have very low propensities for the α_L region, and are even lower in the γ_L region. Fig5(B) shows the propensities at α_R , α_L^m and γ_L for all 20 amino acids. Apart for glycine, the propensities at α_R are all relatively small in value ($|Prop(X, \alpha_R)| < 0.2$). This means that, in an α-sheet, which alternates between α_R and α_L, it is the α_L region which discriminates most strongly amongst the amino acids.

Three especially significant findings emerge from Figs 4 and 5. One is that glycine has a high propensity for the γ_L region, and a negative propensity for the α_L region, whereas non-glycine has a small positive propensity for the α_L region and a negative propensity for the γ_L region. Another is that charged or polar amino acids favour the α_L region, particularly asparagine and, to a lesser degree, aspartate. The only polar amino acids with no positive propensity anywhere in Fig 5(A) are threonine and tyrosine. Thirdly, the β-branched amino acids, threonine, isoleucine and valine, have very low propensities for both α_L and γ_L regions.

The two main types of motifs in which α_L or γ_L conformations regularly occur in proteins are NESTs (the upper case means here a nest that is broadly defined, to include α_L or γ_L residues) and β-turns (of types I′ and II; (Hutchinson and Thornton, 1994; Venkatachalam, 1968; Wilmot and Thornton, 1990)). 74% of amino acids with these conformations occur in such situations; of these 66% are in NESTs and 34% are in β-turns. In Table 1 the effect of having glycines or L-amino acids at the α_L/γ_L position of NEST dipeptides in proteins on the main chain ϕ,ψ angles of the dipeptides is shown. In the left-hand columns of Table 1 it is seen, as expected from the propensities in Fig 5, that, for both RL and LR dipeptides, glycines favour the γ_L conformation, giving rise to nests, while L-amino acids favour the α_L conformation, forming α-strands. Two β-turn types incorporate α_L or γ_L residues: type II and type I′; for type II it is one residue; for type I′ there are two such residues. Data for the average ϕ,ψ angles are given for glycines and L-amino acids for these three residue positions in the right-hand columns of Table 1. As before a decided tendency for glycines to have higher ϕ angles is evident. With regard to residue 2 of type I′ β-turns, which has a conformation near to $\alpha_{L,}$ the proportion of glycines is fairly low (14%), which is also consistent with previous findings. Returning to Fig 2, visual comparison of the α_L and γ_L regions shows that glycines occupy areas at higher ϕ values than alanines do, which is also consistent with the results in Table 1.

Fig 6 shows a contour plot for the radius of curvature of strands formed from repeating dipeptide conformations, with $(-62°,-42°)_i(\phi°,\psi°)_{i+1}$; this means the residue at i is fixed at α_R and the residue at i+1 is allowed to vary position. As the residue at i+1 moves upwards along the line in Fig2(A), the radius of curvature increases rapidly reaching a maximum near the mirror condition. Inset in Fig 6 are structural models of $\alpha_R\gamma_L$ and of $\alpha_R\alpha_L{}^m$ repeating dipeptide conformations. Indicated in the figure is the line along which structures form rings in the sense that the helical rise is zero (Hayward et al., 2014). As can be seen, all three structures, $\alpha_R\gamma_L$, $\alpha_R\alpha_L{}^e$ and $\alpha_R\alpha_L{}^m$, either lie on this line or are close to it. The $\alpha_R\gamma_L$ ones are ring-shaped nests (the radius of curvature is 5 Å); the NH groups point to the centre such that they have the potential to donate their NH hydrogen to anions. Nests in proteins are partial rings (Hayward et al., 2014) so the ring aspect may not be immediately obvious. The $\alpha_R\alpha_L$ repeating dipeptides have a slight curve; the radius of curvature of the $\alpha_R\alpha_L{}^m$ structure is 34 Å, but slight adjustments in angles at i+1 can give rise to much larger radii (see Fig 6); such structures could self-associate into the α-sheet that we and others suggest has the properties expected for the material of the toxic amyloid precursor.

Conclusions

Most previous workers, considering amino acid preferences within proteins, have grouped together the α_L and γ_L regions of the Ramachandran plot and noted that the area is favoured for glycines. This is true for the two regions taken together but does not apply to the α_L region, which may be less occupied than the γ_L region and be outweighed by it. It would seem that the oversight has led to a commonly held view that the α-sheet alternating α_L/α_R conformation is so energetically unfavourable for peptides of naturally occurring proteins (consisting of many L-amino acids and a few glycines) that it would rarely occur. The idea is reinforced by the lack of α-sheets in native folded proteins; however it is likely that most proteins incorporating α-sheet, the putative toxic component of amyloid, have been eliminated during evolution because of the damage caused to cell membranes (Arispe et al., 1993; Jang et al., 2010). In any case, short α-strands, as opposed to α-sheets, *are* found in modest numbers in proteins.

Our work shows that the amino acid propensities for the α_L and γ_L regions are drastically different. In the α_L^m region 97% of amino acids are L-amino acids, with 3% being glycine, such that glycine is not even the most common of the 20 amino acids. On the other hand, in the γ_L region, 82% are glycines and the remaining 18% are L-amino acids. The idea that α-sheet would not be favoured by peptides rich in L-amino acids is now seen to be false. We are not claiming that α-sheet is as favourable as β-sheet in all circumstances but that it is a sufficiently stable conformation for its adoption by the average polypeptide to be perfectly feasible under appropriate conditions.

Another matter in relation to α-sheet is that, when fitting peptide conformations to electron density maps, it is often unclear to crystallographers whether the α-sheet or the β-sheet conformation is correct since the two are related by a 180° flip without much effect on the rest of the protein. Given that crystallographers tend to assume that, for peptides with L-amino acids, β-sheet is overwhelmingly more likely than α-sheet, it is possible they might be biased in favour of fitting β-sheet. The problem of finding the precise location of peptide bond atoms has been considered by Touw *et al* (Touw et al., 2015) and Keedy *et al* (Keedy et al., 2015); they examined protein crystallography data and recommended the reversal of tens of thousands of peptide planes.

Our studies also reveal that asparagine, and to a lesser degree aspartate, has a particular tendency to adopt the α_L conformation. As seen in Fig 5 for those that have a positive propensity for the α_L conformation, the ordering in their propensity from highest to lowest is: asparagine, aspartate, histidine, glutamine, lysine, cysteine, arginine and glutamate. This proclivity of asparagine has been noted (Deane and Blundell, 1999; Hovmoller et al., 2002; Richardson, 1981; Swindells et al., 1995) with regard to the α_L plus γ_L regions taken together. However, the finding that it applies so strongly for the α_L position is novel and deserves further investigation as to the cause. The other relevant finding, which has been shown previously, is that the three β-branched amino acids, threonine, valine and isoleucine, are outstandingly unlikely to adopt the α_L conformation. Of course, proline is the least likely amino acid of all to adopt this conformation. In summary, a high proportion of amino acids in the α_L conformation are either polar or charged; the proportion of NDHQKCRES amino acids with the $\alpha_L{}^m$ conformation being

81% while the proportion of PTVI amino acids is 2%. Distributions of amino acid sets in both the α_L and γ_L regions, are illustrated in Fig 7. An alternative to the assertion that L-amino acids are favoured by the α_L conformation becomes apparent; perhaps what is more important is the predominance of polar over non-polar amino acids. Two points can be made in reply. One is that the proportion of α_L^m glycines, at 3%, is low, leaving 97% L-amino acids. Secondly, in proteins, the α_L conformation is almost always situated at exterior positions of folded domains, so we conclude that L-amino acids are in general favoured at α_L conformations, which happen mostly to be polar ones. Whatever the underlying explanation, these results suggest the sorts of sequences that favour α-sheet. Also, even with an almost total absence of α-sheet structures, they indicate the feasibility of predicting α-sheet from sequence alone.

Looking at the issue from a different viewpoint, we have also examined the effect of a glycine or L-amino acid being at the α_L/γ_L position in situations of proteins where they recur. 74% are in two types of motif: **1.** RL ($\alpha_R\gamma_L$ or $\alpha_R\alpha_L$) or LR ($\alpha_L\alpha_R$ or $\gamma_L\alpha_R$), NEST dipeptides, and **2.** type II and type I′ β-turns. In all cases glycines occupy residue positions with higher ϕ and lower ψ values than L-amino acids do. For the nest-like motifs, glycines usually occur, giving rise to curved main chain conformations, while L-amino acids are mostly associated with α-strand conformations in which the main chain atoms are extended or slightly curved. The two repeating dipeptide structures are illustrated in Fig 6.

Propensities, as defined here, are useful for analysis of amino acids in different conformations, as shown in Fig 3 and Fig S1. These Ramachandran propensity plots, which are worth studying, give an idea of the probability of each amino acid for a region relative to the collective probabilities of all amino acids for that region. For a full appreciation of such plots readers should keep in mind the frequencies of the various conformations, as shown in the Ramachandran plot of Fig 2.

The main message of this paper is to point out that the α_L conformation is inherently favourable for L-amino-acid-containing peptides from ordinary proteins, such that α-sheet is a likely conformation under appropriate conditions. The pervasive idea that it is unfavourable would seem to have emerged from three mutually supportive issues. The first is the main focus of this article, the second is that, just because a conformation is rarely seen in folded proteins, does not mean it is necessarily uncommon in the unfolded state is the possibility of confusion between α- and β-sheet by X-ray crystallographers mentioned earlier. A further point regarding biophysical techniques is that α-sheet, due to its alternating α_R/α_L state, is expected to exhibit weak CD and ROA signals. These points, together with the new experimental and theoretical findings, suggest it is time to take α-sheet more seriously.

Glossary

α_L /γ_L/α_R: are main chain conformations of individual amino acid residues. The ϕ,ψ values chosen for γ_L are (78°,2°). Two alternative pairs of ϕ,ψ values, given below, are chosen for α_L. To avoid undue repetition the term α_R used here encompasses both α_R (when defined as the right-handed α-helix conformation) and γ_R.

α_L^e, α_L^m: are alternative α_L conformations. α_L^e has ϕ,ψ = (62°,42°), the main-chain enantiomeric form of α_R the right-handed α-helix conformation with ϕ,ψ = (-62°,-42°). α_L^m has ϕ,ψ = (42°,62°), the main-chain mirror form of α_R. The meaning of "mirror" is described below; it does not generally imply an enantiomeric relationship between two conformations.

α-sheet: resembles β-sheet except that the conformation of individual strands (α-strands) have alternating α_L and α_R conformations. α-sheet has more polarity than β-sheet. See Fig 1.

Nest: is normally defined as a peptide motif incorporating two or more residues with alternating (α_L or γ_L) and (α_R or γ_R) conformations. In this article, however, we wish to distinguish between α_L and (γ_L or α_L) conformations, so a **NEST,** in upper case, is used for the broadly defined (γ_L or α_L) nests while **nest** in lower case is reserved for γ_L nests, which have a concavity, as in Fig 6, that typically bind anionic or δ- atoms. α_L NESTs are straight and lack a concavity, and are described here as α-strands, as in Fig 6, whether or not they belong to α-sheet.

Mirror: describes a conformation B that is related to a previous conformation A on the Ramachandran plot. B lies on the opposite side of the diagonal $\phi = -\psi$ from A, such that (ϕ_B, ψ_B) = ($-\psi_A$, $-\phi_A$), i.e. they are reflections of each other in this diagonal.

Probability of an amino acid, often for a particular conformation, is the number of instances of that amino acid divided by the number of instances of all 20 amino acids.

Propensity: log likelihood ratio that quantifies the probability of occurrence of an amino acid in a region, relative to the probability of occurrence of all amino acids in that region. Note that other definitions of propensity are often used.

Acknowledgments

There are no acknowledgments.

Supplemental Information

Figure S1 showing propensity plots for all 20 amino acids.

Data S1: List of PDB files used.

References

Afzal, A.M., Al-Shubailly, F., Leader, D.P., Milner-White, E.J., 2014. Bridging of anions by hydrogen bonds in nest motifs and its significance for Schellman loops and other larger motifs within proteins. Proteins-Structure Function and Bioinformatics 82, 3023-3031.

Arai, H., Glabe, C., Luecke, H., 2012. Crystal structure of a conformation-dependent rabbit IgG Fab specific for amyloid prefibrillar oligomers. Biochimica Et Biophysica Acta-General Subjects 1820, 1908-1914.

Arispe, N., Rojas, E., Pollard, H.B., 1993. Alzheimer-disease amyloid beta-protein forms calcium channels in bilayer-membranes - blockade by tromethamine and aluminum. Proc. Natl. Acad. Sci. U. S. A. 90, 567-571.

Armen, R.S., Alonso, D.O.V., Daggett, V., 2004a. Anatomy of an amyloidogenic intermediate: Conversion of beta-sheet to alpha-sheet structure in transthyretin at acidic pH. Structure 12, 1847-1863.

Armen, R.S., DeMarco, M.L., Alonso, D.O.V., Daggett, V., 2004b. Pauling and Corey's alpha-pleated sheet structure may define the prefibrillar amyloidogenic intermediate in amyloid disease. Proceedings of the National Academy of Sciences of the United States of America 101, 11622-11627.

Babin, V., Roland, C., Sagui, C., 2011. The alpha-sheet: A missing-in-action secondary structure? Proteins-Structure Function and Bioinformatics 79, 937-946.

Balasco, N., Smaldone, G., Vigorita, M., Del Vecchio, P., Graziano, G., Ruggiero, A., Vitagliano, L., 2019. The characterization of Thermotoga maritima Arginine Binding Protein variants demonstrates that minimal local strains have an important impact on protein stability. Scientific Reports 9.

Balupuri, A., Choi, K.E., Kang, N.S., 2019. Computational insights into the role of alpha-strand/sheet in aggregation of alpha-synuclein. Scientific Reports 9.

Balupuri, A., Choi, K.E., Kang, N.S., 2020. Aggregation Mechanism of Alzheimer's Amyloid beta-Peptide Mediated by alpha-Strand/alpha-Sheet Structure. International Journal of Molecular Sciences 21.

Beck, D.A.C., Alonso, D.O.V., Inoyama, D., Daggett, V., 2008. The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. Proc. Natl. Acad. Sci. U. S. A. 105, 12259-12264.

Berkholz, D.S., Shapovalov, M.V., Dunbrack, R.L., Karplus, P.A., 2009. Conformation Dependence of Backbone Geometry in Proteins. Structure 17, 1316-1325.

Bi, T.M., Daggett, V., 2018. The Role of alpha-sheet in Amyloid Oligomer Aggregation and Toxicity. Yale Journal of Biology and Medicine 91, 247-255.

Bleem, A., Francisco, R., Bryers, J.D., Daggett, V., 2017. Designed alpha-sheet peptides suppress amyloid formation in Staphylococcus aureus biofilms. Npj Biofilms and Microbiomes 3.

Carugo, O., Carugo, K.D., 2013. Half a century of Ramachandran plots. Acta Crystallographica Section D-Structural Biology 69, 1333-1341.

Childers, M.C., Daggett, V., 2019. Drivers of alpha-Sheet Formation in Transthyretin under Amyloidogenic Conditions. Biochemistry 58, 4408-4423.

Childers, M.C., Daggett, V., 2020. Edge Strand Dissociation and Conformational Changes in Transthyretin under Amyloidogenic Conditions. Biophysical Journal 119, 1995-2009.

Chiti, F., Dobson, C.M., 2017. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade, p. 27-68, in: R. D. Kornberg, (Ed.), Annual Review of Biochemistry, Vol 86.

Daggett, V., 2006. alpha-sheet: The toxic conformer in amyloid diseases? Accounts of Chemical Research 39, 594-602.

Davis, I.W., Arendall, W.B., Richardson, D.C., Richardson, J.S., 2006. The backrub motion: How protein backbone shrugs when a sidechain dances. Structure 14, 265-274.

Deane, C.M., Blundell, T.L., 1999. Examination of the less favoured regions of the Ramachandran plot, p. 196-208, in: M. Vijaya, et al., Eds.), Perspectives in Structural Biology, Indian Academy of Sciences.

Deshmukh, M., Evans, M.L., Chapman, M.R., 2018. Amyloid by Design: Intrinsic Regulation of Microbial Amyloid Assembly. J.Mol.Biol. 430, 3631-3641.

Efimov, A.V., 1993. Standard Structures in Proteins. Progress in Biophysics & Molecular Biology 60, 201-239.

Eisenberg, D., Jucker, M., 2012. The Amyloid State of Proteins in Human Diseases. Cell 148, 1188-1203.

Erskine, E., MacPhee, C.E., Stanley-Wall, N.R., 2018. Functional Amyloid and Other Protein Fibers in the Biofilm Matrix. J.Mol.Biol. 430, 3642-3656.

Fandrich, M., Fletcher, M.A., Dobson, C.M., 2001. Amyloid fibrils from muscle myoglobin - Even an ordinary globular protein can assume a rogue guise if conditions are right. Nature 410, 165-166.

Grillo-Bosch, D., Carulla, N., Cruz, M., Sanchez, L., Pujol-Pina, R., Madurga, S., Rabanal, F., Giralt, E., 2009. Retro-Enantio N-Methylated Peptides as beta-Amyloid Aggregation Inhibitors. Chemmedchem 4, 1488-1494.

Hayward, S., 2001. Peptide-Plane Flipping. Protein Science 10, 2219-2227.

Hayward, S., Milner-White, E.J., 2008. The geometry of alpha-sheet: Implications for its possible function as amyloid precursor in proteins. Proteins-Structure Function and Bioinformatics 71, 415-425.

Hayward, S., Milner-White, E.J., 2011. Simulation of the beta- to alpha-sheet transition results in a twisted sheet for antiparallel and an alpha-nanotube for parallel strands: Implications for amyloid formation. Proteins-Structure Function and Bioinformatics 79, 3193-3207.

Hayward, S., Leader, D.P., Al-Shubailly, F., Milner-White, E.J., 2014. Rings and ribbons in protein structures: Characterization using helical parameters and Ramachandran plots for repeating dipeptides. Proteins-Structure Function and Bioinformatics 82, 230-239.

Hilaire, M.R., Ding, B., Mukherjee, D., Chen, J.X., Gai, F., 2018. Possible Existence of alpha-Sheets in the Amyloid Fibrils Formed by a TTR(105-115)Mutant. J.Am.Chem.Soc. 140, 629-635.

Ho, B.K., Thomas, A., Brasseur, R., 2003. Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. Protein Science 12, 2508-2522.

Hollingsworth, S.A., Karplus, P.A., 2010. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol Concepts 1, 271-283.

Hopping, G., Kellock, J., Barnwal, R.P., Law, P., Bryers, J., Varani, G., Caughey, B., Daggett, V., 2014. Designed alpha-sheet peptides inhibit amyloid formation by targeting toxic oligomers. Elife 3.

Hovmoller, S., Zhou, T., Ohlson, T., 2002. Conformations of amino acids in proteins. Acta Crystallogr. Sect. D-Biol. Crystallogr. 58, 768-776.

Hutchinson, E.G., Thornton, J.M., 1994. A revised set of potentials for beta-turn formation in proteins. Protein Science 3, 2207-2216.

Jang, H., Arce, F.T., Ramachandran, S., Capone, R., Lal, R., Nussinov, R., 2010. beta-Barrel Topology of Alzheimer's beta-Amyloid Ion Channels. Journal of Molecular Biology 404, 917-934.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577-2637.

Kayed, R., Head, E., Thompson, J.L., McIntire, T.M., Milton, S.C., Cotman, C.W., Glabe, C.G., 2003. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science 300, 486-489.

Keedy, D.A., Fraser, J.S., van den Bedem, H., 2015. Exposing Hidden Alternative Backbone Conformations in X-ray Crystallography Using qFit. Plos Computational Biology 11.

Kellock, J., Hopping, G., Caughey, B., Daggett, V., 2016. Peptides Composed of Alternating L- and D-Amino Acids Inhibit Amyloidogenesis in Three Distinct Amyloid Systems Independent of Sequence. J.Mol.Biol. 428, 2317-2328.

Kim, J.D., Pike, D.H., Tyryshkin, A.M., Swapna, G.V.T., Raanan, H., Montelione, G.T., Nanda, V., Falkowski, P.G., 2018. Minimal Heterochiral de Novo Designed 4Fe-4S Binding Peptide Capable of Robust Electron Transfer. J.Am.Chem.Soc. 140, 11210-11213.

Laskowski, R.A., Furnham, N., Thornton, J. 2013. Biomolecular forms and functions : a celebration of 50 years of the Ramachandran map. World Scientific Pub. Co. Inc., Hackensack, NJ.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK - A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26, 283-291.

Leader, D.P., Milner-White, E.J., 2009. Motivated Proteins: A web application for studying small three-dimensional protein motifs. BMC Bioinformatics 10.

Leader, D.P., Milner-White, E.J., 2012. Structure Motivator: A tool for exploring small three-dimensional elements in proteins. BMC Structural Biology 12.

Lee, S.J.C., Nam, E., Lee, H.J., Savelieff, M.G., Lim, M.H., 2017. Towards an understanding of amyloid-beta oligomers: characterization, toxicity mechanisms, and inhibitors. Chemical Society Reviews 46, 310-323.

Maris, N.L., Shea, D., Bleem, A., Bryers, J.D., Daggett, V., 2018. Chemical and Physical Variability in Structural Isomers of an L/D alpha-Sheet Peptide Designed To Inhibit Amyloidogenesis. Biochemistry 57, 507-510.

McDonald, I.K., Thornton, J.M., 1994. Satisfying hydrogen-bonding potential in proteins. J.Mol.Biol. 238, 777-793.

Meng, F.H., Lu, T., Li, F., 2019. Stabilization of Solvent to alpha-Sheet Structure and Conversion between alpha-Sheet and beta-Sheet in the Fibrillation Process of Amyloid Peptide. Journal of Physical Chemistry B 123, 9576-9583.

Milner-White, E.J., Watson, J.D., Qi, G., Hayward, S., 2006. Amyloid formation may involve alpha- to beta- sheet interconversion via peptide plane flipping. Structure 14, 1369-1376.

Pal, D., Chakrabarti, P., 2002. On residues in the disallowed region of the Ramachandran map. Biopolymers 63, 195-206.

Paranjapye, N., Daggett, V., 2018. De Novo Designed alpha-Sheet Peptides Inhibit Functional Amyloid Formation of Streptococcus mutans Biofilms. J.Mol.Biol. 430, 3764-3773.

Pauling, L., Corey, R.B., 1951. The Pleated Sheet, a New Layer Configuration of Polypeptide Chains. Proceedings of the National Academy of Sciences of the United States of America 37, 251-256.

Porter, L.L., Rose, G.D., 2011. Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints. Proc. Natl. Acad. Sci. U. S. A. 108, 109-113.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations. J.Mol.Biol. 7, 95-&.

Ravikumar, A., Ramakrishnan, C., Srinivasan, N., 2019. Stereochemical Assessment of (phi,psi) Outliers in Protein Structures Using Bond Geometry-Specific Ramachandran Steric-Maps. Structure 27, 1875-+.

Richardson, J.S., 1981. The Anatomy and Taxonomy of Protein Structure, p. 167-339, in: C. B. Anfinsen, et al., Eds.), Advances in Protein Chemistry, Academic Press.

Richardson, J.S., Richardson, D.C., 1989. Principles and patterns of protein conformations, p. 1-98, in: G. D. Fasman, (Ed.), Prediction of protein structure and principles of protein conformation. , Plenum Press, New York.

Richardson, J.S., Keedy, D.A., Richardson, D.C. 2013. Biomolecular forms and functions : a celebration of 50 years of the Ramachandran map. World Scientific Pub. Co. Inc., Hackensack, NJ.

Shea, D., Hsu, C.C., Bi, T.M., Paranjapye, N., Childers, M.C., Cochran, J., Tomberlin, C.P., Wang, L.B., Paris, D., Zonderman, J., Varani, G., Link, C.D., Mullan, M., Daggett, V., 2019. alpha-Sheet secondary structure in amyloid beta-peptide drives aggregation and toxicity in Alzheimer's disease. Proceedings of the National Academy of Sciences of the United States of America 116, 8895-8900.

Shortle, D., 2002. Composites of local structure propensities: Evidence for local encoding of long-range structure. Protein Science 11, 18-26.

Swindells, M.B., Macarthur, M.W., Thornton, J.M., 1995. Intrinsic phi,psi propensities of amino-acids, derived from the coil regions of known structures. Nature Structural Biology 2, 596-603.

Touw, W.G., Joosten, R.P., Vriend, G., 2015. Detection of trans-cis flips and peptide-plane flips in protein structures. Acta Crystallographica Section D-Structural Biology 71, 1604-1614.

Venkatachalam, C.M., 1968. Stereochemical criteria for polypeptides and proteins .v. conformation of a system of 3 linked peptide units. Biopolymers 6, 1425-+.

Wang, S., Meng, F.H., Hao, R.J., Wang, C.Y., Li, F., 2020. Study on the structure and membrane disruption of the peptide oligomers constructed by hIAPP(18-27) peptide and its D,L-alternating isomer. Biochimica Et Biophysica Acta-Biomembranes 1862.

Watson, J.D., Milner-White, E.J., 2002a. The conformations of polypeptide chains where the main-chain parts of successive residues are enantiomeric. Their occurrence in cation and anion-binding regions of proteins. J.Mol.Biol. 315, 183-191.

Watson, J.D., Milner-White, E.J., 2002b. A novel main-chain anion-binding site in proteins: The nest. A particular combination of phi,psi values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. J.Mol.Biol. 315, 171-182.

Wilmot, C.M., Thornton, J.M., 1990. Beta-turns and their distortions - a proposed new nomenclature. Protein Eng. 3, 479-493.

Xu, S.H., 2007. Aggregation drives "misfolding" in protein amyloid fiber formation. Amyloid-Journal of Protein Folding Disorders 14, 119-131.

Yang, M.F., Lei, M., Yordanov, B., Huo, S.H., 2006. Peptide plane can flip in two opposite directions: Implication in amyloid formation of transthyretin. Journal of Physical Chemistry B 110, 5829-5833.

Zhou, A.Q., O'Hern, C.S., Regan, L., 2011. Revisiting the Ramachandran plot from a new angle. Protein Science 20, 1166-1171.

Figure Legends

Figure 1: α-sheet hydrogen bond arrangement in parallel sheet displaying main-chain atoms only. (A) α-sheet model of Pauling and Corey (Pauling and Corey, 1951) with perfectly straight strands. (B) α-sheet model determined using torsion angle driving from β-sheet (Hayward and Milner-White, 2011).

Figure 2: Ramachandran frequency plots for (A) alanine and (B) glycine. Frequencies were calculated for each cell in grid of 4˚×4° cells. The colour of the bin indicates a frequency above which X% (colour vs X% given in Table key) of the amino acids of that type are located. Added to (A) is a sliding 36˚×36˚ window that moves along the line passing through or close to γ_L, α_L^e and α_L^m, indicated by the cyan, green and red spots, at (78˚, 2˚), (62˚, 42˚) and (42˚, 62˚), respectively.

Figure 3: Ramachandran propensity plots generated using a 36˚×36˚ window centred on each 4˚×4˚ grid cell. The colour in each 4˚×4˚ grid cell is for the calculation of the propensity in a 36˚×36˚ window it is centred on - see Methods for details. (A) alanine, (B) glycine, (C) cysteine, and (D) asparagine. Green means the amino acid is like the average amino acid for that region, black that it strongly favours the region, and white (<-0.9) that it strongly disfavours the region. Ramachandran propensity plots for all amino acids are found in Fig S1.

Figure 4: (A) Blue line is total number of amino acids of any type in each position of the sliding 36˚×36˚ window centred on ϕ of the diagonal line in Fig 2(A). Red and black lines give the numbers for glycine and non-glycine, respectively. (B) Probability of X at a

given window position. The continuous red line is for when X is glycine, the continuous black line for when X is all non-glycine, and the broken black lines are when X represents individual non-glycine amino acids.

Figure 5: (A) Propensity against position of the sliding 36˚×36˚ window centred on ϕ of the diagonal line in Fig 2(A). The continuous red line is glycine and the continuous black line is the propensity of the sum of all non-glycine amino acids. The broken magenta lines are β-branched amino acids T, V, and I. The broken blue lines are charged or polar amino acids (apart from T): S, Y, E, D, H, Q, R, N and K. All except S and Y have $Prop > 0$ at α_L^m. The remaining amino acids are shown with broken black lines. In order of increasing propensity at α_L^m the amino acids are: P, I, V, T, L, W, G, F, M, A, Y, S, E, R, C, K, Q, H, D, N. (B) Propensities at α_R (black dots) and α_L^m (red dots).

Figure 6: Contour plot for radius of curvature (Å) for strand formed from repeating dipeptide conformations $(-62°,-42°)_i(\phi°,\psi°)_{i+1}$. The black spot for residue i is at α_R: (-62°,-42°). For residue i+1, the cyan spot is at γ_L: (78°, 2°), the green spot at α_L^e: (62°,42°) and the red spot at α_L^m: (42°,62°). The radius of curvature increases as the i+1 point moves along the sampling line (see Fig2(A)) from γ_L to α_L^m and beyond up to about (37°, 71°) where the radius of curvature reaches a maximum of about 80 Å. Note that the sampling line runs almost perpendicular to the contour lines indicating that it is on the path along which there is maximum change in the radius of curvature. The broken pink line indicates structures that form perfect rings, i.e. there is no helical rise. Inset: structural models of repeating dipeptides (side-chains omitted) with $\alpha_R\gamma_L$ and $\alpha_R\alpha_L^m$ conformations.

Figure 7: Schematic illustrating the distributions of amino acid groupings in the α_L and γ_L regions. The area is proportional to the probability of occurrence calculated at the α_L^m and γ_L points.

Tables

Table 1: Average dihedral angles for $\alpha_{L/}\gamma_L$ residues in various motifs: NEST dipeptides and type I′ or type II β-turns in proteins.

	RL Nest (49%) residue 2		**LR Nest (22%)** residue 1		**Type I′ (9%)** residue 2		**Type I′ (9%)** residue 3		**Type II (20%)** residue 3	
		ϕ,ψ		ϕ,ψ		ϕ,ψ		ϕ,ψ		ϕ,ψ
Glycine	64%	85°,13°	54.5%	84°,4°	14%	56°,30°	67.5%	85°,10°	67%	88°,-8°
Non-glycine	36%	60°,34°	45.5%	60°,34°	86%	48°,44°	33.5%	60°,25°	33%	66°,20°

For each motif the percentage gives the total number of $\alpha_{L/}\gamma_L$ residues in the motif compared to that in all proteins. The glycine/non-glycine percentages are those compared to all amino acids at that position within the motif. For numbering of motif positions, NESTs have two residues and β-turns (here of types I′ and II) have four**.**

Figures

Figure 1

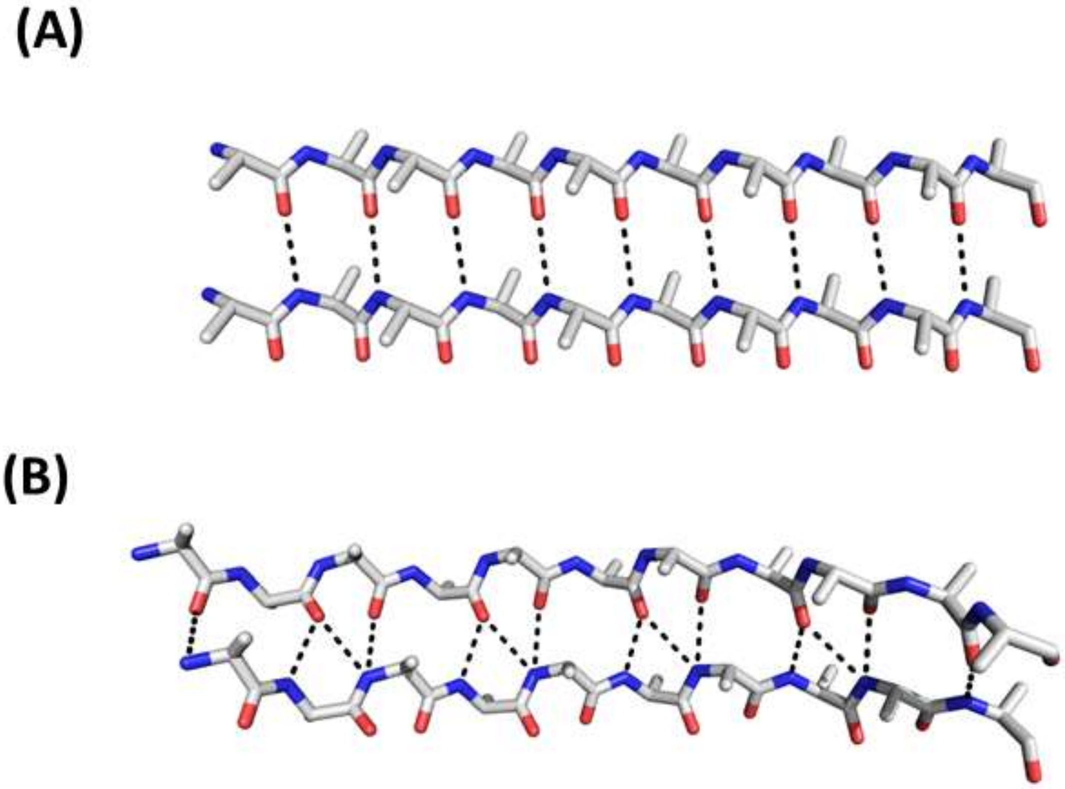

Figure 1: α-sheet hydrogen bond arrangement in parallel sheet displaying main-chain atoms only. (A) α-sheet model of Pauling and Corey (Pauling and Corey, 1951) with perfectly straight strands. (B) α-sheet model determined using torsion angle driving from β-sheet (Hayward and Milner-White, 2011).

Figure 2

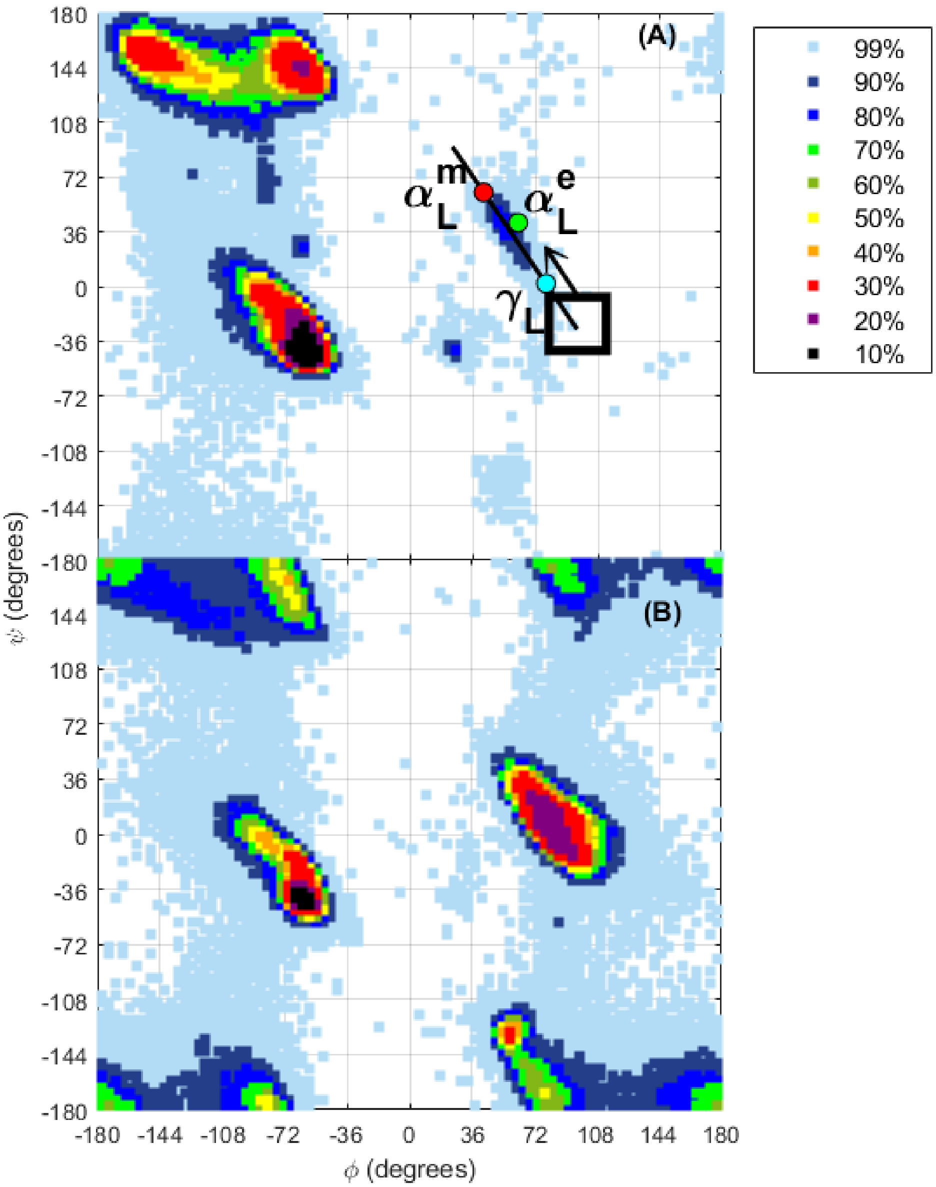

Figure 2: Ramachandran frequency plots for (A) alanine and (B) glycine. Frequencies were calculated for each cell in grid of 4˚×4° cells. The colour of the bin indicates a frequency above which X% (colour vs X% given in Table key) of the amino acids of that type are located. Added to (A) is a sliding 36˚×36˚ window (square with thick black lines) that moves along the line passing through or close to γ_L, α_L^e and α_L^m, indicated by the cyan, green and red spots, at (78˚, 2˚), (62˚, 42˚) and (42˚, 62˚), respectively. Quantities presented in Figures 4 and 5 are generated from samples taken along this sliding window.

Figure 3

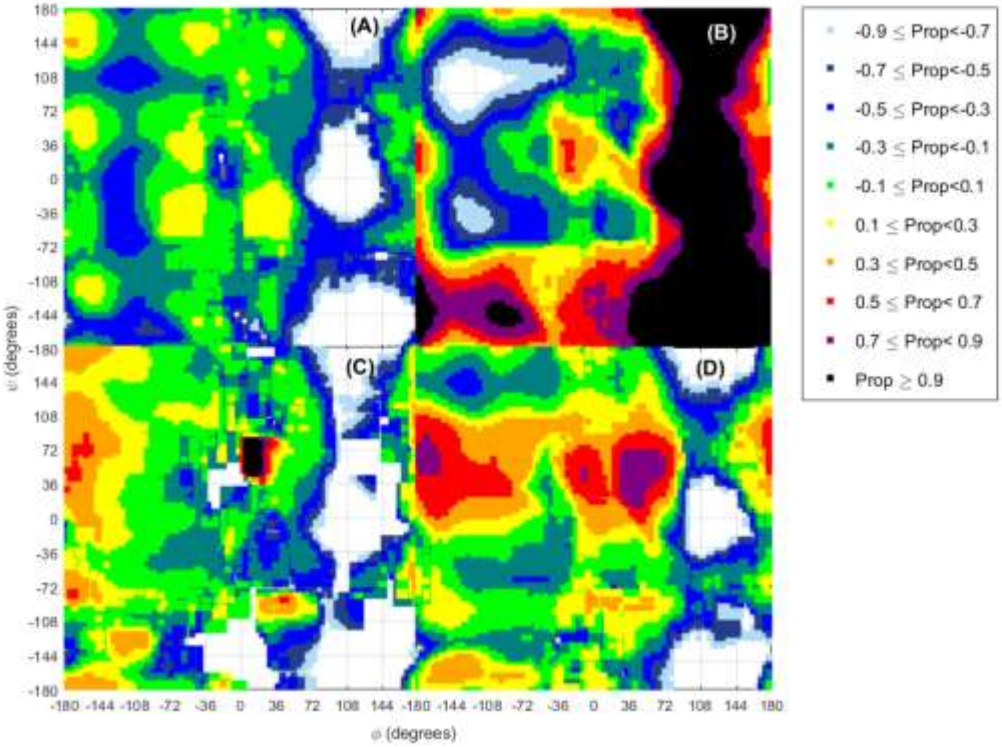

Figure 3: Ramachandran propensity plots generated using a 36˚×36˚ window centred on each 4˚×4˚ grid cell. The colour in each 4˚×4˚ grid cell is for the calculation of the propensity in a 36˚×36˚ window it is centred on - see Methods for details. (A) alanine, (B) glycine, (C) cysteine, and (D) asparagine. Green means the amino acid is like the average amino acid for that region, black that it strongly favours the region, and white (<-0.9) that it strongly disfavours the region. Ramachandran propensity plots for all amino acids are found in Fig S1.

Figure 4

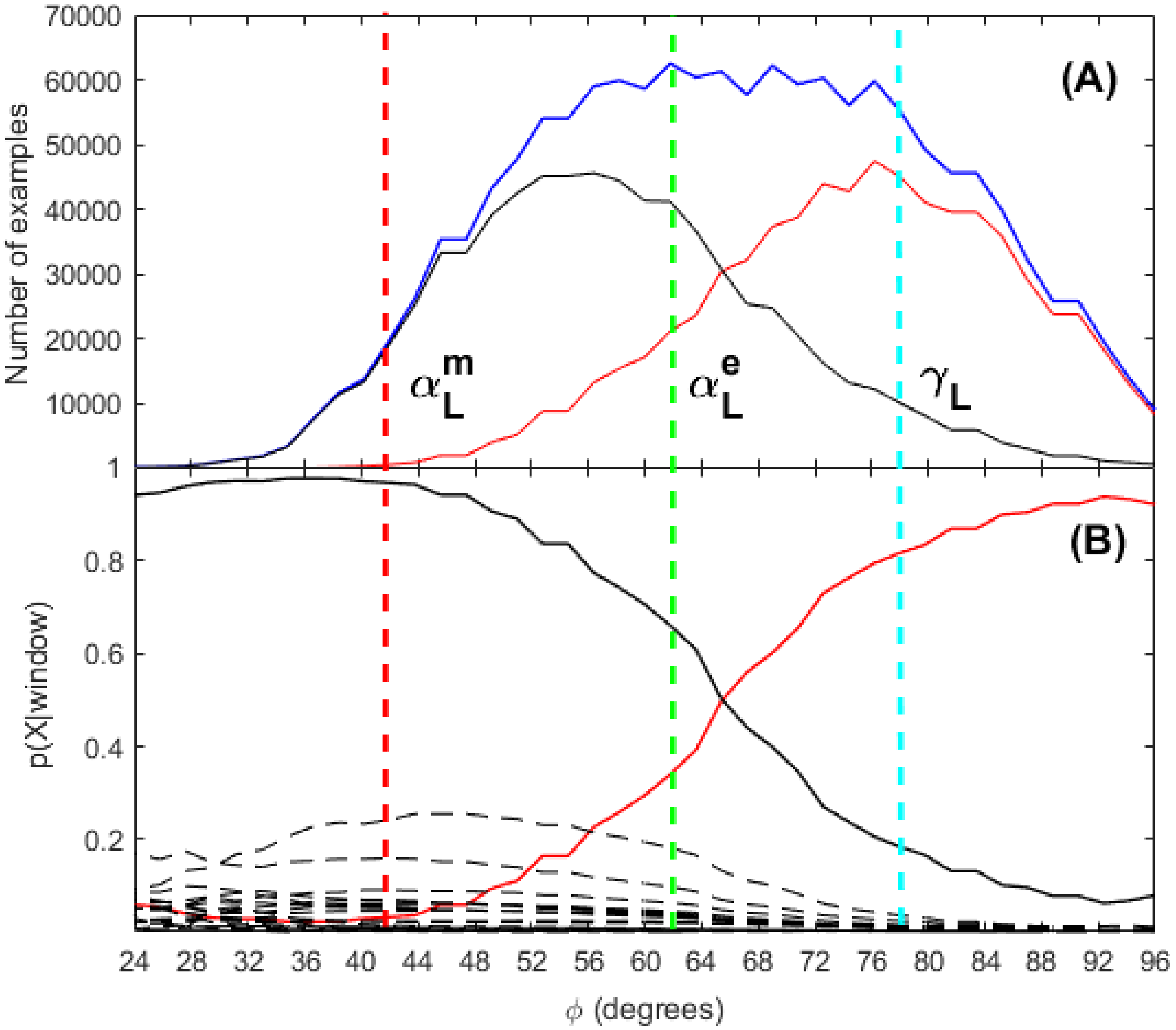

Figure 4: (A) Dark blue line is total number of amino acids of any type in each position of the sliding 36˚×36˚ window centred on ϕ of the diagonal line in Fig 2(A). Red and black lines give the numbers for glycine and non-glycine, respectively. (B) Probability of X at a given window position. The continuous red line is for when X is glycine, the continuous black line for when X is all non-glycine, and the broken black lines are when X represents individual non-glycine amino acids.

Figure 5

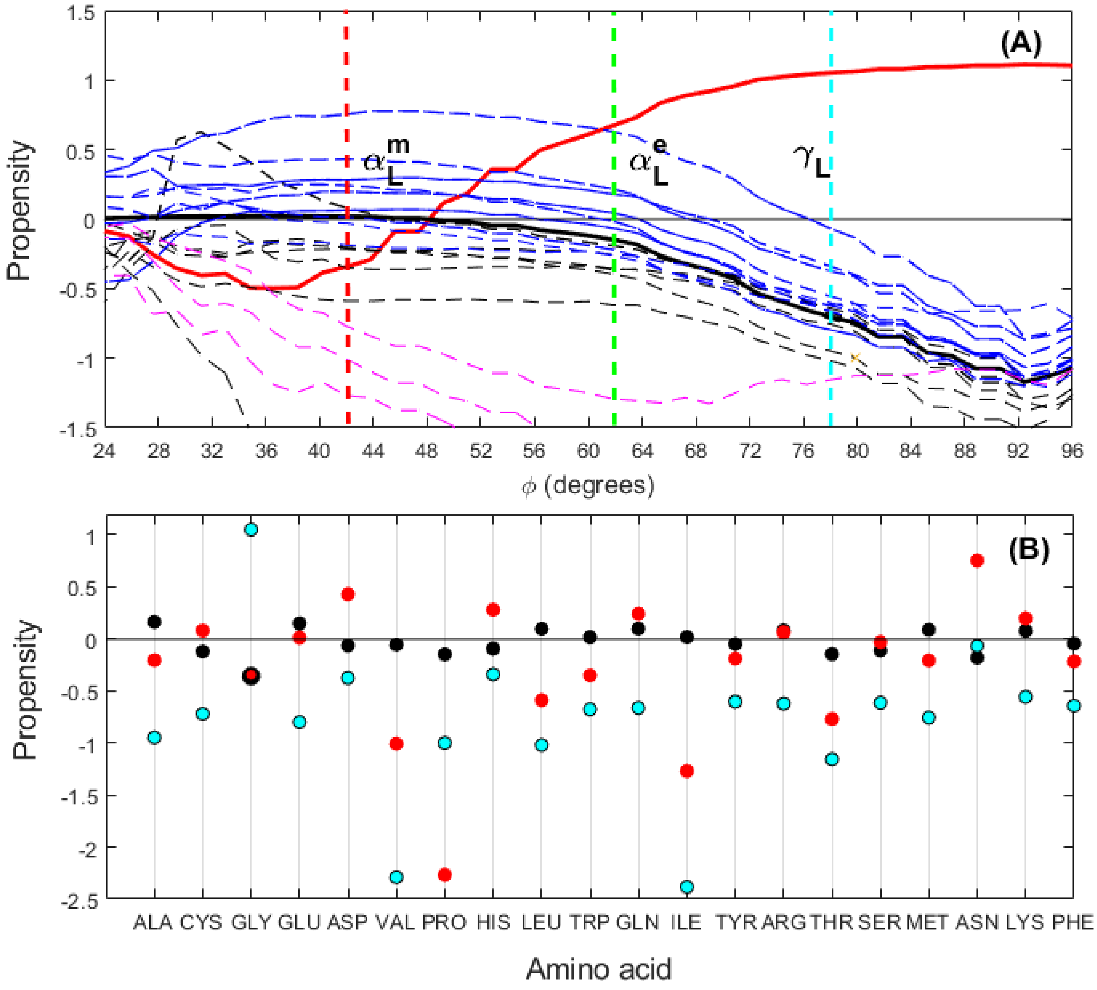

Figure 5: (A) Propensity against position of the sliding 36˚×36˚ window centred on ϕ of the diagonal line in Fig 2(A). The continuous red line is glycine and the continuous black line is the propensity of the sum of all non-glycine amino acids. The broken magenta lines are β-branched amino acids T, V, and I. The broken blue lines are charged or polar amino acids (apart from T): S, Y, E, D, H, Q, R, N and K. All except S and Y have $Prop > 0$ at α_L^m. The remaining amino acids are shown with broken black lines. In order of increasing propensity at α_L^m the amino acids are: P, I, V, T, L, W, G, F, M, A, Y, S, E, R, C, K, Q, H, D, N. (B) Propensities at α_R (black dots), α_L^m (red dots) and γ_L (cyan dots).

Figure 6

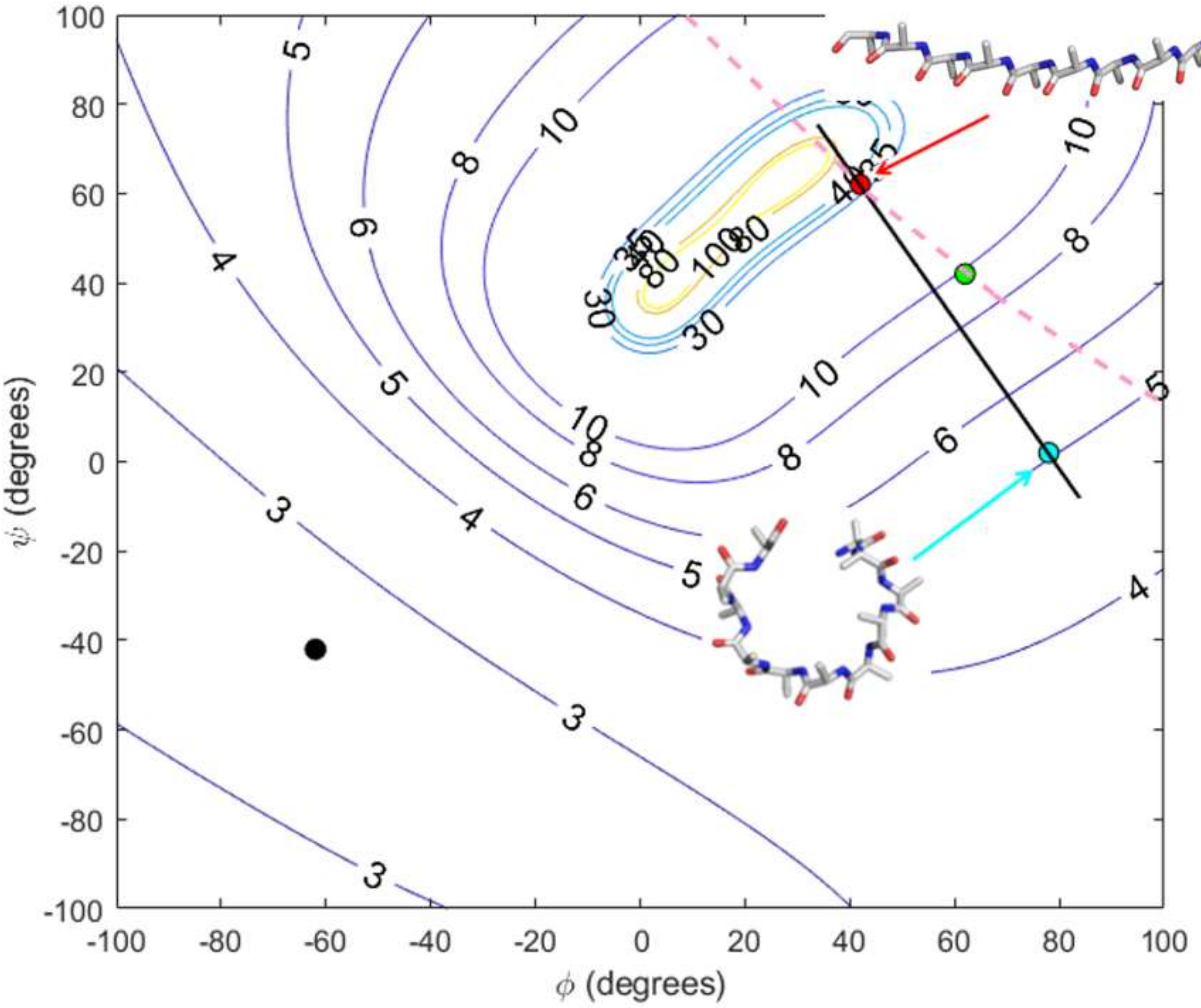

Figure 6: Contour plot for radius of curvature (Å) for strand formed from repeating dipeptide conformations $(-62°,-42°)_i(\phi°,\psi°)_{i+1}$. The black spot for residue i is at α_R: (-62°,-42°). For residue i+1, the cyan spot is at γ_L: (78°, 2°), the green spot at α_L^e: (62°,42°) and the red spot at α_L^m: (42°,62°). The radius of curvature increases as the i+1 point moves along the sampling line (see Fig2(A)) from γ_L to α_L^m and beyond up to about (37°, 71°) where the radius of curvature reaches a maximum of about 80 Å. Note that the sampling line runs almost perpendicular to the contour lines indicating that it is on the path along which there is maximum change in the radius of curvature. The broken pink line indicates structures that form perfect rings, i.e. there is no helical rise. Inset: structural models of repeating dipeptides (side-chains omitted) with $\alpha_R\gamma_L$ (cyan spot) and $\alpha_R\alpha_L^m$ (red spot) conformations.

Figure 7

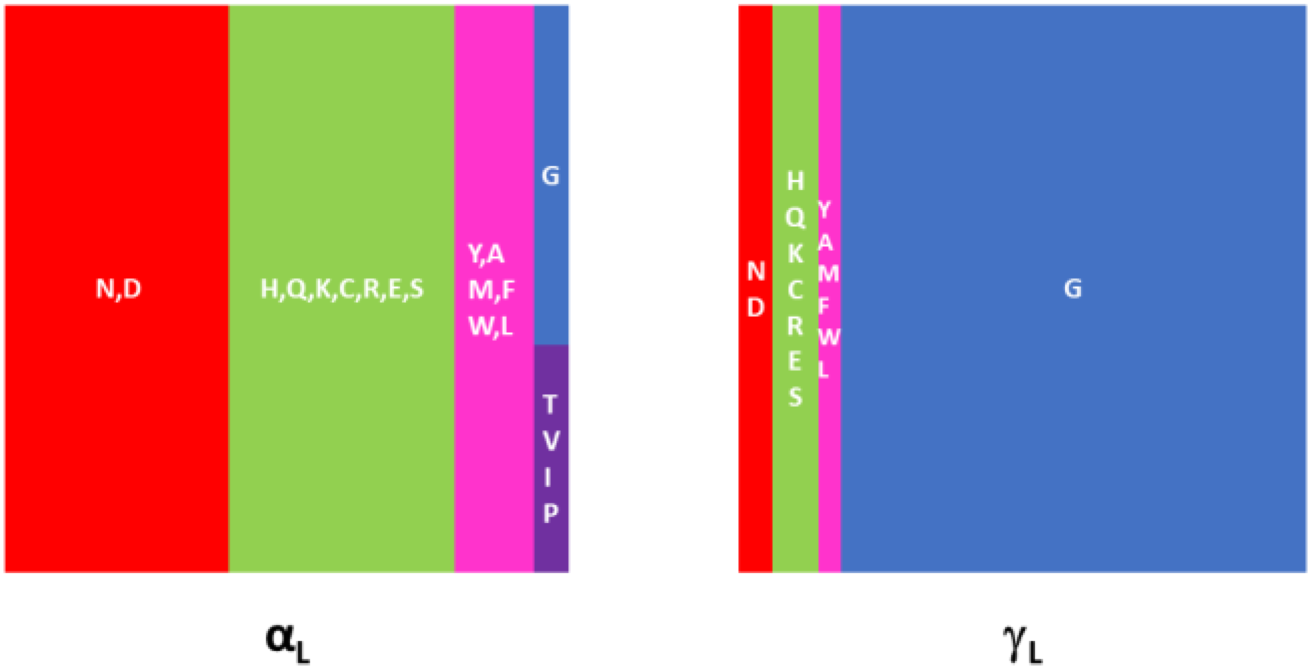

Figure 7: Schematic illustrating the distributions of amino acid groupings in the α_L and γ_L regions. The area is proportional to the probability of occurrence at the α_L^m and γ_L points.