LOOKING FOR L\(\beta\)H

Oral exam introduction
Youval Dar
UC Davis Physics Department
April 2009

PURPOSE:
To review my research with Daniel Cox’s group.

I have developed software that helps in identifying protein regions that might be able to assume a specific structure called left hand \(\beta\) helix (L\(\beta\)H). This software has two goals; the first is, assuming a region in a protein has L\(\beta\)H structure, to suggest a thread of the sequence on to the structure, which will most likely be a stable one. And second, to look at proteins with known or unknown structures and identify regions which might adopt L\(\beta\)H structure.

BACKGROUND:
Amyloids are insoluble fibrous protein aggregates. Besides sharing some structural properties, amyloids are believed to play a major role in many neurodegenerative diseases, including Mad-Cow disease, Alzheimer’s disease, Type II diabetes, Huntington and many others (1). Our group focuses on the prion protein, which is the only known infectious protein and is abundant around nerve cells and the brain.

Protein structure is significant to the tasks the protein performs. The structure is formed by folding of the amino acid chain, which makes the protein, in a specific way. The structure is a result of free-energy minimization. This energy landscape, of possible protein conformations, can be one with several minima and barriers separating those minima. In the process of protein folding, the protein gets to the lowest energy state by thermal fluctuation, physical interaction within the protein, between the protein and its surroundings and sometimes with the help of other proteins. There might be some space of conformations, the protein might assume, that are close in free energy terms. Different conformations, of the same protein might differ greatly in functionality and as a result misfolding can lead to diseases.

The prion in its normal form performs some tasks in the cell, but the prion has an isoform, a misfolded form, that is not only stable but also induces reconformation of normal form prion to this isoform. This isoform has very different functional properties. It aggregates to create fibrils and crystallites and it is destructive to the cell.
**Motivation:**

Knowing the prion isoform structure will help understand what it does to the cell and the infection mechanism. It might help to understand and find structures of other amyloids\(^1\) with unknown structures, and it can help find other proteins that can potentially form amyloids.

**LβH Structure:**

The prion shares several properties with other amyloids, insolubility and cross-β\(^2\) structure (2) (3). The misfolded prion is found in forms of fibrils and little crystallites which make it hard to find its structure since solution NMR can’t be used and the crystals are not large enough for X-ray diffraction. Low resolution images were obtained using electron crystallography and more structural data using fiber diffraction. Several papers were published proposing the LβH as a good candidate structure for the prion isoform (4) (5). Our group is involved in the research of LβH structure as a candidate for amyloid formation. The group proposed a model for prion oligomer\(^3\) that were grown in a lab, using the LβH structure (6) and proposed a structure stabilizing model, using domain swapping, for trimeric LβH prion fibrils structure (7).

**Method:**

**The Problem**

When trying to fit an amino acid sequence, that makes a protein, onto a specific structure (thread the sequence onto a structure), we can propose a thread and then try this thread in some molecular dynamic (MD) software to check if the proposed thread is indeed stable. The MD software will run for many hours simulating only a few nano seconds of dynamical evolution so it is important to start with a thread or small collection of threads that are most likely to be stable.

There are many ways a sequence can be threaded onto the LβH structure; so many that we can’t try all threads in a reasonable time.

There are millions of known protein sequences with unknown structures

A tool that can quickly provide a list of possible stable threads and screen protein sequences for regions with a tendency to LβH structure can be very useful.

---

1. Amyloids are insoluble protein aggregates that share typical structural properties and involved in protein related diseases.
2. Cross-β structure is a structure typical to protein fibers made of β sheets, where the β strands in the β shits are perpendicular to the direction of the fiber.
3. An Oligomer consists of a limited number of monomers (not unlimited number as polymer is)
SOLUTION IDEA
I have adopted a method of dynamic programming (8) where I create a scoring matrix based on the alignment of a protein sequence to the LβH unique structure. The scoring function is not based on statistical data, like in many other sequence alignment applications, because there are only 15 known proteins with LβH structure. Instead, I have created a scoring function that is based on physical quantities in a very qualitative way. I am taking into account the hydrophobic, polar and charge residues properties, the volume enclosed in the structure, size of residues, residues side chain interaction, loops, corner cuts and residue position bias.

In the dynamic programming method we solve a problem by overlaying multiple sub-problems, representing all possible solutions, in a scoring matrix and then extract from the scoring matrix selected solutions without the need to actually evaluate directly all sub-problems.

SCORING FUNCTION
My scoring function has 12 unknown parameters. I have used the known LβH structures and the Powel method to calibrate my software and fix those parameters. I did it in a way that minimizes the difference between the known LβH threads and the thread my software produced for the same sequence.

SOFTWARE INPUT/OUTPUT
There are several modes of operation in my current software version.

STRUCTURE PREDICTION
The input is an amino acid sequence, a portion from a protein sequence:

NGVSFVNPEATYIDVEIASEVQIEANVTLKQQTGKAGETVLTNGTYVDSTIGAGAVITNNSMIEESSVA

And the output is:

86.4 :: NGVSFV|N(PE)AYID|DVEIA|SEVQIE|ANVTLK|GQT|G|ETVLT|NGT|Y|VVDSTIG|AGAVIT|NNSMIE|ESSVA

---

4 Hydrophobicity, polarity, charge, residue weight, H-Bond side chain interaction, Disulfide bond side chain interaction, loop penalty, loop Hydrophobicity weight, corner cut penalty, full turn weight and special biases.
The thread with a score per amino acid indicates the quality of the thread as LβH structure and the plot gives a sense of how the threads conformation space appears. (The software allows controls of the range of scores collected both in the scoring process and the result collection.)

Sequence scanning

In the sequence scanning mode of operation, the input is the full protein amino acid sequence. The program scans the sequence in steps of six amino acids and in each step calculates a score of a sequence segment, indicating the likelihood of LβH structure. Those plots are sensitive to the length of segment and the allowed abnormalities in the structure. I repeat this procedure with different parameters and produce a plot for each run and a list of threads for any segment with a score above a desired score.

CURRENT STATUS

We are working on two projects. First, we want to check if the LβH structure might appear in other amyloid, disease related, proteins like the Curli protein (E-Coli), the SOD1 (ALS disease), the Alpha-synuclein (Parkinson’s disease) and others. And second, we want to compare the predicted LβH threads across different species, since different species react differently to prion PrPsc infection, we want to explore the thread – sequence relations.

We have collected 52 different prion sequences and 15 disease related proteins (we will gather several more) and produced lists of threads and plots for those proteins.
Next we need to evaluate the stability of selected threads using MD software. And we hope reach to some publishable results.

**REFERENCE**


