Chapter 1

Introduction

1.1 Proteins

Proteins nowadays receive full attention from the scientific community. When one tries "protein" as a search term in Google Scholar[1] one finds in excess of 13 million hits. By the same method one finds 9 million hits on "metals", and less on other classes of (solid state) materials (see table 1.1). The same picture is given by a search on Web of Science[2] in which one can find 93 thousand hits on protein publications during 2004 alone*. Although one could argue that apparently the protein science community writes a lot, it is a good indication of the interest people have in the subject. Part of the interest in proteins originates from its omnipresence in living organisms, including ourselves. Proteins play a crucial role in many biochemical processes on a cellular level, for instance catalysis (enzymes), structural or mechanical functions (e.g. cytoskeleton, transport), cell signaling, and immune responses (e.g. references [3–7]).

Protein molecules consist of a chain of L-α-amino acids, linked together by peptide bonds (fig. 1.1). The length of this chain ranges typically from 100 to 27,000 amino acids, thus ranking proteins in the class of biological macro-

---

*2004 is the last year for which the Web of Science database returns less than 10^5 hits on "protein", which is the maximum number of hits shown by the search engine.

Table 1.1: Google Scholar and Web of Science "experiment" for different classes of solid state materials: Number of Google Scholar hits at February 10, 2007; Web of Science database hits for the year 2004.

| Search term	# of hits ($\times 10^6$) Google Scholar	# of hits ($\times 10^4$) Web of Science, 2004
protein	13.6	9.3
metal	9.6	3.0
semiconductor	2.5	0.64
ceramic	1.3	0.37
oxides	1.0	0.59
superconductor	0.2	0.15

molecules together with molecules like nucleic acids and polysaccharides[8]. Although proteins contain many amino acids, the variation in type is limited to a set of 20 different amino acids, which differ in the composition of their side chains, or "residues". The specific sequence of amino acids in a protein is called its primary structure and is usually given by a list of three-letter abbreviations of the amino acid type (fig. 1.2a).
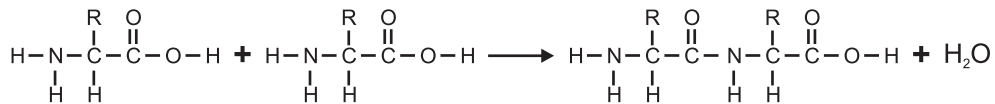
Figure 1.1: The peptide bond.

The peptide bonds between the amino acids form the backbone of the protein molecule. Within the backbone the peptide bonds can occur in different conformations which can have a repetitive character. Typical local repetitive structures are the α-helix and the β-sheet (fig. 1.2b), the occurrence of which is called the secondary structure of a protein. The three-dimensional structure of a protein as a whole is called its tertiary structure (fig. 1.2c). In general, a protein molecule seeks to minimize its free energy by folding up into a specific

three-dimensional structure, which is stabilized through salt bridges, hydrogen bonds, disulfide bonds and van der Waals forces. Here, the side chains play an important role as their (non-)polarity determines the type of interactions possible and thus the 3D structure. For instance, globular proteins tend to form a hydrophobic core with the hydrophobic residues on the inside of the molecule and the hydrophilic ones at the surface[9].
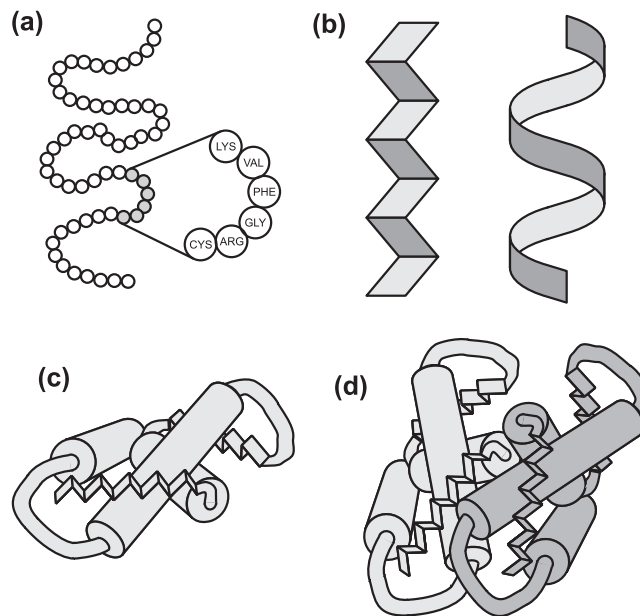
Figure 1.2: Structure of protein molecules. (a) Primary structure; the sequence of amino acids in the protein. (b) Secondary structure; β-sheets and α-helices, formed by a sequence of amino acids linked by hydrogen bonds. (c) Tertiary structure; the folding of the amino acid chain. The cylinders represent α-helices, the zigzags β-sheets (d) Quaternary structure; a protein complex consisting of more than one amino acid chain.

The biological function of a protein usually cannot be understood from the amino-acid sequence alone. Often, the biological potency of proteins depends on their 3D structure[10, 11], while the primary structure for the different conformations is the same. Another example for which knowledge of the 3D

Table 1.2: Number of structures in the Protein Data Bank[12] divided by type of biological macromolecule and the methods used to obtain their structures, as of the 10th of February 2007.

Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	**Total**	Deposited in 2006
XRD	32723	941	1526	28	**35218**	5650
NMR	5208	738	128	7	**6081**	872
EM†	95	10	38	0	**143**	34
Other	78	4	3	0	**85**	12
Total	**38104**	**1693**	**1695**	**35**	41527	6568

structure is required is drug design. To design a drug which enters and blocks the active site of a malfunctioning protein molecule, knowledge of the atomic positions of this site is required. Thus, to understand the functionality or, in the case of diseases, dysfunctionality of a protein, one would like to know the tertiary structure of the protein molecule.

Various methods exist to determine the three-dimensional structure of a protein molecule down to atomic resolution. Solved structures usually are deposited in the Protein Data Bank[12], whose statistics show that the main method for structure determination is X-ray diffraction (XRD), see table 1.2. For small protein molecules (<20.000 Da) atomic resolution can be achieved by nuclear magnetic resonance (NMR). Neutron scattering and electron microscopy form a minority in solved structures.

Any structure determination starts with the acquisition of the protein from biological material (fig. 1.3a). Tissue containing the protein of interest is dissolved, and subsequently purified by methods like centrifugation and high precision liquid chromatography (HPLC) [13]. For a structure determination by XRD, a single crystal is required (fig. 1.3b) to acquire diffraction patterns like the one shown in figure 1.3c. The positions of the peaks in these patterns supply information on the crystal lattice, i.e. the axis lengths and angles of the

---

†Electron Microscopy

unit cell. In addition, the intensities of the peaks supply information on the electron density in three dimensions[14, 15], thus enabling one to form a 3D picture of the structure of the protein molecule (fig. 1.3d). Both the signal-to-noise ratio and the spatial extent of the diffraction pattern are of importance for the level of structural detail one can obtain by XRD[16]. These factors depend on the quality of the crystal, i.e. the better the crystalline order, the better the XRD pattern. Thus, a prerequisite for structure determination by XRD is to obtain a high quality single crystal of the protein of interest. However, proteins do not crystallize easily, and structure determinations seldomly reach details of 1 Å, whereas resolutions of 2∼3 Å are more common[12], which for macromolecules usually is insufficient to obtain essential structural information. In the process of 3D structure determination of biological macromolecules, growing XRD quality crystals remains the bottle-neck.

1.2   Protein crystal growth

Because most proteins are water soluble, they are usually crystallized from aqueous solutions. For crystals to be formed in the solution, the protein concentration must be higher than the equilibrium concentration, i.e. the protein concentration in a saturated solution. From a thermodynamical point of view, the chemical potential difference between the dissolved phase and the crystalline phase, $\Delta\mu$, indicates which of these two phases is the most energetically favourable, being positive for crystallization and negative for dissolution. In the case of crystal growth from solution, the chemical potential difference is related to the solute concentration, c, by

$$\Delta\mu = kT \ln \frac{f\,c}{f_{eq}\,c_{eq}} \ ,$$
(1.1)

in which k is Boltzmann's constant, T is the temperature, f (f_{eq}) is the (equilibrium) activity coefficient of the protein in solution, c is protein concentration, and c_{eq} is the equilibrium protein concentration. In this equation the activity coefficient of the dissolved protein is assumed to be independent of concentration. In equation 1.1 one can see that to create a supersaturated
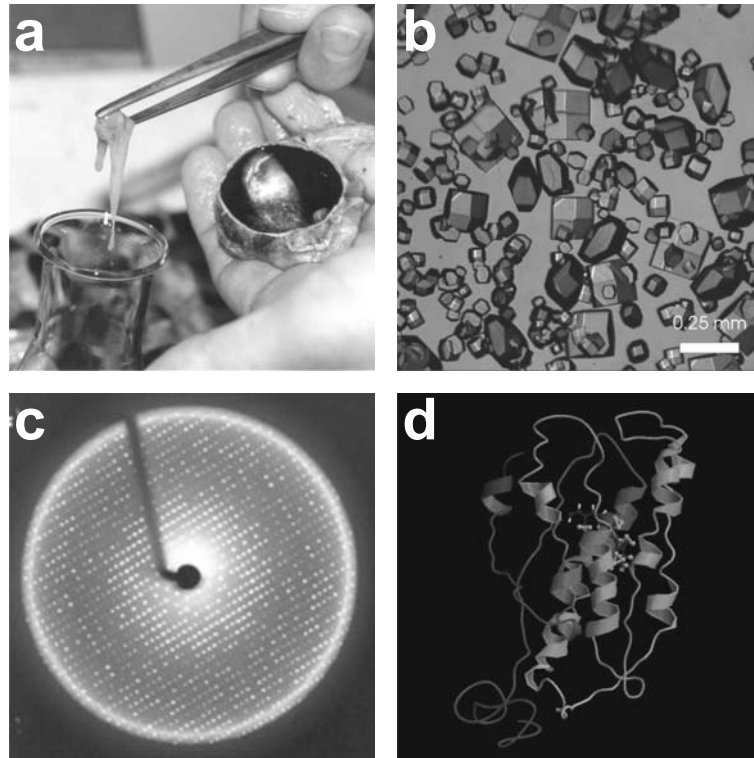
Figure 1.3: The process of determination of the 3D molecular structure of proteins. (a) Acquisition of proteins from biological material, in this case bovine rhodopsin from the retina of cow's eyes (photo by courtesy of Petra Bovee-Geurts, NCMLS). (b) Growth of protein single crystals. Shown is a sample of tetragonal hen egg-white lysozyme (HEWL) crystals. (c) Example of an X-ray diffraction pattern of a protein. (d) 3D structure of bovine rhodopsin (PDB entry 1jfp), showing the folding of the protein.
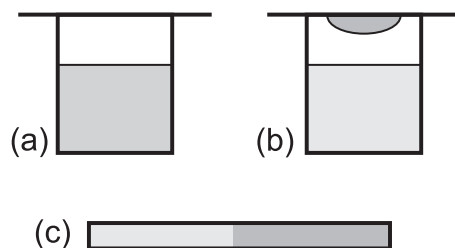
Figure 1.4: Crystallization methods for proteins. (a) In the batch method a supersaturated mixture is prepared and left to nucleate and crystallize. (b) The vapour diffusion method is based on vaporization of the undersaturated mother liquor due to a salt concentration difference with the larger reservoir. (c) For the free interface diffusion method a protein solution and a salt solution are carefully brought into contact in a capillary. Due to diffusion a range of different supersaturation conditions is formed.

solution, one can either increase the protein concentration or decrease the equilibrium concentration. The equilibrium concentration depends on many parameters including temperature, pH, and solution composition[17–19]. These parameters make up the complex multidimensional phase diagram of the protein in solution.

Traditional macromolecular crystallization techniques include the batch method, vapour diffusion methods, and free interface diffusion (fig. 1.4). These methods have in common the use of salts to reduce the protein solubility, while also polyethylene glycol (PEG) is often used as an additional precipitating agent. The methods differ in the starting point and pathways through the phase diagram (fig. 1.5). Every protein has its own specific phase diagram, and a crystallization experiment should be designed to fit this phase diagram. When crystallizing a protein for XRD purposes, usually its phase diagram is unknown. A first attempt thus can be nothing more than an educated guess based on experience with similar proteins.

To find the protein crystallization conditions, one screens the solution composition, i.e. type and concentration of additives, by trail-and-error. A typical protein crystallization can take days up to months. To reduce the time to find
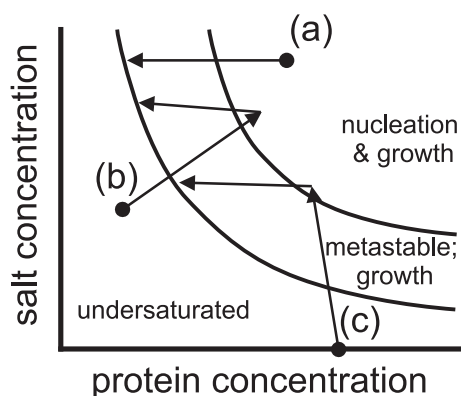
Figure 1.5: Pathways through the phase diagram for three of the most common macromolecular crystallization techniques. As an example, salt concentration is taken as solubility parameter. (a) Batch crystallization. (b) Vapour diffusion. The solution approaches supersaturation from undersaturated conditions. (c) free interface diffusion, or counter diffusion. In this method a whole range of conditions is tested at the same time. Shown is a pathway starting at zero salt concentration

the proper conditions for XRD-quality, crystals screenings are performed in multi-well plates. Experiments are often automated by robotics[20]. An advantage of the trail-and-error approach is that it increases the rate at which XRD-quality protein crystals can be produced with little knowledge of the complex machinations of the crystal growth process. However, the educated guess on the correct crystallization conditions becomes more educated if one acquires knowledge on the fundamental aspects of the protein crystallization processes.

1.3   Surface kinetics versus mass transport

Protein crystals grow mainly via the same mechanisms as small molecules[9, 21], i.e. normal (or rough) growth, birth-and-spread, and spiral growth. Surface kinetics processes, like surface diffusion, step bunching, and impurity blocking also occur for macromolecular crystals [22–24] albeit quantitatively

different compared to small molecular crystals. For instance, the step kinetic coefficient β_s for proteins is in the order of 10^{-4} $cm\,s^{-1}$, while for small molecules it is in the order of 10^{-2} $cm\,s^{-1}$ [25].

Solution crystal growth is a sequence of two processes. The first process is transport of mass toward the surface of the crystal, the second is the incorporation of growth units into the crystal (fig. 1.6). In other words, without a fresh supply of growth units to the solution-crystal interface the crystal cannot grow, and a crystal will not grow as well in case it is difficult to attach growth units to the surface. The overall growth rate of the crystal is determined by the slowest of these two processes.
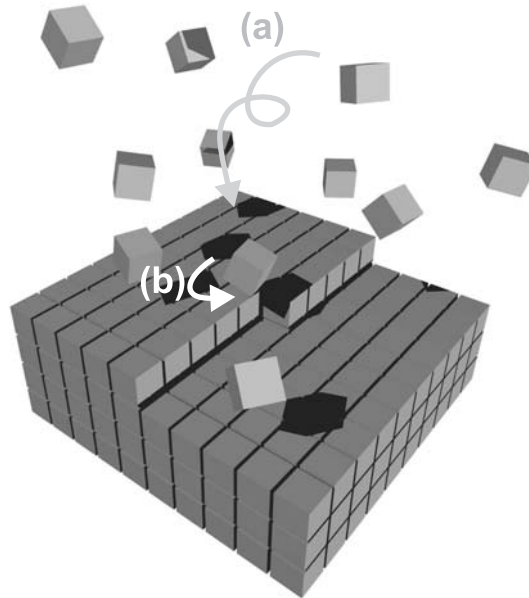
Figure 1.6: Schematic, simplified, representation of the crystallization process, consisting of a mass transport part (a) and a surface incorporation part (b). The crystal growth rate is determined by the slowest of these two parts.

Mass transport itself can be subdivided into two different processes; con-

vection and diffusion. Natural convection is a consequence of density differences in the solution, created by the growing crystal, and gravity. Convection is an effective means of mass transport, strong enough to drag along large impurities. Upon reaching the crystal surface, these impurities can become incorporated into the crystal and thus reduce the crystal quality. Therefore, mass transport is not only important for the crystal growth rate, but also for the crystal quality[25–28]. Research focuses on various methods to influence the transport of impurities to the surface, all of them aiming to cancel convectional mass transport. These methods include the use of gels[29–31], experiments in microgravity (i.e. in space) [28, 32], using nanoliter volumes[33, 34], and, as described in this thesis, the use of inhomogeneous high magnetic fields. With convection cancelled, diffusion remains as the sole means of transport, and is a slow process for large impurities compared to the diffusion of protein molecules. Another beneficial effect of cancelling convection is the reduced local supersaturation during growth. For nucleation a high $\Delta\mu/kT$ is required, but for further growth a low supersaturation is preferred because experience shows low $\Delta\mu/kT$ to give higher quality crystals. Thus, it is thought that the cancellation of convection should improve crystal quality[25–28]. However, if the growth is very much limited by the rate of diffusion, the morphology of a crystal can become unstable or dendritic[35]. To grow a good quality crystal, one needs to find the right balance between all processes.

1.4   So, this thesis....

The work in this thesis focuses on the crystallization of the protein hen egg-white lysozyme (HEWL). Hen egg-white lysozyme is a well-known protein, and serves as the guinea pig of protein crystallization. Although much has been written about this protein, its crystallization process is still not well understood. For instance, in 1937 the protein was crystallized for the first time[‡], but only since the 1990's the role of liquid-liquid phase separation for the crystallization process became more apparent[36, 37]. In this thesis, the lyso-

---

[‡]By P. Abraham and R. Robinson.

zyme crystallization process is investigated from a physico-chemical point of view, mainly by optical microscopy, aiming to contribute to the understanding of mass transport and surface kinetics in protein crystal growth. The first few chapters focus on the surface processes during crystallization, revealing anisotropic growth spirals by atomic force microscopy, and showing growth inhibition, probably due to a self-poisoning mechanism, by optical microscopy. Subsequently, the focus shifts to mass transport, showing its influence in the formation of spherulites, and in crystal growth from a liquid-liquid phase separated system. Control of mass transport is the topic of chapters 6 and 7, in which the use of high magnetic fields to cancel convection during crystal growth is explored. In the last chapter, surface kinetics and mass transport find each other in an experiment in which both processes compete to dominate, resulting in morphologically unstable crystals.

References

[1] http://scholar.google.com.

[2] Web of Science, Thompson, http://isiknowledge.com.

[3] Kerssemakers, J.; Munteanu, E.; Laan, L.; Noetzel, T.; Janson, M.; Dogterom, M. *Nature* **2006,** *442,* 709-712.

[4] Uemura, T.; Kashiwagi, K.; K.Igarashi, *The Journal of Biological Chemistry* **2007,** *282,* 7733-7741.

[5] Rautsi, O.; Lehmusvaara, S.; Salonen, T.; Häkkinen, K.; Sillanpää, M.; Hakkarainen, T.; Heikkinen, S.; Vähäkangas, E.; Ylä-Herttuala, S.; Hinkkanen, A.; Julkunen, I.; Wahlfors, J.; Pellinen, R. *The Journal of Gene Medicine* **2007,** *9,* 122-135.

[6] Boyman, O.; Kovar, M.; Rubinstein, M.; Surh, C.; Sprent, J. *Science* **2007,** *311,* 1924-1927.

[7] Pawson, T. *Nature* **1995,** *373,* 573-580.

[8] Hennessy, D.; Buchanan, B.; Subramanian, D.; Wilkosz, P.; Rosenberg, J. *Acta Crystallographica D* **2000,** *56,* 817-827.

[9] McPherson, A. *Crystallization of Biological Macromolecules;* Cold Spring Harbor Laboratory Press: New York, 1999.

[10] Svensson, M.; Håkansson, A.; Mossberg, A.-K.; Linse, S.; Svanborg, C. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** *97,* 4221-4226.

[11] Kang, T.; Radic, Z.; Talley, T.; Jois, S.; Taylor, P.; Kini, R. *Biochemistry* **2007,** *46,* 3338-3355.

[12] RCSB-Protein Data Bank, http://www.pdb.org.

[13] Lorber, B.; Giegé, R. Biochemical aspects and handling of macromolecular solutions and crystals. In *Crystallization of Nucleic Acids and Proteins*; Oxford University Press: Oxford, 1999.

[14] Drenth, J. *Principles of Protein X-Ray Crystallography;* Springer: New York, 1999.

[15] Sawyer, L.; Turner, M. X-ray analysis. In *Crystallization of Nucleic Acids and Proteins*; Oxford University Press: Oxford, 1999.

[16] Weiss, M. *Applied Crystallography* **2001,** *34,* 130-135.

[17] Howard, S.; Twigg, P.; Baird, J.; Meehan, E. *Journal of Crystal Growth* **1988,** *90,* 94-104.

[18] Cacioppo, E.; Pusey, M. *Journal of Crystal Growth* **1991,** *114,* 286-292.

[19] Riès-Kautt, M.; Ducruix, A. From solution to crystals with a physico-chemical aspect. In *Crystallization of Nucleic Acids and Proteins*; Oxford University Press: Oxford, 1999.

[20] Stevens, R. *Current Opinion in Structural Biology* **2000,** *10,* 558-563.

[21] Land, T.; Malkin, A.; Kuznetsov, Y.; McPherson, A.; Yoreo, J. D. *Physical Review Letters* **1995,** *75,* 2774-2777.

[22] Malkin, A.; Kuznetsov, Y.; McPherson, A. *Journal of Crystal Growth* **1999,** *196,* 471-488.

[23] Astier, J.; Bokern, D.; Lapena, L.; Veesler, S. *Journal of Crystal Growth* **2001,** *226,* 294-302.

[24] Plomp, M.; McPherson, A.; Malkin, A. *Proteins* **2003,** *3,* 486-495.

[25] Chernov, A. *Journal of Structural Biology* **2003,** *142,* 3-21.

[26] Carter, D.; Limb, K.; Hoa, J.; Wright, B.; Twigg, P.; Miller, T.; Chapman, J.; Keeling, K.; Rublea, J.; Vekilov, P.; Thomas, B.; F. Rosenberger, A. C. *Journal of Crystal Growth* **1999,** *196,* 623-637.

[27] Snell, E.; Judge, R.; Crawford, L.; Forsythe, E.; Pusey, M.; Sportiello, M.; Todd, P.; Bellamy, H.; Lovelace, J.; Cassanto, J.; G.E.O.Borgstahl, *Crystal Growth & Design* **2001,** *1,* 151-158.

[28] Thomas, B.; Chernov, A.; Vekilov, P.; Carter, D. *Journal of Crystal Growth* **2000,** *211,* 149-156.

[29] Robert, M.; Lefaucheux, F. *Journal of Crystal Growth* **1988,** *90,* 358-367.

[30] Biertümpfel, C.; Basquin, J.; Suck, D.; Sauter, C. *Acta Crystallographica D* **2002,** *58,* 1657-1659.

[31] Garcia-Ruiz, J.; Moreno, A.; Viedma, C.; Coll, M. *Materials Research Bulletin* **1993,** *28,* 541-546.

[32] Kundrot, C.; Judge, R.; Pusey, M.; Snell, E. *Crystal Growth & Design* **2001,** *1,* 87-99.

[33] Carter, D.; Rhodes, P.; McRee, D.; Tari, L.; Dougan, D.; Snell, G.; Abolac, E.; Stevens, R. *Journal of Applied Crystallography* **2005,** *38,* 87-90.

[34] Cherezov, V.; Caffrey, M. *Journal of Applied Crystallography* **2003,** *36,* 1372-1377.

[35] Nanev, C. *Progress in Crystal Growth and Characterization of Materials* **1997,** *35,* 1-26.

[36] Berland, C.; Thurston, G.; Kondo, M.; Broide, M.; Pande, J.; O. Ogun, G. B. *Proceedings of the National Academy of Sciences of the United States of America* **1992,** *89,* 1214-1218.

[37] Muschol, M.; Rosenberger, F. *Journal of Chemical Physics* **1997,** *107,* 1953-1962.