DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF COPENHAGEN

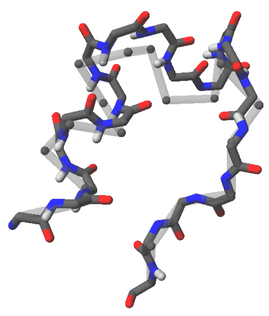
Fitting an All-atom Protein Model to a C_α-trace

Martin Dybdal, Anders Boesen Lindbo Larsen and Esben Skaarup
dybber@dybber.dk, abll@diku.dk and sben@diku.dk

Supervisors: Pawel Winter and Rasmus Fonseca

Summary

In this work we investigate a strategy for predicting the structure of proteins. Given a so-called *C_α-trace*, we wish to fold the protein to match the trace.



Our problem

Currently, we are only able to predict protein structures at the C_α-trace level. Our goal with this project is to extend the C_α-trace with the remaining atoms to get an all-atom model. The difference between the two is shown in Figure 1.

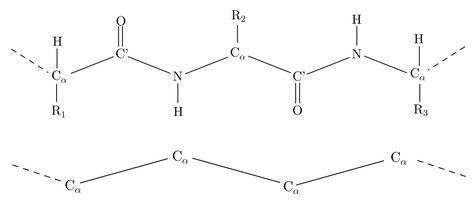


Figure 1: Top: All-atom protein backbone, with R_1, R_2 and R_3 representing side-chains. **Bottom**: C_α-trace.

Our approach

The folding should be conducted, such that it minimizes the number of clashes and at the same time minimizes the deviation from the target C_α-trace.
We consider our fitting problem as two somewhat separate problems. First, we fold the backbone to the C_α-trace. Hereafter, the amino acid side-chains are added to the backbone.

Protein Geometry

- Proteins are built from unbranched chains of amino acids.
- All amino acids share the same basic structure and a variable side-chain.
- The structure of an amino acid can be described by bond lengths, bond angles and rotational angles.
- The bond lengths and bond angles only displays small variations between amino acids (see tables).
- Only three different rotational angles occurs along an amino acid chain. These are named ϕ, ψ and ω and are shown on Figure 2.

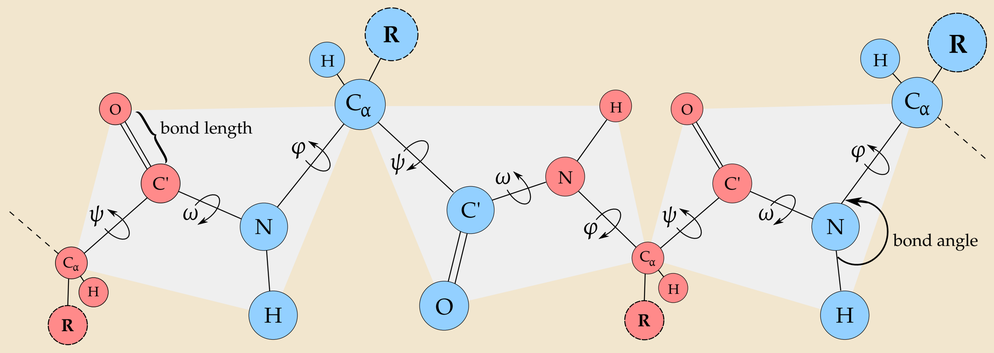


Figure 2: Rotational angles in a protein backbone

Bond	Avg. length	Std.dev.
C-O	1.2260 Å	0.0188 Å
CA-C	1.5272 Å	0.0191 Å
N-CA	1.4680 Å	0.0237 Å
C-N	1.3234 Å	0.0215 Å

Table 1: Average bond lengths (in ångstrøm)

Angle	Avg. angle	Std.dev.
H-N-CA	118.9553°	1.9979°
N-CA-C	110.6099°	2.4668°
CA-C-N	116.7804°	1.7682°
C-N-CA	121.4547°	1.9946°

Table 2: Average bond angles

- The ω-angle is almost always at 180°, except for a few occurrences where it is 0°. To simplify the problem, we have assumed that it is always locked at 180°.
- The ϕ and ψ angles are the most variable parts of the protein backbone and many models use these as the only parameters when performing protein structure prediction.
- The ϕ and ψ are the only parameters we modify when folding the backbone.

Backbone folding

Given a C_α-trace, we wish to fold the protein backbone (only the ϕ and ψ angles) to match the trace as closely as possible. The backbone fitting problem can be regarded as an *inverse kinematics* problem.
To solve this problem, we have devised an extension to the *cyclic coordinate descent* (CCD) algorithm:

- CCD works by adjusting angles one by one in a greedy manner.
- We adjust each angle with the goal of minimizing the mean distance between the three forthcoming amino acids C and their corresponding C_α-targets T.
- The angles are adjusted iteratively in both directions interchangably until the deviation has converged.

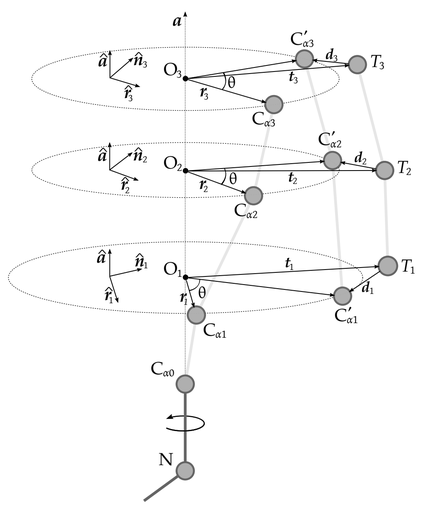


Figure 3: Adjusting an angle according to the C_α-targets with the CCD algorithm.

Rotamer selection

- After the backbone is folded in place and fitted to the C_α-trace, we remain with the problem of adding side chains to each amino acid.

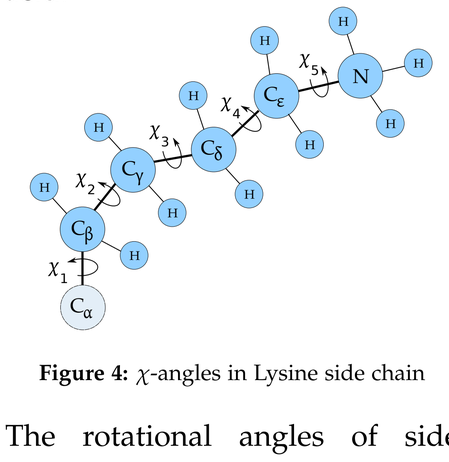


Figure 4: χ-angles in Lysine side chain

- The rotational angles of side chains are named χ_1-χ_5, but many amino acids only have one or two of these angles.
- Each side chain tends to have certain configurations of its χ-angles. These often occuring configurations are called rotamers of the side-chain.
- There exists rotamer libraries containing these common configurations together with their likelihood.
- We have developed a rotamer search algorithm that minimizes the number of occurences, which uses the rotamer-probabilities as a good initial guess.