Biological Background

**Protein structure and folding**

Proteins constitute the base for all biological activity, for example as enzymes catalyzing chemical reactions, they are involved in cell signaling and as structural components of cells. Although the high diversity among proteins, they are all linear polymers made up of the same 20 amino acid building blocks [1]. The amino acids are linked together by peptide bonds forming between the carboxyl group of the first amino acid and the amino group of the next, resulting in a long chain with hundreds of residues, the protein primary structure[2]. Driven by the formation of hydrogen bonds, van der Waals interactions, electrostatic and hydrophobic interactions between residues, preferences in backbone angles and a minimal chain entropy, the amino acid chain folds into a native protein.

// Vill skriva ungefär det här. Hur? // protein structure can be described by a hierarchical system, with levels corresponding to primary sequence (covalent bonding of amino acids), secondary structure (segments of recurring arrangement of amino acids consecutive in the sequence), tertiary structure (mutual arrangement of secondary structures in a protein domain), and quaternary structure (mutual arrangement of domains within a multi-domain protein or different subunits in a multi-protein complex).  [3]

The 3D structure determines the protein properties and functions, and the understanding of how different structural details affect protein– protein interactions and enzyme activity is of high importance in for example drug and enzyme design[4], [5].

**Protein structure predictions**

X-ray crystallography and nuclear magnetic resonance (NMR) are the main approaches for experimental determination of tertiary structures [2]. However, the methods are also expensive, both in terms of running costs, time, and labour, resulting in a large gap between the number of determined 3D structures and the number of known primary structure sequences. But as the techniques for primary structure determination have become more effective and sequence and structure databases have grown, so has the possibility for computational predictions of protein structure[6].

There are three main approaches in computational protein structure prediction, homology modeling, fold recognition or threading and *ab initio* methods, also known as first principle methods. *Ab initio* methods are based on Anfinsen’s dogma [REF], that all information needed to predict a proteins 3D structure is embedded in its amino acid sequence. This is often done by searching for the protein conformation with the lowest free energy. Database information, such as general rules of folding, can be used as complement. MEN DÅ ÄR DET JU INTE HELT AB INITIO.

Model Quality Assessment

Even a low-resolution model, only showing residue positions, can be useful. Kihara et al. (2009) states that the major reason for not applying predictions in practical work is that the quality of the model is unknown, not that it is inaccurate. Even somewhat inaccurate models, or models with low resolution, can be used in for example early stages of drug development, as long as the estimated error is known, entailing high accuracy model quality assessments for ranking predicted models.

Deep learning

3DCNN

[1] D. Whitford, *Proteins: structure and function*. Hoboken, NJ: J. Wiley & Sons, 2005.

[2] J. M. Berg, J. L. Tymoczko, L. Stryer, and L. Stryer, *Biochemistry*, 5th ed. New York: W.H. Freeman, 2002.

[3] M. J. Gajda, M. Pawlowski, and J. M. Bujnicki, ‘Protein Structure Prediction: From Recognition of Matches with Known Structures to Recombination of Fragments’, in *Multiscale Approaches to Protein Modeling*, A. Kolinski, Ed. New York, NY: Springer New York, 2011, pp. 231–254.

[4] A. P. Joseph and A. G. de Brevern, ‘From local structure to a global framework: recognition of protein folds’, *J. R. Soc. Interface*, vol. 11, no. 95, pp. 20131147–20131147, Apr. 2014.

[5] S. Wang, J. Peng, J. Ma, and J. Xu, ‘Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields’, *Sci. Rep.*, vol. 6, no. 1, May 2016.

[6] M. Dorn, M. B. e Silva, L. S. Buriol, and L. C. Lamb, ‘Three-dimensional protein structure prediction: Methods and computational strategies’, *Comput. Biol. Chem.*, vol. 53, pp. 251–276, Dec. 2014.