



Bachelor's thesis
Bachelor's Programme in Physical Sciences
Theoretical physics

Analytical method for calculating geometrical metrics for protein structures

Sakari Johannes Pirnes

January 14, 2022

Supervisor(s): Dr. Waldemar Kulig, Dr. Giray Enkavi

Examiner(s): Prof. Ilpo Vattulainen

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE

PL 64 (Gustaf Hällströmin katu 2a)
00014 Helsingin yliopisto

Tiedekunta — Fakultet — Faculty Faculty of Science	Koulutusohjelma — Utbildningsprogram — Degree programme Bachelor's Programme in Physical Sciences Theoretical physics
Tekijä — Författare — Author Sakari Johannes Pirnes	
Työn nimi — Arbetets titel — Title Analytical method for calculating geometrical metrics for protein structures	
Työn laji — Arbetets art — Level Bachelor's thesis	Aika — Datum — Month and year January 14, 2022
Sivumäärä — Sidantal — Number of pages 39	
Tiivistelmä — Referat — Abstract The α -helix and β -strand are the most common motifs in the secondary structure of proteins. The tilt and rotation of transmembrane α -helices are key parameters in cell signaling, where the signals are presumably transmitted mechanically through the membrane by dimerization of two transmembrane α -helices. The tilt is quantified by the angle between a chosen reference vector and the axis of the helical structure. For a bent or kinked helical structure the axis is defined piecewise. In this thesis, an analytical method to define the axis piecewise is derived and shown to hold for α -helix and β -strand like secondary structures. The rotation describes which amino acid residues of a tilted helical structure are pointing towards the chosen reference. The rotation is quantified by the rotation angle, which for a fixed tilt describes how the helical structure rotates about its own axis. In this thesis, a local rotation angle is constructed to define how a single residue of a helical structure is pointing towards the chosen reference vector. An analytical expression of the local rotation angle is derived and proven to be unique. The rotation angle is constructed from the local rotation angles. The end result is a robust analytical method to analyze the tilt and rotation of straight, bent and kinked α -helix and β -strand like secondary structures. Two example analyses are provided, where the method is used to analyze the tilt, rotation and kink of transmembrane α -helices. First example considers a kinked transmembrane α -helix and in the second one glycophorin A transmembrane α -helix monomer and dimer are compared.	
Avainsanat — Nyckelord — Keywords Tilt, Rotation, Alpha helix, Beta strand, Membrane proteins	
Säilytyspaikka — Förvaringsställe — Where deposited	
Muita tietoja — Övriga uppgifter — Additional information	

Tiedekunta — Fakultet — Faculty Faculty of Science	Koulutusohjelma — Utbildningsprogram — Degree programme Bachelor's Programme in Physical Sciences Theoretical physics
Tekijä — Författare — Author Sakari Johannes Pirnes	
Työn nimi — Arbetets titel — Title Analytical method for calculating geometrical metrics for protein structures	
Työn laji — Arbetets art — Level Bachelor's thesis	Aika — Datum — Month and year January 14, 2022
Sivumäärä — Sidantal — Number of pages 39	
Tiivistelmä — Referat — Abstract Proteiinien yleisimmät sekundaarirakenteet ovat α -heliksi ja β -laskos. Solukalvossa olevien α -heliksien kallistus ja rotaatio ovat olennaisia parametreja soluviestinnässä, jossa viestit välittyvät solukalvon läpi oletettavasti mekaanisesti kahden solukalvossa sijaitsevan α -heliksin dimerisoitumisella. Kallistus kvantifioidaan kallistuskulmana helikaalisen rakenteen akselin ja valitun vertailukohdevektorin välillä. Taipuneelle tai mutkaiselle helikaaliselle rakenteelle akseli määritellään paloittain. Tässä opinnäytetyössä johdetaan analyyttinen menetelmä akselin paloittain määrittelemiseksi ja sen osoitetaan toimivan α -heliksin ja β -laskoksen kaltaisille sekundaarirakenteille. Rotaatio kuvailee, mitkä kallistuneen helikaalisen rakenteen aminohapot osoittavat vertailukohdevektorin suuntaan. Rotaatio kvantifioidaan rotaatiokulmana, joka kiinnitettylle kallistuskulmalle kertoo kuinka helikaalinen rakenne pyörii oman akselinsa ympäri. Tässä opinnäytetyössä konstruoidaan lokaali rotaatiokulma, joka kuvailee helikaalisen rakenteen yksittäinen aminohappo osoittaa vertailukohdevektorin suuntaan. Lokaalin rotaatiokulman analyyttinen ratkaisu johdetaan ja osoitetaan uniikkiseksi. Rotaatiokulma muodostetaan lokaalien rotaatiokulmien avulla. Lopputuloksena on luotettava analyyttinen menetelmä suorien, taipuneiden ja mutkaisten α -heliksin ja β -laskoksen kaltaisten sekundaarirakenteiden kallistuksen ja rotaation analysoimiseen. Lisäksi esitetään kaksi esimerkkianalyysiä, joissa menetelmää käytetään analysoimaan solukalvossa sijaitsevien α -heliksien kallistusta, rotaatiota ja mutkaa. Ensimmäisessä esimerkissä tarkastellaan solukalvossa sijaitsevaa, mutkalla olevaa α -heliksiä ja toisessa verrataan glykoforiini A:n α -heliksi-monomeeriä ja -dimeeriä.	
Avainsanat — Nyckelord — Keywords Tilt, Rotation, Alpha helix, Beta strand, Membrane proteins	
Säilytyspaikka — Förvaringsställe — Where deposited	
Muita tietoja — Övriga uppgifter — Additional information	

Contents

1	Introduction	2
2	Background for the method	3
2.1	Helical secondary structures	3
2.2	Tilt and rotation	7
3	Analytical method for the tilt and rotation	10
3.1	Tilt	10
3.1.1	Piecewise definition of helical structure and its local axes	11
3.1.2	The axis as an average of local axes	11
3.1.3	The kink angle	12
3.1.4	The tilt vector and local tilt	12
3.1.5	Finding the local axis	13
3.2	Rotation	18
3.2.1	Definition of the local rotation angle	18
3.2.2	Rotation angle as an average of local rotation angles	20
3.2.3	Difference in rotation	22
3.2.4	Constructing the midpoint	22
3.2.5	Finding the local rotation angle	23
4	The method in practice	27
4.1	Example 1: Kinked transmembrane α -helix	27
4.2	Example 2: Difference between glycophorin A monomer and dimer	34
5	Conclusions	37
	Bibliography	39

1. Introduction

In this thesis an analytical method is presented to compute tilt and rotation for α -helix and β -strand like secondary structures. The α -helix and β -strand are the most common motifs in the secondary structure of proteins, they both have a helical arrangement. Tilt describes how a helical structure is tilted from a chosen reference. Rotation describes which side of a tilted helical structure is pointing towards the chosen reference. The tilt and rotation of transmembrane α -helices are key parameters in cell signaling, where the signals are presumably transmitted mechanically through the membrane by dimerization of two transmembrane α -helices.

The tilt and rotation are geometric quantities and thereby it is necessary to interpret the geometry of the helical structures in order to quantify them. Secondary structure is the local three dimensional structure of an amino acid sequence. Each amino acid has a central carbon atom, α -carbon. The α -helix and β -strand like helical structures can be geometrically characterized by a helix curve going through the α -carbon atoms from each amino acid residue. But not all helical structures are straight, they can be bent or kinked, and so, all α -carbon atoms don't lie in the same helix curve and thereby piecewise characterization of helical structure is needed. After the robust method is constructed to handle straight, bent and kinked helical structures, two example analyses about the tilt, rotation and kink of transmembrane α -helices are presented. We begin with the background of α -helix and β -strand like helical structures and the tilt and rotation.

2. Background for the method

In this chapter, we go through the background for the tilt and rotation of protein secondary structures with helical arrangements. We build our way from amino acids to secondary structures and then finally discuss the tilt and rotation.

2.1 Helical secondary structures

Amino acids

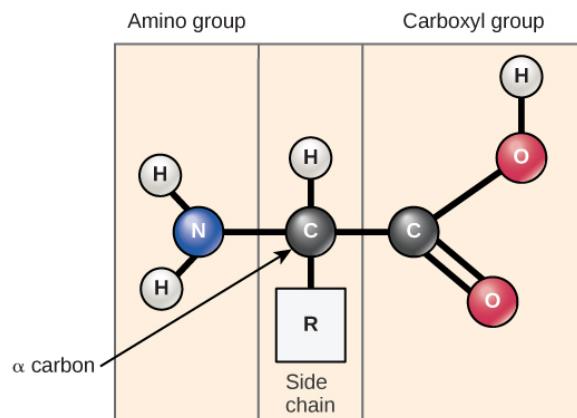


Figure 2.1: General structure of an amino acid. A carboxyl group (COOH), an amino group (NH₂), hydrogen (H) and a side chain (R) are bonded to the α -carbon. (Image credit: OpenStax Biology.)

Amino acids are the building blocks of proteins. An amino acid consists of an amino group, a carboxyl group, hydrogen, and a side chain bonded to a central carbon atom, alpha carbon (α -carbon) (Fig 2.1). The standard genetic code includes 20 different side chains and hence 20 different amino acid. The side chains differ by their charge, polarity and size. [1]

Primary structure

A polypeptide chain is formed by polymerization of amino acids. A carboxyl group of an amino acid is attached to a amino group from another amino acid in a condensation

reaction. The condensation reaction forms a peptide bond between the two amino acids and produces a water molecule (Fig 2.2). The linear order of amino acid residues along the polypeptide chain is the primary structure or sequence. The residues of the primary structure are usually indexed with integers. [1]

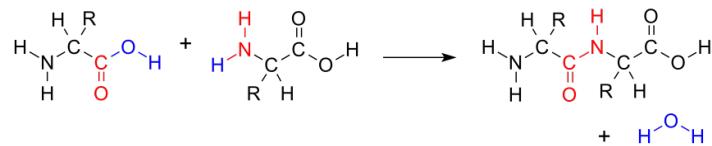


Figure 2.2: Condensation reaction between two amino acids to form a peptide bond.

Secondary structure

The local three dimensional shape of a primary structure is called a secondary structure. The basic secondary structures are the alpha helix (α -helix), the beta strand (β -strand) and turns, while all other structures are variations of these basic structures [1]. We will focus on the helical structures, α -helix and β -strand like structures.

Alpha helix

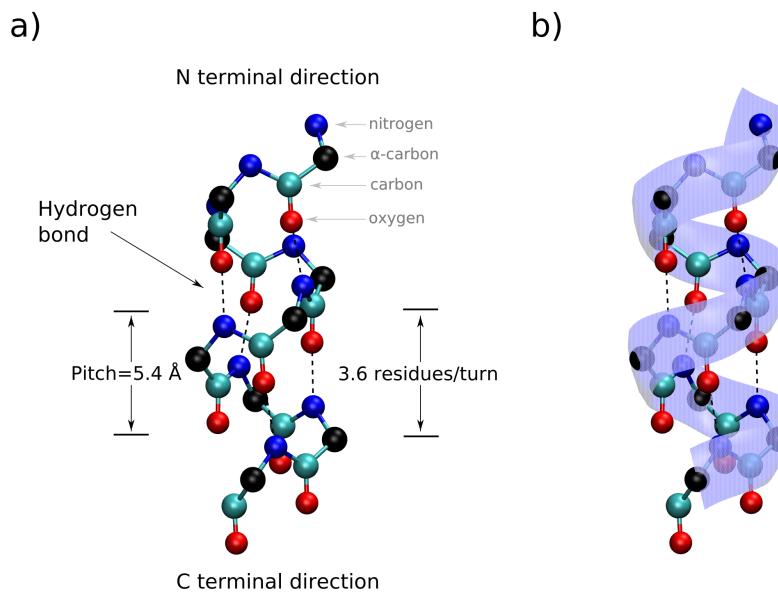


Figure 2.3: The ideal α -helix. Only heavy atoms of the backbone are shown for clarity. Figure b) is same as figure a) but with additional cartoon representation of a α -helix which is commonly used.

The α -helix is the most common structural motif and is a right-handed helical arrangement with 3.6 residues per turn and a translation distance of 1.5 Å[1]. The translation distance is the distance along the axis of the helix from a residue to the preceding

residue. Hence for α -helix, the height of one complete turn, the pitch, is $3.6 \cdot 1.5 \text{ \AA} = 5.4 \text{ \AA}$. In Fig 2.3 only heavy atoms of the backbone are shown. The side chains of the α -helix are pointing outwards away from the backbone [1]. Hydrogen bonds in the backbone between the amine hydrogen from residue i and the carbonyl oxygen of residue $i + 4$ keeps the structure very stable (Fig 2.3) [1]. Less stable structural variations of the α -helix are 3_{10} -helix and π -helix [1]. The dominant hydrogen bonds of 3_{10} -helix and π -helix are between residues $(i, i + 3)$ and $(i, i + 5)$, respectively [1]. Therefore, when compared to the α -helix, the 3_{10} -helix is more elongated with a pitch of 6.1 \AA and the π -helix less elongated with a pitch of 5.2 \AA [1].

Beta strand

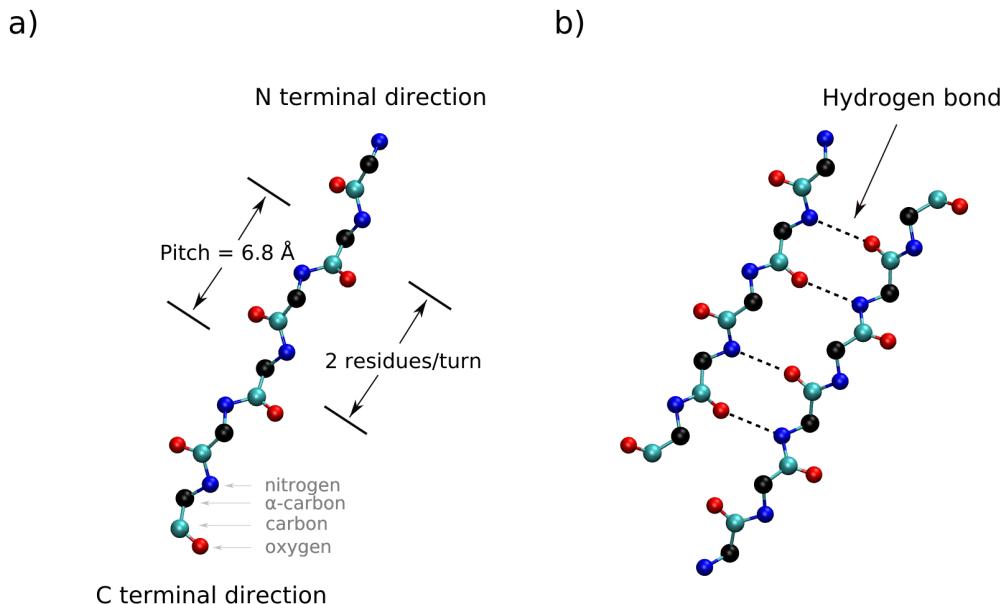


Figure 2.4: The ideal β -strand is shown in figure a). A small element of a β -sheet is shown in figure b). Only heavy atoms of the backbone are shown for clarity.

The β -strand (Fig 2.4a) is a helical arrangement with 2 residues per turn, 3.4 \AA translation distance and 6.8 \AA pitch [1]. Compared to the α -helix, the β -strand is very elongated. Therefore there are not enough local interactions to make the β -strand largely stable [1]. A stable sheet-like structure, called β -sheet (Fig 2.4b), is formed when two or more β -strands are connected with hydrogen bonds [1].

Helix curve through alpha carbons

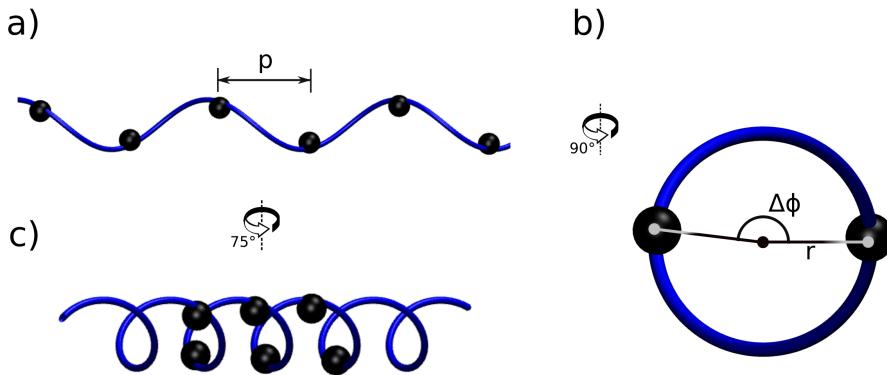


Figure 2.5: A blue helix curve fitted into black α -carbons of a β -strand. a) A parallel side view and the translation distance p between two consecutive α -carbons. b) Top view of the turn angle $\Delta\phi$ between two consecutive α -carbons and the radius r of the helix curve. c) Rotated view of a).

All four structures, α -helix, 3_{10} -helix, π -helix and β -sheet, have a helical arrangement. The helical arrangement can be characterized by a helix curve going through the α -carbon atoms. The reason why α -carbons have been chosen is their stability, α -carbon is the central heavy atom of a amino acid (Fig 2.1). The helix curve is parameterized by its radius r and the pitch $P = p2\pi/\Delta\phi$, where p is the translation distance, $2\pi/\Delta\phi$ number of residues per turn and $\Delta\phi$ is the turn angle between between two consecutive α -carbons (Fig 2.5). [2]

Table 2.1: The normal distribution parameters (μ, σ) for r , p , $\Delta\phi$ presented in [2]. The secondary structures are determined and divided into four groups (H, G, I, E) by DSSP. The helix curve parameters r and p are in ångström and the turn angle $\Delta\phi$ is in degree.

	α -helix (H)			3_{10} -helix (G)			π -helix (I)			β -strand (E)		
	r	p	$\Delta\phi$	r	p	$\Delta\phi$	r	p	$\Delta\phi$	r	p	$\Delta\phi$
μ	2.314	1.516	100.1	2.109	1.829	107.4	2.779	1.196	82.8	0.96	3.34	177.9
σ	0.061	0.086	2.56	0.118	0.138	5.90	0.086	0.056	2.80	0.146	0.158	2.12

In Table 2.1 the normal distribution mean value μ and standard deviation σ are listed for the helix curve parameters r , p and $\Delta\phi$ of α -helix, 3_{10} -helix, π -helix and β -sheet. These parameters are computed by finding the helix curve that best fits four consecutive α -carbon atoms [2]. The secondary structures are identified by DSSP. The DSSP algorithm calculates the most likely secondary structure assignment from atom positions and estimated hydrogen bond energies between all atoms of a given 3D structure [3]. The characterization of the helical arrangement by the helix curve going through the

α -carbon atoms is the geometrical foundation for the analytical solutions of the tilt and rotation in chapter 3.

2.2 Tilt and rotation

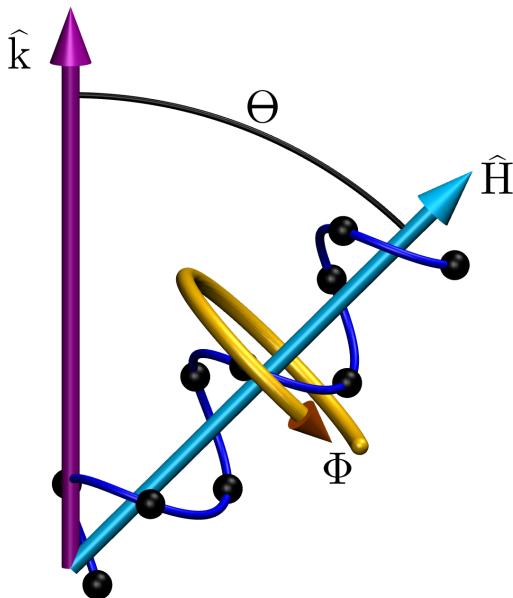


Figure 2.6: A blue helix curve fitted into black α -carbons, a cyan vector representing the axis $\hat{\mathbf{H}}$ of the helix curve, tilt angle Θ between the axis $\hat{\mathbf{H}}$ and the purple reference vector $\hat{\mathbf{k}}$ and orange circular arrow representing the rotation angle Φ .

The signaling mechanism of receptor tyrosine kinases (RTKs) is a great example of the relevance of tilt and rotation. The RTKs are a family of single pass transmembrane proteins, and they have a key role in cellular processes, such as cell growth, apoptosis, division and differentiation, metabolism and migration [4]. Also faults in RTKs are known to lead to malignancies [5]. Receptor tyrosine kinases transmit signals from outside the cell through the membrane into the cell. The signals are presumably transmitted mechanically through the membrane by dimerization of two transmembrane α -helices, that is, two α -helices embedded into the membrane. The differences in the configurations of the active and inactive states are characterized alongside translation by the tilt and rotation [6]. The activity of the state refers to whether a signal is transmitted or not.

The tilt and rotation of a single transmembrane α -helix are determined by its amino acid sequence. Interior of the membrane is hydrophobic and the water both above and below the membrane is polar. To minimize the energy caused by the orientation of the transmembrane α -helix, the number of polar and charged residues in or pointing towards the hydrophobic interior of the membrane and the number of hydrophobic residues in or

pointing towards the polar water has to be minimized. For a straight and rigid transmembrane α -helix these two numbers are adjusted by the tilt, rotation and translation along the membrane normal [6]. The number of residues inside the membrane is adjusted by the tilt angle Θ relative to the membrane normal. The rotation angle Φ relative to the membrane normal adjusts which residues in the boundary of the membrane and water point into the membrane and which into the water.

For a secondary structure with a helical arrangement, tilt and rotation are defined from the helix curve characterized by the α -carbons (Fig 2.6). Tilt describes how the helical structure is tilted away from a chosen reference of interest [6]. The tilt is naturally quantified by the tilt angle $\Theta \in [0, \pi]$ which is the angle between the axis $\hat{\mathbf{H}}$ of the helix curve and the chosen reference vector $\hat{\mathbf{k}}$ (Fig 2.6). Rotation describes which side of the helix, i.e. which residues, point in the direction of a chosen reference vector $\hat{\mathbf{k}}$ [6]. Hence, if $\hat{\mathbf{H}} = \pm\hat{\mathbf{k}}$, then rotation is not defined, since all residues around the helix point similarly towards the reference $\hat{\mathbf{k}}$. Therefore, when we talk about rotation (relative to $\hat{\mathbf{k}}$) we are always assuming that $\hat{\mathbf{H}} \neq \pm\hat{\mathbf{k}}$, i.e $0 \neq \Theta \neq \pi$. Rotation is quantified by the rotation angle $\Phi \in]-\pi, \pi]$, which for a fixed tilt angle defines how the structure rotates about its own axis $\hat{\mathbf{H}}$. Note that the tilt angle Θ and rotation angle Φ are measured relative to the chosen reference vector $\hat{\mathbf{k}}$ and thereby they are invariant under rotations around the reference vector $\hat{\mathbf{k}}$. Therefore, a way to see that for a single transmembrane α -helix the membrane normal is a natural choice for the reference vector is that the membrane has a symmetry around it. For an individual α -helix of a transmembrane α -helix dimer there is no such symmetry, since the other α -helix is right next to it. If one is interested in which residues point from one of the dimer's two α -helices towards the other α -helix, it is quantified by the rotation relative to a reference vector pointing from one of the α -helices towards the other.

There are two kind of main motions to change the rotation angle Φ . First one can be seen from Fig 2.6. For a fixed tilt angle $\Theta \in]0, \pi[$, rotating the structure about its axis $\hat{\mathbf{H}}$ is the only way to change the rotation angle Φ . This means that the tilt angle Θ and the rotation angle Φ are independent, since we can always tilt the helix towards or away from the reference vector $\hat{\mathbf{k}}$ without changing the side of the helix that points towards $\hat{\mathbf{k}}$ as long as $0 \neq \Theta \neq \pi$. The other motion to change the rotation follows from the fact that the rotation is not defined when $\hat{\mathbf{H}} = \pm\hat{\mathbf{k}}$. We can tilt $\hat{\mathbf{H}}$ closer and closer to $\hat{\mathbf{k}}$ without changing the rotation angle Φ , once $\hat{\mathbf{H}}$ is aligned to $\hat{\mathbf{k}}$ the rotation is not defined. Now, we can start tilting in any direction we want, that is, we can choose the rotation angle Φ freely.

Some of the transmembrane α -helices are not even close to being straight. A transmembrane α -helix can be bent or it can have a kink, where the kink means a sharp enough bend (Fig 3.1b). In the following chapter, we will define and derive a robust method to

compute the tilt and rotation for straight, bent and kinked helical structures. After that we will come back to the transmembrane α -helices in chapter 4 with an example analysis of a kinked transmembrane α -helix and another example related to the transmembrane α -helix dimers.

3. Analytical method for the tilt and rotation

3.1 Tilt

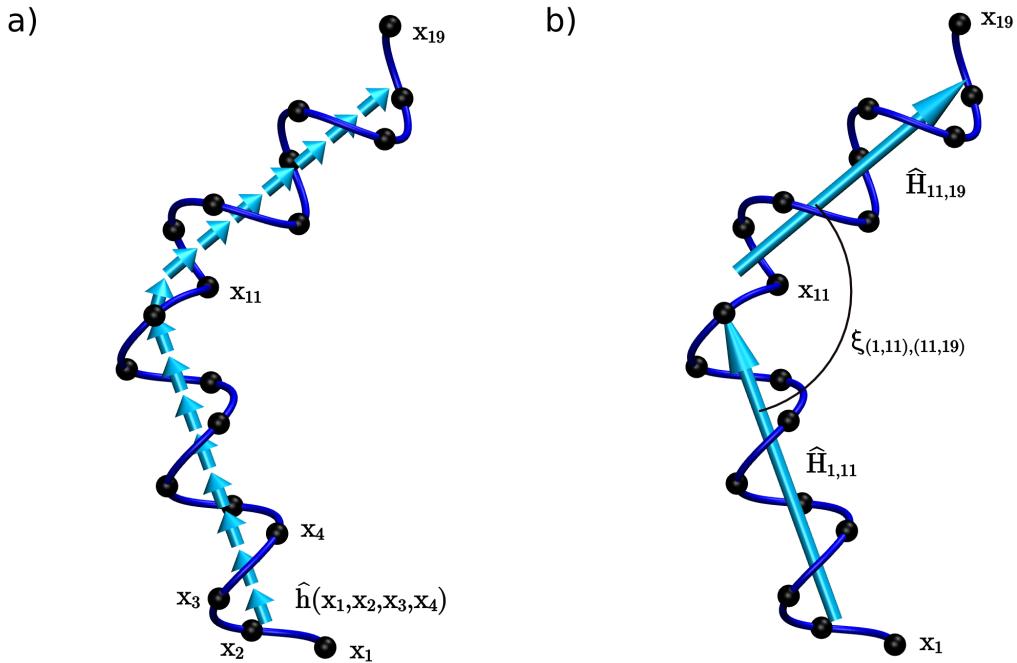


Figure 3.1: A blue kinked helix curve fitted into 19 black α -carbons, with positions denoted by x_i . a) Local axes $\hat{h}(x_i, x_{i+1}, x_{i+2}, x_{i+3})$ for all $i = 1, 2, \dots, 16$ are shown as cyan vectors. b) The axes $\hat{H}_{1,11}$ and $\hat{H}_{11,19}$ of residues $\{1, \dots, 11\}$ and $\{11, \dots, 19\}$, respectively, are shown as cyan vectors. The black arc represents the kink angle $\xi_{(1,11),(11,19)}$ between the axes $\hat{H}_{1,11}$ and $\hat{H}_{11,19}$.

In the previous section 2.2, we discussed the tilt of a helical structure and how it is naturally quantified by the tilt angle. Assume that residues are labeled by the integers $m, m + 1, \dots, n$ form a straight helical structure, i.e. a helix curve fits well into the α -carbons of the helical structure. We denote the axis of the residues m, \dots, n by $\hat{H}_{m,n}$. Then the tilt angle $\Theta_{m,n} \in [0, \pi]$ relative to a chosen reference vector \hat{k} is the angle

between the unit vectors $\hat{\mathbf{H}}_{m,n}$ and $\hat{\mathbf{k}}$,

$$\Theta_{m,n} = \arccos(\hat{\mathbf{k}} \cdot \hat{\mathbf{H}}_{m,n}), \quad (3.1)$$

where " \cdot " is the dot product and $\arccos : [-1, 1] \rightarrow [0, \pi]$ is the inverse of cosine. To be able to compute the tilt angle $\Theta_{m,n}$ from the given helical structure we have to find the axis $\hat{\mathbf{H}}_{m,n}$ as a function of the α -carbon atom positions. We also shortly discussed other shapes than straight helical structures, namely bent and kinked helical structures (section 2.2). An α -helix-like kinked helical structure is shown in Fig 3.1. Our goal is to find a robust analytical solution that works generally for helical structures without requiring the information whether it is straight or not. Natural way to do this is to define the helical structure and its axis piecewise (Fig 3.1a).

3.1.1 Piecewise definition of helical structure and its local axes

We call the quadruples of four consecutive α -carbon atom positions $(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})$ as local segments. A structure containing $N \geq 4$ residues has $N - 3$ local segments $\{(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4), (\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5), \dots, (\mathbf{x}_{N-3}, \mathbf{x}_{N-2}, \mathbf{x}_{N-1}, \mathbf{x}_N)\}$. In this thesis we call a structure a helical structure if a helix curve is a good fit for each local segment $(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})$ of the structure.

The motivation for this piecewise definition for a helical structure is that it has been shown to hold well for the structures of our interest, α -helix, 3_{10} -helix, π -helix and β -sheet (Table 2.1) [2]. Another reason is that we can find a really good approximation of the axis $\hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})$ for the local segments $(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})$. We can't do the same with only three consecutive α -carbon atoms and we don't want to take more than we need. We call the axis $\hat{\mathbf{h}}$ of a local segment a local axis (Fig 3.1a). Before we derive the approximation of the local axis $\hat{\mathbf{h}}$ we will go through how we can use it to define other quantities of our interest, such as the axis $\hat{\mathbf{H}}_{n,m}$ (Fig 3.1b).

3.1.2 The axis as an average of local axes

Now, we can use the local axes $\{\hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})\}$ to construct the axis $\hat{\mathbf{H}}_{n,m}$ of residues n, \dots, m , when we assume that $m \leq n - 3$. We define $\hat{\mathbf{H}}_{m,n}$ to be the arithmetic average of the local axes constructed from α -carbon atoms $\mathbf{x}_m, \dots, \mathbf{x}_n$ and normalized to unity, that is,

$$\hat{\mathbf{H}}_{m,n} = \frac{\sum_{i=m}^{n-3} \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})}{|\sum_{i=m}^{n-3} \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})|}. \quad (3.2)$$

Note that in (3.2) the sum runs from m to $n - 3$, since we sum over all local segments $(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})$ constructed from α -carbon atoms $\mathbf{x}_m, \dots, \mathbf{x}_n$. Hence, $\hat{\mathbf{H}}_{i,i+3} = \hat{\mathbf{h}}(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \mathbf{x}_{i+3})$.

3.1.3 The kink angle

The kink between two consecutive sets of residues $\{m_1, m_1 + 1, \dots, n_1\}$ and $\{m_2, m_2 + 1, \dots, n_2\}$ is quantified by the kink angle

$$\xi_{(m_1, n_1), (m_2, n_2)} = \arccos(\hat{\mathbf{H}}_{m_1, n_1} \cdot \hat{\mathbf{H}}_{m_2, n_2}) \quad (3.3)$$

between the axes $\hat{\mathbf{H}}_{m_1, n_1}$ and $\hat{\mathbf{H}}_{m_2, n_2}$ (Fig 3.1b).

3.1.4 The tilt vector and local tilt

Other useful quantities along with the axis $\hat{\mathbf{H}}_{m,n}$, tilt angle $\Theta_{m,n}$ and kink angle $\xi_{(m_1, n_1), (m_2, n_2)}$ are the axis and tilt at residue i . The axis at residue i is necessary later when we construct and define rotation angle. The tilt at residue i has turned out to be useful in analysis shown later in this thesis. As the name suggests, the axis at residue i describes the axis of the helical structure at residue i . Therefore we want the axis at residue i to be a function of as few residues around the residue i as possible. In this sense, either local axis $\hat{\mathbf{h}}(\mathbf{x}_{i-2}, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})$ or $\hat{\mathbf{h}}(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2})$ would be good choice for the axis at residue i . However, this would be an ambiguous choice, since then the axis at residue i would depend on the direction of indexing, of which the helical structure is unaware of. Therefore, we take the average of these two local axes, that is, $\hat{\mathbf{H}}_{i-2,i+2}$. Another reasoning for using the axis $\hat{\mathbf{H}}_{i-2,i+2}$ of residues $i-2, \dots, i+2$ is that the residue i is in the middle of the five residues $i-2, \dots, i+2$, and thereby $\hat{\mathbf{H}}_{i-2,i+2}$ is independent of the indexing direction. We denote the axis at residue i by

$$\hat{\boldsymbol{\eta}}_i = \hat{\mathbf{H}}_{i-2,i+2} \quad (3.4)$$

and we call it a tilt vector to distinguish it from the more general form of the axis $\hat{\mathbf{H}}_{m,n}$. Then

$$\theta_i = \Theta_{i-2,i+2} = \arccos(\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\eta}}_i) \quad (3.5)$$

represents the tilt angle at residue i , which we call a local tilt angle.

3.1.5 Finding the local axis

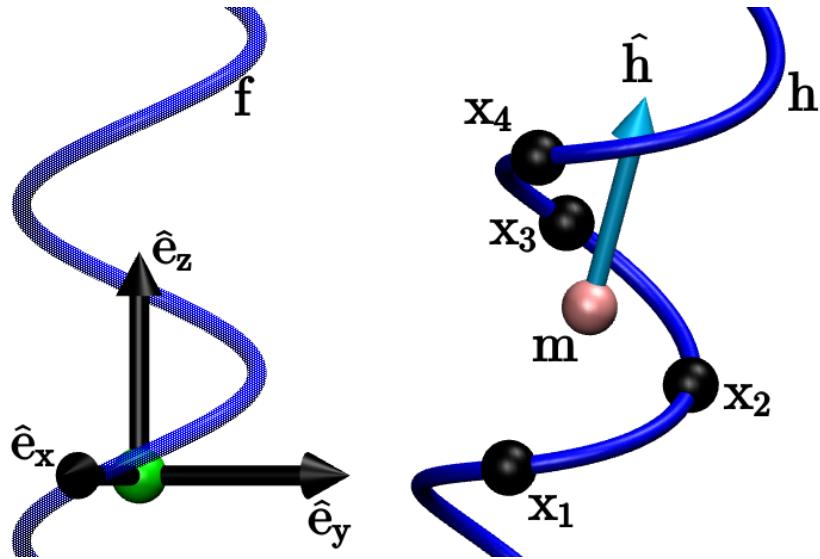


Figure 3.2: The black vectors are the Cartesian basis vectors \hat{e}_x , \hat{e}_y and \hat{e}_z and the green ball is the origin. Blue helix curves f and h , black local segment (x_1, x_2, x_3, x_4) and its pink center point m and cyan local axis \hat{h} .

For all the quantities defined above, it remains to find the local axis \hat{h} as a function of the local segment (x_1, x_2, x_3, x_4) . Let (x_1, x_2, x_3, x_4) be a local segment from a helical structure and let h be the helix curve, where the α -carbon atoms x_i lie in and $\{\hat{e}_x, \hat{e}_y, \hat{e}_z\}$ be the standard Cartesian basis of \mathbb{R}^3 . Let R be a rotation matrix such that $\hat{h} = R\hat{e}_z$. The α -carbon atoms lie in the helix curve h . Hence, for all $i = 1, 2, 3, 4$ there exist $\varphi_i \in \mathbb{R}$, such that

$$x_i = h(\varphi_i), \quad (3.6)$$

where

$$h(\varphi) = Rf(\varphi) + m, \quad (3.7)$$

$$f(\varphi) = r \cos \varphi \hat{e}_x + r \sin \varphi \hat{e}_y + \frac{P}{2\pi} \varphi \hat{e}_z, \quad (3.8)$$

here m is the center point of the local segment and f is a helix curve with axis \hat{e}_z , radius r and pitch P (Fig 3.2). Hence, in (3.7) R rotates f into a helix curve with \hat{h} as its axis and m translates it to the local segment (x_1, x_2, x_3, x_4) .

Solving the local axis

We assume that we can find two unit vectors \hat{s}_1 and \hat{s}_2 orthogonal to \hat{h} . Then the local axis \hat{h} can be constructed as

$$\hat{h} = \frac{\hat{s}_1 \times \hat{s}_2}{|\hat{s}_1 \times \hat{s}_2|}. \quad (3.9)$$

To find $\hat{\mathbf{s}}_i$, we start by showing that $(\mathbf{x}_{i+1} - \mathbf{x}_i) \cdot \hat{\mathbf{h}}/\Delta\phi_{i+1,i} = P/2\pi$ for all $i = 1, 2, 3$, where $\Delta\phi_{i+1,i} = \varphi_{i+1} - \varphi_i$ is the turn angle between residues $i + 1$ and i . We do this by straight forward computation:

$$\begin{aligned} \frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{\Delta\phi_{i+1,i}} \cdot \hat{\mathbf{h}} &= (R\mathbf{f}(\varphi_{i+1}) + \mathbf{m} - R\mathbf{f}(\varphi_i) - \mathbf{m}) \cdot \hat{\mathbf{h}}/\Delta\phi_{i+1,i} \\ &= R(\mathbf{f}(\varphi_{i+1}) - \mathbf{f}(\varphi_i)) \cdot \hat{\mathbf{h}}/\Delta\phi_{i+1,i} \\ &= R(\mathbf{f}(\varphi_{i+1}) - \mathbf{f}(\varphi_i)) \cdot R\hat{\mathbf{e}}_z/\Delta\phi_{i+1,i} \\ &= (\mathbf{f}(\varphi_{i+1}) - \mathbf{f}(\varphi_i)) \cdot \hat{\mathbf{e}}_z/\Delta\phi_{i+1,i} \\ &= \frac{P}{2\pi}(\varphi_{i+1} - \varphi_i)/\Delta\phi_{i+1,i} \\ &= \frac{P}{2\pi}, \end{aligned}$$

where we used (3.6), $\hat{\mathbf{h}} = R\hat{\mathbf{e}}_z$, invarince of dot product under rotation matrix and (3.8).

Then, for both $i = 1$ and $i = 2$

$$\begin{aligned} \left(\frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{\Delta\phi_{i+1,i}} - \frac{\mathbf{x}_{i+2} - \mathbf{x}_{i+1}}{\Delta\phi_{i+2,i+1}} \right) \cdot \hat{\mathbf{h}} &= \frac{P}{2\pi} - \frac{P}{2\pi} = 0. \\ \implies \left((\mathbf{x}_{i+1} - \mathbf{x}_i) - \frac{\Delta\phi_{i+1,i}}{\Delta\phi_{i+2,i+1}}(\mathbf{x}_{i+2} - \mathbf{x}_{i+1}) \right) \cdot \hat{\mathbf{h}} &= 0 \end{aligned}$$

Hence

$$\hat{\mathbf{s}}_i = \frac{(\mathbf{x}_{i+1} - \mathbf{x}_i) - \frac{\Delta\phi_{i+1,i}}{\Delta\phi_{i+2,i+1}}(\mathbf{x}_{i+2} - \mathbf{x}_{i+1})}{|(\mathbf{x}_{i+1} - \mathbf{x}_i) - \frac{\Delta\phi_{i+1,i}}{\Delta\phi_{i+2,i+1}}(\mathbf{x}_{i+2} - \mathbf{x}_{i+1})|} \quad (3.10)$$

is orthogonal to $\hat{\mathbf{h}}$ for both $i = 1$ and $i = 2$. Thus, we have found suitable unit vectors $\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2$ for (3.9). We have solved $\hat{\mathbf{h}}$ as a function of the local segment $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$ and the ratios $\Delta\phi_{i+1,i}/\Delta\phi_{i+2,i+1}$ between the consecutive turn angles. From a given structure we know the α -carbon atom positions but we don't know the precise turn angles. Therefore we want to solve the ratios $\Delta\phi_{i+1,i}/\Delta\phi_{i+2,i+1}$ in terms of the α -carbon positions which will require approximation.

Approximating the local axis

In this section, three approximations called APPROXa, APPROXb and APPROXc are presented. If a structure with helical arrangement has ideal bond lengths between all atoms, then the turn angles will be the same between all residues. Therefore the simplest approximation is $\Delta\phi_{i+1,i}/\Delta\phi_{i+2,i+1} \approx 1$. We call this approximation by APPROXa and denote

$$\hat{\mathbf{s}}_{i,a} = \frac{(\mathbf{x}_{i+1} - \mathbf{x}_i) - (\mathbf{x}_{i+2} - \mathbf{x}_{i+1})}{|(\mathbf{x}_{i+1} - \mathbf{x}_i) - (\mathbf{x}_{i+2} - \mathbf{x}_{i+1})|}, \quad (3.11)$$

$$\hat{\mathbf{h}}_a = \frac{\hat{\mathbf{s}}_{1,a} \times \hat{\mathbf{s}}_{2,a}}{|\hat{\mathbf{s}}_{1,a} \times \hat{\mathbf{s}}_{2,a}|}. \quad (3.12)$$

But we can do better with $\Delta\phi_{i+1,i}/\Delta\phi_{i+2,i+1} \approx |\mathbf{x}_{i+1} - \mathbf{x}_i|/|\mathbf{x}_{i+2} - \mathbf{x}_{i+1}|$. The intuition behind it is that when $\Delta\phi_{i+1,i}$ increases (decreases) then $|\mathbf{x}_{i+1} - \mathbf{x}_i|$ increases (decreases). This approximation has been stated in [7]. We call this approximation by APPROXb and denote

$$\hat{\mathbf{s}}_{i,b} = \frac{(\mathbf{x}_{i+1} - \mathbf{x}_i) - \frac{|\mathbf{x}_{i+1} - \mathbf{x}_i|}{|\mathbf{x}_{i+2} - \mathbf{x}_{i+1}|}(\mathbf{x}_{i+2} - \mathbf{x}_{i+1})}{|(\mathbf{x}_{i+1} - \mathbf{x}_i) - \frac{|\mathbf{x}_{i+1} - \mathbf{x}_i|}{|\mathbf{x}_{i+2} - \mathbf{x}_{i+1}|}(\mathbf{x}_{i+2} - \mathbf{x}_{i+1})|}, \quad (3.13)$$

$$\hat{\mathbf{h}}_b = \frac{\hat{\mathbf{s}}_{1,b} \times \hat{\mathbf{s}}_{2,b}}{|\hat{\mathbf{s}}_{1,b} \times \hat{\mathbf{s}}_{2,b}|}. \quad (3.14)$$

Both approximations APPROXA and APPROXb do work for α -helix like structures but both of them will break for β -strand like structures. This is because we didn't study the case in which $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ are not linearly independent. Note that (3.9) is not defined if $\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 = \pm 1$, since then $\hat{\mathbf{s}}_1 \times \hat{\mathbf{s}}_2 = \mathbf{0}$. For beta sheets $\Delta\phi$ is close to 180° (Table 2.1), when $\Delta\phi_{2,1} = \Delta\phi_{3,2} = \Delta\phi_{4,3} = \pi = 180^\circ$, the four coordinates in the local segment are in a plane and this yields $\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 = -1$. This means that, in this state, we need to approximate $\hat{\mathbf{h}}$ differently. Let

$$\hat{\mathbf{h}}_c = \begin{cases} \frac{\mathbf{x}_4 - \mathbf{x}_2 + \mathbf{x}_3 - \mathbf{x}_1}{|\mathbf{x}_4 - \mathbf{x}_2 + \mathbf{x}_3 - \mathbf{x}_1|}, & \hat{\mathbf{s}}_{1,b} \cdot \hat{\mathbf{s}}_{2,b} < -0.9 \\ \hat{\mathbf{h}}_b, & \text{else.} \end{cases}, \quad (3.15)$$

be called APPROXc. The reasoning for $\mathbf{x}_4 - \mathbf{x}_2 + \mathbf{x}_3 - \mathbf{x}_1$ is simple, if all turn angles $\Delta\phi$ are equal to π , then $\phi_{i+2} - \phi_i = 2\pi$ and hence

$$\begin{aligned} \mathbf{x}_{i+2} - \mathbf{x}_i &= \mathbf{h}(\phi_{i+2}) - \mathbf{h}(\phi_i) \\ &= R\mathbf{f}(\phi_i + 2\pi) + \mathbf{m} - R\mathbf{f}(\phi_i) - \mathbf{m} \\ &= R(r(\cos(\phi_i + 2\pi) - \cos(\phi_i))\hat{\mathbf{e}}_x + r(\sin(\phi_i + 2\pi) - \sin(\phi_i))\hat{\mathbf{e}}_y + \frac{P}{2\pi}(\phi_i + 2\pi - \phi_i)\hat{\mathbf{e}}_z) \\ &= PR\hat{\mathbf{e}}_z \\ &= P\hat{\mathbf{h}} \\ \implies \mathbf{x}_4 - \mathbf{x}_2 + \mathbf{x}_3 - \mathbf{x}_1 &= 2P\hat{\mathbf{h}}. \end{aligned} \quad (3.16)$$

Unlike APPROXb, the APPROXc works well when $\hat{\mathbf{s}}_{1,b} \cdot \hat{\mathbf{s}}_{2,b}$ is close to -1 , e.g. the turn angles $\Delta\phi$ are close to 180° . This is the reasoning for using $\hat{\mathbf{s}}_{1,b} \cdot \hat{\mathbf{s}}_{2,b} < -0.9$ in (3.15) instead of $\hat{\mathbf{s}}_{1,b} \cdot \hat{\mathbf{s}}_{2,b} = -1$.

Justifying the approximations of the local axis

We have defined three approximations, $\hat{\mathbf{h}}_a$, $\hat{\mathbf{h}}_b$ and $\hat{\mathbf{h}}_c$, of the local axis $\hat{\mathbf{h}}$. We will show that one of these approximations works best for the helical structures α -helix, 3_{10} -helix,

π -helix and β -strand. Our hypothesis is that $\hat{\mathbf{h}}_b$ is better than $\hat{\mathbf{h}}_a$ and that $\hat{\mathbf{h}}_c$ is better than $\hat{\mathbf{h}}_b$ for β -strand but equal for others.

Let $\delta_a, \delta_b, \delta_c \in [0, \pi]$ be the angles between the local axis $\hat{\mathbf{h}}$ and the approximations $\hat{\mathbf{h}}_a, \hat{\mathbf{h}}_b, \hat{\mathbf{h}}_c$, respectively, e.g. $\delta_c = \arccos(\hat{\mathbf{h}} \cdot \hat{\mathbf{h}}_c)$. The better the approximation $\hat{\mathbf{h}}_c$ is the closer the angle δ_c is to 0, since if $\delta_c = 0$, then $\hat{\mathbf{h}} = \hat{\mathbf{h}}_c$. If $\delta_a - \delta_c \geq 0$ and $\delta_b - \delta_c \geq 0$, then $\hat{\mathbf{h}}_c$ is better or equally good than $\hat{\mathbf{h}}_a$ and $\hat{\mathbf{h}}_b$. Therefore, in order for our hypothesis to be true for a helical structure, e.g. α -helix, it remains to show that δ_c is close to 0, $\delta_a - \delta_c \geq 0$ and $\delta_b - \delta_c \geq 0$. We are showing this for helical structures α -helix, 3_{10} -helix, π -helix and β -strand by making histograms of $\delta_a - \delta_c$, $\delta_b - \delta_c$ and δ_c .

The angles δ_a, δ_b and δ_c are functions of local segment $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$, and so, to make the histograms $\delta_a - \delta_c$, $\delta_b - \delta_c$ and δ_c for a secondary structure we are going to use Table 2.1 and equation (3.6) to sample local segments for the secondary structure. Note that the angles δ_a, δ_b and δ_c are invariant under translation and rotation of the local segment. Hence, in equation (3.6) we are free to fix $\varphi_1 = 0$, $\mathbf{m} = (0, 0, 0)$ and R to be the identity matrix, i.e. $\hat{\mathbf{h}} = R\hat{\mathbf{e}}_z = \hat{\mathbf{e}}_z$. Then equation (3.6) becomes $\mathbf{x}_i = \mathbf{f}(r, P, \varphi_i)$. Since $\varphi_1 = 0$ and $\Delta\phi_{i+1,i} = \varphi_{i+1} - \varphi_i$, we get $\varphi_2 = \Delta\phi_{1,2}$, $\varphi_3 = \varphi_2 + \Delta\phi_{2,3}$ and $\varphi_4 = \varphi_3 + \Delta\phi_{3,4}$. Remember that for the pitch we have $P = p2\pi/\Delta\phi$. Thus, for one local segment $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$, we need one radius r , one translation distance p , one turn angle $\Delta\phi$ for the pitch P , and three turn angles $\Delta\phi_{1,2}, \Delta\phi_{2,3}$ and $\Delta\phi_{3,4}$ for φ_2, φ_3 and φ_4 . Hence we can write the angles $\delta_a - \delta_c$, $\delta_b - \delta_c$ and δ_c as functions of a point $\mathbf{q} = (q_1, \dots, q_6) = (r, p, \Delta\phi, \Delta\phi_{1,2}, \Delta\phi_{2,3}, \Delta\phi_{3,4})$.

We use the secondary structure's normal distribution parameters μ_j, σ_j for each quantity q_j from Table 2.1 to generate a set Q of points \mathbf{q} . For each coordinate q_j , we are going to sample values that lie within 5 standard deviations σ_j of the mean μ_j . Five standard deviations are enough for us since for normal distribution 99.99994% of values lie within it. Let Q_j be a set of 100 evenly spaced numbers from $\mu_j - 5\sigma_j$ to $\mu_j + 5\sigma_j$, then $Q = Q_1 \times Q_2 \times \dots \times Q_6$ is a set of 100^6 points \mathbf{q} . And for the histogram the weight $\omega(\mathbf{q})$ of the angles $\delta_c(\mathbf{q}), (\delta_a - \delta_c)(\mathbf{q})$ and $(\delta_b - \delta_c)(\mathbf{q})$ is

$$\omega(\mathbf{q}) = \prod_{j=1}^6 g(q_j, \mu_j, \sigma_j), \quad (3.17)$$

where g is the probability density function of normal distribution.

According to Fig 3.3, our hypothesis is true, or at least true in the set Q , since APPROXb is better than APPROXA and APPROXc is better than APPROXb for β -strand but equal for others. The averages $0.16^\circ, 0.43^\circ, 0.14^\circ$ and 0.17° and standard deviations $0.09^\circ, 0.23^\circ, 0.08^\circ$ and 0.13° (Fig 3.3) respectively for α -helix, 3_{10} -helix, π -helix and β -strand of $\delta_c(Q)$ are not biologically significant. Therefore, there is no difference whether we use the exact solution of the local axis $\hat{\mathbf{h}}$ or just use the APPROXc local

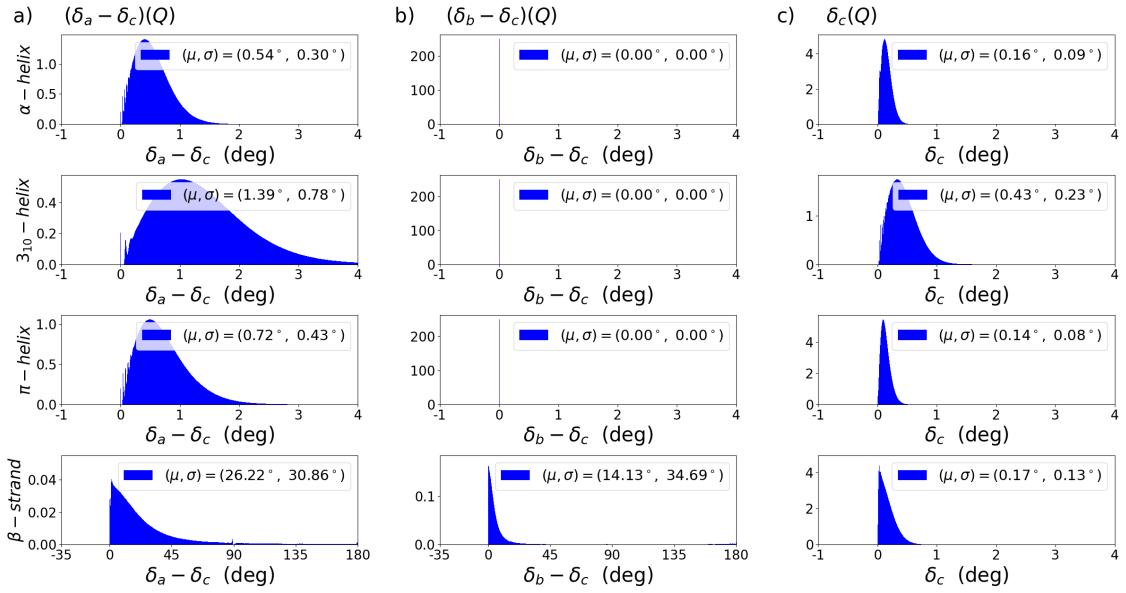


Figure 3.3: Normalized histograms for the helical structures α -helix, 3_{10} -helix, π -helix and β -strand made from the image sets $(\delta_a - \delta_c)(Q)$, $(\delta_b - \delta_c)(Q)$ and $\delta_c(Q)$ in figures a), b) and c) respectively. The normal distribution parameters μ and σ are in the legend for each histogram. Note that the range of the x-axis for β -strand in figures a) and b) is $[-35^\circ, 180^\circ]$ while for other histograms it is $[-1^\circ, 4^\circ]$.

axis $\hat{\mathbf{h}}_c$. Hence, we have found the local axis $\hat{\mathbf{h}} = \hat{\mathbf{h}}_c$ for the helical structures α -helix, 3_{10} -helix, π -helix and β -strand.

Center of the local segment

We will now find the center point \mathbf{m} (Fig 3.2) of the local segment $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$, since it is needed for the rotation. The center point \mathbf{m} is a point in the helix axis, such that its distance $\mathbf{m} \cdot \hat{\mathbf{h}}$ along the local axis $\hat{\mathbf{h}}$ equals the average of the distances $\mathbf{x}_i \cdot \hat{\mathbf{h}}$, i.e. $\mathbf{m} \cdot \hat{\mathbf{h}} = (\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_4) \cdot \hat{\mathbf{h}} / 4$. Then the point $(\mathbf{m} \cdot \hat{\mathbf{h}})\hat{\mathbf{h}}$ and the local axis $\hat{\mathbf{h}}$ define a plane which has the center point \mathbf{m} in it. The local segment $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$ is in the helix curve (3.6), whose local axis is $\hat{\mathbf{h}}$. Hence, the map $V : \mathbf{x}_i \mapsto \mathbf{x}_i + (\mathbf{x}_i \cdot \hat{\mathbf{h}} - \mathbf{m} \cdot \hat{\mathbf{h}})\hat{\mathbf{h}}$ projects the α -carbon atoms of the local segment $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$ into a circle in the plane and the circle's center is the midpoint \mathbf{m} . Only three points from a circle are needed to calculate the circle's center. We have four points $(V\mathbf{x}_1, V\mathbf{x}_2, V\mathbf{x}_3, V\mathbf{x}_4)$ in the circle and therefore we can calculate the center point four times, we take their average and define it to be the center point \mathbf{m} of the local segment.

3.2 Rotation

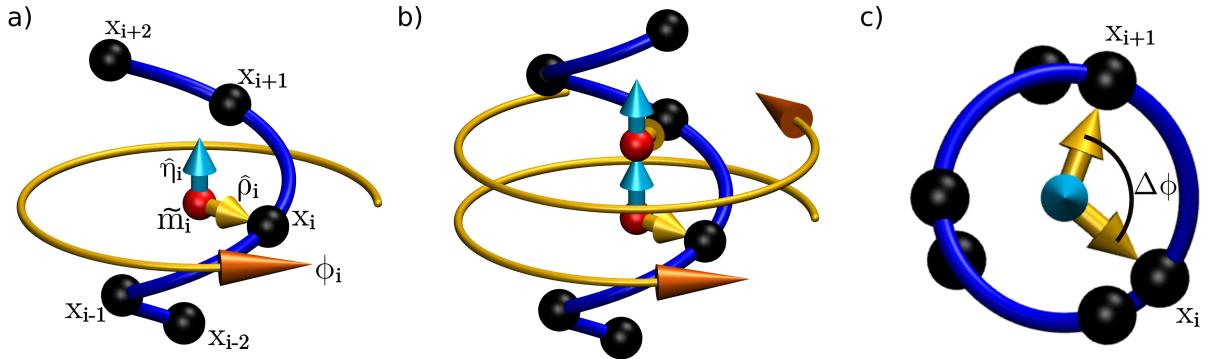


Figure 3.4: a) Five black α -carbon atoms $\mathbf{x}_{i-2}, \dots, \mathbf{x}_{i+2}$ on a blue helix curve, cyan tilt vector $\hat{\eta}_i$, orange rotation vector $\hat{\rho}_i$, red midpoint $\tilde{\mathbf{m}}_i$ and an orange circular arrow representing rotation angle ϕ_i . b) Six black α -carbon atoms $\mathbf{x}_{i-2}, \dots, \mathbf{x}_{i+3}$ on a blue helix curve, cyan tilt vectors $\hat{\eta}_i$ and $\hat{\eta}_{i+1}$, orange rotation vectors $\hat{\rho}_i$ and $\hat{\rho}_{i+1}$, red midpoints $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}_{i+1}$ and orange circular arrows represents rotation angles ϕ_i and ϕ_{i+1} . c) Top view of figure b without the circular arrows. The black arc represents the turn angle $\Delta\phi$ between residues i and $i + 1$.

We assume throughout section 3.2 that α -carbon positions \mathbf{x}_i are from helical structure, e.g. α -helix. As discussed in the section 2.2, the rotation angle $\Phi_{m,n} \in]-\pi, \pi]$ of residues $m, m + 1, \dots, n$ describes, which residues point in the direction of the chosen reference vector $\hat{\mathbf{k}}$. We start by defining local rotation angle $\phi_i \in]-\pi, \pi]$ (Fig 3.4), which defines how the residue i is pointing in the direction of the chosen reference vector $\hat{\mathbf{k}}$. After this, we will construct the rotation angle $\Phi_{m,n}$ as an average of the local rotation angles $\{\phi_i\}$.

3.2.1 Definition of the local rotation angle

Let midpoint $\tilde{\mathbf{m}}_i$ of five consecutive α -carbon atoms $\mathbf{x}_{i-2}, \dots, \mathbf{x}_{i+2}$ be a point on the helix axis, such that $\mathbf{x}_i - \tilde{\mathbf{m}}_i$ is orthogonal to the tilt vector $\hat{\eta}_i$ (Fig 3.4a). Then rotation vector $\hat{\rho}_i$ (Fig 3.4a) of residue i , defined as

$$\hat{\rho}_i = \frac{\mathbf{x}_i - \tilde{\mathbf{m}}_i}{|\mathbf{x}_i - \tilde{\mathbf{m}}_i|}, \quad (3.18)$$

is orthogonal to the tilt vector $\hat{\eta}_i$, i.e. $\hat{\eta}_i \cdot \hat{\rho}_i = 0$. The local tilt angle $\phi_i \in]-\pi, \pi]$ defines how the residue i points at the chosen reference vector $\hat{\mathbf{k}}$, that is, how the rotation vector $\hat{\rho}_i$ points at chosen reference vector $\hat{\mathbf{k}}$. We define the local rotation angle $\phi_i = \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{\mathbf{k}})$ as a function of $\hat{\eta}_i$, $\hat{\rho}_i$ and $\hat{\mathbf{k}}$ which satisfies following conditions:

- i If R is a rotation matrix, then $\phi(R\hat{\eta}_i, R\hat{\rho}_i, \hat{\mathbf{k}}) = \phi(\hat{\eta}_i, \hat{\rho}_i, R^{-1}\hat{\mathbf{k}})$, where R^{-1} is the inverse rotation matrix of R .

- ii If $R_{\hat{\mathbf{k}}}$ is a rotation matrix about the reference vector $\hat{\mathbf{k}}$, then

$$\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \phi(R_{\hat{\mathbf{k}}}\hat{\boldsymbol{\eta}}_i, R_{\hat{\mathbf{k}}}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}).$$
- iii If $R_{\hat{\boldsymbol{\eta}}_i}$ is a rotation matrix about the tilt vector $\hat{\boldsymbol{\eta}}_i$ by an angle $\Delta\varphi$, then

$$\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \phi(R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) - \Delta\varphi.$$
- iv If $R_{\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}}$ is a rotation matrix about the vector $\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}$, then

$$\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \phi(R_{\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}}\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}).$$
- v $\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = 0$ if and only if $\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}_i = \max\{\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}'_i \mid \hat{\boldsymbol{\rho}}'_i \text{ orthogonal to } \hat{\boldsymbol{\eta}}_i\}.$

Next, we will go through how the conditions (i)-(v) characterize the rotation angle's properties discussed in section 2.2. The condition (i) assures that the local rotation angle ϕ_i depends only on how the tilt and rotation vectors are relative to a chosen reference $\hat{\mathbf{k}}$. The condition (ii) characterizes the symmetry around the reference vector $\hat{\mathbf{k}}$. The condition (iii) characterizes that the local rotation angle ϕ_i describes how $\hat{\boldsymbol{\rho}}_i$ rotates about $\hat{\boldsymbol{\eta}}_i$ when $\hat{\boldsymbol{\eta}}_i$ is fixed. Note that $\hat{\boldsymbol{\eta}}_i$ is fixed under the rotation matrix about itself, i.e. $R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\eta}}_i = \hat{\boldsymbol{\eta}}_i$.

In section 2.2, we noticed that the rotation is not defined if the axis is parallel to the reference vector, i.e. $\hat{\mathbf{H}} = \pm\hat{\mathbf{k}}$. This follows from conditions (ii) and (iii). In (iii) take $\Delta\varphi \neq 0$ and $\hat{\boldsymbol{\eta}}_i = \pm\hat{\mathbf{k}}$. This leads to $\phi(\pm\hat{\mathbf{k}}, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \phi(R_{\pm\hat{\mathbf{k}}}(\pm\hat{\mathbf{k}}), R_{\pm\hat{\mathbf{k}}}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) - \Delta\varphi$. Both $R_{\hat{\mathbf{k}}}$ and $R_{-\hat{\mathbf{k}}}$ are rotation matrices about $\hat{\mathbf{k}}$. Then, from (ii) we get that $\phi(R_{\pm\hat{\mathbf{k}}}(\pm\hat{\mathbf{k}}), R_{\pm\hat{\mathbf{k}}}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \phi(\pm\hat{\mathbf{k}}, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$. Hence, $\phi(\pm\hat{\mathbf{k}}, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \phi(\pm\hat{\mathbf{k}}, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) - \Delta\varphi$, and so, $\Delta\varphi = 0$. This yields a contradiction, since we took $\Delta\varphi \neq 0$. This means that $\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$ can't be defined when $\hat{\boldsymbol{\eta}}_i = \pm\hat{\mathbf{k}}$, i.e. when $\theta_i = 0$ or $\theta_i = \pi$. Therefore, always when we discuss about the local rotation angle ϕ_i , we assume that $\hat{\boldsymbol{\eta}}_i \neq \pm\hat{\mathbf{k}}$, $0 \neq \theta_i \neq \pi$. Hence, when $\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$ is defined, we have $\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}} \neq \mathbf{0}$, and so, the rotation matrix $R_{\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}}$ in condition (iv) is defined.

The condition (iv) characterizes the fact noticed in the section 2.2 that the tilt angle can be changed without changing the rotation angle. Similarly, the condition (iii) tells us that the rotation angle can change freely for a fixed tilt. The condition (v) defines what the possible values of the local rotation angle $\phi_i \in]-\pi, \pi]$ mean. The greater $\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}_i$ is the more $\hat{\boldsymbol{\rho}}_i$ is aligned with $\hat{\mathbf{k}}$. The condition (v) means that if the residue i points as much as possible in the direction of the reference vector $\hat{\mathbf{k}}$, then $\phi_i = \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = 0$ and vice versa. Later, we will find out that $\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}_i = \sin \theta_i \cos \phi_i$. We have $\sin \theta_i > 0$ for all $\theta_i \in]0, \pi[$, and so, $\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}_i > 0$ if and only if $|\phi_i| < \pi/2 = 90^\circ$ and respectively $\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}_i < 0$ if and only if $|\phi_i| > \pi/2 = 90^\circ$. This means that if $|\phi_i| < \pi/2 = 90^\circ$, then we know that the residue i points in the direction of the reference vector $\hat{\mathbf{k}}$.

To understand why the conditions (i)-(v) form a useful definition of the local rotation angle ϕ_i , take two states $(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i)$ and $(\hat{\boldsymbol{\eta}}'_j, \hat{\boldsymbol{\rho}}'_j)$ with $\phi_i = \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$ and $\phi'_j = \phi(\hat{\boldsymbol{\eta}}'_j, \hat{\boldsymbol{\rho}}'_j, \hat{\mathbf{k}})$

being, respectively, their local rotation angles. The main benefit of the orthogonality of the tilt and rotation vectors is that we have $\hat{\eta}_i \cdot \hat{\rho}_i = \hat{\eta}'_j \cdot \hat{\rho}'_j = 0$. Thereby, there exists rotation matrix R such that $(\hat{\eta}'_j, \hat{\rho}'_j) = (R\hat{\eta}_i, R\hat{\rho}_i)$. The vectors $\hat{\eta}_i$, $\hat{\mathbf{k}}$ and $\hat{\eta}_i \times \hat{\mathbf{k}}$ are linearly independent, that is, they form a basis for \mathbb{R}^3 . Hence, there exists the rotation matrices: $R_{\hat{\mathbf{k}}}$ about $\hat{\mathbf{k}}$, $R_{\hat{\eta}_i \times \hat{\mathbf{k}}}$ about $\hat{\eta}_i \times \hat{\mathbf{k}}$ and $R_{\hat{\eta}_i}$ about $\hat{\eta}_i$ by an angle $\Delta\varphi$, such that $R = R_{\hat{\eta}_i} R_{\hat{\eta}_i \times \hat{\mathbf{k}}} R_{\hat{\mathbf{k}}}$. Then, from conditions (i)-(v) it follows that $\Delta\varphi = \phi'_j - \phi_i$. This means that $\phi'_j - \phi_i$ describes how much more the residue j in the primed state points towards the reference vector $\hat{\mathbf{k}}$ than the residue i in the non-primed state. This is exactly what we want from the local rotation angle.

Before we construct the midpoint $\tilde{\mathbf{m}}_i$ from the α -carbon atoms $\mathbf{x}_{i-2}, \dots, \mathbf{x}_{i+2}$ and find the unique function ϕ satisfying the conditions (i)-(v), we will construct the rotation angle $\Phi_{m,n}$ and discuss how the local rotation angles can be used to compute the difference in rotation between two helical structures.

3.2.2 Rotation angle as an average of local rotation angles

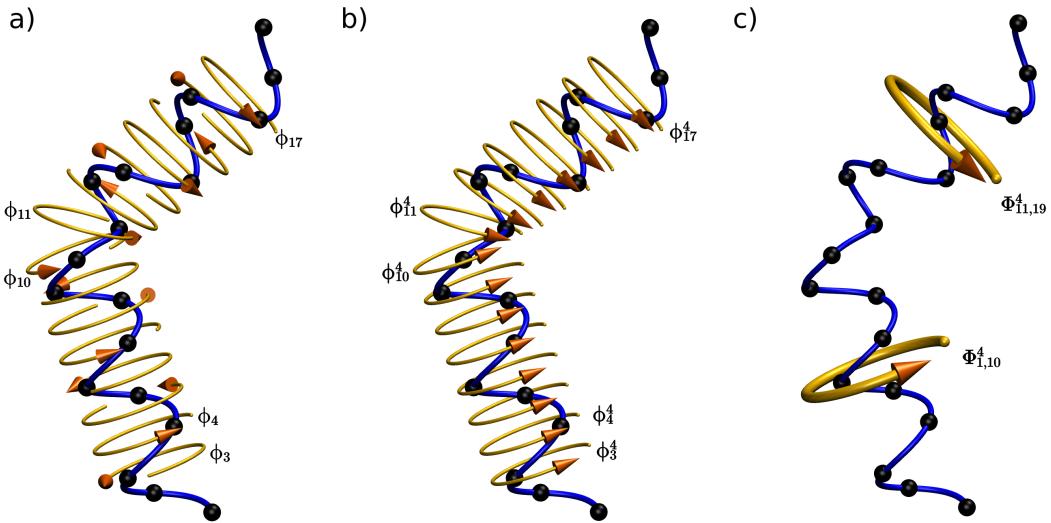


Figure 3.5: A blue kinked helix curve fitted into 19 black α -carbons, with positions denoted by \mathbf{x}_i . a) Local rotation angles ϕ_i are shown as orange arrows for all residues $i = 3, \dots, 17$. b) Local rotation angles ϕ_i^4 in phase of residue 4 are shown as orange arrows for all residues $i = 3, \dots, 17$. c) Rotation angles $\Phi_{1,10}^4$ and $\Phi_{11,19}^4$ of residues $\{1, \dots, 11\}$ and $\{11, \dots, 19\}$, respectively, are shown as orange arrows.

To define $\Phi_{m,n}$, we assume that the turn angle $\Delta\phi$ is roughly the same between all residues m, \dots, n (Fig 3.4c). Then, from the conditions (i)-(iv) we get that the local rotation angles ϕ_i and ϕ_l have a phase difference of $(l - i)\Delta\phi$. Hence, the local rotation angle $\phi_i^l \in]-\pi, \pi]$ of residue i in phase of residue l is

$$\phi_i^l = \phi_i + (l - i)\Delta\phi, \quad (3.19)$$

where the equal sign means that $\phi_i^l \in]-\pi, \pi]$ is equivalent to $\phi_i + (l-i)\Delta\phi$ as an angle, i.e. $\phi_i^l - (\phi_i + (l-i)\Delta\phi) = 2\pi k$ for some integer k .

The angles ϕ_i^l have the same phase, i.e. $\phi_i^l \approx \phi_j^l$ for all residues i and j (Fig 3.5b). If $\Delta\phi$ is precisely the turn angle between all residues, we have $\phi_i^l = \phi_j^l$. We define $\Phi_{m,n}^l$ to be an average of the angles $\{\phi_{m+2}^l, \dots, \phi_{n-2}^l\}$. The local rotation angle is 2π -periodic. Hence, we use the circular average, that is, we take the polar angle of a \mathbb{R}^2 vector obtained by taking arithmetic average of vectors $\{(\cos \phi_{m+2}^l, \sin \phi_{m+2}^l), \dots, (\cos \phi_{n-2}^l, \sin \phi_{n-2}^l)\}$. Thus

$$\Phi_{m,n}^l = \arctan2\left(\sum_{i=m+2}^{n-2} \sin \phi_i^l, \sum_{i=m+2}^{n-2} \cos \phi_i^l\right), \quad (3.20)$$

where $\arctan2 : \mathbb{R}^2 \setminus \{(0,0)\} \rightarrow]-\pi, \pi]$,

$$\arctan2(y, x) = \begin{cases} \arctan(y/x), & x > 0 \\ \arctan(y/x) + \pi, & x < 0 \text{ and } y \geq 0 \\ \arctan(y/x) - \pi, & x < 0 \text{ and } y < 0 \\ \pi/2, & x = 0 \text{ and } y > 0 \\ -\pi/2, & x = 0 \text{ and } y < 0 \end{cases}$$

transforms the Cartesian coordinates x, y into the polar angle and $\arctan : \mathbb{R} \rightarrow]-\pi/2, \pi/2[$ is the inverse of the tangent. Note that in (3.20), the sums run over all local rotation angles constructed by $\mathbf{x}_m, \dots, \mathbf{x}_n$. We need five consecutive α -carbons $\mathbf{x}_{i-2}, \dots, \mathbf{x}_{i+2}$ for $\hat{\boldsymbol{\eta}}_i$ and $\hat{\boldsymbol{\rho}}_i$, and respectively for ϕ_i and ϕ_i^l . Hence, in (3.20) the sums run from $m+2$ to $n-2$ instead from m to n .

We have now defined the rotation angle of residues m, \dots, n in a phase of a chosen residue l . Different phases are related by $\Phi_{m,n}^{l+1} = \Phi_{m,n}^l + \Delta\phi$, since $\phi_i^{l+1} = \phi_i + (l+1-i)\Delta\phi = \phi_i^l + \Delta\phi$. We have assumed that $\Delta\phi$ is roughly the turn angle between all residues, and thereby $\Phi_{m,n}^l \approx \phi_l$. From (3.19) it follows that $\Phi_{m,n}^l \approx \phi_l = \phi_l^l \approx \phi_i^l = \phi_i + (l-i)\Delta\phi$. Hence, the local rotation angle $\phi_i \approx \Phi_{m,n}^l - (l-i)\Delta\phi$ can be extracted from the rotation angle $\Phi_{m,n}^l$. When comparing rotation angles one must take into account in which phases they are computed. When doing analysis, it's a good practice to fix l to a specific residue. If the indexing of residues is shifted, then one needs to remember to shift l accordingly.

In the definition of the rotation angle $\Phi_{m,n}^l$, we have assumed that the residues m, \dots, n have roughly the same helical arrangement. If the helical arrangement of residues m, \dots, n is broken, e.g. due to a kink, the rotation angle $\Phi_{m,n}^l$ might be misleading, since it is an average of the local rotation angles $\phi_{m+2}^l, \dots, \phi_{n-2}^l$. In the chapter 4, we will see an example analysis of a transmembrane helix with a kink, such that the residues before and after the kink don't have the same rotation.

The advantage of approaching rotation angle the way we have here is the possibility to first work with the local rotation angles $\{\phi_i\}$ at the residue level. Then take suit-

able averages $\Phi_{m,n}^l$ to reduce the number of angles, if possible. The same applies when comparing rotation angles between different states.

3.2.3 Difference in rotation

Assume that we have two states, non primed $\mathbf{x}_m, \dots, \mathbf{x}_n$ and primed $\mathbf{x}'_m, \dots, \mathbf{x}'_n$, of residues m, \dots, n . The difference between the local rotation angles ϕ_i and ϕ'_j from the non-primed state and primed state is

$$d(\phi'_j, \phi_i) = \phi'_j - \phi_i. \quad (3.21)$$

We assume that the turn angle $\Delta\phi$ is the same for the non-primed and primed states, e.g. that both of them are α -helical. Then $d(\phi'^l_i, \phi^l_i) = (\phi'_i + (l-i)\Delta\phi) - (\phi_i + (l-i)\Delta\phi) = d(\phi'_i, \phi_i)$, for all l and i . Hence, the difference in rotation angle between the two states can be calculated as an average of the angles $d(\phi'_i, \phi_i) \in]-\pi, \pi]$,

$$D(\phi'_i, \phi_i) = \arctan2\left(\sum_{i=m+2}^{n-2} \sin d(\phi'_i, \phi_i), \sum_{i=m+2}^{n-2} \cos d(\phi'_i, \phi_i)\right). \quad (3.22)$$

Similarly, as for the rotation angles, we can first work with the angles $d(\phi'_i, \phi_i)$ and then take suitable averages $D(\phi'_i, \phi_i)$, if possible.

3.2.4 Constructing the midpoint

In the beginning of this chapter, we defined the midpoint $\tilde{\mathbf{m}}_i$ (Fig 3.4a) of α -carbon positions $\mathbf{x}_{i-2}, \dots, \mathbf{x}_{i+2}$ to be a point on the helix axis, such that $\mathbf{x}_i - \tilde{\mathbf{m}}_i$ is orthogonal to the tilt vector $\hat{\boldsymbol{\eta}}_i$. By the definition of the center point $\mathbf{m}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$ of the local segment $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$, we have that the average $\bar{\mathbf{m}}_i = (\mathbf{m}(\mathbf{x}_{i-2}, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}) + \mathbf{m}(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2})) / 2$ is a point in the axis of the helix close to the midpoint $\tilde{\mathbf{m}}_i$. Then $\tilde{\mathbf{m}}_i$ is obtained by translating $\bar{\mathbf{m}}_i$ along the axis $\hat{\boldsymbol{\eta}}_i$ so that $\mathbf{x}_i - \tilde{\mathbf{m}}_i$ is orthogonal to $\hat{\boldsymbol{\eta}}_i$. From the Gram-Schmidt method, we obtain

$$\tilde{\mathbf{m}}_i = \bar{\mathbf{m}}_i + ((\mathbf{x}_i - \bar{\mathbf{m}}_i) \cdot \hat{\boldsymbol{\eta}}_i) \hat{\boldsymbol{\eta}}_i. \quad (3.23)$$

Now by definition (3.18), we have the desired orthogonality of the tilt vector $\hat{\boldsymbol{\eta}}_i$ and the rotation vector $\hat{\boldsymbol{\rho}}_i$,

$$\begin{aligned} \hat{\boldsymbol{\rho}}_i \cdot \hat{\boldsymbol{\eta}}_i &= \frac{\mathbf{x}_i - \tilde{\mathbf{m}}_i}{|\mathbf{x}_i - \tilde{\mathbf{m}}_i|} \cdot \hat{\boldsymbol{\eta}}_i \\ &= \frac{1}{|\mathbf{x}_i - \tilde{\mathbf{m}}_i|} ((\mathbf{x}_i - \bar{\mathbf{m}}_i) \cdot \hat{\boldsymbol{\eta}}_i - (\mathbf{x}_i - \bar{\mathbf{m}}_i) \cdot \hat{\boldsymbol{\eta}}_i) \\ &= 0. \end{aligned}$$

3.2.5 Finding the local rotation angle

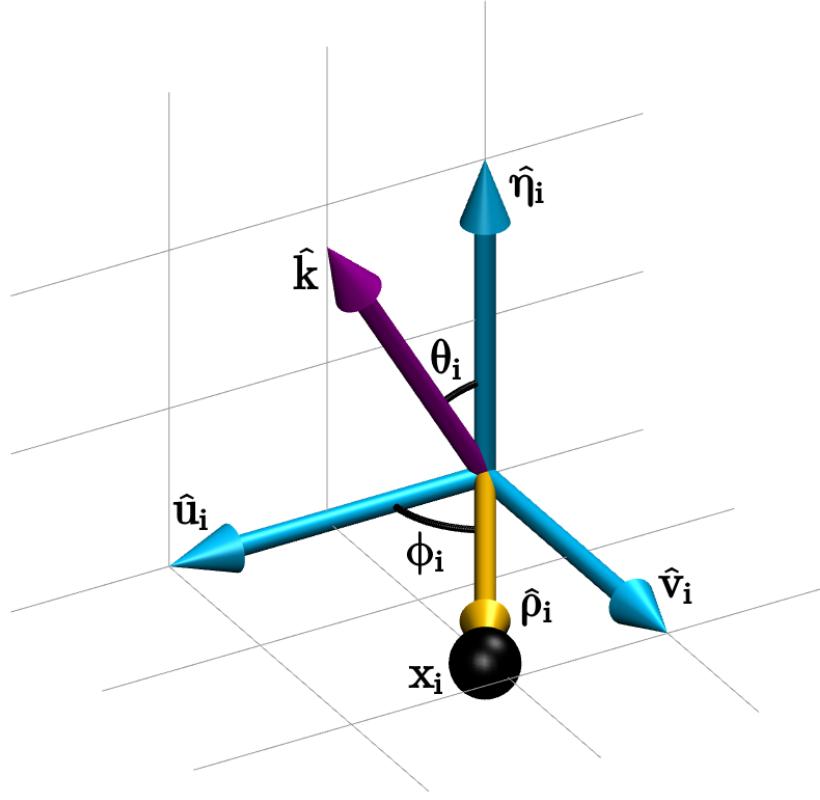


Figure 3.6: The purple arrow is the reference vector $\hat{\mathbf{k}}$, orange the rotation vector $\hat{\boldsymbol{\rho}}_i$, cyan the tilt vector $\hat{\boldsymbol{\eta}}_i$ with the two other cyan basis vectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ and black sphere the α -carbon \mathbf{x}_i . The black arcs represent the local tilt angle θ_i and local rotation angle ϕ_i .

Now, we will find the final missing piece of this method, the unique local rotation angle function $\phi_i = \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$ defined by the conditions (i)-(v). We start by fixing the reference vector $\hat{\mathbf{k}}$. Let $(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i)$ be an arbitrary tilt and rotation vector pair, that is, $\hat{\boldsymbol{\eta}}_i \cdot \hat{\boldsymbol{\rho}}_i = 0$. For the local rotation angle to be defined, we assume $\hat{\boldsymbol{\eta}}_i \neq \pm \hat{\mathbf{k}}$, i.e. $0 \neq \theta_i \neq \pi$. Then $|\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}| \neq 0$ which yields that the unit vectors

$$\hat{\mathbf{v}}_i = \frac{\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}}{|\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}|}, \quad (3.24)$$

$$\hat{\mathbf{u}}_i = \hat{\mathbf{v}}_i \times \hat{\boldsymbol{\eta}}_i = \frac{\hat{\mathbf{k}} - \hat{\boldsymbol{\eta}}_i(\hat{\boldsymbol{\eta}}_i \cdot \hat{\mathbf{k}})}{|\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}|} \quad (3.25)$$

are well-defined. The sequence $(\hat{\mathbf{u}}_i, \hat{\mathbf{v}}_i, \hat{\boldsymbol{\eta}}_i)$ forms a right handed orthonormal basis for \mathbb{R}^3 (Fig 3.6). Therefore, the rotation vector can be written as

$$\hat{\boldsymbol{\rho}}_i = (\hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{u}}_i)\hat{\mathbf{u}}_i + (\hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{v}}_i)\hat{\mathbf{v}}_i + (\hat{\boldsymbol{\rho}}_i \cdot \hat{\boldsymbol{\eta}}_i)\hat{\boldsymbol{\eta}}_i$$

and from the orthogonality $\hat{\rho}_i \cdot \hat{\eta}_i = 0$, (3.25) and (3.24) it follows that

$$\hat{\rho}_i = \frac{\hat{\rho}_i \cdot \hat{k}}{|\hat{\eta}_i \times \hat{k}|} \hat{u}_i + \frac{\hat{\rho}_i \cdot (\hat{\eta}_i \times \hat{k})}{|\hat{\eta}_i \times \hat{k}|} \hat{v}_i. \quad (3.26)$$

Since $\hat{\rho}_i$ is a unit vector in the $\hat{u}_i \hat{v}_i$ -plane, there exists an angle $\vartheta_i \in]-\pi, \pi]$, such that

$$\hat{\rho}_i = \cos \vartheta_i \hat{u}_i + \sin \vartheta_i \hat{v}_i. \quad (3.27)$$

Note that $\hat{k} = \sin \theta_i \hat{u}_i + \cos \theta_i \hat{\eta}_i$, where θ_i is the local tilt angle (Fig 3.6). We are aiming to use the condition (v) here. Therefore, we want to find the unit vector $\hat{\rho}_i''$ orthogonal to $\hat{\eta}_i$ for which $\hat{k} \cdot \hat{\rho}_i'' = \max\{\hat{k} \cdot \hat{\rho}'_i \mid \hat{\rho}'_i \text{ orthogonal to } \hat{\eta}_i\}$. Similarly, as for $\hat{\rho}_i$ there is ϑ_i , there exists $\vartheta'_i \in]-\pi, \pi]$, such that $\hat{\rho}'_i = \cos \vartheta'_i \hat{u}_i + \sin \vartheta'_i \hat{v}_i$. Then, $\hat{k} \cdot \hat{\rho}'_i = \sin \theta_i \cos \vartheta'_i$. Since $\theta_i \in]0, \pi[$, we have that $\sin \theta_i > 0$. Hence, $\hat{k} \cdot \hat{\rho}'_i$ has its maximum when $\cos \vartheta'_i$ has its maximum, that is, when $\vartheta'_i = 0$. Thus $\hat{\rho}_i'' = \cos \vartheta''_i \hat{u}_i + \sin \vartheta''_i \hat{v}_i = \hat{u}_i$, where $\vartheta''_i = 0$.

Assume that there exists a function ϕ satisfying the conditions (i)-(v). Let $R_{\hat{\eta}_i}$ be a rotation matrix about $\hat{\eta}_i$ by the angle $-\phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k})$. Then

$$R_{\hat{\eta}_i} \hat{\rho}_i = \cos(\vartheta_i - \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k})) \hat{u}_i + \sin(\vartheta_i - \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k})) \hat{v}_i. \quad (3.28)$$

From the condition (iii) it follows that $\phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k}) = \phi(R_{\hat{\eta}_i} \hat{\eta}_i, R_{\hat{\eta}_i} \hat{\rho}_i, \hat{k}) - (-\phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k}))$, and so, $\phi(R_{\hat{\eta}_i} \hat{\eta}_i, R_{\hat{\eta}_i} \hat{\rho}_i, \hat{k}) = 0$. Hence, from the condition (v) we get that $R_{\hat{\eta}_i} \hat{\rho}_i = \hat{\rho}_i''$. Which yields $\vartheta_i - \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k}) = \vartheta''_i = 0$. Using the equalities (3.26) and (3.27), orthogonality $\hat{u}_i \cdot \hat{v}_i = 0$, and substituting $\vartheta_i = \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k})$, we obtain

$$\begin{aligned} \cos \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k}) &= \hat{\rho}_i \cdot \hat{u}_i = \frac{\hat{\rho}_i \cdot \hat{k}}{|\hat{\eta}_i \times \hat{k}|}, \\ \sin \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k}) &= \hat{\rho}_i \cdot \hat{v}_i = \frac{\hat{\rho}_i \cdot (\hat{\eta}_i \times \hat{k})}{|\hat{\eta}_i \times \hat{k}|}. \end{aligned}$$

Since arctan2 is the map from the Cartesian coordinates of \mathbb{R}^2 into the polar angle, we have that $\text{arctan2}(\sin \varphi, \cos \varphi) = \varphi$ for all $\varphi \in]-\pi, \pi]$. Thus

$$\begin{aligned} \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k}) &= \text{arctan2}(\sin \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k}), \cos \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k})) \\ &= \text{arctan2}(\hat{\rho}_i \cdot \hat{v}_i, \hat{\rho}_i \cdot \hat{u}_i) \\ &= \text{arctan2}\left(\frac{\hat{\rho}_i \cdot (\hat{\eta}_i \times \hat{k})}{|\hat{\eta}_i \times \hat{k}|}, \frac{\hat{\rho}_i \cdot \hat{k}}{|\hat{\eta}_i \times \hat{k}|}\right) \\ &= \text{arctan2}(\hat{\rho}_i \cdot (\hat{\eta}_i \times \hat{k}), \hat{\rho}_i \cdot \hat{k}), \end{aligned} \quad (3.29)$$

where in the last step we used the fact the polar angle doesn't depend on the length of the vector, i.e $\text{arctan2}(ry, rx) = \text{arctan2}(y, x)$ for all $r > 0$ and $(x, y) \neq (0, 0)$.

In order for (3.29) to be the local rotation angle $\phi_i = \phi(\hat{\eta}_i, \hat{\rho}_i, \hat{k})$, it remains to prove that it satisfies the conditions (i)-(v).

(i) Let R be arbitrary rotation matrix and I the identity matrix. From $I = RR^{-1}$ it follows that $\hat{\mathbf{k}} = RR^{-1}\hat{\mathbf{k}}$. Hence, $R\hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{k}} = R\hat{\boldsymbol{\rho}}_i \cdot RR^{-1}\hat{\mathbf{k}} = \hat{\boldsymbol{\rho}}_i \cdot R^{-1}\hat{\mathbf{k}}$, since rotation matrix preserves the dot product. From the rotational invariance of cross product we get $R\hat{\boldsymbol{\rho}}_i \cdot (R\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}) = R\hat{\boldsymbol{\rho}}_i \cdot (R\hat{\boldsymbol{\eta}}_i \times RR^{-1}\hat{\mathbf{k}}) = R\hat{\boldsymbol{\rho}}_i \cdot R(\hat{\boldsymbol{\eta}}_i \times R^{-1}\hat{\mathbf{k}}) = \hat{\boldsymbol{\rho}}_i \cdot (\hat{\boldsymbol{\eta}}_i \times R^{-1}\hat{\mathbf{k}})$, where we used in the last step again the invariance of dot product under rotation matrix. Thus $\phi(R\hat{\boldsymbol{\eta}}_i, R\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \arctan2(R\hat{\boldsymbol{\rho}}_i \cdot (R\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}), R\hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{k}}) = \arctan2(\hat{\boldsymbol{\rho}}_i \cdot (\hat{\boldsymbol{\eta}}_i \times R^{-1}\hat{\mathbf{k}}), \hat{\boldsymbol{\rho}}_i \cdot R^{-1}\hat{\mathbf{k}}) = \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, R^{-1}\hat{\mathbf{k}})$.

(ii) If $R_{\hat{\mathbf{k}}}$ is a rotation matrix about $\hat{\mathbf{k}}$, then $R_{\hat{\mathbf{k}}}^{-1}\hat{\mathbf{k}} = \hat{\mathbf{k}}$. Hence, the condition follows from condition (i).

(iii) Let $R_{\hat{\boldsymbol{\eta}}_i}$ be a rotation matrix about $\hat{\boldsymbol{\eta}}_i$ by an angle $\Delta\varphi$. Then $R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\eta}}_i = \hat{\boldsymbol{\eta}}_i$, and so, $\phi(R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \phi(\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \arctan2(R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{v}}_i, R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{u}}_i)$. Similarly as in (3.28), we have that $R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\rho}}_i = \cos(\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) + \Delta\varphi)\hat{\mathbf{u}}_i + \sin(\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) + \Delta\varphi)\hat{\mathbf{v}}_i$. Thus $\phi(R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \arctan2(\sin(\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) + \Delta\varphi), \cos(\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) + \Delta\varphi)) = \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) + \Delta\varphi$.

(iv) Note that $\hat{\mathbf{v}}_i \propto \hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}$, and so, a rotation matrix about $\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}$ is a rotation matrix about $\hat{\mathbf{v}}_i$ and vice versa. Let $R_{\hat{\mathbf{v}}_i}$ be a rotation matrix about $\hat{\mathbf{v}}_i$ by an angle $\Delta\theta$. In the $(\hat{\mathbf{u}}_i, \hat{\mathbf{v}}_i, \hat{\boldsymbol{\eta}}_i)$ basis we have

$$\hat{\mathbf{k}} = \begin{pmatrix} \sin \theta_i \\ 0 \\ \cos \theta_i \end{pmatrix}, \quad R_{\hat{\mathbf{v}}_i} = \begin{pmatrix} \cos \Delta\theta_i & 0 & \sin \Delta\theta_i \\ 0 & 1 & 0 \\ -\sin \Delta\theta_i & 0 & \cos \Delta\theta_i \end{pmatrix}, \quad \hat{\boldsymbol{\eta}}_i = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \hat{\boldsymbol{\rho}}_i = \begin{pmatrix} \cos \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) \\ \sin \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) \\ 0 \end{pmatrix}.$$

Then by straight forward computation we obtain $R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\rho}}_i \cdot (R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}) = \sin(\theta_i - \Delta\theta)\sin\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$ and $R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{k}} = \sin(\theta_i - \Delta\theta)\cos\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$. Assuming that $\phi(R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\eta}}_i, R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$ is defined, we have $0 \neq \theta_i - \Delta\theta \neq \pi$, and so, $\sin(\theta_i - \Delta\theta) \neq 0$. Thus

$$\begin{aligned} \phi(R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\eta}}_i, R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) &= \arctan2(R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\rho}}_i \cdot (R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}), R_{\hat{\mathbf{v}}_i}\hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{k}}) \\ &= \arctan2(\sin(\theta_i - \Delta\theta)\sin\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}), \sin(\theta_i - \Delta\theta)\cos\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})) \\ &= \arctan2(\sin\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}), \cos\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})) \\ &= \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}). \end{aligned}$$

(v) We have $\hat{\mathbf{k}} = \sin \theta_i \hat{\mathbf{u}}_i + \cos \theta_i \hat{\boldsymbol{\eta}}_i$ and if $\hat{\boldsymbol{\eta}}_i \cdot \hat{\boldsymbol{\rho}}'_i = 0$ then there is a unique $\vartheta'_i \in]-\pi, \pi]$, such that $\hat{\boldsymbol{\rho}}'_i = \cos \vartheta'_i \hat{\mathbf{u}}_i + \sin \vartheta'_i \hat{\mathbf{v}}_i$. Then $\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}'_i = \sin \theta_i \cos \vartheta'_i$ and

$$\begin{aligned} \max\{\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}'_i \mid \hat{\boldsymbol{\eta}}_i \cdot \hat{\boldsymbol{\rho}}'_i = 0\} &= \max\{\sin \theta_i \cos \vartheta'_i \mid -\pi < \vartheta'_i \leq \pi\} \\ &= \sin \theta_i \max\{\cos \vartheta'_i \mid -\pi < \vartheta'_i \leq \pi\} \\ &= \sin \theta_i \max\{\cos \vartheta'_i \mid -1 \leq \cos \vartheta'_i \leq 1\} \\ &= \sin \theta_i. \end{aligned}$$

On the other hand $\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}_i = \sin \theta_i \cos \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$ and $\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) \in]-\pi, \pi]$. Thus,

$$\begin{aligned} \hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}_i &= \max\{\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}'_i \mid \hat{\boldsymbol{\eta}}_i \cdot \hat{\boldsymbol{\rho}}'_i = 0\} \\ \iff \sin \theta_i \cos \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) &= \sin \theta_i \\ \iff \cos \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) &= 1 \\ \iff \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) &= 0. \end{aligned}$$

We have proven that the function $\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \arctan2(\hat{\boldsymbol{\rho}}_i \cdot (\hat{\boldsymbol{\eta}}_i \times \hat{\mathbf{k}}), \hat{\boldsymbol{\rho}}_i \cdot \hat{\mathbf{k}})$ satisfies the conditions (i)-(v). To prove the uniqueness of the local rotation angle $\phi_i = \phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$, we assume that $\tilde{\phi}$ is another function satisfying the conditions (i)-(v). Take arbitrary reference vector $\hat{\mathbf{k}}$ and tilt rotation vector pair $(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i)$. Assuming that the local rotation angle is defined, we have $\hat{\boldsymbol{\eta}}_i \neq \pm \hat{\mathbf{k}}$. Let $R_{\hat{\boldsymbol{\eta}}_i}$ be a rotation matrix about $\hat{\boldsymbol{\eta}}_i$ by an angle $\Delta\varphi$, such that it maximizes the inner product $\hat{\mathbf{k}} \cdot R_{\hat{\boldsymbol{\eta}}_i} \hat{\boldsymbol{\rho}}_i$, i.e. $\hat{\mathbf{k}} \cdot R_{\hat{\boldsymbol{\eta}}_i} \hat{\boldsymbol{\rho}}_i = \max\{\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\rho}}'_i \mid \hat{\boldsymbol{\eta}}_i \cdot \hat{\boldsymbol{\rho}}'_i = 0\}$. Then from the condition (v) we have that $\phi(\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i} \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = 0 = \tilde{\phi}(\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i} \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$. On the other hand, from the condition (iii) we have that $\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) - \Delta\varphi = \phi(\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i} \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$ and $\tilde{\phi}(\hat{\boldsymbol{\eta}}_i, R_{\hat{\boldsymbol{\eta}}_i} \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \tilde{\phi}(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) - \Delta\varphi$. Hence, $\phi(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}}) = \tilde{\phi}(\hat{\boldsymbol{\eta}}_i, \hat{\boldsymbol{\rho}}_i, \hat{\mathbf{k}})$. Thus the conditions (i)-(v) define the local rotation angle uniquely.

4. The method in practice

To demonstrate the capabilities of the method, two simple example analyses of transmembrane α -helices are presented in this chapter. For studying transmembrane proteins, classical molecular dynamic simulations are a widely used technique, where the Newton's equation of motion for each atom of the system is solved by numerical integration. Our main source for the α -helix structures used in the examples are given by the trajectories obtained from classical molecular dynamic simulations.

In the first example, we study the tilt, rotation and kink of a kinked transmembrane α -helix. In the second example, we compare the tilt and rotation between glycophorin A transmembrane α -helix monomer and dimer. We do this by comparing five trajectories of the glycophorin A transmembrane α -helix monomer to experimentally obtained glycophorin A transmembrane α -helix dimer. In these examples our focus is on the method presented in the previous chapter for the tilt and rotation. Therefore, we take the trajectories obtained from molecular dynamic simulations and the experimental structure as given. The first example is like a sightseeing tour of the method. We are looking on how the tilt, rotation and kink evolve during the given trajectory and what their values do and do not describe. In the second example, we answer relatively simple question about how the tilt and rotation differ between the glycophorin A experimental dimer and the monomer trajectories.

To utilize the method to perform actual analysis, it has been written in a Python package HelixTiltRot [8]. The presented analysis and all the figures in the following examples are done with this Python package, and these Python scripts are also included in the Github [8].

4.1 Example 1: Kinked transmembrane α -helix

We study the tilt, rotation and kink of a kinked transmembrane α -helix in a 5242 ns long trajectory obtained from molecular dynamic simulations. The amino acid sequence of the α -helix is given in Table 4.1, where single letter codes of amino acids are used and the residues are labeled from N-terminus to C-terminus with the integers 0, ..., 44.

Table 4.1: Amino acid sequence of the transmembrane α -helix. Charged, polar uncharged and hydrophobic residues are colored blue, purple and green, respectively. Other residues, so called special cases, are colored white.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
D	Y	L	D	V	P	S	N	I	A	K	I	I	I	G	P	L	I	F	V	F	L	F	S	V	V	I	G	S	I	Y	L	F	L	R	K	R	Q	P	D	G	P	L	G	P

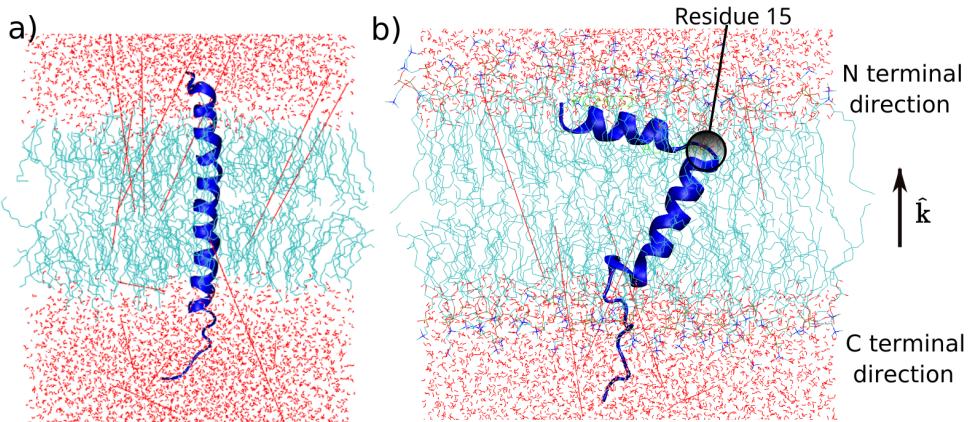


Figure 4.1: The transmembrane α -helix showed in blue cartoon representation. The membrane is colored with light blue and the water above and below the membrane is colored with red. The black arrow represents the membrane normal $\hat{\mathbf{k}}$. a) Beginning of the trajectory. b) End of the trajectory and position of the residue 15 is highlighted.

System of the transmembrane α -helix is shown in Fig 4.1. The transmembrane α -helix is straight in the beginning of the trajectory (Fig 4.1a) and kinked in the end of the trajectory (Fig 4.1b). We fix the reference vector $\hat{\mathbf{k}}$ to be the membrane normal (Fig 4.1b).

In Fig 4.2, we see the local tilt angles $\{\theta_i\}$ at each time step. In Fig 4.2a the local tilt angles $\{\theta_i\}$ are computed from all residues. It's important to remember that the local tilt angle θ_i is a function of the five α -carbon atoms from residues $i - 2, \dots, i + 2$. Therefore we can't compute the local tilt angle for the first two and last two residues, which is why they are colored white in Fig 4.2a. In Fig 4.2b only helical residues are used. Here, by a helical residue, we mean a residue, such that DSSP identifies it as an α -helix, π -helix or 3_{10} -helix. Hence, if a point (t, i) has a nonwhite color in Fig 4.2b, then DSSP has identified all the five residues $i - 2, \dots, i + 2$ needed for the local tilt angle $\theta_i(t)$ to be helical at the time step t . For example, since residue 17 has nonwhite colors through the whole trajectory, the residues 15 and 16 are both helical through the whole trajectory although they are mainly white. This way Fig 4.2b can be used to study the evolution of secondary structure.

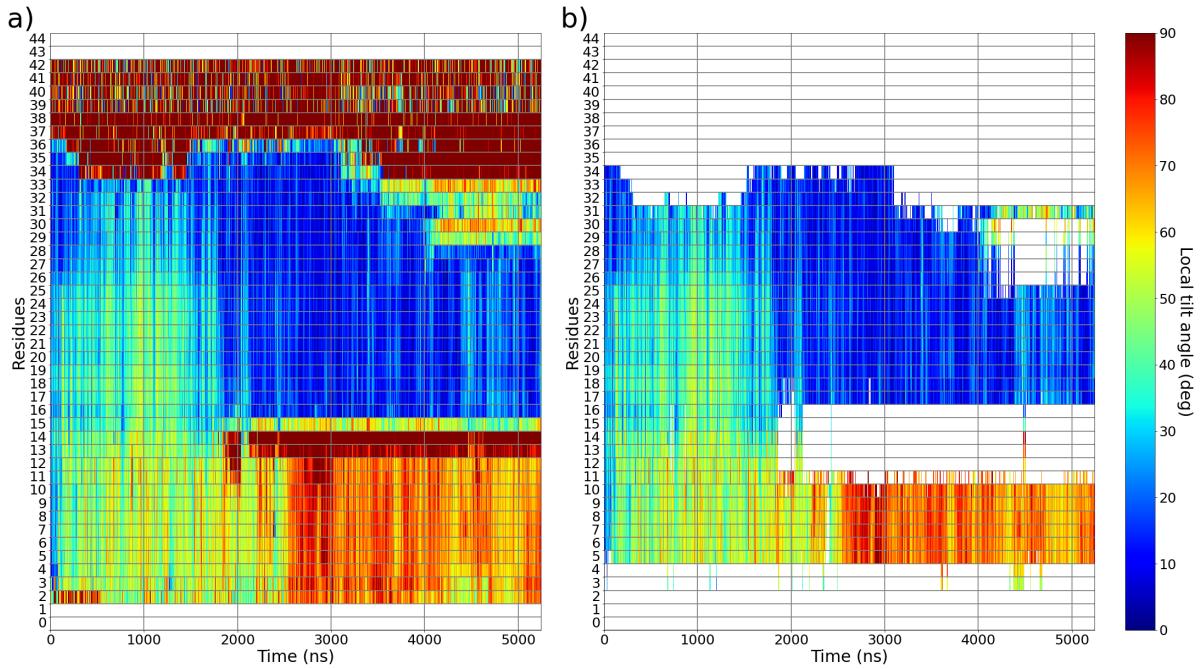


Figure 4.2: The local tilt angles $\{\theta_i\}$ for each time step of the trajectory. Time is in the x-axis, residues are in the y-axis and the local tilt angle is shown with colors. The presented color map is used in both figures a) and b). The local tilt angle $\theta_i(t)$ of residue i at time t is shown as a color at a point (t, i) . If the local tilt angle doesn't have a value at a point (t, i) , then the point (t, i) is colored white. a) All residues are used to compute local tilt angles. b) Only helical residues are used to compute local tilt angles.

The Fig 4.2 tells how the initial straight α -helix (Fig 4.1a) has evolved into the kinked α -helix at the end of the trajectory (Fig 4.1b). The residues starts to tilt so fast at the beginning of the trajectory, that one can barely see $\theta_i \approx 0^\circ$ (dark blue) for all residues i at $t = 0$ ns (Fig 4.2). In a few hundred nanoseconds, $\theta_i \approx 50^\circ$ (green) for almost all residues i . In the first ~ 1500 ns at almost every time step, all residues have roughly the same local tilt angle. This indicates that the α -helix is roughly straight and rigid during the first ~ 1500 ns. After ~ 1500 ns the α -helix starts to kink, the residues $i > 15$ tend to go back to small tilt (dark blue) and the residues $i < 15$ tilt even more (Fig 4.2). According to DDSP, after ~ 1800 ns the residue 14 is not helical anymore and the residue 13 oscillates between helical and non-helical (Fig 4.2). After ~ 1800 ns the local tilt angle settles down roughly into range $[0^\circ, 30^\circ]$ (blue) for the residues $i > 15$ (Fig 4.2). After 2500 ns, the local tilt angle settles down roughly into range $[50^\circ, 80^\circ]$ (orange-red) for the residues $i < 15$ (Fig 4.2). Hence there is a kink at residue 15.

The residue 15 is proline (Table 4.1). Proline in the middle of a transmembrane α -helix is known to be a common motif for a kink [9]. The kink angle $\xi_{(5,14),(16,26)}$ and the tilt angles $\Theta_{5,14}$ and $\Theta_{16,26}$ are shown in Fig 4.3. After ~ 2300 ns the kink angle

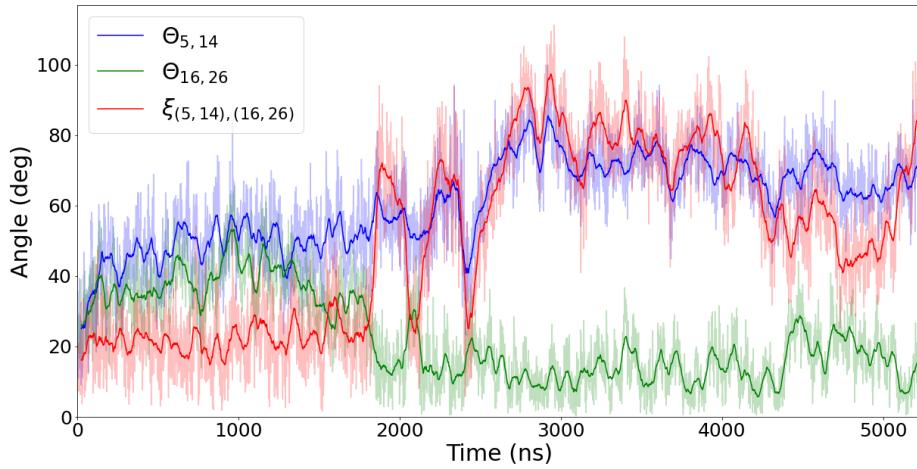


Figure 4.3: The kink angle $\xi_{(5,14),(16,26)}$ and the tilt angles $\Theta_{5,14}$ and $\Theta_{16,26}$ are shown in red, blue and green, respectively. For each angle rolling average over 50 ns is shown with a darker line and the actual values are shown with a transparent line.

$\xi_{(5,14),(16,26)}$ and the the tilt angle $\Theta_{5,14}$ are roughly the same (Fig 4.3). This is because the arithmetic average of the axes $\{\hat{\mathbf{H}}_{16,26}(t) \mid t > 2300 \text{ ns}\}$ normalized to unity is $(\hat{\mathbf{H}}_{16,26})_{\text{avg}} = -0.002\hat{\mathbf{i}} - 0.025\hat{\mathbf{j}} + 0.960\hat{\mathbf{k}} \approx \hat{\mathbf{k}}$, where $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ is an orthonormal basis and $\hat{\mathbf{k}}$ is still the membrane normal (Fig 4.1). Remember that $\xi_{(5,14),(16,26)}$ is the angle between the axes $\hat{\mathbf{H}}_{5,14}$ and $\hat{\mathbf{H}}_{16,26}$ and $\Theta_{5,14}$ is the angle between $\hat{\mathbf{H}}_{5,14}$ and $\hat{\mathbf{k}}$. Hence $\xi_{(5,14),(16,26)} \approx \Theta_{5,14}$ follows from $(\hat{\mathbf{H}}_{16,26})_{\text{avg}} \approx \hat{\mathbf{k}}$.

As described in the section 2.2, the rotation is not defined if $\hat{\mathbf{H}} = \pm\hat{\mathbf{k}}$, since then all sides of the helical structure are pointing equivalently into the reference vector $\hat{\mathbf{k}}$. Hence $(\hat{\mathbf{H}}_{16,26})_{\text{avg}} \approx \hat{\mathbf{k}}$ indicates that the residues $i > 15$ don't have a specific rotation. Since the helix $i > 15$ is on average parallel to the membrane normal $\hat{\mathbf{k}}$, there are most likely no residues causing a specific rotation by reaching into the hydrophobic membrane or into the polar water. The rotation of residues $i > 15$ might not be interesting, but it is a good way to clarify the definition of rotation and the two motions described to change rotation in the section 2.2. Therefore, we will come back to the rotation of residues $i > 15$ later.

On the contrary, the rotation of residues $i < 15$ is very interesting. Their tilt is roughly 70° (Fig 4.3) and their position is on the boundary of the membrane and the water (Fig 4.1b). The local rotation angles $\{\phi_6, \dots, \phi_{11}\}$ are shown in Fig 4.4. From the figures and standard deviations in Fig 4.4, we can conclude that the residues $6, \dots, 11$ stay in a specific rotation during the whole trajectory. Remember that from the conditions (iii) and (v) it followed that the closer ϕ_i is to 0 the more aligned the rotation vector $\hat{\rho}_i$ is to the reference vector $\hat{\mathbf{k}}$, and if $|\phi_i| < 90^\circ$, then $\hat{\rho}_i$ points in the direction of $\hat{\mathbf{k}}$. Therefore, from Fig 4.4, we get that the residues 6, 7, 10 and 11 point at the direction

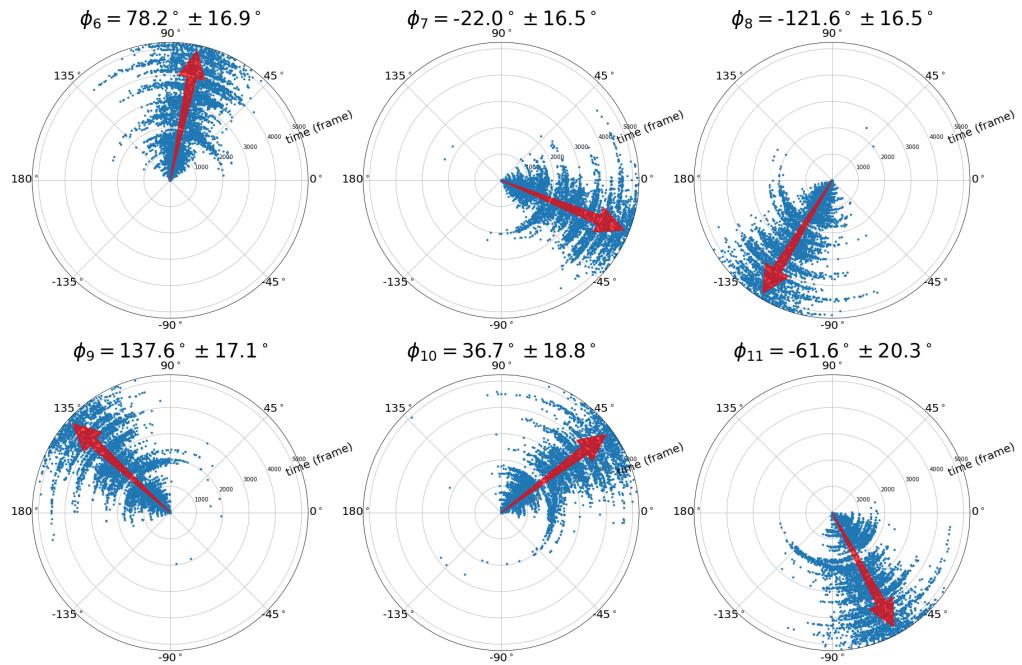


Figure 4.4: For each local rotation angle ϕ_6, \dots, ϕ_{11} a polar figure is shown. Time is in the radius and the local rotation angle in the polar angle. The angle of the red arrow is the circular average μ_i and the greater its length the narrower the distribution. The average μ_i and the circular standard deviation σ_i are shown in the title of each figure as $\phi_i = \mu_i \pm \sigma_i$.

of the membrane normal $\hat{\mathbf{k}}$ and the residues 8 and 9 point at the opposite direction of $\hat{\mathbf{k}}$ (Fig 4.4). The residues are on the boundary of the membrane and the water and on this boundary $\hat{\mathbf{k}}$ points into the water (Fig 4.1b). Hence the residues 6, 7, 10 and 11 point into water and the residues 8 and 9 point into the membrane. This result is perfectly compatible with the amino acid sequence (Table 4.1) of the α -helix. The residues 6, 7 and 10 are polar or charged, which is why they want to have contact with polar water. The residues 8 and 9 are hydrophobic, which is why they want to point into the hydrophobic interior of the membrane.

The turn angle $\Delta\phi$ of the α -helix is 100.1° (Table 2.1). The local rotation angles $\{\phi_i^{10}\}$ in the phase of residue 10 are shown in Fig 4.5, where we have used $\Delta\phi = 100.1^\circ$ to compute them. For the local rotation angle we had that $\phi_i - \phi_{j+1} = \Delta\phi$. This can be seen in Fig 4.4 for residues $6, \dots, 11$ and in Fig 4.5 separately for residues $i < 15$ and $i > 15$ after ~ 2000 ns. To see this in Fig 4.5, note that $\phi_i^{10} = \phi_{i+1}^{10}$ follows from $\phi_i - \phi_{j+1} = \Delta\phi$. For almost every time step $t > 2000$ ns, we have $\phi_i^{10} = \phi_{i+1}^{10}$ for all residues $i < 15$ and $i > 15$ (Fig 4.5). This means that reducing the local rotation angles of the α -helix into two rotation angles $\Phi_{5,14}^{10}$ and $\Phi_{16,26}^{10}$ is reasonable (Fig 4.6).

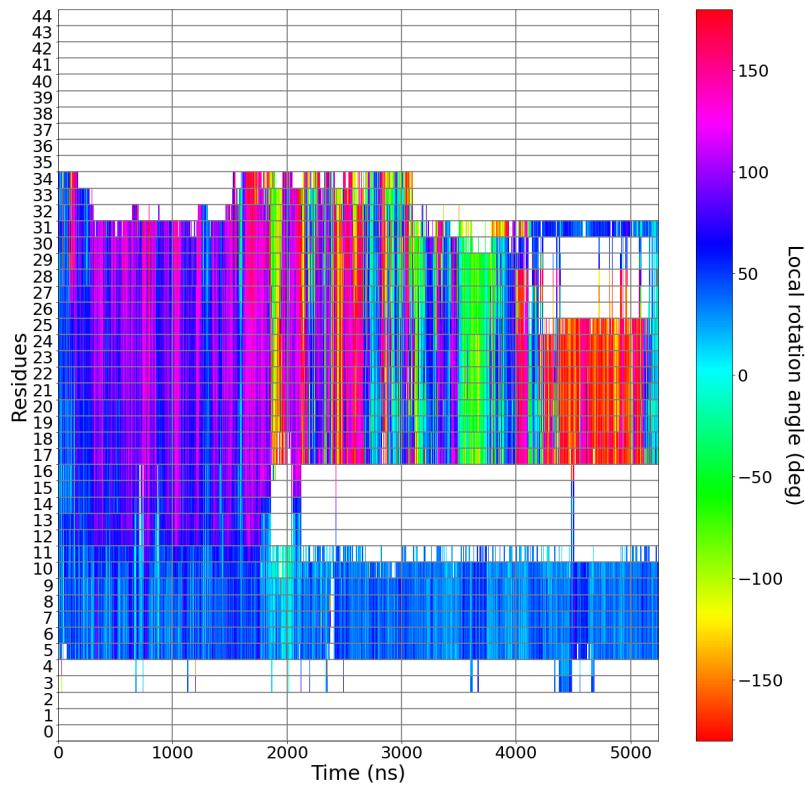


Figure 4.5: Otherwise equivalent figure for local rotation angle as (Fig 4.2b) is for local tilt angle but the used color map is periodic.

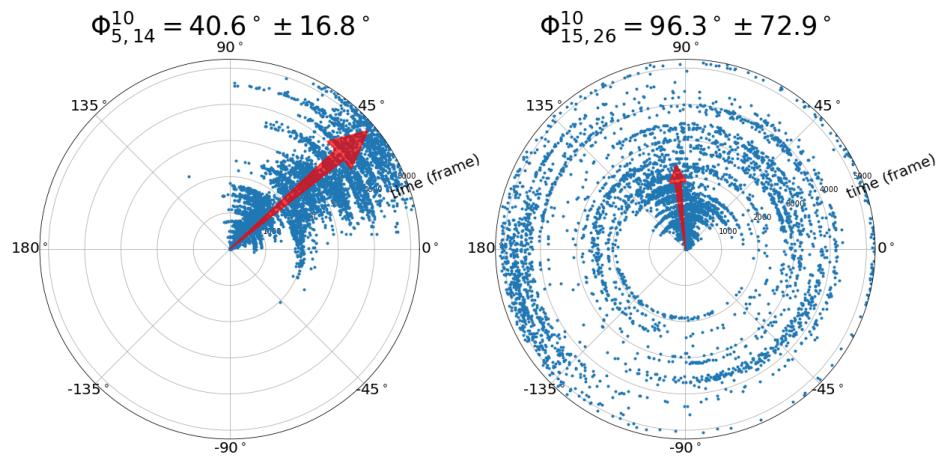


Figure 4.6: The rotation angles $\Phi_{5,14}^{10}$ and $\Phi_{15,26}^{10}$ in polar coordinates, where time is on the radius.

The residues $i < 15$ have really specific rotation but residues $i > 15$ don't, in fact the rotation angle $\Phi_{16,26}^{10}$ has values from -180° to 180° (Fig 4.6). Does it mean that the residues $i > 15$ do rotate widely around the axis $\hat{\mathbf{H}}_{16,26}$? No. As described in the section 2.2, the rotation angle Φ describes which residues point in the direction of the reference vector $\hat{\mathbf{k}}$. As long as the axis $\hat{\mathbf{H}}$ is not close to $\pm\hat{\mathbf{k}}$, the rotation angle Φ describes how the helical structure rotates around its axis $\hat{\mathbf{H}}$. But the axis $\hat{\mathbf{H}}_{16,26}$ of residues $i > 15$ frequently aligns to our reference $\hat{\mathbf{k}}$, since the tilt angle $\Theta_{16,26}$ frequently goes to 0 (Fig 4.3). Hence, the Fig 4.6 tells us only that the residues $i < 15$ do not have specific rotation, i.e. the residues pointing towards the membrane normal $\hat{\mathbf{k}}$ are changing constantly. This leaves two options for how the helix is moving in this trajectory as discussed in the section 2.2. First one is the widely rotating helix around its own axis $\hat{\mathbf{H}}_{16,26}$ and second one is that the axis $\hat{\mathbf{H}}_{16,26}$ wiggles around the membrane normal $\hat{\mathbf{k}}$, so the residues pointing towards the membrane normal $\hat{\mathbf{k}}$ are changing.

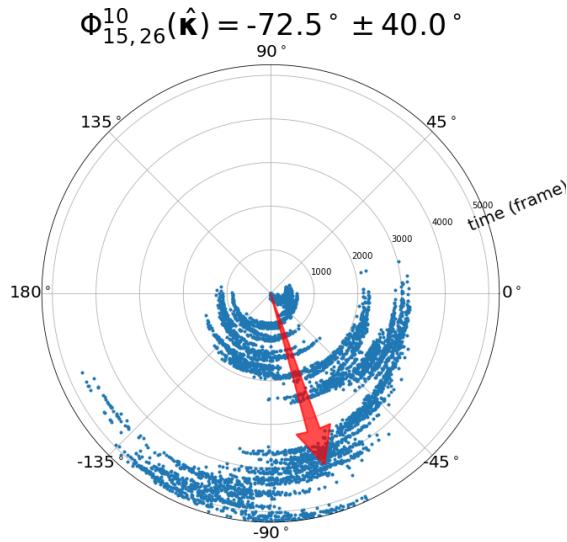


Figure 4.7: The rotation angle $\Phi_{15,26}^{10}(\hat{\kappa})$ in polar coordinates, where time is on the radius.

We can figure out which one of these two motions we have by taking a new reference vector $\hat{\kappa}$, such that it is roughly orthogonal to $\hat{\mathbf{H}}_{16,26} \approx \hat{\mathbf{k}}$ throughout the simulation. This way we know that the rotation angle $\Phi_{15,26}^{10}(\hat{\kappa})$ relative to $\hat{\kappa}$ describes how the helix is rotating around its axis $\hat{\mathbf{H}}_{16,26}$. We got $\Phi_{15,26}^{10}(\hat{\kappa}, t) \in [-180^\circ, 0^\circ]$ for basically every time step t (Fig 4.7). Hence, the residues $i > 15$ of the α -helix do not rotate widely around $\hat{\mathbf{H}}_{16,26}$, i.e. the axis $\hat{\mathbf{H}}_{16,26}$ wiggles around the membrane normal $\hat{\mathbf{k}}$. The takeaway is that if the tilt is frequently around 0° or 180° and the rotation angle has very wide range of values, then one can't straight away conclude how the helix is moving, i.e. whether it is wiggling or rotating widely.

4.2 Example 2: Difference between glycophorin A monomer and dimer

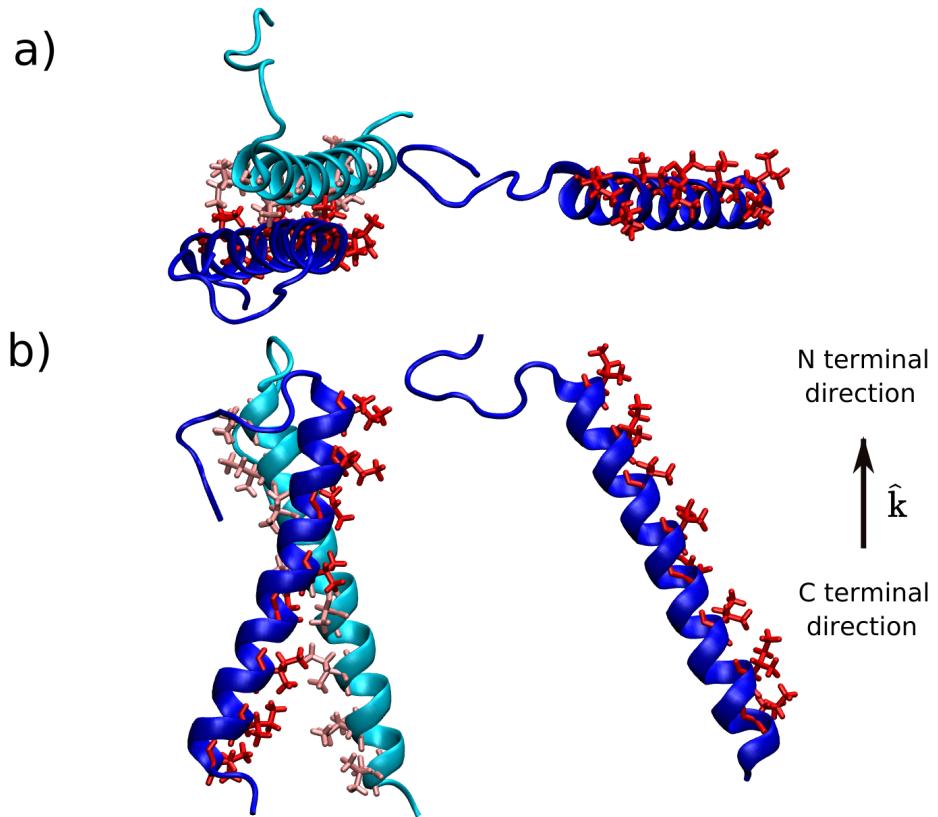


Figure 4.8: The experimental dimer is on the left and the monomer from last frame of the fourth repeat is on the right. The chain B of the dimer and the monomer are colored blue while the chain A of the dimer is colored cyan. The side chains are shown for residues 11, 15, 18, 22, 25, 29, 33 and 36 and are colored red for the monomer and chain B and pink for chain A. a) Top view, the membrane normal \hat{k} is pointing out of the screen. b) Side view, the membrane normal \hat{k} is pointing up.

We will now study how the tilt and rotation differ between a dimer and a monomer of glycophorin A transmembrane α -helix. For the dimer we are going to use an experimental structure, model 1 of the family of structures [10] with the Brookhaven Protein Databank accession number 1AFO. The dimer is a homodimer, that is, both of the transmembrane α -helices, chain A and chain B, have the same amino acid sequence (Table 4.2). The dimer has a C2 symmetry, that is, the chain A can be aligned with chain B by rotating it about the membrane normal and around their center of mass by an angle 180° . For the α -helix monomer, we are using five ~ 5000 ns long trajectories with the same sequence as chain A and chain B (Table 4.2). For each of these five trajectories, the initial state is obtained by inserting the α -helix into the membrane parallel to the membrane normal. Therefore, we will use the words trajectory and repeat equivalently.

Table 4.2: Amino acid sequence of the glycophorin A transmembrane α -helices labeled from N-terminus to C-terminus with the integers 0, …, 39. Charged, polar uncharged and hydrophobic residues are colored blue, purple and green, respectively. Other residues, so called special cases, are colored white.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
V	Q	L	A	H	H	F	S	E	P	E	I	T	L	I	I	F	G	V	M	A	G	V	I	G	T	I	L	L	I	S	Y	G	I	R	R	L	I	K	K

For the analysis, we are going to use only the last 2000 ns from the trajectories, as it takes some time for the system to find an equilibrium. The dimer has a C2 symmetry and thereby the chains A and B have the same tilt and rotation. Therefore, it is enough to compare only one of the chains to the monomer, we will choose the chain B. We fix the reference vector $\hat{\mathbf{k}}$ to be the membrane normal. The distributions of the tilt angle differences $\{\Theta^{M_j}(t) - \Theta^B \mid t \text{ in last 2000 ns}\}$ are shown in Fig 4.9 for each repeat $j = 1, 2, 3, 4, 5$. Here Θ^B is the tilt angle of chain B from the dimer and $\Theta^{M_j}(t)$ is the tilt angle of the monomer at time step t from repeat j . To compute the tilt angles Θ^B and $\Theta^{M_j}(t)$, only those residues are used that DSSP identifies as an α -helical. The tilt angle difference distributions are nicely aligned in the five repeats and on average the monomer has 15.5° greater tilt angle than the dimer (Fig 4.9).

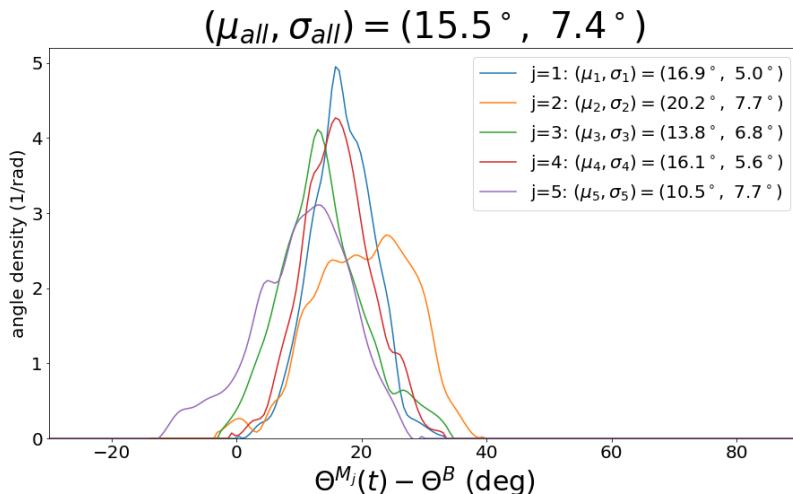


Figure 4.9: Distributions of tilt angle differences $\Theta^{M_j}(t) - \Theta^B$ for each repeat $j = 1, 2, 3, 4, 5$. The arithmetic average μ_j and the standard deviation σ_j for each repeat are in the legend. The arithmetic average μ_{all} and standard deviation σ_{all} from values obtained in all five repeats are in the title.

The tilt of the chain B is $\Theta^B = 18.9^\circ$, and so, chain B has a rotation relative to the membrane normal. The smallest and largest values for the difference $\Theta^{M_j}(t) - \Theta^B$ are roughly -10° and 40° (Fig 4.9). Hence, the smallest and largest values for $\Theta^{M_j}(t)$ are roughly $-10^\circ + 18.9^\circ = 8.9^\circ$ and 58.9° . Therefore, the rotation relative to membrane normal is also defined for the monomer through the trajectories. The distributions of

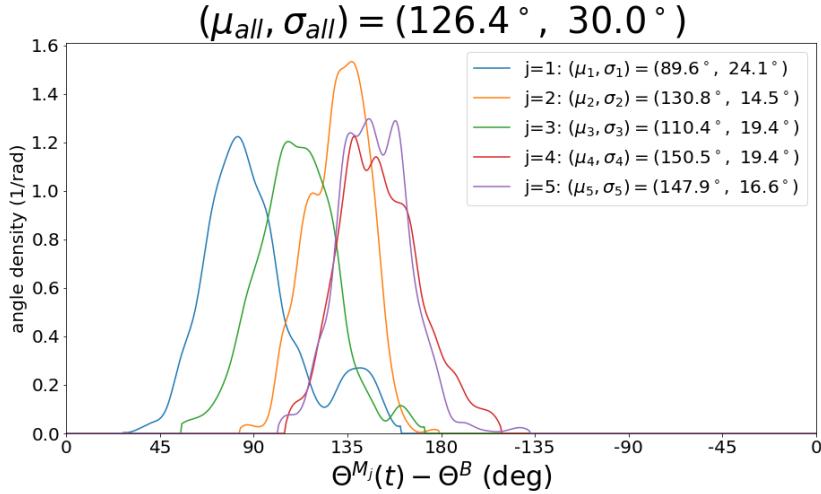


Figure 4.10: Distributions of rotation angle differences $D(\phi_i^{M_j}(t), \phi_i^B)$ for each repeat $j = 1, 2, 3, 4, 5$. The arithmetic average μ_j and the standard deviation σ_j for each repeat are in the legend. The arithmetic average μ_{all} and standard deviation σ_{all} computed from all values obtained in all five repeats are in the title.

the rotation differences $\{D(\phi_i^{M_j}(t), \phi_i^B) \mid t \text{ in last 2000 ns}\}$ are shown in (Fig 4.10) for each repeat. The distributions are overlapping nicely, although the first repeat seems to be an outlier. Interestingly the rotation difference between the dimer and monomer is on average 126.4° (Fig 4.10). For many time steps, the difference is $\sim 180^\circ$ (Fig 4.10), which means that the dimer and monomer have opposite rotations. Such time step of the monomer is shown next to the dimer in Fig 4.8, where the shown red side chains are pointing up for the monomer and down for the dimer. Thus according to these trajectories the glycophorin A monomer has on average greater tilt angle by 15.5° and rotation angle by 126.4° than the experimentally obtained glycophorin A dimer.

5. Conclusions

I developed a new robust analytical method for calculation of the tilt, rotation and kink of various biological secondary structures. These geometrical quantities are important to characterize the orientation of a helical structure relative to its environment. For example, in the signaling process of RTKs the dimerization of two transmembrane α -helices presumably transmits the signal through the membrane. The tilt and rotation alongside translation characterizes the active and inactive states of the dimer. Similarly, the orientation of a single transmembrane α -helix in the membrane is characterized by the tilt, rotation and translation. Therefore, to study the signaling process of RTKs, being able to quantify the tilt and rotation is essential.

The method can be used to analyze experimental and theoretical helical structures and molecular dynamic simulations of helical structures. To make it easy for others to use the method in their work, I have implemented the method in a Python package HelixTiltRot [8].

I presented a detailed mathematical derivation of the calculations and approximations. The main building block of the method is the approximation of local segment's local axis. I derived the approximation of the local axis stated in [7] and showed that it is biologically equivalent to the exact local axis for the α -helix, 3_{10} -helix and π -helix. I also added a condition to this approximation to make it work for β -strand-like secondary structures. The main result of this thesis is the developed method to describe a rotation of straight, bent and kinked helical structures. The definition of the rotation is built on the local rotation angle, which is uniquely defined by the postulated conditions (i)-(v). The local rotation angle makes it possible to analyze rotation of a single residue, the whole structure and to compare rotation between two single residues and between two structures.

Although the thesis focuses on helical secondary structures of proteins, the method is not limited to proteins, it can be used to other helical structures as well, perhaps for DNA. When applying the method for a helical structure different from the α -helix, 3_{10} -helix, π -helix and β -strand, one has to keep in mind the two approximations of the method. The first approximation is the assumption that a helix curve is a good fit for each local segment containing four consecutive α -carbon atoms and the second one is that the APPROXc is

a good enough approximation for each local segment's local axis. These approximations have to be checked especially if one uses this method for a helical structure, such that its turn angle $\Delta\phi$ deviates from the ones listed in Table 2.1 or otherwise the method might break down severely.

To demonstrate the method in practice, I showed two case studies using the method in molecular dynamic simulation studies of transmembrane α -helices. In the first example, a trajectory of a kinked transmembrane α -helix was studied. The local tilt angle turned out to be useful to spot the kink and to see how it formed and evolved during the trajectory. It was shown how the local rotation angle can be used to analyze single residues. The first example was also used to clarify, why one has to be careful before jumping into conclusions about the structure's motion from its rotation, if the tilt is frequently going to zero. The second example demonstrated how the method can be used to compare the tilt and rotation between two states of a protein. Interestingly, it turned out that the used glycophorin A experimental dimer has noticeably different tilt and almost opposite rotation than the glycophorin A monomer in the given trajectories.

Bibliography

- [1] David Whitford. *Proteins: Structure and Function*. John Wiley & Sons Ltd, 2005.
- [2] Chen Cao, Shutian Xu, and Lincong Wang. An algorithm for protein helix assignment using helix geometry. *PLOS ONE*, 10:1–20, 07 2015.
- [3] Sander C Kabsch W. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.
- [4] J. Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 103(2):211–225, 2000.
- [5] M. A. Lemmon and J. Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 141(7):1117–1134, 2010.
- [6] J. M. Westerfield and F. N. Barrera. Membrane receptor activation mechanisms and transmembrane peptide tools to elucidate them. *Journal of Biological Chemistry*, 295(7):1792–1814, 2020.
- [7] Peter C. Kahn. Defining the axis of a helix. *Computers & Chemistry*, 13(3):185–189, 1989.
- [8] Sakari Pirnes. HelixTiltRot, 1 2022. <https://github.com/SakariPirnes/helixtiltrot>.
- [9] Gunnar von Heijne. Proline kinks in transmembrane α -helices. *Journal of Molecular Biology*, 218(3):499–503, 1991.
- [10] Kevin R. MacKenzie, James H. Prestegard, and Donald M. Engelman. A transmembrane helix dimer: Structure and implications. *Science*, 276(5309):131–133, 1997.