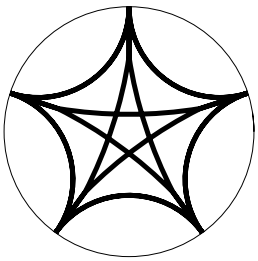


On representation of protein backbones with (framed) space curves

Peter Røgen and Henrik Bohr

DEPARTMENT OF MATHEMATICS

TECHNICAL UNIVERSITY OF DENMARK

Mat-Report No. 2002–14

August 2002

DTU

Title of Paper:

On representation of protein backbones with (framed) space curves

Authors:

Peter Røgen* and Henrik Bohr†

Addresses:

*Department of Mathematics, Technical University of Denmark, Matematiktorvet, Building 303, DK-2800 Kongens Lyngby, Denmark

†QUP, Technical University of Denmark, Building 309, DK-2800 Kongens Lyngby, Denmark

Email:

Peter.Roegen@mat.dtu.dk
hbohr@fysik.dtu.dk

Abstract:

We investigate fundamental questions concerning representations of protein backbones by framed (ribbon) or unframed space curves. Maximal sets of protein structures on which elastic rod like models of proteins can be constructed are found. It is concluded that such models can model deformations of α-helices but not more general shaped proteins. A consequence of our general result that: "Protein backbones cannot be represented by framed space curves (ribbons) in a "time" continuous way", is that any notion of torsion of protein backbones should be avoided.

Key words: Differential geometry of protein backbones, Local protein structure description, Approximation of protein backbones by helices, Proteins as elastic rods, Bending and torsional profiles of proteins.

1991 Mathematics Subject Classification: Primary 92E10; Secondary 53A04 92C40;

*The work was supported by a DTU-grant and by a grant from Carlsbergfondet.

1 Introduction

The importance of determining protein structures from spectroscopic data, and making predictions from sequence information has been eminent in the last decade due to the fact that many developments in bio-technology for instance drug design, rely on the 3-dimensional structure of proteins. The understanding of protein folding lies in the heart of these matters and still remains an unsolved problem. It has been suggested[1] that the process of protein folding from synthesis until the establishing of eventual folding intermediates may be modelled as an elastic rod. To establish such a continuous protein backbone model, a potential energy as function of space curve geometry has to be defined. The question of to which extend this is possible, is the main motivation of the work presented here.

It is also relevant to ask how well-defined mathematically a space curve geometry of the protein backbone should be for being of use to construct the complete 3-dimensional structure of the backbone. Evidently full knowledge of all the dihedral angles suffices to construct the essential 3-D backbone structure. However, the opposite case of deriving a continuous space curve geometry from 3-D protein structure data has ambiguities build in, a fact that is due to the discreteness of the latter data. Since protein structures in 3-D for most problems is the final answer, to e.g. drug design, one might ask what the relevance of constructing a space curve geometry is. The relevance is understood from the fact that the protein folding problem becomes easier to solve if the space curve geometry of a protein, i.e., the overall shape or fold, is given, and in turn the curve geometry, e.g. in terms of torsion and bending profiles, are easier to predict from sequence data than the full 3-D structure of the protein. The predictions are often made from knowledge based systems such as neural networks and hidden Markov models [2, 3, 4]. These prediction tools are trained from existing 3-D structures, such as X-ray crystallographic diffraction data. Therefore, it is important to understand the mathematical problem of constructing space curve geometry from 3-D protein structures. We shall be concerned mostly with the mathematical problem of representing protein structure by space curves in this first paper, while in the succeeding paper we shall deal with global geometric descriptions of protein backbone chains.

The shape of a protein backbone seen from a distance is often represented or merely visualized by a continuous space curve that eg. joins the C_α atoms. Backbones are equipped with an orientation that goes from nitrogen to oxygen. This orientation is induced to curves representing backbones. Furthermore, the curve is often equipped with an orthogonal unit vector field to model the planar links in backbone chains. Hence, a ribbon or a framed curve is used to model or visualize the shape of a backbone. The first question one has to address is: "*What is the space curve representation going to be used for?*". In the case of structural analyses consider a tube around the protein backbone. The structural point of view is to emphasize the shape of the tube rather than what is inside it. That is to say that two backbones have equivalent large scale structure if the lie in the same tube even if their dihedral angles etc not are the same. Collapsing this tube to a curve, it is seen that large scale shape of protein backbones is precisely captured by space

curve geometry. In the case of protein dynamics we show that such a large scale structure does not contain enough information to define a potential energy, that is space curve evolution does not apply to protein dynamics. An interesting question is thus: "*Is there a 1-1 map between protein geometry and space curve geometry?*". This question is answered by calculating the determinant of the Jacobian between protein geometry and space curve geometry and demanding that it is zero free. From this analysis it follows that ribbons or elastic rods can not be used to model protein dynamics as have been successful in the case of DNA [5].

In this paper we have analysed local properties of backbone geometry. In the subsequent paper we shall develop and perform analysis for classifying protein structures when considering global geometric properties.

2 On ribbons

Let $\mathbf{r} : I \to \mathbb{R}^3$ be a space curve parametrized by arch-length s, i.e. $\frac{d\mathbf{r}}{ds} = \mathbf{r}'$ is an unit tangent vector $\mathbf{t}$ at each point of the curve. Let $\mathbf{d}_2 : I \to \mathbb{S}^2$ be a unit vector field along $\mathbf{r}$ that is orthogonal to $\mathbf{r}$. The scalar product $\mathbf{d}_2 \cdot \mathbf{t}$ vanishes thus for all s. The pair $\{\mathbf{r}, \mathbf{d}_2\}$ defines a ribbon by the map $F(s,u) = \mathbf{r}(s) + u\mathbf{d}_2(s)$ for $s \in I$ and $|u| < \varepsilon$. Supplying with $\mathbf{d}_3 = \mathbf{t} \times \mathbf{d}_2$ and renaming $\mathbf{t}$ as $\mathbf{d}_1$, the set $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ (of directies) constitutes an orthonormal basis of $\mathbb{R}^3$ for each point of the curve $\mathbf{r}$. By orthonormality Frenet like equations are obtained:

$$\begin{array}{rcrrr} \mathbf{d}_1' &=& & k_1\mathbf{d}_2 & +k_2\mathbf{d}_3 \\ \mathbf{d}_2' &=& -k_1\mathbf{d}_1 & & +k_3\mathbf{d}_3 \\ \mathbf{d}_3' &=& -k_2\mathbf{d}_1 & -k_3\mathbf{d}_2 & \end{array} \tag{1}$$

In case $\mathbf{t}'$ is non-zero the unit principal normal $\mathbf{n}$ and the curvature $\kappa > 0$ may be defined by the equation $\frac{d\mathbf{t}}{ds} = \kappa\mathbf{n}$. Defining the binormal by $\mathbf{b} = \mathbf{t} \times \mathbf{n}$ the usual Frenet frame is obtained which obeys the Frenet equations:

$$\begin{array}{rcrrr} \mathbf{t}' &=& & \kappa\mathbf{n} & \\ \mathbf{n}' &=& -\kappa\mathbf{t} & & +\tau\mathbf{b} \\ \mathbf{b}' &=& & -\tau\mathbf{n} & \end{array}, \tag{2}$$

where τ is the torsion of the space curve. The fundamental theorem of space curve geometry (see eg. [6]) states that up to translation and rotation a space curve with non-vanishing curvature is uniquely given by its curvature and torsion, that is the shape of a space curve is given by its curvature and torsion functions.

To lay down the relation between $\{k_1, k_2, k_3\}$ and $\{\kappa, \tau\}$, let $\theta : I \to \mathbb{R}$ be a continuous choice of the angle from $\mathbf{n}$ to $\mathbf{d}_2$, that is, $\mathbf{d}_2 = \cos(\theta)\mathbf{n} + \sin(\theta)\mathbf{b}$ and $\mathbf{d}_3 = -\sin(\theta)\mathbf{n} + \cos(\theta)\mathbf{b}$. Straight forward calculations give $k_1 = \kappa\cos(\theta)$, $k_2 = -\kappa\sin(\theta)$, and $k_3 =$

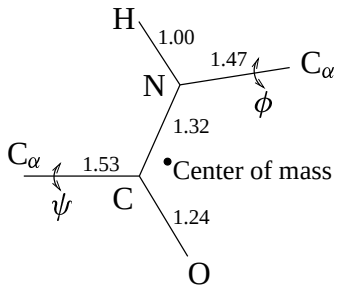


Figure 1: One link of a backbone chain for the usual trans peptide bond as given in [12]. The bond lengths are indicated in Å. The C and the N atoms are planar and the following angles then determine the geometry: $\angle(\mathrm{N}-\mathrm{C}_\alpha-\mathrm{C}) = 70°$, $\angle(\mathrm{C}_\alpha-\mathrm{C}-\mathrm{O}) = 119°$, $\angle(\mathrm{C}_\alpha-\mathrm{C}-\mathrm{N}) = 114°$, and $\angle(\mathrm{C}-\mathrm{N}-\mathrm{H}) = \angle(\mathrm{C}-\mathrm{N}-\mathrm{C}_\alpha) = 123°$. Calculating the center of mass each C_α atom is weighted by one half. ϕ is the dihedral angle given by C-N-C_α-C, ψ is the dihedral angle given by N-C_α-C-N, and both angles are zero in their resp. trans configuration.

$\tau + \theta'$. Hence, $\kappa = \sqrt{k_1^2 + k_2^2}$ and

$$\tau = k_3 - \theta' = k_3 - \frac{\left(\frac{k_1}{\sqrt{k_1^2+k_2^2}}\right)'}{\frac{k_2}{\sqrt{k_1^2+k_2^2}}} = k_3 - \frac{k_2 k_1' - k_1 k_2'}{k_1^2 + k_2^2}.$$

Note, that on a ribbon who's space curve $\mathbf{r}$ has zero curvature at $\mathbf{r}(s_0)$ both τ and θ' may and generally will diverge but k_3 stays finite. This is the reason why the Frenet frame can not be used to study general ribbons, but the Frenet frame is still the natural frame to use when describing what happens when moving a ribbon.

Consider a family of ribbons, one for each time t in some interval, endowed with an elastic energy, the twist, k_3, obeying the twist evolutionary equation of the form

$$\frac{d^2 k_3}{dt^2} - \frac{d^2 k_3}{ds^2} = \frac{d}{dt}\left(\kappa \mathbf{b} \cdot \frac{d\mathbf{t}}{dt}\right),$$

see e.g. ref. [7]. On a fixed curve the solutions to the twist evolutionary equation are standard torsional waves, but in general this equation provides the coupling between curve motion and twist and is studied in e.g. [8, 9, 10, 11].

3 The geometry of protein backbones

A standard simplified model of a protein backbone consists of very stiff planar links that are relatively free to rotate at the C_α-atoms, corresponding to the two dihedral angles

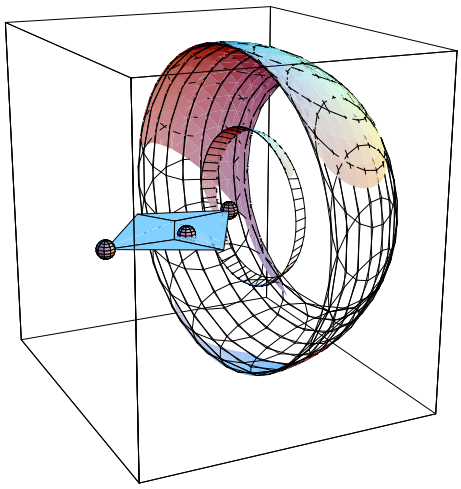


Figure 2: One link of a backbone chain for the usual trans peptide bond. The three spheres indicate two C_α-atoms and the center of mass of this link. The two fully drawn surfaces are the positions of the next C_α-atom resp. the center of mass of the next link corresponding to (ϕ, ψ) in the square with vertices $(-140, -80)$, $(-30, -80)$, $(-50, 210)$, and $(-240, 210)$ and in the square with vertices $(70, -30)$, $(140, -30)$, $(70, 70)$, and $(20, 70)$, containing the most likely dihedral angle pairs, see eg [13]. The two nets are the positions of the next C_α-atom resp. the center of mass of the next link corresponding to all dihedral angle pairs. Both of these nets lie on a sphere.

ϕ and ψ. In this section the large scale shape of a protein backbone is taken to be the polygonal curve connecting either the C_α-atoms or the centers of mass of each link in the backbone chain, cf. Figure 1.

The shape of a simplified protein backbone is given by two discrete functions of its length, namely the pairs of dihedral angles. The shape of a smooth space curve i given by the two functions of its length, curvature and torsion. The following hypothetical situation shows however, that there is a great difference between protein backbones shapes and elastic rods. Consider a protein backbone that is restricted to lie in a plane. All dihedral angles have to be either 0 or 180 degrees. Once the dihedral angles are chosen the protein backbone is totally ridged, when regarding bond lengths and angles as stiff. This is in striking contrast to the flexibility of e.g. a planar elastic rod. In the following this is referred to as the planar problem.

Figure 2 shows the flexibility of the backbone chain. Note, that a given position of the C_α-atom or center of mass on the nets on Figure 2 is obtained by exactly two sets of dihedral angles, except at the boundaries of these nets (corresponding to $\psi = 0$ and $\psi = 180$) where only one set of dihedral angles gives each boundary point. When restricting the dihedral angels to a region relatively free of steric hindrance, this ambiguity still occurs for approximately half of the dihedral angle pairs.

Consider a short piece of elastic rod that is straight when relaxed, clamped at one

end, and free at the other end. The possible positions of the free end describe a spherical cap. From Figure 2 it is clear that such an elastic rod does not serve as a good model of a protein backbone. In case of an elastic rod that prefers a given curvature the possible positions of the free end describe a strip on a sphere as the nets on Figure 2. If such an elastic rod is restricted to lie in a plane it still has the same flexibility as when freely situated in space. Hence, naturally curved elastic rods do not solve the planar problem above.

Elastic rods are up to Euclidean motions given by three (four when extendible) functions of arch-length. Hence, their conformational freedom has to be constrained if rods are to model protein backbones. This could e.g. be done by letting the twist be a function of curvature and torsion. However, Sections 5 and 7 show that elastic rods cannot model protein backbones, making further elaboration on their conformational freedom absolute.

4 1-pointed shapes of protein backbones

Let $\mathbf{r}_n$ denote the position of the n'th C_α-atom. Choose a coordinate system for the zeroth link in the peptide chain as follows: Set $\mathbf{r}_0 = \mathbf{0}$ and let the x-axis be in the direction from the 0'th C_α-atom to the C-atom and choose the y-axis in the plane of the link orthogonal to the x-axis, such that the O-atom has negative y coordinate, see Figure 1. Finally, choose the z-axis (pointing directly upwards from the plane of this paper on Figure 1) to obtain a positive orthonormal basis, as on Figure 2. In this coordinate system the next C_α-atom lies app. at $\mathbf{r}_1 = (3.52,\ 1.44,\ 0) = \mathbf{v}$ in units of Å. The next coordinate system (number 1) is chosen as the first, but with respect to the next link in the chain. Thus the third C_α-atom lies at the same $\mathbf{v}$, but with respect to coordinate system number 1. Hence, $\mathbf{r}_2 = \mathbf{v} + \mathbf{O}\mathbf{v}$, where $\mathbf{O}$ is an orthonormal matrix that can be calculated explicitly in terms of ϕ and ψ[1].

Consider the situation that the pairs of dihedral angles are equal for all links in the chain. By induction, the position of the n'th C_α atom is $\mathbf{r}_n = \mathbf{v}+\mathbf{O}\mathbf{v}+\mathbf{O}^2\mathbf{v}+\cdots+\mathbf{O}^{n-1}\mathbf{v}$. Let $\mathbf{p}$ be the coordinates of the choice of point with resp. to each link. The position of the n'th $\mathbf{p}$-point is thus $\mathbf{r}_n(\mathbf{p}) = \mathbf{v} + \mathbf{O}\mathbf{v} + \mathbf{O}^2\mathbf{v} + \cdots + \mathbf{O}^{n-1}\mathbf{v} + \mathbf{O}^n\mathbf{p}$. A polygonal curve given by $\mathbf{r}_n(\mathbf{p})$ for varying n and fixed $\mathbf{p}$ is referred to as an 1-pointed shape of a backbone since it describes the shape of a backbone when focusing at one point, $\mathbf{p}$, of each residue.

As $\mathbf{r}_{n+1}(\mathbf{p}) - \mathbf{r}_n(\mathbf{p}) = \mathbf{O}^n(\mathbf{v} + \mathbf{O}\mathbf{p} - \mathbf{p})$ the polygonal curve (for given repeated dihedral angle pair) has equal length line segments. This polygonal curves is thus fully described by this common length together with a curvature angle (which is non zero, see Figure 2) and the dihedral angle over each line segment. To calculate these angles let $\mathbf{t} = \frac{\mathbf{v}+\mathbf{O}\mathbf{p}-\mathbf{p}}{|\mathbf{v}+\mathbf{O}\mathbf{p}-\mathbf{p}|}$ be the zeroth unit (tangent) vector, i.e., the unit vector pointing from $\mathbf{r}_0(\mathbf{p})$ to $\mathbf{r}_1(\mathbf{p})$. Note, that the n'th tangent vector is given by $\mathbf{O}^n\mathbf{t}$. The inverse of $\mathbf{O}$ equals

[1] The columns of $\mathbf{O}$ are given by
$(.3378 + .1470\cos\phi,\ .0535 - .9281\cos\phi,\ -.9397\sin\phi)$,
$((-.9281 + .0535\cos\phi)\cos\psi - .1564\sin\phi\sin\psi,\ (-.1470 - .3378\cos\phi)\cos\psi$
$+ .9877\sin\phi\sin\psi,\ -.3420\cos\psi\sin\phi - \cos\phi\sin\psi)$, and
$(-.1564\cos\psi\sin\phi + (-.0535\cos\phi + .9281)\sin\psi,\ .9877\cos\psi\sin\phi$
$+ (.1470 + .3378\cos\phi)\sin\psi,\ \cos\phi\cos\psi + .3420\sin\phi\sin\psi)$.

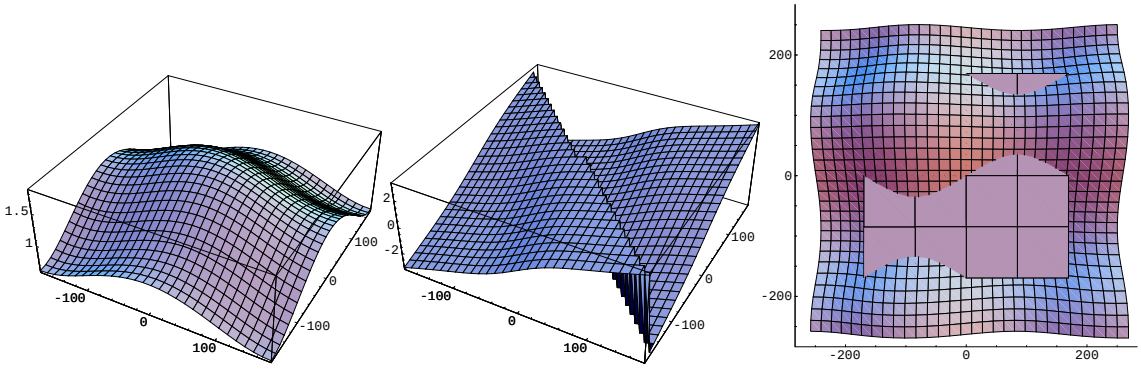


Figure 3: The bending angle ω to the left and in the middle the torsional angle θ of the polygonal curve through the C_α-atoms are shown as functions of the dihedral angles. The discontinuity in the middle figure is artificial as plus and minus π corresponds to the same angle. To the right the graph of the Jacobian of the map $(\phi, \psi) \mapsto (\omega, \theta)$ is intersected with a part of the zero-plane corresponding to a fundamental domain. This intersection is made to visualize the sign of the Jacobian as the map $(\phi, \psi) \mapsto (\omega, \theta)$ is 1 to 1 only when restricted to a set of dihedral angles giving constant sing on this figure.

the transposed of $\mathbf{O}$ denoted by $\mathbf{O}^T$. Hereby the triple of successive tangents giving the easiest possible calculations is $\left\{\mathbf{O}^T\mathbf{t}, \mathbf{t}, \mathbf{Ot}\right\}$. The curvature angle, $0 < \omega < \pi$, is given by

$$\omega = \operatorname{Arccos}\left(\mathbf{t} \cdot (\mathbf{Ot})\right).$$

The (torsional) dihedral angle θ is well-defined up to an integral multiple of 2π, since $\omega \neq 0, \pi$ and can be determined by the equations:

$$\sin(\theta) = \frac{\left[\mathbf{O}^t\mathbf{t}\ \mathbf{t}\ \mathbf{Ot}\right]}{\sin^2(\omega)} = \frac{\left[\mathbf{O}^t\mathbf{t}\ \mathbf{t}\ \mathbf{Ot}\right]}{1 - (\mathbf{t} \cdot (\mathbf{Ot}))^2}$$

$$\cos(\theta) = -\frac{\left(\mathbf{O}^t\mathbf{t} - \left(\left(\mathbf{O}^t\mathbf{t}\right) \cdot \mathbf{t}\right)\mathbf{t}\right) \cdot (\mathbf{Ot})}{\sin^2(\omega)} = \frac{\left(\mathbf{O}^t\mathbf{t} - (\mathbf{t} \cdot (\mathbf{Ot}))\,\mathbf{t}\right) \cdot (\mathbf{Ot})}{1 - (\mathbf{t} \cdot (\mathbf{Ot}))^2}$$

together with a fundamental interval for θ of length 2π.

By the first of the Frenet equations the integral of curvature over an interval equals the length of the curve traced out by its unit tangents on the unit 2-sphere, S^2. The shortest curve between two points $\mathbf{t}_1$ and $\mathbf{t}_2$ on S^2 is a part of a great circle and has length equal to the angle ω between $\mathbf{t}_1$ and $\mathbf{t}_2$. Hence, ω is the smallest integral of curvature for any space curve starting and ending with tangents $\mathbf{t}_1$ resp. $\mathbf{t}_2$. The angle ω thus corresponds to the integrated curvature per residue of the backbone.

The integral of torsion over an interval of a space curve equals the integral of the geodesic curvature of the corresponding image of its tangents on S^2, Theorem 15 in [14]. Considering polygonal curves, their associated tangent images are the piecewise great circle curves connecting the tangents of the polygonal curves. Great circles are geodesics

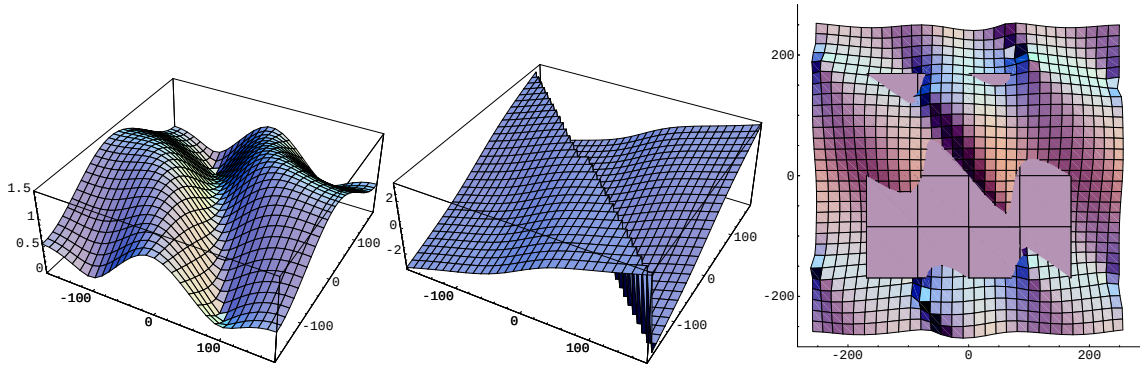


Figure 4: The bending angles ω to the left and in the middle the torsional angle θ of the polygonal curve through the centers of mass are shown as functions of the dihedral angles. To the right the graph of the Jacobian of the map $(\phi, \psi) \mapsto (\omega, \theta)$ is intersected with a part of the zero-plane corresponding to a fundamental domain.

and have zero geodesic curvature. However the discontinuity of the tangents to the great circles at their intersection ($= \theta$) must by the Gauss-Bonnet's Theorem be considered as integrated torsion. The angle θ thus corresponds to the integrated torsion over one residue.

On Figure 3 the two angles, ω and θ, are shown as functions of the dihedral angle pair (ϕ, ψ). Also the Jacobian of the map, sending (ϕ, ψ) into (ω, θ), is shown on this figure. Note that θ traverses a full rotation if either ϕ or ψ does. On a fundamental domain, ν thus completes two full rotations.

By symmetries, the integral of the determinant of the Jacobian determinant shown on Figure 3 over one fundamental period equals zero. From this figure is also seen that the map $(\phi, \psi) \mapsto (\omega, \theta)$ is 2 to 1 at most of the points.

Instead of connecting the C_α-atoms, it could be interesting to focus on e.g. the centers of mass, the positions of the oxygen atoms, or any other fixed point at, above, or below each link. However, such new representations correspond to continuous deformations of the map from dihedral angle pairs to the curvature and tosion angle pair considered above. Hence, they are at best two to one maps.

5 On helical representation of protein geometry

When representing "the overall" protein geometry by space curves the family of circular helices is a natural candidate that has been considered by several authors. This is primarily due to the frequent helical secondary structures. Consider a protein backbone on which all pairs of dihedral angles are equal. Furthermore, for each link in the protein chain choose a point, such that each point lies at the same spot with respect to its link. Such a sequence of points lie on a helix. Such helixes may degenerate to circles and may for almost all dihedral angle pair be chosen in a unique way. We examine to what extent such curves can help bridge the gab between backbone geometry and space curve geometry in order

to transform protein dynamics to space curve dynamics.

If one, e.g. inspired by the twist evolutionary equation from Section 2, intends to model protein dynamics via space curve dynamics it is natural, at first, to seek to endow space curves with an energy density that is independent of arch-length (homogeneous) but dependent on the geometry of the curve. Such models have been successful in the case of DNA and may help understand proteins in the process of folding, as suggested in [1]. An symmetry argument suggests that a given helix always is a stationary point of the energy functional if the curve is constrained to go through equally distributed points on the given helix and the length of the curve is unconstrained. Any homogeneous space curve model will, due to long α-helices, β-strands often be forced to assume shapes close to helixes. This gives yet a other reason to pay attention to helical representations of protein geometry.

Even though the idea of representing protein geometry by helices has a long history the authors have been unable to find an analysis of the representation itself. This section is devoted to this analysis and concentrates on the case of proteins on which all pairs of dihedral angles are equal. Even in this simple case, the conclusion is that a helical representation can, at best, be unique and treatable for small deformations of the α-helix structure and not for more general proteins with repeated dihedral angles. This conclusion thus also holds for generally shaped proteins.

To be able to perform the below analysis we have developed a new way of defining helices from protein geometry based on the fact that such a helix is fully determined by the euclidean motion that brings one link in the protein chain into the the following link together with the choice of points that the helix has to go through. Hence, assuming the ideal backbone geometry given in Figure 1, the helix is given by one pair of dihedral angles and the choice of points with respect to the links. On a generally shaped protein backbone a piecewise helical curve is thus defined by the varying pairs of dihedral angles along the backbone.

In [15] the positions of four C_α-atoms in sequence is used to define a best fitting helix. As mentioned in Section 3, half of the conformational information is thrown away when considering the positions of eg. the C_α-atom instead of the full geometry of the backbone in terms of the dihedral angles. The number of residues needed to define a helix piece here (that is two) is thus smaller than in [15], which gives a higher resolution (in fact the highest possible) when applied to structural analysis.

Consider the euclidean motion that brings one link in the protein chain into the following link. With notation as i Section 3 this Euclidean motion is given by a translation by $\mathbf{v}$ followed by the rotation given by $\mathbf{O}$. Figure 2 shows that $\mathbf{O}$ is always a proper rotation. Let ν denote the angle that $\mathbf{O}$ rotates. The trace of $\mathbf{O}$ equals $1+e^{i\nu}+e^{-i\nu} = 1+2\cos(\nu)$. Hence,

$$\cos(\nu) = \frac{\text{trace}(\mathbf{O}) - 1}{2}.$$

Note, that $\cos(\nu)$ is independent of the choice of points on the link.

The fix point set of the map given by $\mathbf{O}$ is always a line l, since $\mathbf{O}$ is newer the identity, see Figure 2. For any integer n and for any vector $\mathbf{c}$, not in l, the vector $\left(\mathbf{O}^n\mathbf{c} - \mathbf{O}^{n-1}\mathbf{c}\right)$

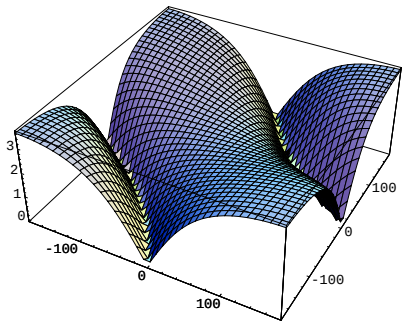


Figure 5: The rise per residue, in units of Ångstrøm, shown as a function of the dihedral angles.

is non zero and orthogonal to l. Hence, a vector $\widetilde{\mathbf{e}_l}$ inside l, may for $\left(\mathbf{O}^{n-1}\mathbf{c}-\mathbf{O}^{n-2}\mathbf{c}\right)\times\left(\mathbf{O}^{n}\mathbf{c}-\mathbf{O}^{n-1}\mathbf{c}\right)\neq\mathbf{0}$ be chosen as $\widetilde{\mathbf{e}_l}=\frac{\mathbf{O}^{n-1}\mathbf{c}-\mathbf{O}^{n-2}\mathbf{c}}{\left|\mathbf{O}^{n-1}\mathbf{c}-\mathbf{O}^{n-2}\mathbf{c}\right|}\times\frac{\mathbf{O}^{n}\mathbf{c}-\mathbf{O}^{n-1}\mathbf{c}}{\left|\mathbf{O}^{n}\mathbf{c}-\mathbf{O}^{n-1}\mathbf{c}\right|}$ independent of n. Since $\mathbf{O}$ is length preserving $\left|\mathbf{O}^{n}\mathbf{c}-\mathbf{O}^{n-1}\mathbf{c}\right|=\left|\mathbf{O}\left(\mathbf{O}^{n-1}\mathbf{c}-\mathbf{O}^{n-2}\mathbf{c}\right)\right|=\left|\mathbf{O}^{n-1}\mathbf{c}-\mathbf{O}^{n-2}\mathbf{c}\right|$. As the inverse of $\mathbf{O}$ equals the transpose of $\mathbf{O}$ it is easiest to calculate $\widetilde{\mathbf{e}_l}$ as

$$\widetilde{\mathbf{e}_l}=\frac{\left(\mathbf{c}-\mathbf{O}^{t}\mathbf{c}\right)\times\left(\mathbf{O}\mathbf{c}-\mathbf{c}\right)}{\left|\mathbf{O}\mathbf{c}-\mathbf{c}\right|^{2}},$$

which is independent of $\mathbf{c}$. Two choices of $\mathbf{c}$ that newer lie in l are $\mathbf{c}=\mathbf{v}$ or $\mathbf{c}=(1,0,0)$. However $\widetilde{\mathbf{e}_l}$ is ill-defined whenever $\mathbf{O}$ rotates 180 degrees (but l is not). A natural choice of a unit vector in l is $\mathbf{e}_l=\pm\frac{\widetilde{\mathbf{e}_l}}{|\widetilde{\mathbf{e}_l}|}$ such that the rise per residue $d=\mathbf{e}_l\cdot\mathbf{v}$ for the C_α curve is positive.

Let $d(\mathbf{p})$ denote the rise per residue when considering a helix through points with coordinates $\mathbf{p}$ with respect to each link. The vector from the n'th point to the $(n+1)$'st point is, as in Section 3, given by $\mathbf{O}^n(\mathbf{v}+\mathbf{O}\mathbf{p}-\mathbf{p})$. The rise per residue is then

$$\begin{aligned}d(\mathbf{p})&=\mathbf{e}_l\cdot\mathbf{O}^n(\mathbf{v}+\mathbf{O}\mathbf{p}-\mathbf{p})\\&=\left(\mathbf{O}^n\mathbf{e}_l\right)\cdot\mathbf{O}^n(\mathbf{v}+\mathbf{O}\mathbf{p}-\mathbf{p})\\&=\mathbf{e}_l\cdot(\mathbf{v}+\mathbf{O}\mathbf{p}-\mathbf{p})\\&=\mathbf{e}_l\cdot\mathbf{v}\\&=d,\end{aligned}$$

where we use that $\mathbf{e}_l$ is invariant under $\mathbf{O}$, $\mathbf{O}$ is angel and length preserving, and finally that the vector $(\mathbf{O}\mathbf{p}-\mathbf{p})$ is orthogonal to $\mathbf{e}_l$. Note, that $d(\mathbf{p})=d$ is independent of the choice of points on the link. The rise per residue is shown on Figure 5. Note, that $d=0$ along a curve in the (ϕ,ψ)-torus. At this curve in the parameter space, the helix is degenerated to a circle and crossing this curve causes an inflection of $\mathbf{e}_l$, corresponding to a change of orientation of the helical axis.

To determine the angel ν one can, in addition to the equation $\cos(\nu)=\frac{\mathrm{trace}(\mathbf{O})-1}{2}$, use that $\sin(\nu)=\widetilde{\mathbf{e}_l}\cdot\mathbf{e}_l$ or $\sin(\nu)=\mathrm{sign}(\widetilde{\mathbf{e}_l}\cdot\mathbf{e}_l)\sqrt{1-\cos^2(\nu)}$, where the last expression has

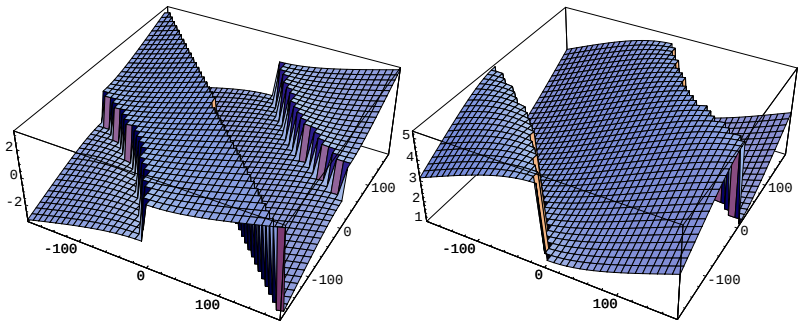


Figure 6: The representative, within the interval $[-\pi, \pi[$ to the left and within the interval $[0, 2\pi[$ to the right, of the angle ν as a function of the dihedral angles.

turned out to be the easier for computational purposes. On Figure 6 the angle ν is shown as a function of the dihedral angles. The angle ν, mod 2π, changes sign along the curve corresponding to $d = 0$ (see Figure 5). As $\mathbf{e}_l$ inflects when crossing this curve, there is no discontinuity of the chosen helixes for $d = 0$. On Figure 6 on the left hand side the jump for ν between plus and minus π corresponds to a jump between a right handed helix resp. a left handed helix.

The angel ν together with the rise per residue is all the information on helixes that can be extracted from $\mathbf{O}$ alone. The remaining information depends on the choice of points on each chain link.

Consider a projection to a plane orthogonal to l and let $\mathbf{w} = (\mathbf{v} + \mathbf{Op} - \mathbf{p}) - ((\mathbf{v} + \mathbf{Op} - \mathbf{p}) \cdot \mathbf{e}_l)\, \mathbf{e}_l$ be the projection of the vector between the two first points in the helix, $(\mathbf{v} + \mathbf{Op} - \mathbf{p})$. From Figure 7 it follows that

$$|\mathbf{w}|^2 = |(r\cos(\nu), r\sin(\nu)) - (r, 0)|^2 = 2r^2\,(1 - \cos(\nu))\,,$$

giving the radius of the helix as

$$r = \sqrt{\frac{\mathbf{w}\cdot\mathbf{w}}{2\,(1 - cos(\nu))}} = \sqrt{\frac{\mathbf{w}\cdot\mathbf{w}}{3 - \text{trace}\,(\mathbf{O})}}.$$

Consider the helix given by $\mathbf{q}(t) = (r\cos(\nu t), r\sin(\nu t), dt)$. The set $\{\mathbf{q}(z)\}_{z\in\mathbb{Z}}$ equals the set $\{\mathbf{r}_z\}_{z\in\mathbb{Z}}$ up to an Euclidean motion. We say that $\mathbf{q}$ interpolates the discrete helix given by the $\mathbf{r}_n$'s. There are infinitely many helixes that interpolates the $\mathbf{r}_n$'s (corresponding to replacing ν by $\nu + 2\pi z$ for any integer z, thus changing the pitch). Curvature and torsion of the curve $\mathbf{q}$ equal $\frac{\nu^2 r}{d^2+\nu^2 r^2}$ resp. $\frac{d\nu}{d^2+\nu^2 r^2}$ for each choice of ν. The length of the helix segment used to interpolate two points is $L = \int_0^1 \left|\mathbf{q}'(t)\right| dt = \sqrt{r^2\nu^2 + d^2}$.

Figure 8 shows data on helical curves through the C_α-atoms when choosing $\nu \in [-\pi, \pi[$. When $\nu = (\pm)\pi$ both a right handed and a left handed helix interpolate equally well. The discontinuity of the torsion occurs thus when the interpolating helix changes handedness. This discontinuity may be moved by changing the fundamental interval of

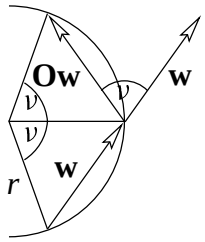


Figure 7: Planar projection of a helix. The angle between **w** and **Ow** equals ν as the triangles are isosceles.

ν, as done on Figure 9. The line of discontinuity, on Figure 8 goes right through the set of dihedral angles that corresponds to β-strands. The geometry of β-strands is thus very badly captured by the helixes considered on Figure 8.

The determinant of the Jacobian of the map from pairs of dihedral angles to pairs of curvature and torsion concerning the C_α-atoms is shown on Figure 8 at the bottom left. On a connected set of pairs of dihedral angles, on which the determinant of the Jacobian has constant sign, the map from pairs of dihedral angles to pairs of curvature and torsion is invertible. The set of dihedral angle pairs that corresponds to α-helixes contains a serious sign change of the Jacobian. Consider, a space curve representation that goes through the C_α-atoms and describes a helix in case of a protein in an α-helix shape. By the sign change of the determinant of the Jacobian, the shape (i.e. the curvature and torsion) of the space curve can not be translated to dihedral angles in a unique way. Geometrically this representation is not unique and it is e.g. impossible to define a potential energy of the space curve from its shape even when concerning models that only has to be valid for small variations of the α-helix shape!

Figure 10 shows data on helical curves through the centers of mass, which are natural to consider if one is interested in dynamics. In the case of $\nu \in [-\pi, \pi[$ there is a rather thin area with positive determinant of the Jacobian which is close to both an α-helix and a β-strand structure but none of these structures are covered by this region.

The determinant of the Jacobian coming from interpolating the oxygen atoms with helixes is shown on Figure 11 at the bottom left. It shows that the oxygen atom lies at a geometrically significant position. There is a great area with positive Jacobian that contains both α-helix as well as β-strand in the case $\nu \in [0, 2\pi[$. This opens the possibility for building a representation of protein backbones that contains both types of the main secondary structures. The problem with this representation is that the length of the helix piece representing one residue varies from 1.5Å to 20Å.

When reconstructing protein geometry from space curve geometry, one, travelling along a space curve, has to decide when to reconstruct the next link in the peptide chain. For given constant curvature and torsion values one knows how far to travel along the curve before the next link is to be (re-)constructed. However, corresponding to a generally shaped protein, the curvature and torsion of its space curve representative varies with arch-length. Hence, the length of the curve piece that corresponds to a link in the peptide chain

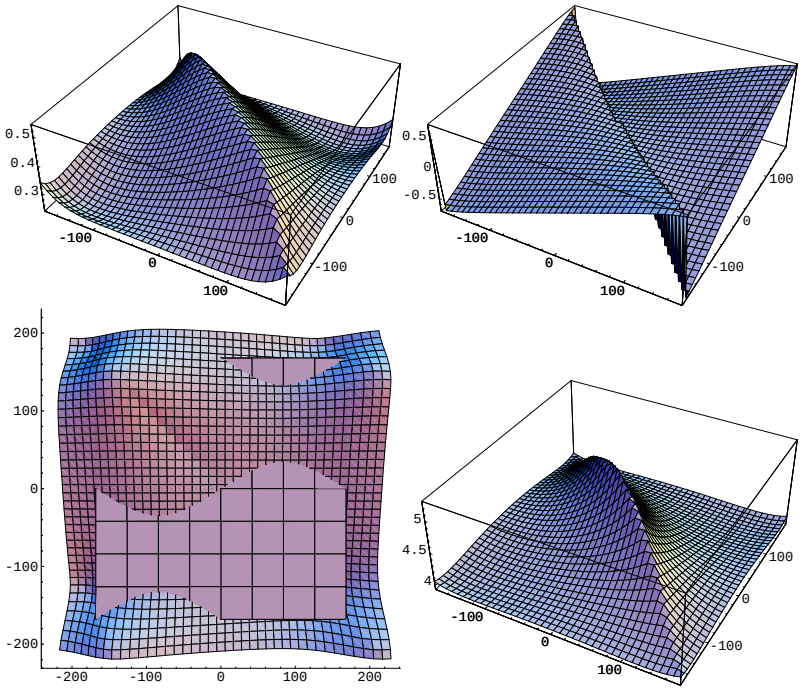


Figure 8: On the top left and right: Curvature resp. torsion of the helix interpolating the C_α-atoms. At the bottom left: The determinant of the Jacobian of the map $(\phi, \psi) \mapsto (\kappa, \tau)$. At the bottom right: The length $(= \sqrt{r^2\nu^2 + d^2})$ of the helical piece interpolating two neighbouring C_α-atoms. The angle ν is chosen in $[-\pi, \pi[$ and in the grid on the plane at the bottom left the lines lie 45 degrees apart.

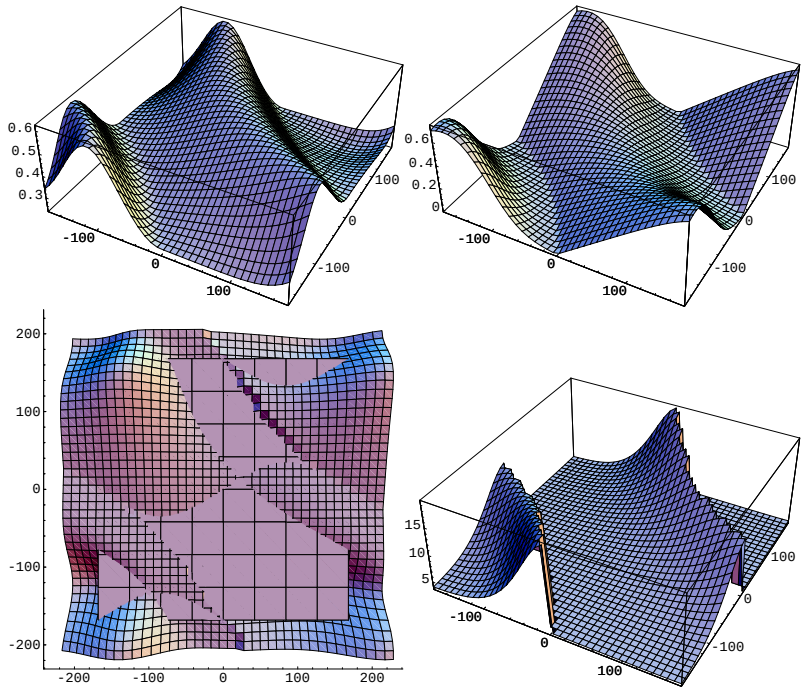


Figure 9: This figure shows the helical data for helixes through the C_α-atoms as on Figure 8, but with the angle ν between zero and 2π.

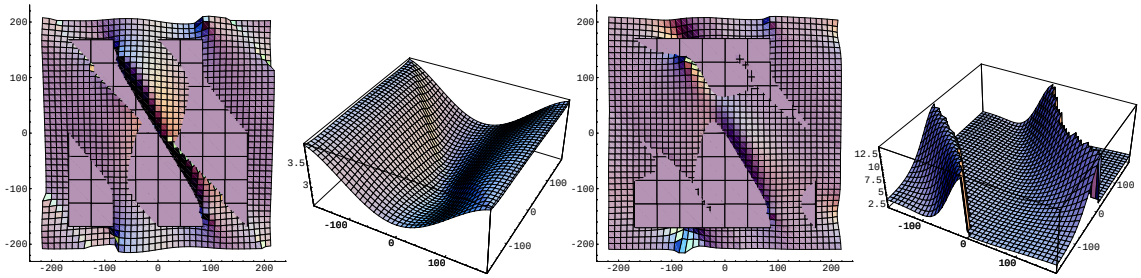


Figure 10: From left to right: Determinant of the Jacobian of the map $(\phi, \psi) \mapsto (\kappa, \tau)$, and length of helixes through the centers of mass. In the two most left graphs the angle $\nu \in [-\pi, \pi[$ and in the two most right graphs $\nu \in [0, 2\pi[$.

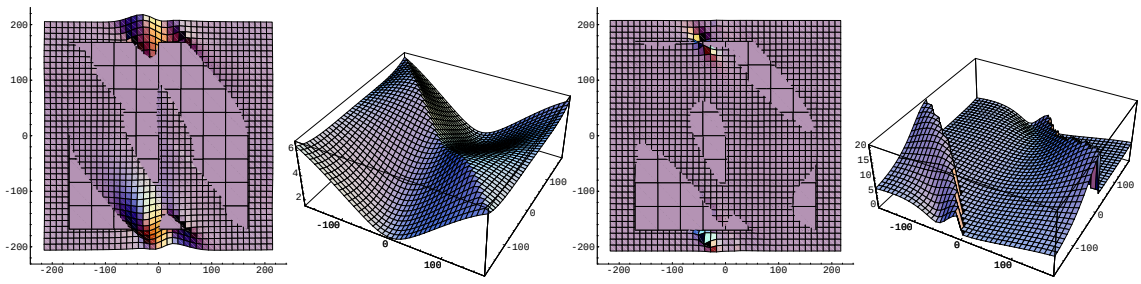


Figure 11: From left to right: Determinant of the Jacobian of the map $(\phi, \psi) \mapsto (\kappa, \tau)$, and length of helixes through the Oxygen atoms. In the two most left graphs the angle $\nu \in [-\pi, \pi[$ and in the two most right graphs $\nu \in [0, 2\pi[$.

varies when moving along a space curve and it is thus not clear that a transformation from space curve to peptide chain can be constructed. Even if such a transformation exists on the ideal model considered until now, the fact that protein backbones do not have ideal geometry (as bond lengths and angles vary) implies that such a transformation will be inaccurate. In the case of interpolating the oxygen atoms, the relative variation of the length of the curve pieces that corresponds to one chain link is so large, see Figure 11, that even with a very accurate transformation the chain links may be misplaced several chain links when reconstructing e.g. a 150 residue chain.

6 Analysing protein structure

When considering large scale protein structure instead of dynamics, the non-invertability of maps from protein geometry to space curve geometry is, as argued in the Introduction, desirable instead of a "no go" result. Several authors have considered the best helical approximation of four to five successive C_α-atoms and used curvature and torsion of approximating helixes to analyse protein structure. The method of calculating a helix from two links in a protein backbone chain given in the above section makes it possible to get a higher (actually the highest possible) resolution. Due to the finite resolution of the atomic coordinates and the not ideal geometry of each link the choice of coordinate systems taken in the above section is not appropriate. The choice taken here is as follows: Let P be the plane through two neighbouring C_α-atoms that minimizes the sum of the squared plane-point distances for the remaining atoms within that link in the protein chain. The first axis is in the direction from the first C_α-atom to the next, the third axis orthogonal to the plane P, and the second axis is supplied to give an positively oriented orthonormal basis.

The axis of the helical piece is given by the direction $\mathbf{e}_l$ and a point on the axis. As seen from Figure 12, this point may be chosen as

$$\mathbf{r}_n + \frac{1}{2}\mathbf{w} + r\,|\cos(\nu/2)|\,\mathbf{e}_l \times \frac{\mathbf{w}}{|\mathbf{w}|} + \frac{1}{2}d = \frac{1}{2}(\mathbf{r}_n + \mathbf{r}_{n+1}) + r\,|\cos(\nu/2)|\,\mathbf{e}_l \times \frac{\mathbf{w}}{|\mathbf{w}|}.$$

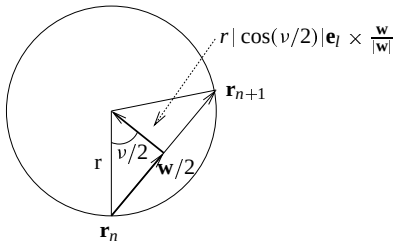


Figure 12: This figure is similar to Figure 7 and shows a path from the C_α-atom, at $\mathbf{r}_n$, to the point on the helix axis that lies at the "same height" as the C_α-atom.

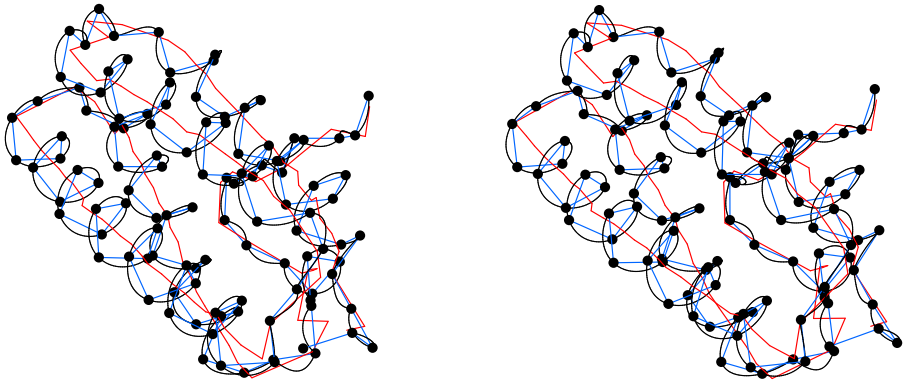


Figure 13: This figure shows the polygonal curve through the C_α-atoms of the protein 2HMQ:A together with the piecewise helical approximation and the polygonal curve through the chosen points on the helix axes. The helical approximation is chosen such that the angle ν lies between plus and minus 180 degrees.

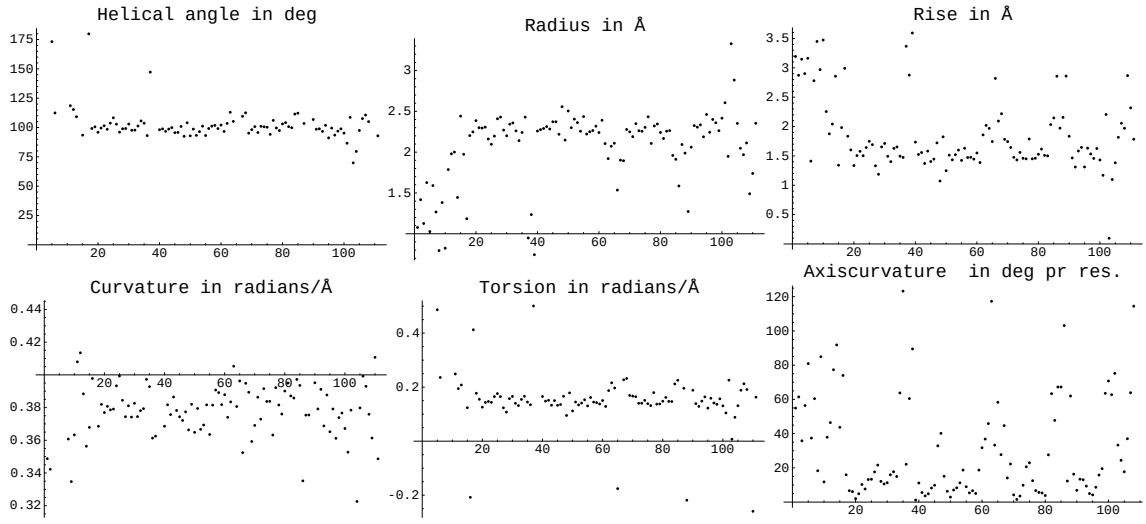


Figure 14: As function of the residue number is from top left to bottom right shown: The helical angle ν, radius of the helix, rise per residue, curvature of the helix, torsion of the helix, and the bending angle of the axis curve of the helical approximation shown on Figure 13.

On Figure 14 is shown some of the data from applying the above method to the protein 2HMQ:A. It is not surprising that the data shown on Figure 14 are more noisy than when interpolating over longer pieces of the protein backbone especially on the regular α-helices, where the helix axis bend up to 20 degrees. Alone the fact, that the N-C_α-C-angles vary within plus and minus 6 degrees from the ideal angle of 70 degrees, has to give some noise. However, when considering the non-regular regions, like turns, the more local definition of helical approximation given here seems to be more natural than to interpolate over 4 to 5 C_α-atoms, which corresponds to a full turn. It has been tried to vary the choice of the local coordinatesystems to the three axes of inertia[2] of each link in the chain using masses resp. Van der Wahl radi as weights to emphasis dynamics resp. structure, but the results are almost independent of these changes. The noisiness of the data on Figure 14, especially on the α-helices of 2HMG:A, suggests that protein geometry is not suited for local description.

7 Topology and framing

In this section some fundamental questions regarding any representation of backbone geometry by (framed) space curves are addressed. To do this some notation is introduced in order to formalize considerations.

Definition 1

[2]One of these axes is perpendicular to the plane of the link, see [16].

1 *A differential geometric representation of protein backbone geometry is a map R from a set of sequences of dihedral angles into the set of oriented smooth space curve shapes, that is,*

$$R : \left(\mathbb{S}^1 \times \mathbb{S}^1\right)^{n-1} \to C^k\left\{I \to \mathbb{R}^3\right\} / \text{Euclidean motions},$$

where n is the number of residues of the protein, $k \geq 3$ gives the smoothness of the representation, and two space curves are considered equivalent if one of them can be brought into the other by an euclidean motion preserving orientation.

2 *A differential geometric representation of protein backbone geometry is said to be framed if each space curve is equipped with a framing.*

3 *A smooth parametrized (framed) representation of protein backbone geometry is as a differential geometric (framed) representation of protein backbone geometry - except that two (framed) regular space curves are considered equivalent if one of them can be brought into the other by an euclidean motion preserving parametrization. (In this case two identical curves are considered equivalent only if they have the same parametrization.)*

4 *A differential geometric or smooth parametrized (framed) representation of protein backbone geometry is said to be faithful if R is injective.*

5 *A differential geometric or smooth parametrized (framed) representation of protein backbone geometry is said to be continuous if R is continuous.*

Consider the polygonal curve given by the $\cdots - C_\alpha$–C–N–$C_\alpha - \ldots$ bonds. Smoothing the kinks of this polygonal curve gives both a differential geometric continuous and faithful representation as well as a smooth parametrized continuous and faithful representation. Using $\cdots - C_\alpha$–C–N–$C_\alpha - \ldots$ polygons as control polygons will for any piecewise rational interpolating curve family give a continuous representation and, perhaps by inserting additional control points at each bond, also a faithful representation. The not very surprising starting point is thus:

Proposition 2 *Both differential geometric and smooth parametrized representations that are continuous and faithful exists.*

The next result is of a topological nature and came as a bit of a surprise to the authors. The result deals with representations having the property that the change of the shape of the space curve and frame at one point due to a change of the protein shape at another point decreases with the linear distance (the number of residues) between the two points. We say that such a representation has finite persistence length. This is e.g. the case for Bézier curves that have a finite "cut of" distance depending on their degrees.

Theorem 3 *There do not exist a continuous framed representation with finite persistence length.*

Proof: Consider a protein backbone that is so long compared to its persistence length that changes performed at one point have negligible influence at points that lie at least a third of the length of the protein away. Clamp the protein at both ends (allowing rotation of the terminal end) and choose a framed curve c between the two clamped ends to give a total closed framed curve. The curve c may be chosen such that it avoids contact with the protein under the deformation of the protein considered below. The total closed framed curve fulfills the equation $\mathrm{Link}_1 = \mathrm{Wr}_1 + \mathrm{Tw}_{1,p} + \mathrm{Tw}_{1,c}$, where Wr is the writhe of the total closed curve, $\mathrm{Tw}_{1,p}$ is the twist of the frame along the space curve representation of the protein, and $\mathrm{Tw}_{1,c}$ is the twist of the frame along the curve piece c.

Fix the first and the last thirds of the protein and perform a ridged full rotation of the last third. Due to the persistence length of the representation the framed curve near the first end is almost fixed and the framed curve near the last end performs almost a full ridged rotation. To compensate for this rotation of the last end of the protein the framing of the curve c (which is kept fixed) has to perform a full rotation at this end while its frame is fixed at the other end. The twist of c is thus now $\mathrm{Tw}_{2,c} = \mathrm{Tw}_{1,c} \pm 1$, where the sign depends on the handedness of the full rotation applied. The protein may be deformed back to its original position by turning one dihedral angel of the middle third 360 degrees while keeping the first and the last thirds fixed. Recalculating the linking number associated with the total framed closed curve gives $\mathrm{Link}_2 = \mathrm{Wr}_2 + \mathrm{Tw}_{2,p} + \mathrm{Tw}_{2,c}$, where the writhe of the total closed curve, $\mathrm{Wr}_2 = \mathrm{Wr}_1$, is unchanged, $\mathrm{Tw}_{2,p} = \mathrm{Tw}_{1,p}$ is the twist of the frame along the space curve representation of the protein, and $\mathrm{Tw}_{2,c} = \mathrm{Tw}_{1,c} \pm 1$ is the twist of the frame along the curve piece c. Hence, $\mathrm{Link}_2 = \mathrm{Link}_1 \pm 1$. This implies that the performed deformation of the total closed framed space curve has been discontinuous. As there has not been discontinuities at c there has been at least one discontinuity of the framed representation of the protein. Hence, any framed representation is potentially discontinuous at each link when considering proteins that are sufficiently long compared to the representations persistence length. □

Assume, in order to obtain a contradiction, that a differential geometric representation of protein backbones has the property that all curves have non-vanishing curvature. By this they all have a well-defined Frenet frame. For each curve in a family of curves consider the vector field given by the principal normals. The combination of curves and principal normal fields gives a framed representation. By Theorem 3 this representation is discontinuous. The conclusion is that curvature has to vanish at some point of some curve in such a family of curves. A corollary of Theorem 3 is thus

Corollary 4 (of Theorem 3) *There do not exist a continuous representation (or differential geometric representation) with non-vanishing curvature and finite persistence length.*

The vanishing of curvature asserted by Corollary 4 has to happen such that the integral of torsion over an arbitrarily small interval containing the zero curvature point is discontinuous in time. A similar remark counts for the discontinuity of framed representations

asserted by Theorem 3 where the integral of twist, k_3, over an arbitrarily small interval containing the point of discontinuity of the framing is discontinuous. The existence of the discontinuity of the helical representation observed in Section 5 is a consequence of corollary 4. As the problem is of a topological rather than of a geometric nature it will not disappear if one slightly change the class of curves used to represent the backbone geometry. This was also seen as moving the choice of points for the helices to go through did not change the global picture namely that going from dihedral angles to curvature and torsion is at best a two to one map.

In order to apply space curve geometry to proteins one thus has to use different types of representation, e.g. as in the constructive proof of Proposition 2. This is done in the following section. It seems however that the local differential geometric invariants curvature and especially twist and torsion, due to Theorem 3 resp. its corollary, can not be applied to space curve representatives of protein structure.

We conclude this section by pointing out a very serious consequence when trying to describe protein dynamics via framed space curves equipped with an elastic rod like energy. A quadratic twist term of the potential energy has an infinite energy barrier that prevents frame discontinuities. Hence, any elastic model with non zero quadratic twist energy prohibits full rotations of each dihedral angel of the modelled protein.

8 Conclusion

To conclude, our investigation of local space curve representation of proteins have identified the following:

Regarding protein dynamics: It is impossible to represent general protein backbones by space curves, such that the space curve can be endowed with a translationally invariant potential energy corresponding to the part of the potential energy of the protein coming from neighbouring links in the backbone chain. When restricting the configuration space of the protein such that a potential energy functional may be defined, we have found the remaining part of the configuration space too small to be of interest.

Regarding protein structure description: We introduce a new way of defining helix pieces approximating protein backbones, that depends on two neighbouring links in the backbone chain only. This gives the highest possible resolution for a local helical representation, which indicate that protein geometry is not suited for local description.

Under some natural assumptions, we prove the general result that it is impossible to construct a framed representation of protein backbones that is continuous under deformation of the protein. Consequently any notion of torsion or even integrated torsion of protein backbones should be avoided.

References

[1] J. Bohr, H. Bohr and S. Brunak, Protein Folding and Wring Resonances, Biophysical Chemestry, 63 (1997) 97.

[2] N. Qian and T.J. Sejnowski. Predicting the secondary structure of globular-proteins using neural network models, J. Mol. Biol., 202 (1988) 865.

[3] H. Bohr, J. Bohr, S. Brunak R.M.J. Cotterill B. Lautrup L. Nørskov O.H. Olsen and S.B. Petersen, Protein secondary structure and homology by neural networks - the alpha-helices in rhodopsin, FEBS Lett., 241 (1988) 223.

[4] A. Krogh, B. Brown, I.S. Mian, K. Sjolander , and D. Houssler, Hidden markov-models in computational biology - applications to protein modeling, J. mol. biol. 235 (1994) 1501.

[5] W.K. Olson and V.B. Zhurkin, Modeling DNA deformations, Current Opinion in Structural Biology, 10 (2000) 286.

[6] J. McCleary, Geometry from a Differentiable Viewpoint, Cambridge University Press, 1994.

[7] M. Tabor and I. Klapper, Dynamics of twist and writhe and the modeling of bacterial fibers, in: Mathematical approaches to biomolecular structure and dynamics (Minneapolis, MN, 1994), Springer, New York, 1996, p. 139.

[8] A. Goriely and M. Tabor, Nonlinear dynamics of filaments. I. Dynamical instabilities, Phys. D, 105 (1997) 20.

[9] A. Goriely and M. Tabor, Nonlinear dynamics of filaments. II. Nonlinear analysis, Phys. D, 105 (1997) 45.

[10] A. Goriely and M. Tabor, Nonlinear dynamics of filaments. III. Instabilities of helical rods, Proc. Roy. Soc. London Ser. A, 453 (1997) 2583.

[11] A. Goriely and M. Tabor, Nonlinear dynamics of filaments. IV. Spontaneous looping of twisted elastic rods, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 454 (1998) 3183.

[12] L. Pauling, R.B. Corey, and H.R. Branson, The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain, Proc. Nat. Acad. Sci. USA, 37 (1951) 205.

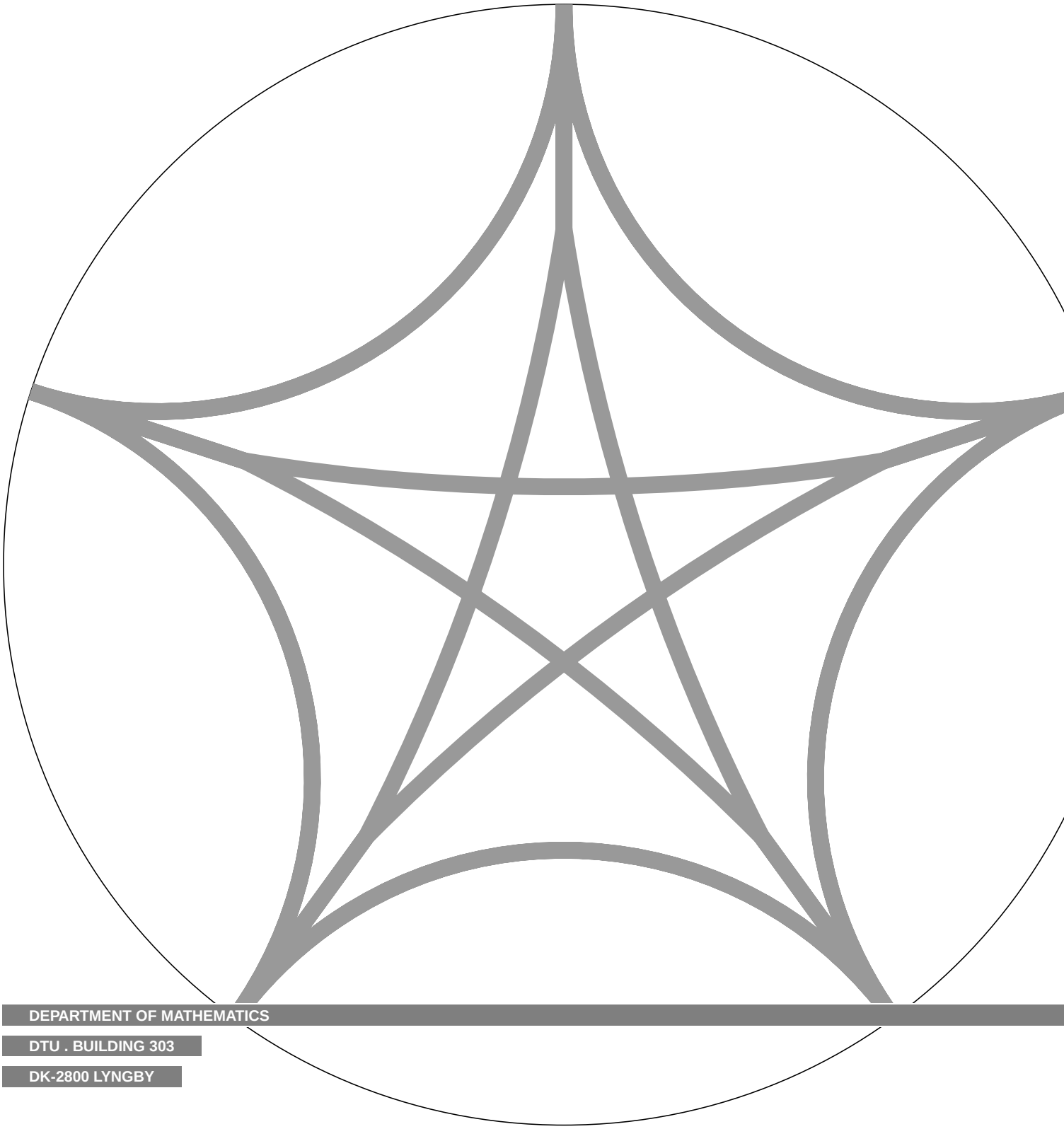
[13] D. Walther and F.E. Cohen, Conformational attractors on the ramachandran map, Acta Cryst., D55 (1999) 506.

[14] P. Røgen, Gauss Bonnet's formula and closed Frenet frames, Geom. Dedicata, 73 (1998) 295.

[15] S. Kumar M. Bansal and R. Velavan, HELANAL: A program to characterize helix geometry in proteins, J. of Biomolecular Structure & Dynamics, 17 (2000) 811.

[16] D.M. Soumpasis, C.-S. Tung, and A.E. Garcia, Rigorous description of DNA structures. II. On the computation of best axes, planes, and helices from atomic coordinates, J. of Biomolecular Structure & Dynamics, 8 (1991) 867.

ISSN 0904–7611



DEPARTMENT OF MATHEMATICS

DTU . BUILDING 303

DK-2800 LYNGBY