A generative, probabilistic model of local protein structure

18. februar 2019 15:44

Key remaining challenge of protein structure prediction:

- An efficient probabilistic exploration of the structural space that correctly reflects the relative conformational stabilities. (So we want to correlate the structure of the protein to whether it is active or inactive or its functionalities?)

What is structural space?

https://www.khanacademy.org/test-prep/mcat/chemical-processes/proteins/v/conformational-stability-protein-folding-and-denaturation
What is conformational stability?

- Conformation referes to the folded 3D structure of the protein chain (ACTIVE)
- Denatured proteins = proteins that have become unfolded (INACTIVE)

So conformational stability is talking about the various forces that help to keep a protein folded in the right way. These various forces are the four different levels of protein structure:

- Primary tructure = sequence of amino acids (determined by peptide bonds)
- Secondary structure = the local substructure in a protein (determined by backbone interactions held together by hydrogen bonds)
- Tertiary structure = the overall 3D structure of a single protein molecule (described by distant interactions between groups within a single protein)
- Quatenary structure = desribes the different interactions between individual protein subunits (folded-up proteins that come together to assemble the completed overall protein.

Task:

- Predict the 3D structure of a protein given its amino acid sequence.

We can simplify things by restricting the angular degrees of freedom in the protein backbone:

- **What is angular degrees in the context of proteins?**

Crystallography:

- Determining the arrangement of atoms in crystalline solids (e.g proteins).

---

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3849216/

What is fragment-assembly?

- Models of protein strucutre are assembled from fragments of known protein structures.

Fragment assembly lack proper statistical foundation.
Fragment assembly normally proceeds by a Markov chain monte carlo

- Is this the same principle as when I implemented a particle filter for robotic positioning (Markov assumption)?

What exactly is an energy function? I can tell it has something to do with guiding the fragments assembly, but what does it include/how does it work?

Stationary probabilistic distribution dictating the transition probabilies of the Markov chain.

Fragment assembly problem:

- Not possible to evaluate the proposal prob. of a given structure, making it difficult to ensure unbiased sampling

Showed that a 1st order markov model forms an efficient probabilistic, generative model of the C_a geometry of proteins in continuos space.

Sequence of dihedral angle pairs:

- **Angles seem very important?**

---

https://www.youtube.com/watch?v=wvTv8TqWC48

Protein basics:

- All proteins are made out of the same 21 building blocks (amino acids)
- Amino acids are made of:
  - Carbon
  - Oxygen
  - Nitrogen
  - Hydrogen
  - Sulfur atoms
- A protein consists of different groups:
  - Amino group
  - Carboxyl group
  - Side chain attached to a central carbon atom
- Protein chains often fold into two types of secondary structures:
  - Alpha helices
  - Beta sheets
- Tertiary structure of a protein is the 3-dimensional shape of the protein chain (determined by the amio acids)
- The functions of many proteins rely on their 3D shapes

- So thats why predicting the folding can become so handy?
