Adriana Pietropaolo  
*Dipartimento di Scienze Chimiche, Università di Catania, Viale A. Doria 6, I-95125 Catania, Italy* 

*Computational Science, Department of Chemistry and Applied Biosciences, ETH Zürich, USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland* 

**Chirality in biochemistry: A computational approach for investigating biomolecule conformations**
Chirality in biochemistry: A computational approach for investigating biomolecule conformations
Abstract

Chirality selection is a key issue in many important biochemical phenomena. The mechanism of protein folding, enzyme recognition, drug design is intimately connected to the chirality of the molecules involved. In this review, the role of chirality in the chemistry of the life sciences is discussed. In particular, a novel methodology for the study of local chirality is presented, which provides one with a deeper understanding of the connection between secondary structure and protein flexibility.
Contents

1 Introduction 1
1.1 Molecular chirality in living systems 1
1.2 Protein secondary structures 4
1.3 Protein secondary structure assignment 5
1.4 Intrinsic chirality and protein secondary structures 6

2 Computational techniques for studying protein dynamics 7

3 Employing Chirality to analyse protein motions 9
3.1 The chirality index 9
3.2 Using chirality to understand protein structure 11
3.3 Chirality index as a tool for monitoring protein dynamics 12
3.4 Chirality and Circular Dichroism 16

4 Perspectives 19

5 Acknowledgments 21

References 23
Introduction

1.1 Molecular chirality in living systems

Since its discovery by Louis Pasteur in 1848 molecular chirality has captured the collective imagination of the chemical community. Pasteur’s realization, that tartaric acid could be observed in number of chemically inequivalent forms (R, S, racemic and meso forms) led to a rich new vein of chemical discovery based on the study of what he christened chirality from the greek, cheir, meaning hand. Any molecule which has no improper rotational symmetry is by definition chiral because it cannot be superimposed on its mirror image. In other words molecules with no improper rotational symmetry can exist in two non-equivalent, enantiomeric forms, named R, S, following the Cahn–Ingold–Prelog priority rule or L, D in the Fisher notation for glyceraldehyde stereocenters (Figure 1.1), being these single notations not correlated each other.

In nature many biomolecules do not have improper rotational symmetry and are thus chiral. All the natural aminoacids, with the exception of glycine, are chiral and therefore have two enantiomeric forms commonly termed L and D. Aminoacids are almost exclusively observed in the L form in biomolecules, while in natural carbohydrates the individual monomers are found in the D form. This selectivity in biomolecules, which is known as homochirality or using the Pasteur notation dissimilarity, has aroused the curiosity of many scientists, many of whom have expended a great deal of effort in attempts to understand why nature is so strict about chirality.

Proteins, the building blocks of the cell, are typically composed of a chain of homochiral L-aminoacids. Only in the cell wall of some microorganisms [1] and various aged human tissues (tooth, bone, brain etc.) [2] are D aminoacids observed. Nanda et al. [3] recently investigated the role of homochirality in the folding free energy landscape of proteins. They showed that syndiotactic chains, in which there is an alternation in the chirality of adyacent monomers, have a large ensemble of accessible conformations (Figure 1.2). Hence, syndiotactic chains have a much greater degree of flexibility in their backbone, than isotactic polypeptides, in which the chiralities of all stereocenters are identical. The fact
that glycine residues, the only aminoacid that lacks the C\textsubscript{\textalpha} stereocenter, can provide flexibility to proteins and can destabilize folded states [4, 5] strengthens the suggestion that chirality plays an important role in protein flexibility. This dependence of protein folding on chirality occurs because the chirality of individual aminoacids can affect local secondary structures which in turn can affect the tertiary structure of the protein. For example exchanging a number of L aminoacids for D aminoacids in Sso7d ribonuclease changes the conformation of the active site and prevents the protein function [6]. Chirality has therefore a deep role in complex systems like proteins in promoting the formation of inter-residue contacts, which stabilize folded configurations.

Fig. 1.1 Graphical representation of enantiomers, due to stereocenter chirality, generated by the reflection through a vertical plane.
Fig. 1.2 Energy landscape representation as described in ref. [3]. Width of each cone represents the number of accessible conformational states, while their depth underlines the energy, linked to structural stability. Unfolded conformations correspond to the highest energy level, here represented by the cone bases, whose perspective is given by the square.
1.2 Protein secondary structures

Protein structures are typically described using a system which separates the various aspects of their structures. So-called primary structure describes the aminoacid sequence of which the protein is composed, while, tertiary structure refers to the global structure adopted by a single protein. This tertiary structure can be thought of in terms of a number of highly-regular structural motifs, or secondary structure elements, the structures of which depend on the underlying primary structure. Proteins can adopt a number of different secondary structures (Figure 1.3) the most common being the right handed helices, which include both the \(\alpha\), \(3_{10}\) and the \(\pi\) helix, the \(\beta\) sheet, the \(\beta\) turn and the left handed helices, an example of which is the poly-L-proline II (PPII).

The characteristics of each of these structures are described below:

**\(\alpha\) helix** This structure, which was discovered by Linus Pauling, typically involves more than four residues, has a pitch of 5.4 Å and is stabilized by hydrogen bonds between the carbonyl oxygen of residue \(i\) and the amide hydrogen of residue \(i+4\).

**\(\beta\) sheets** This structure is flat and the hydrogen bonds it contains are either parallel or antiparallel depending on the orientation of the chain. Often an isolated \(\beta\) sheet is referred to as a bulge.

**\(\beta\) turns** In these structures hydrogen bonds are formed between the \(i\)th and \(i+3\)th residues. Right handed \(3_{10}\) helices are \(\beta\) turns in which more than three residues are involved.

**poly-L-proline II** Is an example of a left-handed helix. It has an elongated structure that involves a pitch of 9.3 Å.

**coil** A coil is any segment of an aminoacid chain which has no secondary structure. Any given protein segment is said to have no secondary structure if every pair of \((\phi, \psi)\) angles is independent of all the other pairs of \((\phi, \psi)\) angles within the segment [7].

The secondary structure of a protein is very important as it is thought that the global folding event is driven by local folding. Furthermore, as already mentioned, the tertiary structure, which controls the protein function, can be thought of as a collection of secondary elements. One incorrect folded (misfolded) secondary structural element can give rise to a different tertiary structure and a prevention of the protein function or even a protein with different properties. What is more, introduction of induced fit concepts surpassed rigid lock and key ideas in the theories of protein docking. This means that, now more than ever, it is necessary to understand the dynamical structures of proteins as well their static structures, since there is a direct link between protein conformational change and protein function.
1.3 Protein secondary structure assignment

A number of methods for secondary structure assignment have been proposed (see e.g. [8, 9, 10, 11]) and there is about 80% agreement between structural assignments made with different methods. The first key contribution was probably by Ramachandran [12], who correlated the native distribution of the $-$N–C$_{\alpha}$- and $-$C$_{\alpha}$–C$-$ protein dihedral angles ($\phi$, $\psi$) with the secondary structure [13]. This approach fails though when there is a great deal of conformational flexibility, which is common for peptides, because the flexibility gives rise to non-standard backbone angles [14, 15, 16].

One of the most commonly used programs for secondary structure determination is the “Dictionary of Protein Secondary Structures” (DSSP) [17]. The DSSP is reliant on an algorithm which recognises hydrogen bond patterns involving the C=O and N–H backbone atoms and neglects the $\phi$ and $\psi$ dihedral angles. Structural classification is thus in terms of eight qualitatively, different classes and neglects the small deviations in the backbone dihedral angles from the ideality which are often important when it comes to the proteins biological function. Moreover, DSSP analysis is known to be error–prone when it comes to the exact detection of the edges of a given motif [18].

There are various other methods which improve on the original DSSP method. The first of these is the so called STRIDE algorithm [19] in which a consideration of the contributions from the dihedral angles is included in addition to the descriptions of the hydrogen bond patterns in the identifiers for the various classes. Recently a further algorithm for classifying turn structures [20] works by screening and clustering a large dataset taken from a protein data bank. From this analysis the three normal, four open and five reverse turn families emerged along with a number of new turn types.

Despite all this algorithmic sophistication, the assignment of secondary structure to conformations which, like poly-L-proline II structures [21, 22], depart strongly from any ideal backbone structure is still challenging. What makes structural assignment in this “twilight zone” of protein structure [8] so difficult is the high degree of flexibility, which makes the coil state become a kind of catch–all structural assignment unless high resolution data are obtained.
1.4 Intrinsic chirality and protein secondary structures

In section 1.1, only chirality related to stereocenters has been discussed. Such stereocenters have a chirality due to the asymmetry arising from a non–coplanar arrangement of atoms in which one atom is connected to “at least” three further non–equivalent atoms. The most ubiquitous example of this is a carbon bonded to four atoms that lie on the vertices of a tetrahedron centered on the carbon (Figure 1.1). This however is not the only form of chirality that can exist as there are countless other three dimensional structures, which have no stereocenter and are also chiral. An example of an intrinsically chiral structure which does not necessarily need to have any stereocenters is a helicoidal structure. Proteins tend to adopt more chiral (α helix, 3_{10} helix, β turn, PPII) than achiral (β sheets) local, secondary structures.

Fig. 1.3 The most common protein secondary structures: α helix is shown in violet, β sheets in yellow, turn in green and coil in white. H–bonds are shown by dotted lines, each represented with the colour of the secondary structures.
Computational techniques for studying protein dynamics

Several computational techniques are available to study the motions of biological macromolecules, some commonly used examples of which are molecular dynamics, Monte Carlo and NMA approaches. All of these methods generally use the equations of classical mechanics combined with a forcefield which is used to calculate potential energies or forces. These forcefields can be either based on terms which describe the interactions between atoms (atomistic models) or terms which describe the interactions between group of atoms (meso–scale models). If the goal is to analyse the fine detail of local arrangements, atomistic scale simulations are preferred. On the other hand, if the purpose is to obtain an understanding of the spatial arrangements of the macromolecule, coarse grained (meso–scale) approaches are preferred as the longer timescales accessible make it possible to examine huge rearrangements of the system.

The simplest approach to study protein folding within a coarse grained model is provided by the Gō potential [23]. The basic idea of which is to devise a forcefield consistent with the experimentally established structure of the native state, in which only the interactions between Cα–Cα are taken into account. These pairs are within some small cutoff distance (usually 7.5 Å [24] ) in the native state structure. As only Cα atoms are included, the model provides only a very coarse determination of the protein motions.

Another commonly used approach to describe large motions in proteins is Normal Mode Analysis (NMA), which is based on the harmonic approximation of the potential energy. Here the dynamics of the molecule is then described in terms of a collection of independent harmonic oscillators on a basis of normal modes eigenvectors (an orthonormal set of directional vectors that represent the uncoupled motions of the system). Low–frequency oscillations represent large–amplitude, collective motions, which often correlate well with the experimentally observed conformational changes associated with protein function. To perform an NMA one first must minimize the potential energy of the system, calculated using either an atomistic or coarse–grained approach, after which one performs a diagonalization of the matrix of second–derivatives of the potential energy [25] (the Hessian).
The method that can in theory provide the greatest amount of detail about protein motion is Molecular Dynamics (MD), which is made possible by the fact that forces are expressed as the gradient of a potential energy function. MD is a computational technique in which the time evolution of a set of interacting particles (generally atoms or molecules) is followed step by step by integrating their equations of motion. In contrast to Monte Carlo simulations, which employ stochastic dynamics to explore phase space, MD employs deterministic dynamics, which means, that given an initial set of positions and velocities, the subsequent time evolution is completely determined and in principle reversible. The major problem with this technique is that it is currently only possible to run at most a few microseconds of simulation, while most processes of interest in nature take place on the order of seconds. However, more advanced variants of the MD techniques have been developed [26, 27, 28, 29], in which longer timescales can be investigated, and used to obtain intriguing results [30].
3

Employing Chirality to analyse protein motions

3.1

The chirality index

So far, we have discussed the main concepts used in the study of protein structure. However, studying the structural data that can be gleaned from molecular simulation is usually far from simple, because, proteins are not static entities and instead exist as dynamical entities. Typically the dynamical motions of proteins involve local rearrangements around the average structure; namely, that which can be probed experimentally in techniques like NMR. A natural way to describe such local fluctuations in protein structures would involve an examination of changes in the local secondary structure of the protein. However a description of this is beyond most of the standard tools like DSSP [17] or STRIDE [19] that have been described. The exception to this is local chirality which is highly sensitive to local conformational motions and can thus provide insight into local conformational changes and the flexibility of particular protein segments [31].

The chirality method works since it is possible to calculate individual indices for fragments of the protein backbone. For each of these fragments a chirality index can be computed using a method that was proposed by [32, 33] for low mass molecules and which succeeded in analysing facial diastereoselectivity [34]. Here, though instead of calculating the chirality index for the entire molecule, it is calculated for each individual fragment of the molecule separately (as shown in Figure 3.1). The value of the index for each protein fragment, is expressed as follows:

\[ G^{a}N_{b} = \frac{4!}{3N_{a}^{4}} \sum_{P_{i,j,k,l}}^{\text{all}} g_{ijkl} \]  

(3.1)

\[ g_{ijkl} = \begin{cases} \frac{(r_{ij} \times r_{kl}) \cdot r_{il}}{(r_{ij} \cdot r_{kl})^{2} r_{il}} & \text{if } r_{ij}, r_{kl}, r_{ij} \cdot r_{ij} < r_{c}, \\ 0 & \text{otherwise} \end{cases} \]  

\[ \sum_{i,j,k,l \leq N_{b} + a - 1} \]
where a is the first (closest to the N–terminal) atom of a given sequence of $N_a = 15$ consecutive backbone atoms with coordinates $\mathbf{r}$, and $r_c = 12 \text{ Å}$ is a cutoff radius, added to avoid the computation of negligible long–range terms.

Fig. 3.1 A backbone composed of 15 atoms. The $G$ index is calculated for every permutation of four atoms, inside this 15 atoms fragment. Then, second $G$ index is calculated ahead of one atom, here underlined by a line starting from the $C_{\alpha}$.

The efficacy of this method is that the instantaneous value of the chirality indices can be calculated from only a knowledge of the atomic coordinates. What is more, because it is possible to calculate the chirality index in idealized secondary structures (here reported in Table 3.1, Figure 3.2) we can determine which secondary structure a local region of the protein has and how much the structure deviates from that structure during its dynamical motions.

Table 3.1 Average $G$ values and relative standard deviations of $G$ for ideal secondary structures, involving at least $N_R$ residues. Each structure was built by sampling $\phi$ and $\psi$ angles from a gaussian distribution, centered on the ideal $\phi$ and $\psi$ values with sigma=15 degree (see reference [31]).

<table>
<thead>
<tr>
<th>Structure</th>
<th>$\langle G \rangle$</th>
<th>$\sigma_G$</th>
<th>$N_R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$ helix</td>
<td>-.04</td>
<td>0.02</td>
<td>$&gt;3$</td>
</tr>
<tr>
<td>$\beta_{10}$ helix</td>
<td>-.07</td>
<td>0.01</td>
<td>$&gt;3$</td>
</tr>
<tr>
<td>$\beta$ Turn I</td>
<td>-.07</td>
<td>0.01</td>
<td>2,3</td>
</tr>
<tr>
<td>$\beta$ Sheets</td>
<td>+.00</td>
<td>0.01</td>
<td>$\geq 2$</td>
</tr>
<tr>
<td>PPII</td>
<td>+.10</td>
<td>0.03</td>
<td>$&gt;3$</td>
</tr>
<tr>
<td>$\pi$ helix</td>
<td>-.01</td>
<td>0.02</td>
<td>$&gt;3$</td>
</tr>
</tbody>
</table>
3.2 Using chirality to understand protein structure

Any function that purports to be able to determine secondary structure in proteins must provide a range of values for each of the possible secondary structures. To assign these ranges ideal secondary structures were analysed. To do this, the periodicity in the backbone angles ($\phi, \psi$) was used to generate ideal segments of protein structures for which chirality indices were computed. The periodicities employed were: $(-67^\circ, -41^\circ)$ for $\alpha$ helix [35], $(-49^\circ, -26^\circ)$ for $3_{10}$ helix [36], $(-67^\circ, -59^\circ)$ for $\pi$ helix [35], $(-60^\circ, -30^\circ, -90^\circ, 0^\circ)$ for type I $\beta$ turns [37], $(-75^\circ, 147^\circ)$ for PPII helix [38] and $(-130^\circ, 130^\circ)$ for sheets regions [39]. In all cases the $\omega$ angles were kept fixed to the trans value of $180^\circ$. The type I $\beta$ turn conformation is not periodic in proteins, as generally it only ever involves 4 consecutive residues, for the sake of simplicity however it was considered periodic in [31] so that comparison with the other structural motifs is possible.

Figure 3.2 shows the behavior of the $G$ index for the ideal structures along the backbone. Here the index was calculated using $N_a = 15$ and $r_c = 12$ Å. The opposite chirality indices of the right handed $\alpha$ helix (negative) and the left handed poly-L-proline II helix (positive) is worth noting along with the observation that $\beta$ sheets, have no chirality, because of their flat shape.

Figure 3.3 shows the chirality index calculated along the backbone of a typical protein taken from the protein data bank of structures. It is possible to note from this figure that in most of this protein the chirality index is in one of the ranges corresponding to the idealized structures. In addition to its ability to recognise secondary structure types, the chirality index is very sensitive to poly-L-proline dihedrals. A positive peak in the $G$ value indicates that at least one aminoacid in a given protein region has the PPII structural motif. Figure 3.4
Employing Chirality to analyse protein motions shows the \( G \) value of a model poly-L-proline peptide fragment (1JMQ, residues 51-60), where there is known to have poly-L-proline dihedral angles. A good overlap between the PPII ideal structure and the PPII model peptide exists in the residues labeled 3-5 in the figure, however after residue 5 the \( G \) value of 1JMQ drops as it must take into account residues 7 and 8 which are not in the PPII conformation. Detection of the PPII structure using DSSP-like algorithms is hampered because prolines do not form hydrogen bonds. Therefore, despite the fact that this structure is observed in other polypeptides, its extended conformation (9.3 Å pitch) is not conducive to the assignment of a well defined hydrogen bond pattern, which means they are commonly misclassified as loops or coils [22]. \( G \)'s sensitivity to these PPII structures makes the \( G \) function an important tool for the better identification of this particular structure class.

**Fig. 3.3** Chirality index, \( G \), along the backbone for hemoglobin protein, from Xray structure. Typical secondary structures, with the negative periodicity concerning the \( \alpha \) helices, and with the \( G \) typical values for the other secondary structures (cf Figure 3.2) are easily identified. The DSSP assignment is also plotted as the numeric code: 3\(_{10}\)=0, turn=1, bend=2, bridge=2.5, \( \alpha \)=3, sheets=4, coil=5.

### 3.3 Chirality index as a tool for monitoring protein dynamics

Because the chirality index can be calculated easily from only the atomic coordinates, it can prove useful in examining protein dynamics, e.g., the dynamics of a transition from an \( \alpha \) helix to a 3\(_{10}\) helix transition or from a coil to a helix (see Figures 3.5 [a] and [b], respectively). It is particularly useful for the examination of folding transitions, as it contains information on both the \( \phi \) and \( \psi \) dihedral angles. Moreover, \( G \) index is stable during dynamics and in response to random thermal fluctuations [31], which proves that it can be used to follow dynamical processes. Furthermore, the average value of the chirality index
3.3 Chirality index as a tool for monitoring protein dynamics

Fig. 3.4 Chirality index, $G$, along the backbone for poly-L-proline II model peptide. The DSSP assignment is also plotted as the numeric code: 3$\text{I}$O=0, turn=1, bend=2, bridge=2.5, $\alpha$=3, sheets=4, coil=5.

can be used to detect protein motions and therefore useful for studying protein flexibility. This is an important issue in the study of intrinsically disordered proteins.

Figure 3.6 reports the averages and standard deviations of the chirality indices calculated along the backbone of rigid (a) and flexible structures (b), respectively during the course of an MD trajectory. As previously discussed the average chirality provides information on the average local structure adopted, which, combined with the standard deviations of $G$, provides information on the flexibility of local regions in the protein. The typical method used to measure this quantity is the Root Mean Square Deviations (RMSD), which, in fact measures only the difference in the structure from a given reference configuration and is not particularly sensitive to the changes in dihedral angles that local chirality measures. In particular high RMSD indicates that the structure highly differs from the reference structure. However, the reference structure may not be the true average structure and, even if it is, the RMSD may not represent significant motions particularly if the single configurations examined are far from the average structure. The standard deviation of the chirality index makes sense, however, as it is a deviation about an average in a quantity that measures secondary structures. This average quantity is zero if no secondary structure is present, while, if a secondary structure is present then a constant value for the average chirality index is returned and this depends on the average secondary structure. Meanwhile, the standard deviation measures directly the flexibility of the secondary structure in a local region of the protein.
Fig. 3.5 [a]: $\alpha$ to $3_{10}$ helix transition, followed by chirality. [b]: Coil to helix transition, underlined from the positive-to-negative switch of $G^\alpha$. 

Employing Chirality to analyse protein motions
Fig. 3.6 Standard deviations of $G$ among a) hemoglobin 1-18 and b) 1REI 1-30 configurations. It is worth to note the persistence of the chirality index for a rigid peptide, as the first helix of hemoglobin, and the high values of the standard deviations for a flexible peptide, as the trajectories of the first 30 residues of 1REI immunoglobulin antigen. As comparison the $G$ from PDB and from the trajectories is shown.
3.4 Chirality and Circular Dichroism

Circular Dichroism (CD) experiments are deeply connected with chirality because the L and D enantiomers absorb polarised light differently. These experiments thus provide a useful probe for chiral molecules but not for the achiral ones. The circular dichroism spectra of proteins are strongly related to the backbone dihedral angles, $\phi$ and $\psi$ and can be used to obtain insight into the average structure the proteins adopt in solution. What is more a combination of both the information deriving from CD and molecular dynamics may help in unravelling the conformational ensemble adopted in solution. As an example we consider the CD spectra of the avian prion hexarepeat region as a function of pH [16]. Prion proteins (PrP$^C$) are glycoproteins that in mammals, but not in avians, can cause prion diseases, which involve the incorrect folding (misfolding) of proteins. It is commonly believed that this misfolding involves a conversion from a structure rich in $\alpha$ helices to a $\beta$–sheet enriched pathogenic isoform (PrP$^{Sc}$). Until now, NMR structure determination of the prion N–terminal portion has been hampered because of its flexibility. Mammal and avian proteins show different N–terminal tandem repeats, PHGGGWGQ and PHNPGY respectively. Both of these contain histidine however, only the avian proteins involve tyrosine in their primary sequence. These residues are of particular interest because they are both highly sensitive to pH variations, having average $pK_a$ values of 6.1 and 9.9 respectively [16].

The far UV–CD spectra of the avian tetra–hexarepeat (PHNPGY)$_4$, reported in Figure 3.7, shows a signal varying with pH. At pH 4, a pH at which one would expect all histidyl residues to be protonated (average $pK_a$ 6.1), the spectrum is broad and has a minimum at 203 nm, a weak shoulder around 216 nm and a maximum at 230 nm. In general, this shape indicates an equilibrium between different conformations, suggesting the presence of other secondary structure elements besides the random coil. This spectrum shape is similar to that found for shorter peptide fragments as reported in [40, 15]. For these smaller fragments the presence of both random coil and $\beta$–turn structures was suggested for those primary sequences that encompass a PXXP motif, a sequence that normally favors $\beta$–turn and/or poly-L-proline II structures [41]. When the pH is increased the spectrum of the avian tetra–hexarepeat (PHNPGY)$_4$ changes in a different way as shown in Figure 3.7. Up to pH 7 there is no change to the band centered at 230 nm but there is a shift in the position of the minimum, which takes it towards 200 nm, and a decrease in the bands intensity. Further increases in the pH cause a general broadening of the spectra and a significant decrease of the signal at 230 nm. Finally, at pH 10, a new maximum at 250 nm is observed. It is well known that aromatic side–chains give rise to a contribution to the far UV–CD spectra of peptides and that in phenols this far UV signal is red–shifted in the deprotonated phenolate ion [42, 43]. Hence, the new maximum, observed at 250 nm, is attributable to the deprotonation of...
3.4 Chirality and Circular Dichroism

tyrosine residues. Furthermore, the decrease and subsequent disappearance of the positive maximum in the signal at 230 nm is also related to the protonation state of the tyrosine residue as it is caused by the phenolic group on it. Alongside the evidence the UV–CD spectrum provides about the protonation state of the tyrosine residue changes in the peptide secondary structure, that occur as a function of pH, can also be detected. For instance, the strong positive band at 190 nm (Figure 3.7) and the shoulder found at 216 nm are features typical of β turn like conformations [44, 45, 46], which appear to be predominant at neutral and basic pH. The weakening of these signals at lower pH suggests that the equilibrium configuration is shifted, thus it would seem that the conformation adopted is strongly dependent on the protonation state of the histidine and tyrosine residues.

A chirality analysis of the conformational ensemble obtained from molecular dynamics simulations reveals that there is an enhancement in the number of turn conformations as pH increases, (this is indicated by the negative sharp peaks shown in the bottom side of Figure 3.7). In particular, at acidic pH (LH⁴⁺), the broad negative peak centered at residue 7 (Pro), which includes residues 6-8 (Tyr, Pro and His respectively), confirms the presence of a 3₁₀ helix structure [31], while the negative sharp peak centered on residue 11 (Gly) suggests the presence of a turn which includes residues 9-12 (Asn, Pro, Gly and Tyr respectively). At neutral pH (LH₄) chirality analysis indicates the presence of a turn region which includes residues 4-6 (Pro, Gly, Tyr respectively). This turn is highly flexible however as indicated by the wide standard deviation and the average value for this negative peak is close to the upper bound of values observed for this particular structural motif [-0.1:0.06]. Two further turn regions, which are signaled by much stronger negative peaks, are observed at neutral pH and are centered on residues 11 and 17 (Gly). At basic pH (L⁻) those negative peaks that were observed at neutral pH show smaller values for their standard deviations. This suggests that the stability of turn regions, inside the peptide, is enhanced by the increasing pH. Moreover, by examining the chirality pattern along the backbone, we can observe that on changing the pH the chirality index varies most strongly for residues 4-8 (Pro, Gly, Tyr, Pro and His respectively) and for the C–terminal region. These are the regions in which His8 and Tyr24 are involved and this indicates the pivotal role that these residues play in any conformational change.

The turn structures observed in the molecular dynamics simulations are of course stabilized by the large number of proline residues. However, in this system it would appear that histidine, particularly histidine 8, also plays a key role in stabilizing turn regions. This is evidenced by the fact that at physiological pH, when this residue would be expected to be deprotonated, turn regions become more stable.

The above example has demonstrated how a combination of experimental CD and chirality analysis of simulation data can be used to explain how the conformational states accessible to this molecular system depend on pH. Furthermore, Figure 3.7 shows that at physiological pH the chirality pattern of the
tetra hexarepeat region reflects the periodicity in the primary structure. This finding reinforces the suggestion that local folding events mainly drive the global protein folding toward the tertiary structure.

![Fig. 3.7](image-url)

Fig. 3.7  Up on the left, CD spectra of avian tetra-hexarepeat fragment as a function of pH. At basic pH the shoulder approximately at 216 nm, reveals the formation of type I β turn. Up, On the right, typical conformations of avian tetra-hexarepeat fragment at acidic (LH$_4^{8+}$), neutral (LH$_4$) and basic (L$^-_4$) pH. Bottom, chirality index along the backbone of the avian tetra-hexarepeat trajectories in the three protonation states. The shift toward type I β turn is underlined from the negative sharp peaks, adopted mainly in the basic state, L$^-_4$, as shown in the CD spectra. The peptide structures are shown according to a colour code: blue for turn, violet for 3_10 helix and gray for coil regions. The side chains hydrogen bonds are circled in red. N and C termini are shown respectively from left to right.
4

Perspectives

The connections between chirality and biosystems, particularly proteins, have been discussed here at length. It has been shown that protein structures are particularly amenable to descriptions based on chirality concept since proteins have intrinsically asymmetric structures. In particular, a chirality index has been introduced that can be used to analyse the extent to which proteins are folded and the local flexibility of the protein. This index has well defined values for the typical secondary structure elements and is particularly effective in detecting poly-L-proline II motifs, often misclassified as coils when other techniques are employed. One word of caution regarding coil states must be mentioned here: Coils are often considered to be unstructured regions in the protein and thus on the face of it would be expected to be achiral. However, because proteins are isotactic, regions of proteins that adopt coiled states should present some degree of left handed chirality and thus should not have a chirality index of zero [31].

A particularly interesting class of protein structures are the so–called coiled–coils, examples of which include keratine and the muscle protein tropomiosyn. These structures consist of α helices wound together in a manner that is stabilized by hydrophobic interactions. What makes them intriguing from a chirality point of view is that while the α helices themselves are right handed the structure is often overall left handed. Recently a correlation between the chirality of the coiled coils and their mechanical properties has been found [47]. This correlation links the plasticity of coiled coil structures in a selection of biochemical partners with the partner intrinsic chirality.

A particularly intriguing question related to these discussions is the chirality of amyloid suprastructures. Amyloids, like Aβ_{1–40} and hen lysozyme, are fibrillar aggregates of proteins with a characteristics cross–β conformation that until now were generally accepted to have a structure composed of left–handed helices. Recently however a surprising discovery was made concerning the peptide of serum amyloid A protein (SAA_{1–12}). It was found that structures composed of aminoacids with S–stereocenters, form right handed helices, while those composed of only R–stereocenters forms the left–handed suprastructures [48].
Hence, it would seem that supramolecular chirality depends on the structure organization and it is not necessarily the same for all fibrils.

Fascinating proteins with high symmetries are found in viruses. A virus consists of a fragment of nucleic (DNA) or ribonucleic acid (RNA) enclosed in a highly regular arrangement of capsomer proteins. In spherical viruses this external, capsid shell has its individual capsomer molecules located on the vertex positions of an equilateral triangulation of a sphere [49] with an icosahedral symmetry [50, 51]. Viruses can also adopt non–spherical geometries in which there is a skew in the capsid, which can cause it to become chiral. These chiral shaped capsids are found to require a higher rupture force than the perfect icosahedral capsids [52].

Proteins play a central role in the life sciences as both the molecular workhorses and building blocks of all living things. Simulation and experiment are beginning to provide genuine insight into how these complicated molecules function in isolation but, how they interact or organize themselves to form a cell is an area where a lot of work is still required. Recently though, in a discovery which surprised the community, cells were found to have a chiral organizing principle, which works in terms of asymmetrical compartmentalization and probably originates in the centrosome [53]. Xu et. al. [53] argue that intrinsic functional chirality is a property of eukaryotic cells that probably confers on them a selective advantage during the course of evolution. This finding suggests that chirality is a central organising principle in life that confers order at every length scale up to the full cell. This organisation on the basis of chirality must have emerged at some point in the past as an evolutionary response that protected the growth of life.
Acknowledgments

A number of people have contributed for the realization of the chirality index methodology. In particular, it is worth to note the combined effort of the authors Luca Muccioli, Roberto Berardi and Claudio Zannoni in the development of chirality method applied to protein structures.

Diego La Mendola and Enrico Rizzarelli have made possible the combined experimental–computational approaches for studying the pH dependence of the N–terminal region of avian prion protein, by using Circular Dichroism and Chirality from MD simulations.

I gratefully acknowledge Claudio Zannoni, Enrico Rizzarelli and Michele Parrinello for precious suggestions and discussions.

A particular acknowledgment goes to Gareth Tribello, Meher Prakash Ayalasomayajula and Francesco Mauriello for several useful suggestions in writing this manuscript.
References


References


