

J. D. Gunton
A. Shiryayev
and D. L. Pagan

Protein Condensation

Kinetic Pathways to Crystallization and Disease

CAMBRIDGE

PROTEIN CONDENSATION

Kinetic Pathways to Crystallization and Disease

J. D. GUNTON, A. SHIRYAYEV, AND D. L. PAGAN

*Department of Physics
Lehigh University*



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521851213

© J. Gunton, A. Shiryayev & D. Pagan 2007

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2007

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

ISBN 978-0-521-85121-3 hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to
in this publication, and does not guarantee that any content on such
websites is, or will remain, accurate or appropriate.

PROTEIN CONDENSATION

Kinetic Pathways to Crystallization and Disease

This book deals with the phase transitions, self-assembly, and aggregation of proteins in solution. Its primary purpose is to bring an interdisciplinary audience the state of the art in current research. The book discusses issues related to the production of high quality protein crystals from solution, in which the bottleneck is crystal nucleation. Since protein function is determined by protein structure, it is necessary to grow high quality crystals in order to determine their structure, usually by X-ray crystallography. The main challenge is to determine the initial solution conditions so that optimal crystal nucleation occurs. The book also discusses diseases that occur due to undesired protein condensation, an increasingly important subject. Examples include sickle cell anemia, cataracts, and Alzheimer's disease. Current experimental and theoretical work on these diseases aims to understand the diseases at a fundamental, molecular level, in order to prevent the undesired condensation from occurring. Suitable for graduate students and academic researchers in physics, chemistry, structural biology, protein crystallography, and medicine.

J. D. GUNTON is Joseph A. Waldschmitt Professor of Physics at Lehigh University in Pennsylvania. He is the author of approximately 200 articles in refereed journals on equilibrium and nonequilibrium phase transitions. He is a Rhodes Scholar and a Danforth Fellow, and is a Fellow of the American Physical Society.

A. SHIRYAYEV and D. L. PAGAN received their Ph.D. degrees from Lehigh University in 2005. Both have already published several refereed articles that deal with the condensation of globular proteins.

To our wives, Peggy, Maria, and Martha
for their patience, love, and continued support

Preface

This book deals with a truly interdisciplinary subject: protein condensation from solution. We use “condensation” in this book to denote one of several forms of proteins: a dense, protein-rich fluid phase, an amorphous aggregate, a gel, a crystal, or a polymer fiber. All these forms have been observed experimentally and are important in their own right. The primary purpose of the book is to bring to a wide audience the current status of research in the field, which is still evolving at a rapid rate. The bulk of the book deals with issues related to producing high quality protein crystals from solution, in which the bottleneck is crystal nucleation. Here the main challenge is to determine the initial solution conditions so that optimal crystal nucleation occurs. A second and increasingly important subject that we discuss involves diseases that occur due to undesired protein nucleation. A classic example is the nucleation of polymer fibers of sickle hemoglobin molecules within the red blood cells that distorts the cells and produces sickle cell anemia. Another example is that of age-related cataracts produced by the undesired aggregation of γ -crystallin protein molecules within the vitreous fluid of the eye. A third, somewhat different, example involves the role of amyloid β protein in Alzheimer’s disease. This list is likely to grow as scientists become more aware of the molecular origins of different diseases.

As the field is interdisciplinary, the first part of the book involves several brief reviews of subjects relevant to understanding protein condensation. Readers with an expertise in these topics should omit them and begin with the second part of the book, which treats several examples of globular and membrane proteins. The third part deals with the three diseases mentioned above.

We should note what the book does not deal with. It is not a treatise on the practical art/science of growing protein crystals. The classic work on this subject is A. McPherson’s book, *Crystallization of Biological Macromolecules*. The emphasis in our book is on developing a statistical mechanics theory of the equilibrium and non-equilibrium aspects of protein condensation. We do not

discuss the kinetics of crystal growth since several review articles exist on this topic.

We would like to acknowledge the enormous help and encouragement we have received from our colleagues around the world. These include G. Benedek, F. Bonneté, F. Ferrone, T. Odijk, W. Poon, H. E. Stanley, A. Tardieu, D. Teplow, B. Urbanc, and P. Vekilov. We are particularly indebted to R. Sear for his critical reading of several chapters and his helpful, constructive comments. We also wish to thank I. Dokukina, M. Gunton, and N. Wentzel for their help with the manuscript.

Finally, one of us (JDG) would like to acknowledge funding from the Division of Materials Research of the National Science Foundation and the G. Harold and Leila Y. Mathers Charitable Foundation during the period we were writing this book. Without their support, this book would not have been written.

Contents

<i>Preface</i>	<i>page xi</i>
1 Introduction	1
1.1 Overview	1
1.2 Protein function	2
1.3 Types of proteins	4
1.4 Protein crystallization	5
1.5 Outline of book	6
2 Globular protein structure	9
2.1 Amino acids and primary structure	9
2.2 Secondary structure	11
2.3 Tertiary structure	14
2.4 Quaternary structure	17
3 Experimental methods	19
3.1 Methods to determine three-dimensional protein structure	19
3.2 Solubility measurements	21
3.3 Second virial coefficient	21
3.4 Scattering theory	23
3.5 Dynamic light scattering	26
3.6 Self-interaction and size-exclusion chromatography	28
3.7 Cloud point measurements	28
3.8 Methods of protein crystallization	29
3.9 Circular dichroism spectroscopy	30
3.10 Differential scanning calorimetry	30
3.11 High pressure liquid chromatography	31
3.12 Cross-linking	31
3.13 SDS-PAGE analysis	32

4	Thermodynamics and statistical mechanics	33
4.1	Thermodynamics	33
4.2	Free energies	34
4.3	McMillan–Mayer solution theory	37
4.4	Chemical reactions	47
4.5	The van 't Hoff equation	50
4.6	Introduction to chemical equilibrium approach	51
4.7	Phase transitions	52
5	Protein–protein interactions	61
5.1	Introduction	61
5.2	Excluded volume interactions	63
5.3	DLVO theory	64
5.4	Asakura–Oosawa depletion forces	71
5.5	Hofmeister effect	72
5.6	Ion dispersion forces	75
5.7	Hydration and hydrophobic forces	82
6	Theoretical studies of equilibrium	91
6.1	Simulation studies	92
6.2	Extended law of corresponding states	97
6.3	Thermodynamic perturbation theory	98
6.4	Correlation function approaches	99
6.5	Anisotropic models	101
7	Nucleation theory	109
7.1	Classical nucleation theory	109
7.2	Nucleation theorem	113
7.3	Monte Carlo studies of nucleation	114
7.4	Density functional theory of nucleation	117
7.5	Two-step nucleation theory	121
7.6	Nucleation of systems with anisotropic interactions	128
8	Experimental studies of nucleation	135
8.1	Direct determination of nucleation rates: lysozyme	135
8.2	Direct determination of nucleation rates: HbS	146
8.3	Other measurements of HbS nucleation rate	148
8.4	Heterogeneous nucleation in porous medium	153
9	Lysozyme	156
9.1	Introduction	156
9.2	Solubility of lysozyme	158
9.3	Liquid–liquid phase separation	164

9.4	Protein–protein interactions	172
9.5	Theory for charge and salt effects	181
9.6	Metastability limit for lysozyme solutions	188
9.7	Effects of PEG on protein interactions	192
9.8	Smooth transition from metastability to instability	193
9.9	Equilibrium cluster formation	197
10	Some other globular proteins	205
10.1	Introduction	205
10.2	Glucose isomerase: crystallization via liquid–liquid phase separation	206
10.3	Urate oxidase	208
10.4	BPTI	215
10.5	Alpha-crystallin	217
10.6	ATCase	217
10.7	Apoferitin	218
11	Membrane proteins	221
11.1	Introduction	221
11.2	Three-dimensional crystals of membrane proteins	224
11.3	Role of protein surfactant interactions	225
11.4	Two-dimensional crystallization of membrane proteins	228
11.5	Two-dimensional modified Lennard-Jones model	229
11.6	Two-dimensional square well model with solvent	232
12	Crystallins and cataracts	241
12.1	Introduction	241
12.2	The crystallins	241
12.3	The γ -crystallins	246
12.4	Structure of the γ -crystallins and age-related cataracts	254
12.5	Genetic cataracts	258
12.6	Square well model of γ -crystallins	265
12.7	Anisotropic studies	267
13	Sickle hemoglobin and sickle cell anemia	269
13.1	Experimental results	272
13.2	Theoretical studies	285
14	Alzheimer's disease	299
14.1	Introduction	299
14.2	Amyloids	299

14.3 Kinetics of fibrillogenesis	305
14.4 Soluble amyloid β proteins and neurotoxicity	307
14.5 Tau proteins and neurofibrillary tangles	309
14.6 Modeling the soluble amyloid β protein	309
<i>References</i>	329
<i>Index</i>	361

Color Plate Section between pages 134 & 135

1

Introduction

1.1 Overview

This book deals with the condensation of proteins from solution, including protein crystal nucleation and certain diseases related to undesirable protein condensation.¹ We use the word condensation to describe a variety of possible states of matter, including dense, protein-rich fluids, amorphous aggregates, polymer fibers, gels, and crystals. Much of the book deals with understanding how to grow high quality protein crystals from aqueous solutions of protein molecules. This is of importance in structural biology, which deals with the study of the architecture and shape of biological macromolecules, and in particular with proteins and nucleic acids. Biologists are interested in knowing the structure of proteins, since structure determines function. To determine structure requires high quality protein crystals for use in X-ray crystallography. It is quite difficult to grow high quality protein crystals from solution, however; crystal nucleation is the major bottleneck in protein crystallography. Understanding the dependence of crystal nucleation on the initial conditions of the protein solution is a fundamental problem in statistical physics and is a major theme of this book. Understanding protein crystal nucleation is also important in biomedical research. For example, the sustained release of medications, such as insulin and interferon- α , depends on the slow dissolution rate of protein crystals [1–6]. One can obtain steady medication release rates for longer periods of time by using a dose of a few, larger, equidimensional crystallites than by a dose with a broad crystal size distribution. To obtain such a narrow size distribution requires an almost simultaneous nucleation of the crystals, so that the crystals can grow at the same decreasing supersaturation. In addition, as

¹ G. Benedek used “condensation” in the context of describing apparently unrelated diseases, including cataract, sickle cell anemia, and Alzheimer’s disease [470]. He argued that these are all representations of a broad class of pathologies that he designated as “molecular condensation diseases”; these result from proteins condensing into dense, frequently insoluble phases. We extend this use of “condensation” to include states such as crystals, gels, and other aggregates.

discussed in subsequent chapters, certain diseases, such as sickle cell anemia and human cataracts, result from undesired protein condensation. In such cases one wishes to slow down or prevent the nucleation producing such condensation.

Protein condensation is an intellectually challenging subject in statistical mechanics, as there are many kinetic pathways for condensation to occur. Systems that are evolving toward their equilibrium states, which correspond to states with the lowest free energy, often get stuck in long-lived metastable intermediates. In many cases the outcome depends strongly on the initial position in the phase diagram. Several possible condensation states have been found to occur, such as those noted in the previous paragraph. In order to reach the desired outcome, one must understand the possible kinetic pathways; this is a formidable challenge. To obtain optimal crystallization, one must avoid gel and amorphous aggregate states and, instead, often take advantage of protein-rich liquid droplets via a metastable protein-poor, protein-rich liquid–liquid phase separation. To prevent sickle cell anemia from occurring in patients, one must prevent the nucleation of polymer fibers of sickle hemoglobin molecules from occurring while sickle hemoglobin is in its deoxygenated state in the cells. Although we are far from a detailed solution of the many problems discussed in this book, progress can be made by understanding the free energy landscapes for these systems. Indeed, it has been argued that one can use the free energy landscape of a system, normally used only for calculating its *equilibrium* properties, to predict the possible kinetic pathways that can occur in the course of a phase separation [7]. Although this does not in itself provide guidance on how to choose between various permitted pathways, it does at least limit the possibilities. Some progress has been made in obtaining the theoretical free energy landscape of a system by knowing its equilibrium phase diagram [7].

1.2 Protein function

Why study proteins? What makes them so important to human existence? The answer, of course, is that every living cell and all biological processes depend on proteins. This general statement reflects the fact that proteins are involved in every activity that is undergone in humans, or animals, on every level within the body. To illustrate their importance, consider the proteins involved in catalytic reactions, referred to by their special categorical name, *enzymes*. Examples of enzymes include pepsin, chymotrypsin, and trypsin, which are involved in the digestive process. Like all catalytic agents, enzymes accelerate the rates of chemical reactions while remaining intact themselves. The three aforementioned enzymes are produced in the mucosal lining of the stomach, and act to break down dietary proteins. These enzymes work together to simplify ingested proteins

into their fundamental components, which can then be easily absorbed by the intestinal lining. Interestingly, these digestive enzymes were among the first to be successfully crystallized, confirming an earlier finding that enzymes were indeed proteins. Another example of protein function is the way we protect ourselves from injury or disease. When we are wounded, the protein thrombin, along with other proteins and platelets, is activated in an attempt to form a clot, thereby preventing the loss of blood. Deficiency of these clotting factors is the cause of bleeding disorders such as hemophilia. When we are sick from bacterial or viral infection, our immune system responds by activating antibodies (the proteins of which are referred to as the immunoglobins) to fight off the invasion. Proteins are also involved in supporting the structure of our cells. Examples include collagen, found in tendons and cartilage, and keratin, which is found in hair and fingernails. As another example of their diversity, consider that proteins are needed in order to transport material. A prime example is hemoglobin, which is found in red blood cells and is responsible for transporting oxygen to living cells. Other obvious examples are nutrient proteins such as ovalbumin and casein, required by our bodies for proper growth and development. Thus, a whole consortium of proteins, with various functions and degrees of importance, exist in the body.

The extraordinary diversity of protein function is due to the precise specificity of a given protein's interactions with molecules. Molecules have to "fit" into the protein, which requires a relatively rigid spatial structure of the protein. The structure of a protein determines its function through determining the molecules with which the protein interacts. As a consequence, obtaining protein structure is a high scientific priority. Currently, X-ray crystallography is the primary method to determine structure; this requires high quality protein crystals. The growth of such crystals from supersaturated solutions of protein in solvent depends sensitively on the initial conditions of the solution. Until relatively recently, finding the particular initial conditions requisite for optimal crystal nucleation from solution was a trial and error process. Considerable progress has been made in understanding the role of the initial conditions, however. It turns out that *metastable* fluid–fluid critical points play a key role in determining optimal crystal nucleation, as we discuss in Chapter 7 and elsewhere.

The condensation of globular proteins is also a crucial factor in certain human diseases. The completion of the human genome project has brought with it a huge inventory of information regarding identification of genes, and, in addition, has helped to usher a change in philosophy regarding our outlook on disease and its treatment. Scientists are increasingly considering disease at a molecular level in order to understand the causes and aid in the prevention of certain diseases. Given the abundance and diversity of proteins, it is not unreasonable to view them simultaneously as the cause and treatment of disease. Scientists have begun using

proteins to test how a person's system reacts when foreign proteins are injected into the body. This could be caused by genetic factors and may be involved in the development of such diseases as diabetes mellitus and hypertension. Recent studies on proteins in the body have also shed light on the causes of some diseases. It has been shown (see Chapter 13) that a genetic mutation in the hemoglobin molecule, HbS, is involved in sickle cell anemia. One study linked a liquid–liquid phase transition with the polymerization of the molecule, the precursor of the disease which ultimately gives the red blood cells their signature sickle shape. Other studies (see Chapter 12) have shown that genetic cataracts are also caused by protein crystallization in the eye lens, effectively clouding the transparency of the eye. Alzheimer's disease has also been linked to the crystallization of the protein molecule amyloid β protein (see Chapter 14).

Other uses of proteins stem from applications in the pharmaceutical industry. Protein crystals are being studied for use in vaccine delivery [8]. Drugs are also being designed to attach to protein sites of infected cells to deliver drugs. In another application of drug delivery, proteins themselves are being used to help combat disease such as hepatitis C. The protein molecule interferon, along with the help of a process dubbed pegylation, targets infected cells and delivers medicine to treat the patient. Protein crystals are also becoming useful in biotechnology. For example, the stabilization of enzymes as industrial catalysts involves crystallization of the enzyme followed by a subsequent cross-linking [9].

1.3 Types of proteins

The two major classes of proteins are the fibrous proteins and globular proteins. Fibrous proteins are abundant in cells and perform tasks that require each protein molecule to span a large distance. These have a relatively simple, elongated structure. We will not concern ourselves with this class of proteins in this book. Globular proteins are compact and approximately spherical in shape, with an irregular surface. They are by far the most numerous of cellular proteins and, unlike fibrous proteins, tend to be soluble in aqueous media. Globular proteins comprise most of the structures in the protein data bank. These proteins perform most of the chemical functions of the cell, including synthesis, transport, and body metabolism. Examples of this class include all the enzymes, albumin, globulin, casein, hemoglobin, and protein hormones. Hemoglobin is a respiratory protein contained in red blood cells and carries oxygen throughout the body. There are more than 100 different forms of human hemoglobin, including hemoglobin S, the cause of sickle cell anemia, which is the subject of Chapter 13.

Membrane proteins form a third class of proteins. These are proteins that are associated with the lipid bilayer of a cell membrane and carry out most of the

membrane functions. Many membrane proteins extend through the bilayer, with both hydrophobic and hydrophilic regions. Their hydrophobic regions lie in the interior of the bilayer, in contact with the hydrophobic tails of the lipid molecules. Their hydrophilic regions are exposed to the water environment on either side of the membrane. Other membrane proteins are located completely outside the bilayer, being attached to the bilayer only by one or more covalently attached lipid groups. Finally, others are attached to the membrane only relatively weakly, via non-covalent interactions with other membrane proteins. Chapter 11 deals with membrane proteins.

1.4 Protein crystallization

Crystallization is usually induced in the laboratory by adding salt, alcohol or polymer to dilute protein solutions [10]. These methods were developed from the first studies of crystallization, which took place over 150 years ago. Surprisingly, the pioneering methods such as dialysis of salt solution and the use of organic solutions as precipitating agents are still used today and comprise the basic tools of crystal growth. These early successes showed that protein crystals behave in much the same way as inorganic crystals and were used to provide “proof” of the purity of the sample from which the crystals were obtained. These also demonstrated that crystals could be obtained from solution. Figure 1.1 shows hemoglobin crystals, first successfully crystallized in 1840. With the advent of X-ray diffraction, and subsequently other techniques such as light scattering and NMR spectroscopy, protein crystals are providing the means by which scientists can determine the structure of proteins. Even with these techniques, however, good crystals that are free from defects are not easily produced. The basic conditions under which good quality (suitable for X-ray crystallography, for example) crystals can be obtained are not understood. Methods of growing crystals currently depend on trial and error tests to see under what conditions a protein solution will crystallize. Also, because each protein is different, conditions for which one particular solution will result in protein crystals do not result in crystals for another protein solution. To overcome these problems, crystallographers implement a “factorial” method. This is a brute force means in which all possible variants of initial conditions are examined. Another promising method that is currently being pursued involves the use of microfluid techniques.

Of the many factors that govern protein crystallization, it is known that a super-saturated solution promotes crystal growth. The most popular of the precipitates available to promote supersaturation is poly-ethanol glycol (PEG), a long flexible polymer chain. This enhances precipitation due to a depletion effect, discussed in Chapter 5. PEG can be grown with various lengths. Crystals grown out of

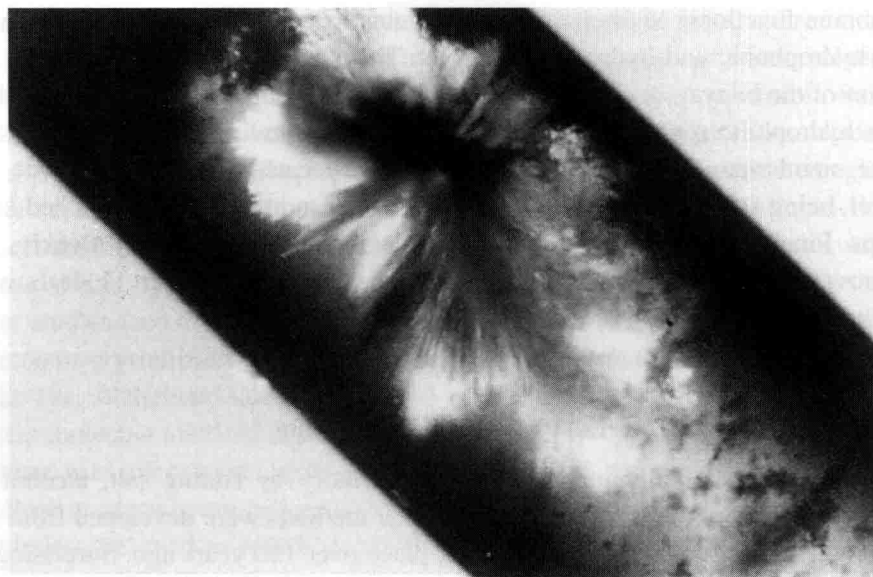


Figure 1.1. Fine needle crystals of hemoglobin, much like those grown by Hünfeld in 1840. Reprinted, with permission, from ref. [10]. For a color image, please see Plate Section.

solutions that contain PEG have been shown by X-ray structure analysis to be the same as those grown by traditional methods. An advantage of PEG over other precipitants is that most macromolecules crystallize within a fairly narrow range of PEG concentration. Recently, many studies have examined the influence of PEG on the phase diagrams of protein solution, as we discuss in later chapters.

1.5 Outline of book

This book is divided (roughly) into three parts. Chapters 2 to 5 review several topics relevant to protein discussion, including protein structure, experimental methods, thermodynamics and statistical mechanics, and protein–protein interactions. The subject of the interactions between protein molecules in solution is fundamental to our ability to calculate phase diagrams and non-equilibrium properties such as nucleation rates. Our theoretical understanding of these interactions, however, is still incomplete, so our discussion is incomplete. There is no doubt that further progress in understanding the role of the solvent (typically water, buffer, and precipitants such as salts or PEG) in determining the interactions between the protein molecules is crucial to our ability to understand and control the relevant equilibrium and kinetic properties of these protein solutions. This is clearly a major subject for future research.