KfK 5220B
Juli 1993

Solution X-ray Scattering Studies of Metalloproteins

G. Großmann
Institut für Toxikologie

Kernforschungszentrum Karlsruhe

Kernforschungszentrum Karlsruhe

Institut für Toxikologie

KfK 5220B

SOLUTION X-RAY SCATTERING STUDIES OF METALLOPROTEINS

Günter Großmann

von der Fakultät für Physik der Universität (TH) Karlsruhe
genehmigte Dissertation

Kernforschungszentrum Karlsruhe GmbH, Karlsruhe

Kernforschungszentrum Karlsruhe GmbH
Postfach 3640, 76021 Karlsruhe

ISSN 0303-4003

Solution X-ray Scattering Studies of Metalloproteins

Abstract. Proteins participate in a multitude of biological processes. The great diversity of proteins is correlated with the immense number of possibilities that exist for the sequence of amino acids, being the building blocks of proteins, and their three-dimensional association. Thus, knowledge of protein conformation, conformational changes or protein assembly is a major step towards understanding the relationship between structure and function of biological molecules. In view of the complexity of such a field, it is obvious that only combinations of techniques along with interdisciplinary co-operation will prove successful. The application of physical, in particular spectroscopic methods plays a significant role in present-day molecular biology.

Solution X-ray scattering, a technique using monochromatic X-rays which are scattered elastically, has been applied to investigate the conformation and conformational changes of proteins. This method is very well suited to elucidate the geometrical structure of macromolecules *in solution*. Its major advantage is the possibility to perform measurements under almost physiological conditions. However, X-ray scattering from such totally disorderd samples implies only low to medium resolution information and thus necessitates model calculations to interpret the experimental scattering results. A reliable and effective data interpretation is particularly possible if scattering pattern simulations are based upon available high resolution crystallographic structures of related proteins. These structures can serve as starting points for molecular modelling studies.

In the framework of this thesis two families of *metalloproteins* have been investigated. This class of proteins is characterized by one or a few metal ions among the several thousand atoms that constitute the molecule. The metal atoms affect critically the structural and functional properties of these proteins. *Transferrins* are the principal iron binding and transport proteins in vertebrates and thus of great importance in iron metabolism. Furthermore, due to their high affinity for iron, transferrins can retard or suppress microbial growth by depriving microorganisms of iron or making iron relatively unavailable. Besides iron, transferrins are capable of binding a great variety of other metals. They are therefore also of considerable toxicological interest. The family of *nitrite reductases*, including copper-containing enzymes, are found in a wide range of bacteria and take part in the denitrification pathway. They are key proteins in the stepwise reduction of nitrite to nitrous oxide and dinitrogen.

The analysis of the solution X-ray scattering data shows that all transferrins undergo a similar conformational change when iron is taken up and that *both* iron-binding halves (the so-called N- and C-lobe) exist in an open conformation when no metal ion is bound. This

observation demonstrates a clear difference between crystal structure and structure in solution. The conformational changes induced by metals other than iron (so-called non-physiological metals) depend strongly on the metal ion coordination. Moreover, the aspartic acid residue, an amino acid involved in metal binding, could be deduced to be a key residue in triggering the closed conformation. In the case of the copper-containing nitrite reductases, the combined approach of X-ray scattering and molecular modelling has allowed to define their trimeric structure in solution. In addition, scattering pattern simulations for the interpretation of experimental data could be improved by considering a hydration layer around the protein.

The experiments reported here have been carried out at the Daresbury Synchrotron Radiation Source (U.K.). The study of proteins under conditions close to physiological environment is desired (e.g. low protein concentrations of about 5 mg/ml were used). However, in view of the relatively weak X-ray scattering arising from these molecules, the presented experiments would be hardly if at all possible with any conventional X-ray source. The use of highly intense and collimated synchrotron radiation allowed to record statistically significant data going beyond the traditional small-angle range. It is precisely this extended scattering region that enables the detection of conformational changes due to different subunit (domain) arrangements in proteins. The conformation of domains plays an important role for both the interpretation of functional aspects and the classification of proteins.

Röntgen-Kleinwinkelstreuung an Metall-Proteinen in Lösung

Zusammenfassung. Eine große Zahl biologischer Prozesse wird von Proteinen getragen. Ihre Vielfalt steht in engem Zusammenhang mit den zahlreichen Möglichkeiten der räumlichen Anordnung der Protein-Bausteine, den Aminosäuren. Ein bedeutender Teil der molekularbiologischen Forschung befaßt sich gegenwärtig mit dem Studium der dreidimensionalen Struktur von Proteinen, um über die Kenntnis ihrer Konformation und ihrer Konformationsänderungen als auch über deren Aggregation mit anderen Protein-Untereinheiten Rückschlüsse auf biologische Funktionen ziehen zu können. Die Komplexität dieses Arbeitsgebietes erfordert interdisziplinäre Zusammenarbeit. Physikalische, insbesondere spektroskopische Techniken zur Strukturbestimmung nehmen dabei einen gewichtigen Rang ein.

Zur Untersuchung der Konformation sowie der Konformationsänderungen von Proteinen wurde im Rahmen der hier vorgestellten Untersuchungen die Methode der *Kleinwinkelstreuung* angewandt. Es handelt sich dabei um eine Technik zur Ermittlung der geometrischen Struktur von Makromolekülen in wässriger Lösung mit Hilfe elastischer Streuung monochromatischer Röntgenstrahlung. Ein besonderer Vorteil dieser Methode besteht darin, daß die Messungen unter angenähert physiologischen Bedingungen durchgeführt werden können. Allerdings bedingt die damit verbundene Unordnung der Moleküle in wässriger Lösung eine geringere räumliche Auflösung, so daß Modellrechnungen zur Analyse der Streuexperimente unerlässlich sind. Sie müssen auf der Grundlage bereits bekannter Information vorgenommen werden, erst dann ist eine schlüssige Dateninterpretation möglich. Die Röntgen-Kleinwinkelstreuung kann besonders wirkungsvoll eingesetzt werden, wenn Atomkoordinaten aus kristallographischen Untersuchungen verwandter Proteine verfügbar sind. Diese können als Ausgangspunkt für molekulare Modellierungsstudien dienen.

In der vorliegenden Arbeit werden zwei Familien von *Metall-Proteinen* vorgestellt. Es sind Proteine, die infolge der Bindung eines oder weniger Metallionen entscheidend in ihrer Struktur und Funktion beeinflußt werden. Die Familie der *Transferrine* erfüllt in Wirbeltieren die wichtige Aufgabe des Eisentransports. Sie ist daher für den Eisenmetabolismus von tragender Bedeutung. Transferrine haben darüberhinaus infolge der Fähigkeit zur spezifischen Eisenbindung die Funktion, schädlichen Mikroorganismen Eisen vorzuenthalten oder zu entziehen, um deren Wachstum zu verlangsamen oder zu unterdrücken. Angesichts der Eigenschaft, neben Eisen noch eine Vielzahl anderer Metalle zu binden, sind die Transferrine für die Toxikologie von außerordentlichem Interesse. Die Familie der *Nitrit Reduktasen*, zu denen Kupfer bindende Proteine gehören, bilden ein

Schlüsselglied in der Enzymkette zur schrittweisen Reduktion von Nitrat zu Stickstoff, d.h. sie sind maßgeblich am Respirationsprozeß (Stickstoff-Zyklus) von Bakterien beteiligt.

Die Analyse der gewonnenen Daten zeigt, daß alle Transferrine nach Eisenbeladung ähnliche Konformationsänderungen eingehen. Zudem konnte im Gegensatz zu kristallographischen Ergebnissen festgestellt werden, daß in physiologischer Umgebung *beide* Eisenbindungsstellen (im sogenannten N- und C-Lobe) vor der Aufnahme des Metallions durch eine offene Konformation charakterisiert sind. Die Experimente mit nichtphysiologischen Metallen ergaben eine deutliche Abhängigkeit der Konformationsänderung von der Koordinationszahl des gebundenen Metallions. Außerdem konnte die Schlüsselrolle einer Aminosäure, die an der Metallbindung beteiligt ist und maßgeblich zur geschlossenen Konformation beiträgt, aufgezeigt werden. Im Falle der Kupfer bindenden Nitrit Reduktasen konnte die Trimer-Struktur des Enzyms in wässriger Lösung festgelegt werden. Die Modellrechnungen zur Interpretation der experimentellen Streudaten lassen sich zudem durch Einbeziehung einer das Protein umgebenden Hydrathülle wesentlich verbessern.

Die in dieser Arbeit beschriebenen Experimente wurden am Synchrotron Radiation Laboratory in Daresbury (U.K.) durchgeführt. Röntgenstrahlen werden an Proteinen grundsätzlich nur äußerst schwach gestreut. Darüberhinaus sollten in den Messungen nahezu physiologische Bedingungen erreicht werden (Proteinkonzentrationen von etwa 5 mg/ml). Die hier beschriebenen Untersuchungen sind daher mit konventionellen Röntgenquellen nur schwerlich, wenn überhaupt möglich. Die hohe Intensität und die gegebene Bündelung der benutzten Synchrotronstrahlung erlauben hingegen die Aufnahme statistisch bedeutsamer Streudaten bis über den sogenannten Kleinwinkelstreubereich hinaus. Es ist gerade dieser auf einige Winkelgrade ausgedehnte Streubereich, der Konformationsänderungen von Unterstrukturen, der sogenannten Domänen der Proteine, erkennbar macht. Die Domänenanordnung in Proteinen ist von entscheidender Bedeutung. Sie kann sowohl zur Interpretation spezifischer Funktionen als auch zur Klassifizierung von Proteinen herangezogen werden.

TABLE OF CONTENTS

Abstract iii

Zusammenfassung v

Table of Contents vii

Chapter 1 : Introduction 1

Chapter 2 : Metalloproteins 3

2.1 Protein functions and structural principles 3
2.2 Flexibility, rigidity and assembly 4
2.3 The role of metal ions 5
2.4 Example 1
Transferrin 6
2.5 Example 2
Nitrite Reductase 10

Chapter 3 : Solution X-ray Scattering 14

3.1 Introduction 14
3.2 General principles 15
3.2.1 Radius of gyration 17
3.2.2 Distance distribution function 17
3.3 Interpretation of scattering profiles 18
3.4 Molecular Modelling 21

Chapter 4 : Experimental 24

4.1 Synchrotron Radiation Sources 24
4.2 Bending Magnet Radiation 25
4.3 X-ray Scattering Station 8.2 at Daresbury 27
4.3.1 Camera system 27
4.3.2 Detection system 28
4.4 Sample Preparation 30
4.4.1 Transferrin 30
4.4.2 Nitrite Reductase 31
4.5 Experimental Protocol 32

Chapter 5 : Data Reduction and Analysis 34

5.1 Normalization 34
5.2 Check on Radiation Damage 34
5.3 Background Subtraction 34
5.4 Division by Detector Response 36
5.5 Removal of TAC hole 36
5.6 Calibration 37
5.7 Radius of Gyration 38
5.8 Distance Distribution Function 39
5.9 Scattering Pattern Calculation and Computer Modelling 40

Chapter 6 : Results and Discussion 42

6.1 Transferrin 42
6.1.1 X-ray scattering from intact molecules 42
6.1.2 X-ray scattering from N- and C-terminal fragments 46
6.1.3 Computer modelling and simulation 47
6.1.4 Transferrins loaded with other non-physiological metals 53
6.1.5 Studies of mutants of the N-terminal fragment 58
6.2 Nitrite Reductase 65
6.2.1 Experimental results 66
6.2.2 Model building 67
6.2.3 Which model is consistent with the scattering results ? 69

Chapter 7: Conclusion and Future Objectives 74

Appendix A : Theoretical Concept of Molecular Model-Building 80

Appendix B : Energy Minimization and Hydration 84

B.1 Computational procedure 85
B.2 Effect of protein on water 86

References 88

Chapter 1 : Introduction

Many proteins that play a crucial role in living systems contain one or a few metal atoms among the several thousand atoms that constitute the biomolecule. These so-called metalloproteins form a large fraction (between one-fourth to one-third) of all known proteins. The metal atoms participate directly in the function of the protein and are therefore among the most interesting parts of the structure to study. The main goal of present-day research in molecular biology is the understanding of the relationship between structure and function of biological molecules such as metalloproteins. In view of the complexity of processes in biological systems, it is obvious that only combinations of techniques along with interdisciplinary co-operation will prove successful. The application of physical, in particular spectroscopic methods plays an outstanding role in these endeavours.

At present, X-ray crystallography [1, 2] and multidimensional NMR in solution [3] are the major experimental techniques used to develop models of biological macromolecules on atomic scale. NMR is a relatively young method compared with X-ray crystallography and is restricted to rather small molecules ($\leq$ 20 000 D[1]). X-ray diffraction from single-crystals remains *the* indisputable method for obtaining high resolution (< 2.5 Å) structures of proteins with molecular mass up to millions of Dalton. However, for such resolution, crystals with a high degree of static long-range order in the arrangement of the molecules are a prerequisite. Furthermore non-trivial questions concerning phase determination and consistency between crystal and solution structures have to be solved.

X-ray scattering offers itself as a further promising method to investigate the structure of biological macromolecules in solution, i.e. in an environment close to physiological conditions [4, 5]. In this case there is no, or very little order in the arrangement of the molecules, since a disordered state will be considered. It is thus a technique of low to medium resolution (> 10 Å). This fact seems to make the method of solution X-ray scattering less attractive, since the structure of a biological system cannot be defined at a sufficient resolution with confidence. But it will be shown that solution X-ray scattering in combination with molecular modelling is able to contribute to the understanding of structure and function of proteins in general and metalloproteins in particular. It will be reported on the application of solution X-ray scattering to two species of metalloproteins :

- the *transferrins*, as the dominant class of iron binding and transport proteins in vertebrates and
- the *nitrite reductases*, proteins involved in the biological electron transfer mechanism of denitrifying bacteria.

---

[1] Throughout the thesis Dalton (D) as atomic mass unit (1 D = $1.6605 \cdot 10^{-27}$ kg) and Ångstrœm (Å) as unit of length (1 Å = 10^{-10} m) will be used.

The major advancement in an understanding of the structure and function of the studied metalloproteins is primarily attributed to

- the use of highly collimated X-rays from a synchrotron radiation source (here: SRS Daresbury, U.K.) including high quality optics, X-ray detectors and efficient data aquisition systems,

and

- the combination of molecular modelling and computer simulations for data interpretation using crystal structures from related protein species as the starting model.

The next two chapters should actually be regarded as a continuation of the introduction so as to describe the biological (protein samples, Chapter 2) and physical aspects (principles of solution X-ray scattering and moleclar modelling, Chapter 3) of the present work. Chapter 4 is dedicated to the experimental section, where some details for the generation of synchrotron radiation and X-ray scattering data collection are given. Chapter 5 will shed light on data reduction and analysis. In Chapter 6, X-ray scattering results for the two different biological molecules are presented and their significance is discussed. A conclusion concerning continuing experiments based on the samples studied here will complete this thesis. Some indication is given for future applications of solution X-ray scattering experiments in the context of structural studies in biology.

A general remark should be added : The study of the interaction of various types of biological matter is becoming an increasingly important branch. Certainly, any results arising therefrom will have far-reaching applications and significant impacts on biotechnology and medicine. Thus the riches of structural biology pose enormous challenges on scientists of different disciplines. The exploration of the three-dimensional structure determines the chemical and physical behaviour and an understanding of the mechanistic basis of the resulting function of biomolecules. It also provides an insight into life on a molecular level. The progress in structure determination, of course, goes hand in hand with biochemical developments such as protein purification, preparation and modification, also genetic engineering has to be mentioned in this context as a novel technique of great importance.

Chapter 2 : Metalloproteins

The purpose of this chapter is not to present a comprehensive overview on proteins, but to give a brief insight into the interesting and important link between the conformation of a protein and its function and to introduce some technical terms. Two metalloproteins have been studied in the framework of this thesis and will be explained in more detail. The methods developed in this work are not restricted to only this class of proteins but are of wider applicability.

2.1 Protein functions and structural principles

Proteins are the most abundant macromolecules in cells, constituting more than 50% of cellular dry weight. Being the product of genes, proteins are the instruments by which the genetic 'machinery' implements cell functions. They are the most versatile of the macromolecules, e.g. they

- catalyze the chemical reactions of metabolism (enzymes),
- transport and exchange nutrient and waste material,
- communicate information between cells (hormones),
- function in the immune system of defense against foreign cells (antibodies),
- determine cell association and adhesion,
- synthesize structural elements or act as structural elements themselves.

The building blocks of proteins are the amino acids which are linearly linked to one another in unique sequences forming non-branched chains (polypeptide chains). Twenty different amino acids are found in proteins, each with a specific side group. It is the chemical properties of these side groups and in particular their non-covalent interactions (exception: the covalent disulphide bonds) which determine the conformation of a protein (see also Appendix A). The number and sequence of amino acids in a polypeptide chain is described by the *primary structure*. Hydrogen bonds, the most specific of the 'weak' non-covalent bonds can lead either to contracted helical (α-helix) or sheetlike (β-sheet) configurations, known as *secondary structure*. These regular regions along with chain segments that show no simple regularity produce a complex, quite often, globular shape, the three-dimensional form of the protein, called the *tertiary structure*. It represents the energetically most favourable arrangement of the polypeptide chain. To complete the picture, large single-chain proteins (monomers) are usually composed of compact folded units, so-called *domains*, that appear separate from the rest and are often stable in solution on their own. Most of the domains or single peptide chains contain between 150 and 250 amino acid residues [6]. This structural level of macromolecules is termed *quarternary structure* and

concerns also interactions by which two or more polypeptide chains associate in a specific manner (oligomers) to form biologically active proteins.

The various properties of globular proteins[2] appear to arise from their modular construction, i.e. the quarternary structure is of critical importance to the proper biological functioning. Moreover the flexibility of parts of a protein which arrange or move substantially are particularly interesting in terms of the requirements for protein activity. The conformational changes and arrangements will be discussed here in more detail since the method of X-ray solution scattering provides an outstanding tool to detect different structural configurations of a protein. Nevertheless, no technique can provide a full description on its own. Although crystallographic results contribute immensely to this subject of structural, molecular biology, different techniques are essential to complete the view of such ingenious molecules as proteins, the building blocks of life.

2.2 Flexibility, rigidity and assembly

It is a thermodynamic requirement that molecules of the size of proteins show substantial transient fluctuations. Protein flexibility involves movements of widely varying magnitudes, ranging from small-scale movements or vibrations of bond lengths and angles or groups of atoms (at the level of amino acid side chains, e.g. ring flipping) up to the other end of the spectrum, where individual domains of large proteins undergo large-scale transitions between different conformations. Here, 'large' means that the motion involves at least 15% of the mass of the protein where single atoms move condsiderably ($\geq$ 10 Å). The phenomenon of protein flexibility is many-sided and complicated, consequently the term *flexibility* cannot be well defined. A detailed review on structural and functional aspects of protein flexibility including selective examples was given by Bennet & Huber (1984) [7] and Huber (1987) [8]. Here only the rigid-body movement of large segments of a molecule relative to one another (domain or hinge flexibility) and the geometric arrangement of subunits (assembly process) will be considered. Two examples of these categories have been studied experimentally during this work.

Domain flexibility and subunit assembly have clear biological functions. Catalytic properties and specifity of proteins (enzymes) are determined by the chemical groups in the region of the protein called the *active site*. The active site, which is often found in a cleft between two domains of a molecule or at the interface between two molecules, has the purpose of substrate binding and/or catalytic action.

The process of substrate (*ligand*) binding to proteins often results in a substantial change in the conformation, i.e. quarternary structure, which alters a protein's activity

[2] Two classes of proteins exist: The often laterally crosslinked, fibrous and water-unsoluble proteins (e.g. α-keratin of hair and skin or collagen of tendons) and the biologically active and water-soluble proteins which have globular and compact structures (e.g. enzymes). Only the latter class will be considered here.

critically. These ligand-induced conformational changes are usually limited to short segments or single peptide units, called hinges, that connect domains changing their relative orientation to each other as rigid bodies. It appears plausible that the trigger for such conformational transitions comes from the interaction of the substrate with a single amino acid occupying a key position. Then the final conformation is stabilized by a multitude of substrate-protein or newly formed protein-protein interactions. In this respect it should be mentioned that already the modification (*mutation*) of single amino-acids (either in naturally occurring variants or engineered by protein chemistry, a technique which will provide a wealth of new information concerning protein structure and function) is able to affect conformational changes considerably [9].

Quartenary structure changes can also occur spontaneously as aggregation of separate molecules. It is the principle of self-assembly that opperates in order to build larger (e.g. oligomeric proteins) or even more complicated structures (e.g. cell membranes or viruses). The development of favourable contacts between constituent molecules drives the association and affects, that segments exposed in the isolated molecules, become buried in the multisubunit aggregate. This drastic change of environment will again alter the biological properties.

Finally it should be mentioned, that proteins are macromolecules in an aqueous environment, thus the interactions with the solvent are of vital importance and increase the complexity of the system. Although the unique configuration in proteins (nonpolar side groups are isolated into internal regions and polar groups are more abundant in surface areas) prevents water from penetrating and converts newly made polypeptide chains into compact shapes during the folding process. The final protein conformation and its activity strongly depends on hydration and water accessible areas around the protein surface.

2.3 The role of metal ions

Metal ions are of particular importance in protein biochemistry. From about 25 elements, that have been recognised as essential and indispensible to life, 15 are metals. At least five of them (Na, K, Mg, Ca and Fe) appear to be crucial to every known form of life, whereas the others (V, Cr, Mn, Co Ni, Cu, Zn, Mo, W and Se) are required only in small amounts by some living organisms [10, 11]. The participation of metal ions is essential in fundamental biological processes such as electron storage and transfer, dioxygen binding, storage and activation, and substrate transport, catalysis and activation. The following roles of metal ions on a molecular level can be specified in connection with protein activity and conformation. They are

- integral to the protein structure,
- involved in determining and stabilizing the structure of a protein,

- part of the active site of enzymes,
- complexed to a substrate or being the substrate themselves.

The metalloproteins are of major interest as they perform largely varying biological functions. The metal sites in these proteins have attracted much interest recently. The extremely elaborate coordination complexes and the relationships which exist between these sites and synthetic model complexes have led to the emergence of the interdisciplinary field of bioinorganic chemistry. The detailed knowledge of coordination chemistry will improve the understanding of the role of metal ions in biological systems. In biological molecules such as proteins the metal ion is often coordinated by oxygen, sulfur, or nitrogen atoms. However, the presence of metal ions does not only determine and affect the environment of the binding site (or active site) but can also influence decisively the conformation, i.e the tertiary and quarternary structure and the stability of proteins.

2.4 Example 1 : Transferrin

Iron is one of the most widespread metals in living systems, it participates in a variety of biological activities. Thus a large amount of iron is required (e.g. the concentration is normally 40-50 mg Fe/kg in the human body). However, there are two major chemical problems associated with the biological use of iron : the poor solubility of iron at physiological pH values and the involvement of iron in the production of extremely reactive radicals of potential toxicity. These complications have led to the evolution of ingenious protein-dependent systems for storage and transport of iron. *Transferrin* is the protein which complexes and transports iron in vertebrates. Apart from its traditional role in iron metabolism (transport and delivery of iron to iron-requiring cells by means of receptor-mediated endocytosis), transferrin acts as inhibitor of bacterial growth (conservation and control of iron levels in body fluids). This versatility is subject of various medical, biochemical and spectroscopical studies so as to understand its functional properties in relation with structural peculiarities (see e.g. Refs. [12-15] for reviews on the physical chemistry and biology of transferrins).

Transferrin is the general term for a family of monomeric and evolutionarily-related glycoproteins with the property of reversibly binding ferric iron (Fe^{3+}). The primary members of the transferrin family are

- *serum transferrin* (ST) in blood plasma and other extracellular fluids,
- *ovotransferrin* (OT), in egg white,
- *lactoferrin* (LF), in milk and other secretory fluids,
- *melanotransferrin* (MT), in the plasma membrane of melanoma cells.

Amino acid sequence determination of a number of transferrins provided clear signs for structural similarities: There is an extensive sequence homology (identical amino acid residues) not only among transferrin species (ranging from about 50% to 60% identical residues in corresponding positions [16]), but also between the N- and C-terminal halves of the molecule. Thus the protein consists of two similar lobes (the so-called *N-* and *C-lobe*) each of which contains one iron binding site. Some structural properties have been summarised in table 2.1 referring to the three transferrin species, in particular, which have been studied in this work.

Recent crystallographic structure analyses of a number of transferrins (see refs. [24-29]) have contributed decisively to the understanding of the structural organization of these proteins. Therefore the polypeptide folding pattern for all transferrins can be characterized by two separate globular lobes of roughly 330 residues, connected by a short bridging peptide. Each lobe is comprised of two dissimilar domains (NI, NII and CI, CII) defining a cleft for the metal binding site at the domain interface (figure 2.1). This interdomain cleft represents a common motif in proteins, facilitating the reversible binding of a substrate [7].

Table 2.1 : Comparison of some structural and compositional properties of three transferrin species.

	Human Serum Transferrin (HST)	Human Lactoferrin (HLF)	Chicken Ovo-transferrin (COT)
molecular mass*	75 700 D	77 200 D	76 800 D
no. of amino acids	679	691	686
N-terminal half (N-lobe)	1 - 332	1 - 333	1 - 332
C-terminal half (C-lobe)	338 - 679	345 - 691	342 - 686
homology between lobes (identical residues)	43 %	37 %	35 %
disulphide bridges			
in N-lobe	8	6	6
in C-lobe	11	10	9
carbohydrate properties#			
molecular mass	4400 D	4700 D	1900 D
no. of glycans	2	2	1
attachment	C-lobe	N- and C-lobe	C-lobe
iron-binding strength¶,‡	1000	261 000	4000
isoelectric point§,†			
apo protein	6.1	9.2	6.7
diferric protein	5.5	8.2	5.8

* calculated from the amino acid composition; # see ref. [17]; ¶ calculated from equilibrium dialysis data, given is the constant K_1 for the binding of the first Fe^{3+} to the corresponding apo-protein; ‡ see refs. [18-20]; § defined by the pH value at which the net charge on the protein is zero; † see refs. [21-23]

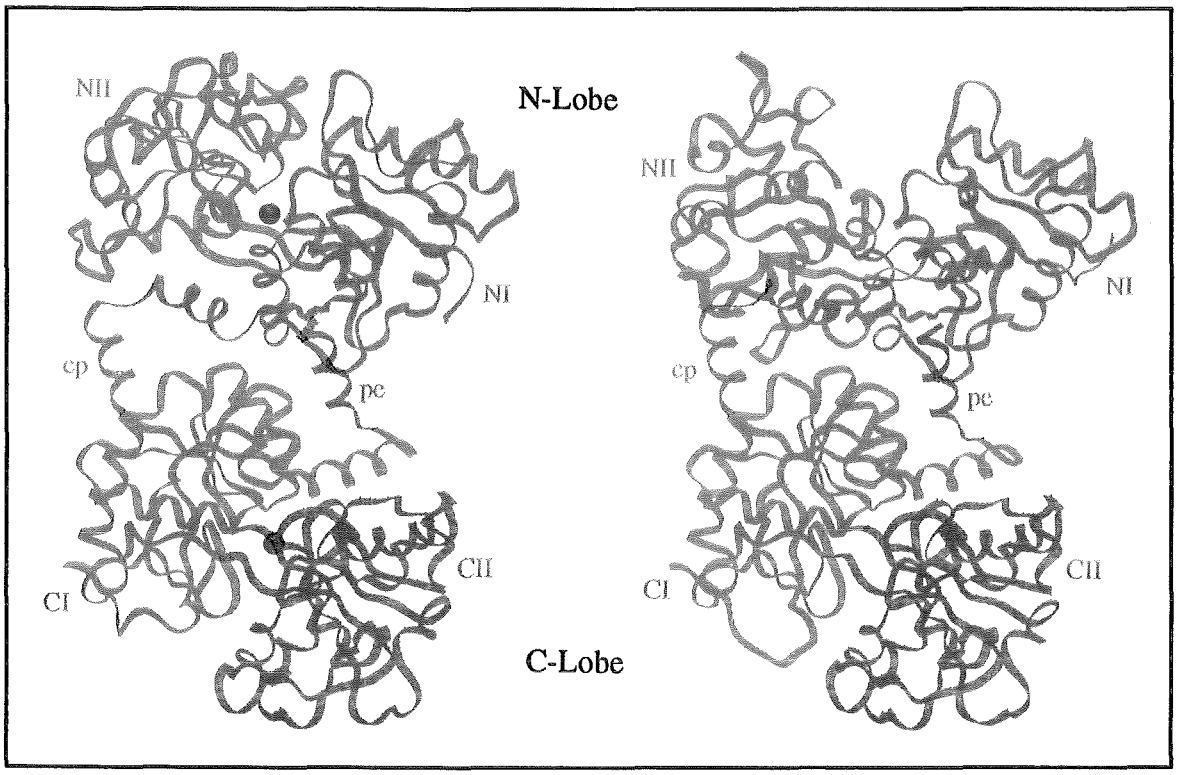


Figure 2.1 :
Ribbon drawings[3] of the molecular organization and conformation of the transferrins. The structures of Fe_2-HLF [25] (on the left) and apo-HLF [26] (on the right) show the characteristic folding into two lobes (N-lobe and C-lobe) joined by a connecting peptide (cp) and four domains (NI, NII, CI and CII). A helix (pe) forms the C-terminal end of the polypeptide chain. Iron atoms are represented as red spheres. The conformational difference between iron-loaded and iron-free state of parts of the N-terminal half is clearly visible.

With the determination of the crystal structure of human apolactoferrin [26], the metal-free state of lactoferrin, a major insight into the substantial conformational change accompanying iron binding was provided. In comparison with diferric lactoferrin, it turned out that the NII domain of the molecule follows a rotation of ~54° relative to the NI domain about a hinge at the back of the iron-binding site. This large-scale conformational change causes the interdomain cleft in the N-Lobe to open wide. Curiously, in this crystal structure representation the structure of apo-HLF revealed that in the C-lobe the iron binding cleft is closed even though no metal iron is present (see figure 2.1).

Although crystallographic findings shed light on a number of open questions regarding the structure-function relationship in the transferrins, so far only limited structural information is available on the conformation of these macromolecules in solution [30-32]. Further studies of the interdomain interactions in the various members of the transferrin family will

[3] The ribbon representation of a structure shows the smoothed path of the polypeptide chain. Important parts of a molecule such as helices are refelcted clearly.

lead to additional structural explanations related to the different tasks of the transferrins. Moreover the crystallographic difference between the two lobes when no metal is bound, represents a surprising feature.

Being glycoproteins, one or more carbohydrate units (sugars) are attached covalently to the transferrins. The carbohydrate chains (*glycans*) are specific for each transferrin and even specific to different species. Glycosylation sites vary in number (see also table 2.1) and are widely distributed over the molecular surface. Crystallographic studies revealed little evidence of defined carbohydrate structure presumably due to its high mobility and spatial disorder. These results together with the fact that the removal of the carbohydrate does not affect the physiological functions, may argue against any structural or functional role for transferrins [33].

The understanding of the tight binding of two Fe^{3+} ions per molecule which is strong enough to resist hydrolysis gained substantially from crystallographic structure determinations. Concomitant with the metal binding of each iron atom, a carbonate anion (CO_3^{2-}) is bound and serves as a fundamental bridging ligand between the metal and the protein. This so-called synergistic anion along with four amino acid residues provide the unique feature of iron-binding in an octahedral geometry (figure 2.2). In addition, unlike other metalloproteins where the metal-binding ligands are contributed by a rather small segment of the

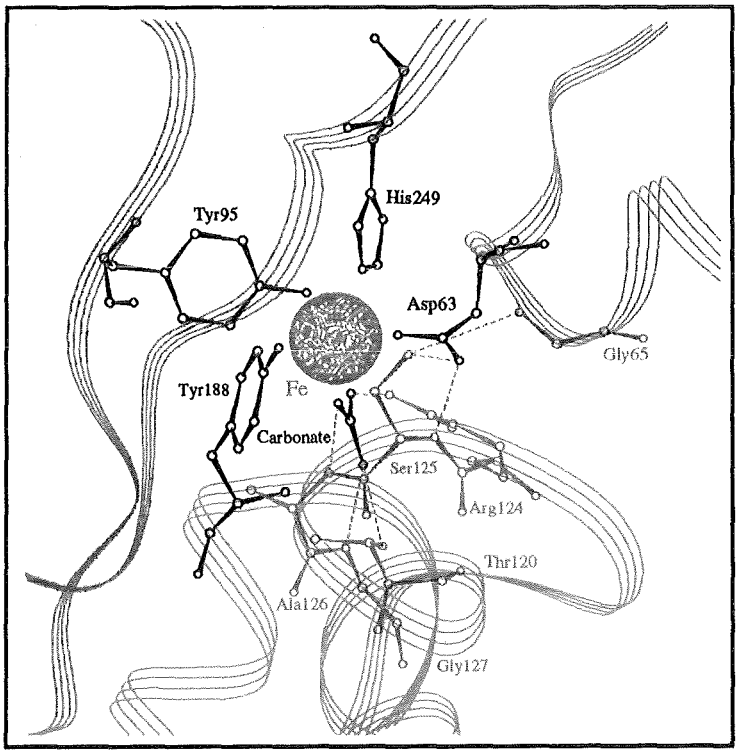


Figure 2.2 :
A view of the iron binding site of the N-lobe of rabbit serum transferrin [29]. Fe^{3+} is bound by four protein ligands (Asp63, Tyr95, Tyr188, His249) and by a carbonate anion in a bidentate fashion, thus providing an octahedral geometry at the iron site. This ligand arrangement (Asp63 is part of domain NI (green ribbon trace), Tyr188 of domain NII (blue ribbon trace), Tyr95 and His249 come (dark blue ribbon trace) from each of the two backbone strands connecting the two domains at the back of the iron site) forms an elegant compensation of the metal ion in concert with the carbonate anion. Hydrogen bonds are represented as green dashed lines and highlight both the anchoring of the carbonate as well as the network of interdomain hydrogen bonds. In this respect the aspartic acid (residue number 63 in human serum transferrin numbering) is not only a ligand to iron but also guarantees the interdomain stability.

polypeptide chain (e.g. in copper proteins [34], calcium-binding [35] or zinc-binding proteins [36]), the four ligating residues in transferrins are widely separated in the amino acid sequence and belong to distinct domains of the protein [37].

Transferrins can bind many other metal ions, including as well the transition metals as many lanthanides and actinides (reviewed in [12] and [13]). It is this property which attributes the biomedical relevance to this serum protein. For instance, the use of Ga^{2+} or In^{2+} labelled transferrins for medical imaging or the identification of transferrins as a carrier of Al^{3+} or heavy metals with the possible implication for toxicity and destructive power to living cells, show the role of transferrins in the biodistribution of markers or impurities in organisms. In this respect the structural information so far available has also raised further questions : If transferrin interacts specifically with a variety of other metal ions, generally with a high positive charge, how does the binding and the likely concomitant structural change of one metallic cation compare with another? What is the trigger for the conformational change after metal uptake and are there metals or other effects which might reduce or even inhibit such a structural change? Here the strongly developing field of protein engineering can also provide outstanding possibilities to study the peculiarities of the characteristic metal binding site in transferrins. Furthermore transferrin binding to specific cell receptors (glycoproteins embedded in the cell membrane) is an essential question with respect to iron delivery to certain tissues and might be related to a particular conformation or structural feature.

Many of these questions cannot be answered promptly by X-ray crystallographic studies. In addition not all modifications of transferrin molecules (e.g. any replacement of iron by other metals or site-specific mutants of transferrin) can be induced to form crystals suitable for X-ray structural analysis. Thus X-ray solution scattering provides a unique tool by studying the samples under conditions close to physiological environment and due to known overall structures of the molecule, models can be built 'easily' to interpret the scattering data.

2.5 Example 2 : Nitrite Reductase

Electron transfer over considerable distances with high specificity and efficiency is central to a large number of biological processes. The reduction and oxidation of key molecules are mediated by a diversity of proteins associated with soluble components of both aerobic and anaerobic energy-generating systems. Copper-containing proteins play an essential role in substrate activation as well as being the redox centre of the electron donors or acceptors. For instance in the global nitrogen cycle, i.e. in the process of denitrification, bacteria use NO_3^- and NO_2^- as terminal electron acceptors to produce gaseous nitrogenous products (NO, N_2O, and N_2). During this course of reduction various copper-containing enzymes are

involved. Understanding the denitrification pathway on a molecular level not only reveals fundamentals of the biological electron transfer mechanism but can also help to gain control of the environmental and health hazard caused by the present-day nitrate pollution. The improper application of fertilizer with a high nitrate content has tremendous negative consequences on the underground water, natural locations, and forest ecosystem. An efficient process of nitrate removal was reported only recently based upon the exploitation of denitrification enzymes which have been immobilized on a solid substrate in order to construct a bioreactor capable of converting nitrate to gaseous dinitrogen [38].

Nitrite Reductase (NiR) is a key enzyme in the stepwise reduction of nitrate. NiRs are classified in copper- and iron-containing enzymes [39, 40] catalyzing the reduction of nitrite ions (NO_2^-) to nitric oxide (*NO*). Many copper-containing NiRs have been purified and characterized from a variety of bacteria and these enzymes show considerable variation in their physicochemical properties such as molecular weight, number of subunits, and number and types of copper centre that they contain. Table 2.2 gives some structural details concerning the copper-containing NiRs of *Achromobacter cycloclastes* (AcNiR), *Achromobacter xylosoixidans* (AxNiR) and *Alcaligenes faecalis* (AfNir), which will be of interest here. The information obtained with standard biochemical techniques is at variance and clearly suggests structural differences among the different bacterial enzymes. However, in all NiRs the enzyme-bound copper ions play an essential role in the catalytic, nitrite reducing activitiy. Decreasing the copper content either by dialysis of the enzyme against cyanide or by use of copper chelating agents, a significant decrease of activities can be observed.

The Cu^{2+} ions in copper proteins have been classified according to characteristics in their electron spin resonance (ESR) spectra. Accordingly, a narrow and sharp hyperfine splitting in the ESR spectrum characterizes the so-called *type I* Cu^{2+}, whereas the typical *type II*

Table 2.2 : Comparison of structural properties of three copper-containing Nitrite Reductases based upon biochemical and spectroscopical investigations.

	colour	molecular mass	subunit composition	Cu content[&] and types[§]	reference
AfNiR	green	110 kD*#	tetramer	4.5 (I + II)	[41, 42]
AcNiR	green	69 kD*	dimer	2 (I + II)	[43-45]
AxNiR	blue	149 kD#	tetramer		[46]
		70 kD*	dimer	1.6 (only I)	[47]

* by gel filtration, # by sedimentation equilibrium centrifugation, & by atomic absorption spectroscopy, § based upon ESR spectra

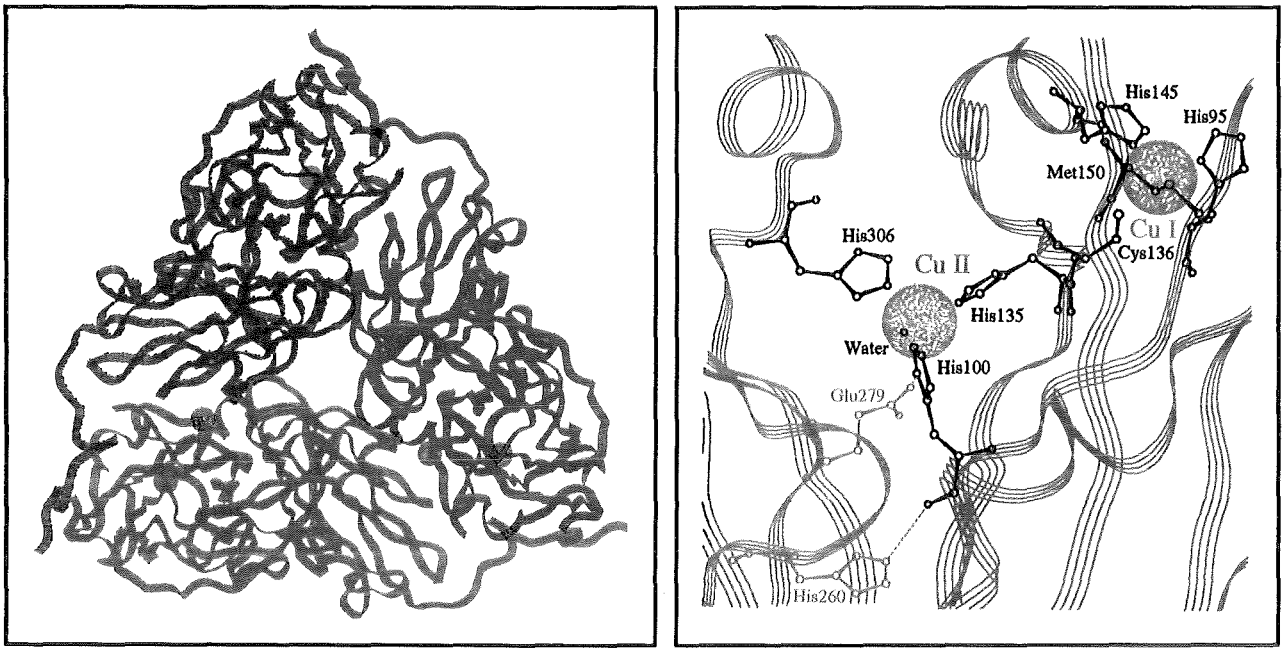


Figure 2.3 :
Ribbon drawing of the trimeric assembly of AcNiR (on the left) and the ligand arrangement around the type I (blue sphere) and type II (pink sphere) copper sites (on the right) based on the 2.3 Å X-ray structure [50]. Hydrogen bonds at the monomer-monomer interface are shown as green dashed lines.

Cu^{2+} exhibits a broader hyperfine splitting [48]. This difference is reflected not only in the atomic absorption spectra but also in the geometric arrangement and type of the copper ligands. A type I copper centre is characteristically ligated by two histidine (His), one methionine (Met) and one cysteine (Cys) residue in a tetrahedral geometry. The strong absorption bands around 450 and 600 nm is mainly associated with the Cu-S (Cys) chromophore. In a type II site, the copper is predominantly bound by three to four histidine ligands.

Since copper ions (Cu^{2+} and Cu^{+}) prefer different ligand coordination, the binding or release of an electron would be expected to produce a change in the protein structure or, conversely, the orientation and position of the coordinating ligands could define either a Cu^{2+} or Cu^{+} state. Thus the effects of copper binding on the protein and the effect of the protein on copper affinity are linked functions. Regarding the incompatible results of subunit composition and copper content, a variation in the activities of these enzymes should be realized. However, all copper-containing NiRs can catalyze the reduction of nitrite to nitric oxide and the necessity of both types of copper being present seems more and more established. In contrast to the previous work, the recent determination of the copper content in AxNiR supports the presence of both type I and type II copper centres in approximately equal proportions [49].

Furthermore the structural insight provided by the crystal structure determination of AcNiR [50] revealed unexpected details about the molecular assembly, copper types and ligation. Surprisingly, the crystals of this enzyme contained *trimeric* molecules (figure 2.3). Since the degree of association in the crystalline state might be biased by crystal packing forces and crystallisation conditions (e.g. high sample concentration), so as to favour trimer formation, a reinvestigation of the molecular weight by sedimentation equilibrium centrifugation gave also clear evidence of a trimeric species in solution [50]. Three obvious kinds of interactions can be specified that would stabilize the trimeric structure in solution :

- The type II copper site is bound by residues not within a single monomer but from each of two monomers of the trimer (see figure 2.3).
- Extensive intermonomer contacts are situated in the vicinity of the copper II site.
- A polypeptide extension in the form of a long arm reaches from one monomer to another monomer in the trimer.

Although these structural details do not support a trimeric association specifically, recent structure determinations of unrelated proteins [51, 52] show that this principle of trimer aggregation is an effective but simple feature for the crucial formation of catalytic sites. In addition, this kind of subunit assembly including polypeptide segments which project away from globular parts of one monomer seems to stabilize and approach a structure which is known from some membrane bound proteins forming diffusion channels within pores of cell walls [53].

Regarding the conflicting results with respect to molecular weight and subunit organization not only among the different species of NiR, the oligomeric conformation of this enzyme in solution is still an open question. On the basis of the known atomic structure of AcNiR, the technique of solution X-ray scattering and molecular modelling should be able to provide additional, unambigous information about the subunit assembly of NiR in solution. Due to a close relationship in the amino acid composition between AcNiR and AxNiR [49], the comparison of solution scattering results and crystallographic findings of different species will be possible.

Chapter 3 : Solution X-ray Scattering

3.1 Introduction

Size and shape of biological macromolecules are the most obvious questions to be answered in order to find a more comprehensive understanding of their functioning and effectivity. There are different techniques available to tackle these structural questions, which provide information at various degrees of reliability and detail.

The most informative method for determining macromolecular size and shape involves working with samples in the 'solid state'. X-ray diffraction from crystalline samples is one of the central sources of structural information for biology at high resolution (at atomic level) and has taught us more about the structure of large molecules than any other method. However, one of the major difficulties in macromolecular crystallography is growing good periodic crystals of proteins and polynucleotides, especially when these biomolecules are composed of several domains or assembled into larger functional units. For this reason, X-ray scattering from totally disordered samples such as solutions of proteins or larger molecular assemblies represents an alternative, though being a low to medium resolution technique, to acquire useful information about macromolecular structure in solution with a characteristic length of the order of tens to hundreds of Ångstrœm.

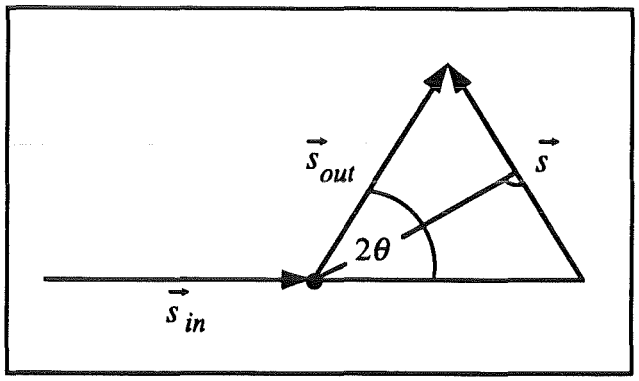


Figure 3.1 :
Basic geometry of an X-ray scattering experiment. The scattering vector $\vec{s}$ for elastic X-ray scattering is the difference vector between the scattered, $\vec{s}_{in}$, and the incident photon wave vector, $\vec{s}_{out}$. In elastic scattering the magnitude of $\vec{s}_{in}$, and $\vec{s}_{out}$ is unchanged. Denoting 2θ the scattering angle and λ the X-ray wavelength, the magnitude of the scattering vecor $\vec{s}$, follows the equation

$$s = |\vec{s}| = 2|\vec{s}_{in}| \sin\theta = \frac{2}{\lambda} \sin\theta .$$

All atoms in a scattering object, independently of their distribution, will scatter in phase along the direction of an impinging X-ray beam. The scattered X-rays are characterized by the scattering vector $\vec{s}$ (as shown in figure 3.1). As the molecules of interest are only a minority component, the measurement consists of the scattering from macromolecules-plus-solvent over that of pure solvent alone. Regarding the form of the strong intensity decay in the forward X-ray scattering, information about the overall shape and size of the scattering molecule can be obtained.

Nevertheless, the advent of synchrotron radiation sources opened up new ways to probe biological structures at the subcellular level. Consequently,

the high intensity and collimation of synchrotron X-rays provide an outstanding possibility to perform X-ray scattering studies during considerably shorter measurement periods than were ever possible with conventional laboratory sources. Moreover, the observation of generally very weak scattering features is possible which are associated with the internal structures of the molecules in solution. Particularly, this latter feature is usually not included in the traditional small-angle scattering region. Thus, in the context of synchrotron radiation the range of the traditional small-angle X-ray scattering (SAXS) has to be extended and defined with the term X-ray scattering from disordered systems or *solution X-ray scattering*.

It is beyond the scope of this thesis to develop the theory of X-ray scattering in detail, the reader is therefore referred to textbooks and in particular to review articles or specific papers concerning applications in biology (see e.g. refs. [54-59]). Only the theoretical basis that is important for the comprehension of this work will be presented. Formulas used for data analysis and interpretation are given. Some emphasis has been put especially on the simulation of solution X-ray scattering spectra in connection with molecular modelling.

3.2 General principles

Since solution X-ray scattering investigates distances that are large compared to interatomic distances, the scattering is described as arising from objects of some electronic density imbedded in a medium of another density. Assuming the elementary scattering objects are molecules, these may be isolated in the sample or distributed according some type of spatial correlations. Hence, the scattering profiles contain information on the size of the particles as well as on the interactions between them. Here, the following restrictions on the scattering system will be dealt with. The system

- is diluted and monodisperse
  (ideal solution of identical molecules with uniform electron density),
- comprises isotropic scatterers (elastic scattering),
- is not subject to multiple and inelastic scattering,
- contains samples with negligible absorption properties.

These assumptions are experimentally acceptable. In order to approach physiological conditions, where most of the water-soluble proteins are found in low concentrations (only a few mg/ml), biological molecules are separated widely enough from each other; thus interparticle effects can be neglected. Inelastic scattering (Compton scattering) is negligible as only small scattering angles are considered, where it is extremely small with respect to elastic scattering. Multiple scattering becomes noticeable only in the case of strong scatterers and a large sample thickness, however, this does not apply to solution X-ray

scattering experiments from proteins reported here. The absorption properties, though being very small, have been taken into account during data reduction.

Thus, each molecule will contribute independently to the elastically scattered intensity. The instantaneous scattering amplitude of one single molecule is just the molecular structure factor given by

$$F(s) = \sum_{n=1}^{N} f_n(s) e^{2\pi i \vec{s} \vec{r}_n}$$

where the sum is taken over all N atoms in the molecule and $f_n(s)$ is the atomic structure factor which is characterized by the number of electrons in individual atoms[4]. Experimentally only the intensity or the square of the amplitude is measured

$$I(s) = F(s) \cdot F^*(s) = \sum_{n=1}^{N} \sum_{m=1}^{N} f_n f_m e^{2\pi i \vec{s} \vec{r}_{nm}} \qquad \text{(E3.1)}$$

with $\vec{r}_{nm} = \vec{r}_n - \vec{r}_m$ defining the difference vector between atom n and m. During the measurement period of an X-ray scattering experiment in solution, each molecule will assume all possible rotational orientations. Moreover, a collection of macromolecules will be observed which will also represent all possible orientations at any time. Therefore what is actually observable is

$$\langle I(s) \rangle = \sum_{n=1}^{N} \sum_{m=1}^{N} f_n f_m \int e^{2\pi i \vec{s} \vec{r}_{nm}} d\Omega$$

The integral takes all possible relative orientations of each vector $\vec{r}_{nm}$ and the scattering vector $\vec{s}$ into account. The computation results in the so-called Debye formula [60]

$$\langle I(s) \rangle = 4\pi \sum_{n=1}^{N} \sum_{m=1}^{N} f_n f_m \frac{\sin 2\pi s r_{nm}}{2\pi s r_{nm}} \qquad \text{(E3.2)}$$

with s and r_{nm} now being scalar quantities. Thus the observed scattering intensity is a function only of the scattering angle 2θ between incident and observed radiation and does not depend on their orientation in space (see figure 3.1).

---

[4] The atomic scattering or structure factor describes the coherent scattering of X-rays from an atom valid for X-ray energies which are large compared with the binding energy of all the electrons in the atom. In the small angle limit ($s \rightarrow 0$) the atomic structure factor corresponds to the electronic charge of each atom.

3.2.1 Radius of gyration

In the limit of very small scattering angles ($\theta \to 0$) equation E3.1 can be expanded as a power series in s

$$\langle I(s)\rangle_{lim\ s\to 0} = 4\pi\left(\sum_{n=1}^{N}\sum_{m=1}^{N} f_n f_m - \sum_{n=1}^{N}\sum_{m=1}^{N} f_n f_m \frac{4\pi^2 s^2 r_{nm}^2}{6} + - \cdots\right) .$$

Selecting the origin so that $\sum_{n=1}^{N} f_n r_n = 0$ and using the expression

$$R_g^2 = \sum_{n=1}^{N} f_n r_n^2 \Big/ \sum_{n=1}^{N} f_n \qquad \text{(E3.3)}$$

which defines the radius of gyration (R_g) of the molecule, the rearrangement yields

$$\langle I(s)\rangle_{lim\ s\to 0} = 4\pi \sum_{n=1}^{N}\sum_{m=1}^{N} f_n f_m\, e^{-\frac{4\pi^2}{3} s^2 R_g^2} = I(0)\, e^{-\frac{4\pi^2}{3} s^2 R_g^2} . \qquad \text{(E3.4)}$$

This representation is known as Guinier's law [61] derived under the assumption of isolated molecules in solution with an intermolecular distance being much larger than their intramolecular size.

The Guinier approximation is widely used for the determination of R_g which is related to the overall size of the molecule. It gives the root mean square distance from the electronic centre of gravity which corresponds to the ordinary centre of gravity in the low angle scattering limit. In addition the molecular weight can also be calculated from $I(0)$. However, it will not be followed up here, since its accuracy is rather limited due to the difficulty in extracting absolute scattering intensities in synchrotron radiation scattering experiments.

3.2.2 Distance distribution function

By looking only at small angles, limited information at very low resolution is obtainable. It is obvious that higher resolution information can be obtained by extending the data to higher scattering angles. A more detailed interpretation of scattering profiles for randomly oriented molecules involves the determination of the distance distribution function $p(r)$. The scattered intensity $I(s)$ (the spatial average denoted by $\langle ... \rangle$ will be omitted from now on) in equation E3.2 can also be formulated as

$$I(s) = 4\pi \int_0^D p(r) \frac{\sin 2\pi sr}{2\pi sr} dr \quad . \tag{E3.5}$$

The distance distribution function represents the probability of finding the distance r between two scattering centres (atoms) within the particle. Consequently, $p(r)$ drops to zero at $r = D$, the maximum intraparticle distance. The distance distribution function is related to the averaged self-convolution of the density distribution of a particle

$$\gamma(r) = p(r)r^2 = \langle \rho(\vec{r}) * \rho(-\vec{r}) \rangle$$

and is also known as spherical averaged Patterson or correlation function [62]. The inverse transform of equation E3.5 can be written as

$$p(r) = \frac{1}{2\pi^2} \int_{s_{min}}^{s_{max}} I(s)\, 2\pi sr \sin 2\pi sr \, ds \tag{E3.6}$$

where the finite s range of the scattering curve has been already taken into account. At very low angles, the range is limited by the presence of the direct beam and at large angles by detector dimensions and low scattering intensities. For this reason the direct Fourier transform is usually replaced by indirect tranform procedures inclusive of effects from counting statistics, restrictions due to finite scattering range and instrumental resolution [63-66].

The mathematical treatment is dependent on the properties of the functions $p(r)$ and $I(s)$. As a consequence of the globular shape of biological macromolecules, $p(r)$ vanishes beyond D. Thus, the sampling theorem of Fourier transform [67] tells us how to measure the scattering curve by use of a certain increment Δs. In practice one will have to stay well below the limit $(2D)^{-1}$.

Finally, it is possible to calculate the radius of gyration also from $p(r)$, using the expression

$$R_g^2 = \int_0^D p(r)r^2 dr \Big/ 2\int_0^D p(r)\, dr \quad . \tag{E3.7}$$

3.3 Interpretation of scattering profiles

Apart from a few cases of distinct symmetry, structural details from solution X-ray scattering experiments have to be deduced indirectly. The Guinier analysis (see equation

E3.4) is valid in the low angle limit and assumes spherical symmetry of the particle. In order to deduce the overall dimension or the shape of the particle, the X-ray scattering data in the extended or wide angle scattering regime is essential for structural interpretations. It is this scattering region which provides information about the internal subunit arrangement of macromolecules in solution. The analysis has to be based on the assumption of various models. Traditionally the scattering curves for simple geometric objects of uniform density (e.g. spheres, ellipsoids, cylinders) are calculated. A trial-and-error technique is applied in optimizing the fit to the experimental curve.

As an example, it should be mentioned that a simplistic molecular shape of human serum transferrin was determined from solution scattering data [30] almost ten years before the crystal structure of rabbit serum tranferrin was solved. This low angle data [30] was in good agreement with an oblate ellipsoid with semi-axes of length 47Å, 47Å and 16Å. It compares fairly well with the maximum dimensions of 95Å x 60Å x 50Å from crystallographic studies of rabbit serum transferrin.

Another example which illustrates the power of solution scattering in detecting conformational changes is given in figure 3.2. It shows three scattering profiles calculated from a simple model of two ellipsoids which define different conformational states of a molecule. It is clear from figure 3.2, that X-ray scattering data at relatively higher angles (2° to 3°, i.e. $s > 0.025$ Å^{-1}) is particularly sensitive for detecting conformational differences.

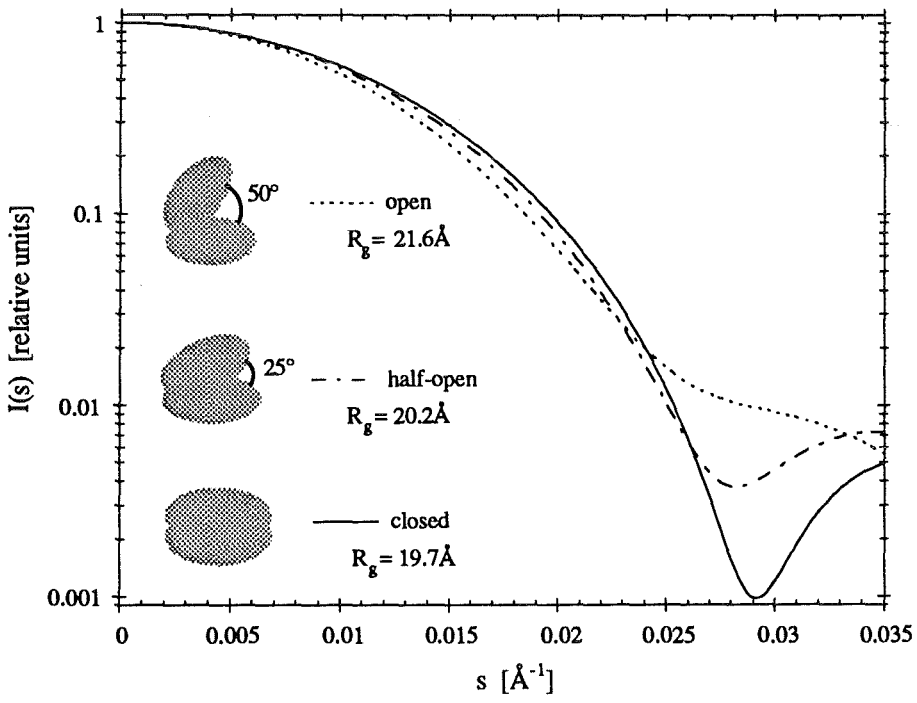


Figure 3.2 :
Two ellipsoids (semi-axes of length 25Å, 25Å and 15Å) have been arranged in an open, half-open and closed state. Overall dimensions correspond approximately to one lobe of transferrin. The differences in the calculated scattering profiles are quite obvious. The most distant parts of the two ellipsoids from the hinge point move considerably. However, the radius of gyration shows only small changes.

Such models are particularly helpful when no other structural information about the molecule is available. However, this can only be a qualitative measure of the molecular

shape and the uniqueness of the solution is not guaranteed. The procedure is based upon the goodness of the fit between experimental and calculated scattering data corresponding to various model structures and thus reflects the potential ambiguity in structure determination. A model-independent approach is based on the multipole expansion method using spherical harmonics which was first proposed by Stuhrmann [68, 69]. The formalism of multipole coefficients is difficult in its practical implementation but the evaluation could be improved now by a special minimization procedure [70]. However, an unambigous determination of more detailed quarternary structures of proteins needs statistically significant scattering data in a wide angular range. This condition can only be provided by the high brilliance of synchrotron radiation sources and will certainly pay off in future applications.

Far more detailed models can be considered when specific information about the structure of the macromolecule is available from other techniques in particular from high resolution protein crystallography. Of special interest are proteins composed of several subunits or domains because the interaction between these subunits can be comparable with the forces in the crystal. Therefore the crystalline state and the solution structure may not necessarily be identical. Solution X-ray scattering gives an accurate scattering pattern in just the region which is sensitive to large internal structural changes or molecular associations. Thus, the crystal structure must be certainly regarded as *the* starting point for studies and interpretations of the conformational behaviour in solution.

The calculation of solution X-ray scattering profiles from atomic coordinates of biological macromolecules can be rather computer intensive. Hence, calculation methods have been proposed which are less time-consuming and allow various molecular models to be tested efficiently. For example, the so-called *cube method* devides the molecule into small cubes for calculating X-ray scattering curves [71, 72]. Subsequent modifications and improvements of the method have been introduced, e.g. consideration of solvent accessible cavities within the molecule has been made [73, 74]. There are a variety of methods based on *Debye's formula* (equation E3.2), including those involving a Monte Carlo procedure that uses a subset of intramolecular distances r_{nm} [75] or a method that involves the expansion of the Fourier transform into spherical coordinates [76]. These provide a more rigorous basis for interpretation and prediction of X-ray scattering profiles of biological molecules in solution.

In the present work, a modification of the so-called *sphere method* [77] has been applied which is also based on Debye's formula. The computation of the scattering profile is performed using the equation

$$I(s) = g(s) \sum_{r_i=0}^{D} p(r_i) \frac{\sin 2\pi s r_i}{2\pi s r_i}$$

where $p(r_i)$ gives the histogramme of distances between every pair of atoms. Distances are binned and weighted according to the product of the number of electrons belonging to the corresonding pair of atoms. Since a molecule is represented by a volume with a certain shape, each scattering centre (given by its atomic coordinates) is associated with a geometrical volume which is introduced mathematically by the shape factor $g(s)$ [78]. Assuming a spherical shape for each atom, $g(s)$ is given by the squared form factor of a sphere of radius R

$$g(s) = \left(\frac{3(\sin\xi - \xi\cos\xi)}{\xi^3} \right)^2 \quad \text{with} \quad \xi = 2\pi sR \quad .$$

A sphere radius of 1.7Å was used which correlates with the average radius of non-hydrogen atoms in proteins. Although this procedure gives a fairly accurate description of the solvent-excluded volume occupied by the macromolecule, tightly bound water molecules on the outer surface are not taken into account unless their positions are resolved in the crystal structure, even though the scattering data in this range ($s \leq 0.035\mathring{A}^{-1}$) contains such information. For this reason, an approach has been developed in order to create solvent water positions around the macromolecule and to test particularly the influence of the water shell on solution X-ray scattering. The work presented here shows that the incorporation of a hydration layer up to about 3Å from the molecular surface improves the scattering pattern simulations significantly.

3.4 Molecular Modelling

The interpretation of solution X-ray scattering profiles benefits from crystallographic structure information. Although solution scattering is only a low to medium resolution technique, its combination with molecular modelling using atomic coordinates from crystal structures can prove advantageous in defining realistic 'high' resolution models of biological macromolecules in solution.

Nowadays molecular structures can be studied easily, clearly and accurately using powerful computer graphics systems. Visualizing biological molecules is essential to an active process of understanding and interpreting macromolecular structures [79]. It is not only helpful for looking at the proteins but in concert with molecular model-building it is an outstanding tool for elucidating the relationship between biological function and molecular structure. Modelling techniques include calculations based on an empirical, essentially classical potential energy function. Such a potential function (see Appendix A) can be used in optimization of geometry (molecular mechanics) or in studies of atomic motion in macromolecules (molecular dynamics). During the last decade the results of these

calculations in concert with experimental data have led to a more complete understanding of biological molecules. The relatively up-to-date review from Karplus *et al.* [80] gives impressive insights in theoretical approaches of protein dynamics.

In the following, three principal concepts will be mentionned where solution X-ray scattering in conjunction with molecular modelling can provide useful structural information of proteins in solution.

- The structural reasons why many proteins change their conformation specifically as they perform their functions (e.g. shape modifications by the action of small molecules or ligands that activate or inhibit them) have been studied extensively in recent years. Energy minimization and molecular dynamics have been employed to explore atomic motions, binding events and enzymatic reactions on the picosecond time scale corresponding to local dynamics such as atomic vibrations or movements of side chains. These fluctuations have been detected by far-infrared and Raman spectroscopy techniques [81] and most recently by inelastic neutron scattering which allowed an improved understanding of the form of the potential energy surface [82]. However, the important tertiary or quarternary structural changes, the phenomenon of protein flexibility, i.e. large-scale motions which represent time scales up to the order of seconds, are still immensely complicated and have yet to be determined. The present understanding is rather limited. Thus, a specific deformation of a protein backbone (e.g. stretching or twisting at a location in the structure) can be modelled by combining information from other related structures. This procedure is purely geometric, foregoing any energetic or steric considerations. It is clear that this kind of model-building goes hand in hand with the interpretation of solution scattering data. In this connection the model-building based on homology of protein sequences and protein docking needs to be mentioned. Also, an understanding of protein-protein-interactions is an important field of research and approaches of molecular modelling and their experimental verification will certainly increase significantly.

- Another topic of considerable current importance in modelling studies is the treatment of solvent effects. The interaction of amino acids with the surrounding water molecules is probably one of the dominant factors in protein folding and thus for the final conformational shape of a protein [83]. In this respect solution X-ray scattering can certainly provide interesting results since water molecules close to the protein contribute significantly to the scattered intensities. The reason for this is the formation of a hydration layer (water and salt) associated with the protein surface which has an electron density very similar to that of proteins. In order to improve the simulation of X-ray scattering profiles with respect to experimental findings this solvent layer has to be taken into account. Unfortunately, X-ray crystallographic experiments do not reveal all the

positions of protein-bound water molecules belonging to the hydration shell due to relatively higher disorder and limited resolution. Thus, the solvent structure must be generated by computational techniques. For the purpose of generating surface solvent molecules, an approach, including energy minimization, was developed and is described briefly in Appendix B.

- The most ambitious goal of molecular modelling is to predict the secondary and tertiary structure of a protein from its amino acid sequence. The occurrence of large significant and non-random arrangements of specific structural motifs has been discovered and can provide ideas about the determinants of tertiary structure and folding [84, 85]. Thus a knowledge-based modelling approach (using the sequence homology with proteins of known three-dimensional structure and empirical energy functions to derive the conformation of minimum potential energy) in conjunction with computer graphics applications is receiving increasing attention. It will be of great importance to test the possible models. In this respect solution X-ray scattering provides a powerful means of shape determination in the absence of structures solved to atomic resolution. In addition X-ray scattering intensities and distance distribution functions calculated on the basis of known crystal structures will make structure interpretations easier. The installation of a corresponding database was proposed only recently [86].

Finally, all developments and improvements in the field of graphical and molecular representation which have certainly increased the general view of understanding protein structures, provide *only* models. Many models are rather accurate, because they are based on substantial quantities of data, but they are still models none the less.

Chapter 4 : Experimental

4.1 Synchrotron Radiation Sources

Synchrotron radiation is the electromagnetic radiation emitted by charged particles accelerated to relativistic energies moving along a curved trajectory under vacuum. A wide spectral range of radiation is covered, from the infrared to the X-ray region. The dependence of the total radiated power P on the particle energy E, its rest mass m_0c^2 and the bending radius R of the trajectory is given by

$$P = \frac{e^2 c}{6\pi\varepsilon_0} \cdot \frac{\gamma^4}{R^2} \qquad \text{with} \quad \gamma = \frac{E}{m_0 c^2} \ .$$

This relationship explains why electrons (positrons) are by far the most important particles generating this kind of radiation since the mass of other particles such as the proton is much larger than that of the electron ($m_0^{electron} / m_0^{proton} \approx 1/1836$).

Until the late 1950's synchrotron radiation was mainly studied because of its negative role in electron accelerator technology. The radiated power of relativistic electrons corresponds to a considerable energy loss which has to be compensated by radio frequency (RF) elements in order to maintain the relativistic velocity of the circulating electrons.

However, its usefulness in many aspects of X-ray physics was recognized early and during the last twenty years synchrotron radiation has become a unique tool for many disciplines (from basic research to technical applications). Until the mid-1970's the radiation was obtained as a by-product from accelerators built for high energy physics. This parasitic use, however, did no longer satisfy the requirements of synchrotron radiation users. Circular electron accelerators were built and dedicated for synchrotron radiation experiments. This type of *synchrotron radiation sources* produced radiation from bending magnets which keep the electrons on a closed orbit. During the last decade the development of so-called insertion devices (wiggler and undulator) led to further improvements particularly concerning the size and divergence of the photon beam.

The components of a synchrotron radiation source are shown in figure 4.1 giving the example of the Synchrotron Radiation Source (SRS) at Daresbury (U.K.). The SRS is the longest established, dedicated, high energy synchrotron radiation source in the world. Although it has been in operation since 1981 a number of improvements and extensions have been made to keep up with the standards of an up-to-date synchrotron radiation source [87]. Closely related with the evolution and exploitation of synchrotron radiation is the development of adequate X-ray optics and highly efficient X-ray detectors and assocoiated data aquisition systems in order to cope with the increasing amount of photon flux. The

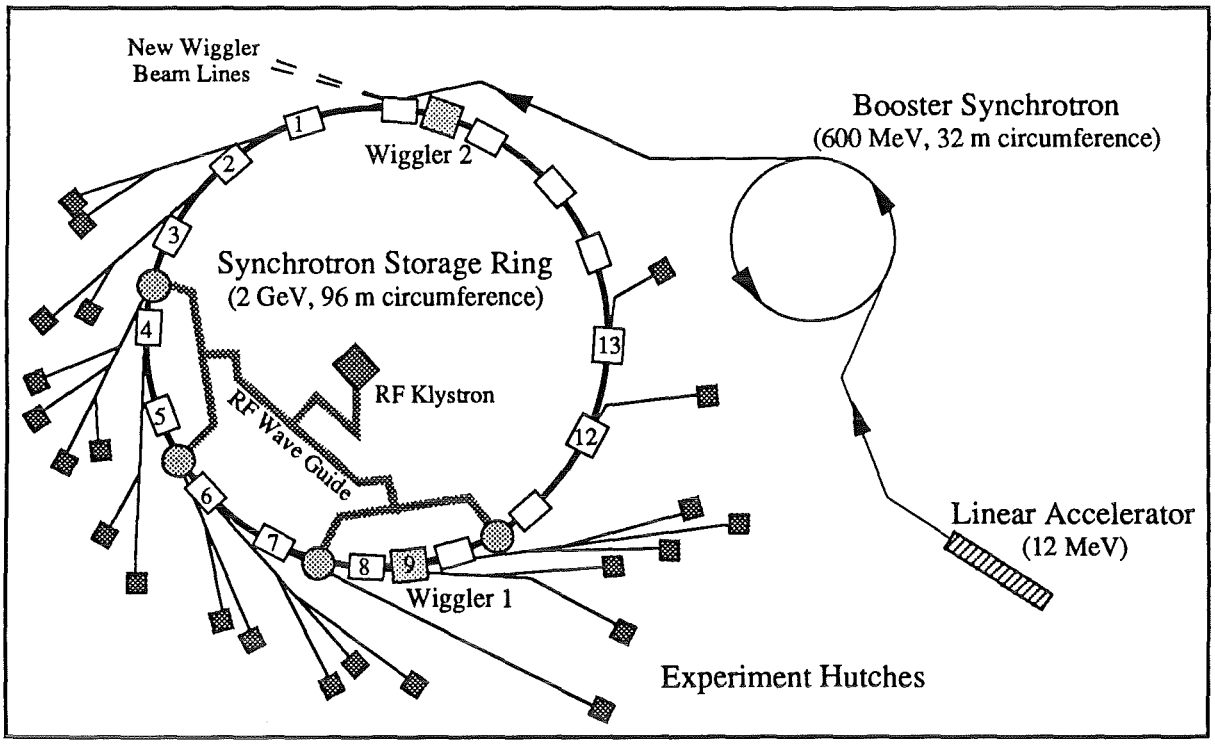


Figure 4.1 :
Schematic diagram of the instrumentation required to produce high energy electrons at the Daresbury SRS (U.K.). 10 out of 16 bending magnets (ports 1-8, 12, 13) are in operation to provide synchrotron radiation for experimental purposes. Port 9 is equipped with a wiggler magnet of 5 Tesla peak field. Another wiggler (6 Tesla peak field) has just been installed (1992) and awaits the construction of additional experimental stations. The energy loss of the circulating electrons is compensated by 4 RF resonant cavities which transfer energy back into the stored beam supplied by the RF klystron.

experiments which will be reported here have been carried out on the bending magnet line on experimental station 8.2 at the SRS, which is well suited for static X-ray scattering measurements of weak scatterers such as biological molecules in solution.

4.2 Bending Magnet Radiation

The spectral distribution[5] of the bending magnet radiation can be described theoretically in terms of the photon flux $\mathcal{F}$ which is a function of the photon wavelength λ and depends on the intensity of the field B of the dipole bending magnet (which defines the bending radius R), the electron energy E, and the beam current I

$$\mathcal{F}\left[\frac{photons}{s\ mrad(horizontal)\ 0.1\%\,bandwidth}\right] = 2.457\cdot10^{13}E[GeV]\,I[A]\,\frac{\lambda_c}{\lambda}\int_{\xi=\lambda_c/\lambda}^{\infty}K_{5/3}(\xi)d\xi$$

5 Equations have been given in practical units (see e.g. [88-90]).

$K_{5/3}$ is a modified Bessel function of second order [91] and λ_c, the so-called *critical wavelength*, is defined as

$$\lambda_c[\mathring{A}] = 5.59 \frac{R[m]}{E^3[GeV^3]} = \frac{18.64}{B[T]\,E^3[GeV^3]} \quad .$$

λ_c divides the spectrum into two parts of equal power, thus representing an essential parameter in defining the spectral output of a synchrotron radiation source. The spectral photon density, i.e. the vertically integrated photon flux of a dipole magnet at the Daresbury SRS is shown in figure 4.2. The emitted radiation is tightly collimated in the forward direction [92] having a mean opening angle $\vartheta = 1/\gamma$ (for the relativistic case where $\gamma >> 1$) between the direction of emission and that of electron motion. The instantaneous cone of radiation is swept out in the electrons' orbital plane generating a fan of radiation as symbolized by a lighthouse.

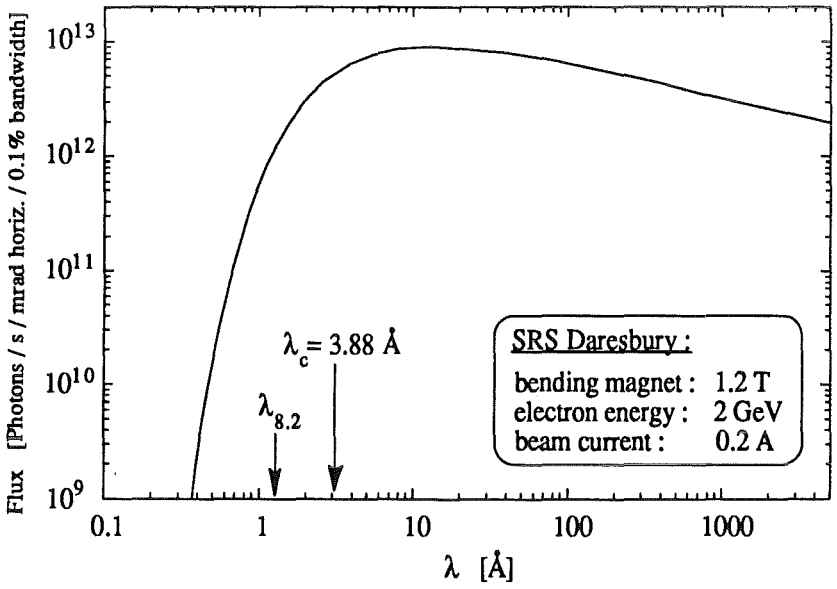


Figure 4.2 : Spectral distribution of the bending magnet radiation at the Daresbury SRS. $\lambda_{8.2}$ denotes the wavelength selected at beamline 8.2.

It is the special combination of synchrotron radiation properties where no conventional source (X-ray tube or rotating anode) can compete with and which have led to far-reaching applications in the fields of biology and many others. High flux, collimation, and spectral continuity as well as time structure and polarization of the photon beam are unique properties that made synchrotron radiation an outstanding tool for research.

Macromolecules in solution are well-known for their weak scattering behaviour. Mainly consisting of low Z atoms (hydrogen, carbon, nitrogen and oxygen), proteins give rise to relatively weak X-ray scattering. The typical scattering power is between 10^{-5} to 10^{-8} times the incident beam intensity [93]. Thus, the large flux and small beam size of

synchrotron radiation make the observation of previously unseen scattering features possible.

4.3 X-ray Scattering Station 8.2 at Daresbury

The availability of synchrotron radiation along with high quality optics and X-ray detectors offers an excellent starting point to perform scattering studies on biological systems. The following two sections will be used to give a few details concerning the instrumentation and are based upon beamline 8.2 at the Daresbury SRS. Most of these details are similar to the SRS station 2.1, where extra intensity enables time-resolved studies (in the *msec* region) to be carried out [94].

4.3.1 Camera system

Despite the remarkable characteristics of the bending magnet radiation, a collimation system is necessary in order to produce a well-defined beam and a low background. This is achieved by focussing optics in combination with slits. Due to the very weak intensities measured in solution X-ray scattering experiments, the slit system has to be designed carefully to keep the background as low as possible.

The following experimental set up of the camera system refers to figure 4.3. Behind the beam splitter a water cooled aperture A and a beam position monitor M_{BP} are present. The aperture consists of vertical and horizontal slits and has the purpose of reducing the

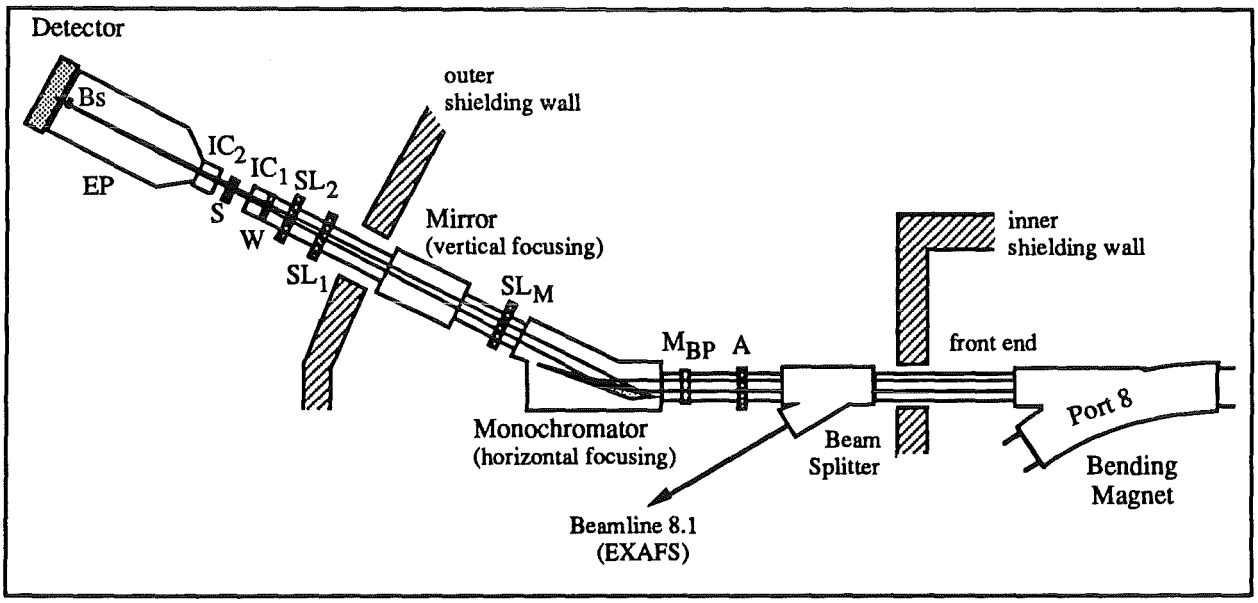


Figure 4.3 :
Schematic layout of a camera system for solution X-ray scattering based on beamline 8.2 at the Daresbury SRS. Focusing optics and slits provide high intensity and collimation of the photon beam combined with a low parasitic background. In addition this monochromator-mirror arrangement produces a very stable focal spot and thus offers profitable experimental conditions (for details see text).

heat load on the monochromator. The flat, triangularly shaped Ge (111) monochromator crystal is aligned such that radiation of 1.54 Å is selected. This wavelength is the best compromise in order to account for absorption of sample and optical system, scattering power of samples, source emission characteristics and spectral response of the detector system. The wavelength resolution of $\Delta\lambda/\lambda \leq 4 \cdot 10^{-3}$ assures that no significant broadening of scattering peaks or overlap of scattering orders occurs. The asymetrically cut monochromator crystal (asymmetry angle of 10.5°) reduces the angular divergence of the X-rays and accepts 5.4*mrad* of horizontal aperture. The radiation monochromator is at a distance of 21*m* from the tangent point (TP) of the bending magnet. The monochromator is equipped with a bending mechanism which allows horizontal focusing (3:1). The slits SL_M cut off scattering and fluoresence from the monochromator. The beam is vertically focussed (3:1) by grazing incidence on a bent, uncoated quartz mirror located 22.2*m* from the TP. The pair of slits SL_1 and SL_2 behind the mirror determine the opening angle for the photon beam and further cut down scattering from components of the beamline. The mica window W marks the back end of the main vacuum system (pressure < 10^{-6}*mbar*) and is required as interface to the sample area. Mica is a very suitable window material because it produces mainly scattering in well-defined directions at high scattering angles due to its crystalline structure. Although the source (TP at the bending magnet) and the specimen S are separated by a distance of 25*m*, this monochromator-mirror configuration with the combination of defining slits provides a beam cross-section of 4*mm* x 1*mm* and a flux of10^{11} photon/s at the sample position (for the SRS running at 2 *GeV* and 200 *mA*).

The sample cell is positioned between two ion chambers IC_1 and IC_2, which are used to monitor the incident intensity and sample absorption, respectively. The necessary air path at the sample position is kept as short as possible (20-30*mm*) to reduce air scatter which contributes to the background. The evacuated pipe EP between sample and detector has a variable length of 1-4*m*. The experiments reported here were carried out with a sample-to-detector-distance of 3*m*. A lead beamstop Bs is used to capture the direct beam thus preventing radiation damage to the detector. The end of the vacuum pipe EP has a diameter of 500*mm* and is sealed by a mylar foil.

4.3.2 Detection system

A multiwire quadrant chamber and data acquisition system [95, 96] were used to record the scattering data. The delay line detector is designed specifically for use with circularly symmetric patterns, such as those arising from solution scattering with a characteristically steep intensity fall off at higher scattering angles. Table 4.1 gives the main specifications of the quadrant chamber and figure 4.4 represents a schematic diagram of the detector composition.

Table 4.1 Quadrant detector characteristics	
operating pressure	1*atm*
active depth	20*mm*
aperture (active length)	200*mm*
angle of quadrant	70°
gas filling	Xe, Ar, CO_2
efficiency	95%
anode wire diameter	10μm
anode wire spacing	1*mm*
cathode spacing	1*mm*
delay line length	100*ns*
spatial resolution (FWHM)	250μm
maximum count rate	850*kHz*

The sector shaped active area increases as a function of distance from the primary beam and thus improves the statistics in the weaker portions of the scattering pattern. The major feature is the special rear cathode that consists of a printed circuit board (P.C.B.) and has concentric tracks forming the pick-up electrodes to match with the circularly symmetric scattering pattern. It replaces the wire plane found in commonly used area chambers for particle detection (see e.g. [97]). The P.C.B. is connected to a delay line where charge pulses induced by the gas ionization process arrive. The charge can travel in either direction along the delay line, the difference in the arrival time of the electric pulses at each end of the line determines the position of the primary photon event.

The two signals from the delay line are preamplified and transmitted to the data acquisition system where the pulses are converted into digital signals by constant fraction discriminators (CFD) in order to allow the time to amplitude converter (TAC) to extract accurate timing information. The TAC generates an analogue pulse whose amplitude is proportional to the time difference between the start and stop pulse provided by the CFDs. Due to the use of a bipolar TAC, information might be lost in the case where the time difference between start and stop pulses is close to zero. For this reason the TAC is adjusted to a few millivolts positive which produces a gap between slightly negative and positive pulses leading to a few blank channels in the centre of the data (TAC hole). The whole active length of the detector is covered by 512 channels.

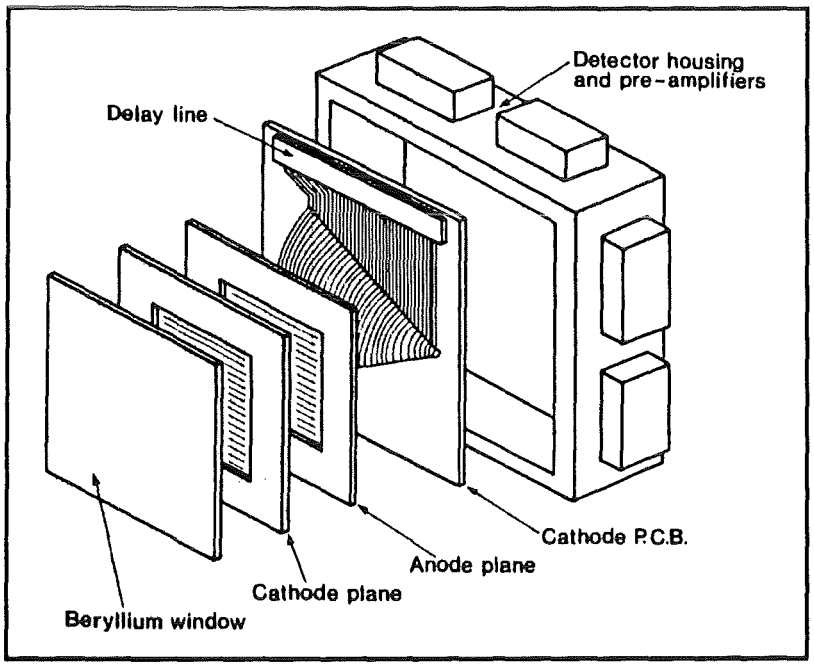


Figure 4.4 : Exploded view of the quadrant detector

The timing information from the TAC which corresponds to a specific scattering angle, is then digitized by an analogue to digital converter (ADC) and stored in a histogramming memory accessable by the station computer (LSI11). A CAMAC time frame generator facilitates the collection of data in time frames. Data pre-selection can be performed by displaying the scattering profile on a graphics screen (i.e. rejection of 'spikey' data caused by non-linearities in the electronics system). For further processing, the raw scattering data is transmitted to another computer system (a central VAX computer).

In data processing, variations in the detector efficiency have to be taken into account. High X-ray fluxes lead to photo-chemical reactions in the detector's gas mixture, which produce impurities that deposit on the wire planes and affect the electric field within the chamber. The detection efficiency not only changes locally but also a general deterioration takes place with time. Most of the detector's non-uniformities and spatial inhomogeneities can be corrected by periodically recording images of the detector response (figure 4.5). This is done by irradiating the detector uniformly with X-rays emitted from a ^{55}Fe source. The $5.9keV$ photon energy is close to the selected synchrotron X-ray energy ($8.1keV$). The detector response is usually recorded during the refilling period of the electron beam into the synchrotron (every 12 or 24 hours).

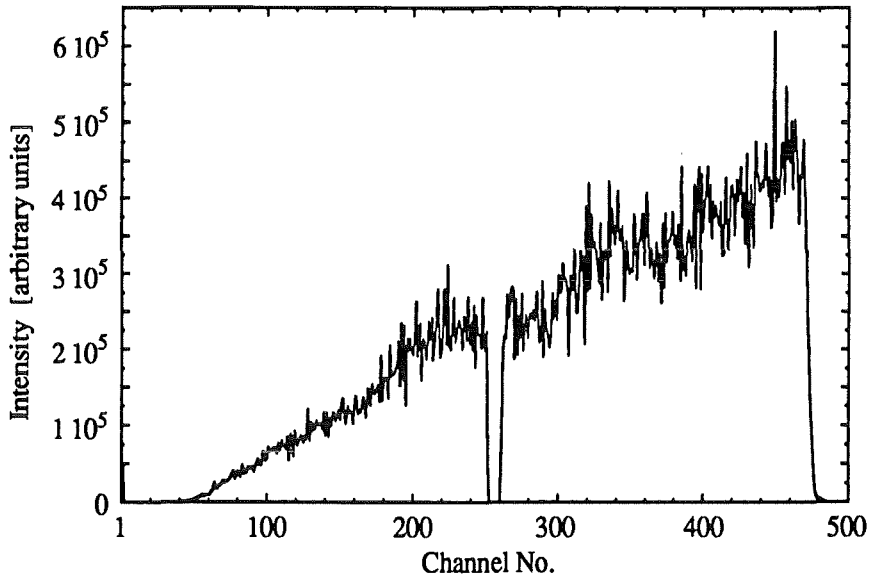


Figure 4.5 :
A typical detector response recorded for two hours by irradiating the detector using a ^{55}Fe source.

4.4 Sample Preparation

4.4.1 Transferrin

Transferrin samples (human serum transferrin HST, human lactoferrin HLF and chicken ovotransferrin COT) were purchased from SERVA (Heidelberg) and used without additional purification. All chemicals were analytical grade. Individual C and N-terminal

half-molecules of COT were prepared following the procedure reported by Oe *et al.* [98]. The protein was dissolved in 20mM $NaHCO_3$, 50mM MES[6], 50mM HEPPS adjusted to pH 7.5. Iron, copper, indium, gallium, aluminium, hafnium and thorium-transferrin were prepared by adding 50% excess of fresh 5*mM* metal solutions (Fe-NTA[7], $CuSO_4$, $InCl_3$, $GaNO_3$, $AlK(SO_4)_2$, Hf-NTA[7], Th-NTA[7], respectively) to the protein. Excess metal was removed by dialysis. Prior to use, the samples were filtered through a 0.2μm syringe filter. The metal saturation was confirmed by UV/Vis spectroscopy.

A low protein concentration of 5*mg/ml* was used to minimize molecular aggregation. It is also similar to the conditions in serum, where the transferrin concentration ranges from 2 to 3*mg/ml* and thus the solution scattering measurements are expected to define the molecular conformation under quasi-physiological conditions.

Recent success in protein engineering provided mutants of the N-terminal fragment of human serum transferrin (HST/2N), comprising 337 residues along with a single iron binding site. Two site-directed mutants, where the aspartic acid 63 is replaced by both a serine (D63S) and a cysteine residue (D63C), and the corresponding wild type have been provided by Prof. Woodworth and Dr. Mason (University of Vermont, U.S.A.). They have been expressed in baby hamster kidney cells and purified to homogeneity [99, 100]. The aspartic acid 63 (Asp63) plays an important role with respect to iron binding and interdomain stability (see also figure 2.2). The protein concentration was 5*mg/ml* in a 100mM HEPES buffer (pH 7.5). The ferric HST/2N molecules were saturated with iron by addition of freshly prepared 5*mM* Fe-NTA solution (see above).

4.4.2 Nitrite Reductase

The copper-containing dissimilatory NiR of *Achromobacter xylosoxidans* was isolated from cells grown anaerobically with nitrite as the terminal electron acceptor and purified to homogeneity on SDS-PAGE, by a combination of $(NH_4)_2SO_4$ fractionation and ion-exchange chromatography [49]. After dialysis against 50*mM* phosphate buffer (pH 7.2) containing 50*mM* NaCl enzyme concentrations of 5 and 10*mg/ml* have been prepared. Due to hydrophobic properties the enzyme is only scarcely soluble in water.

The NiR of *Alcaligenes faecalis* was isolated and purified from algae growing in sludge according to the procedure reported previously [41, 47]. The amount of 5*mg* of the enzyme has been kindly provided by Prof. Beppu (University of Tokyo, Japan). It was dissolved in 10*mM* Tris HCl (pH 8.0) containing 70% $(NH_4)_2SO_4$. In view of the

---

6 The following abbreviations have been used: MES, 4-morpholineethanesulphonic acid; HEPPS, 4-(2-hydroxyethyl)-1-piperazine-propanesulphonic acid; NTA, nitrilotriacetic acid; HEPES, 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid; SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis, which is a rapid and efficient way of removing small amounts of contaminating materials; Tris, Tris(hydroxymethyl)aminoethane.

7 The metal-NTA solutions were prepared by combining equal volumes of a 10mM metal solution ($FeCl_3$, $HfCl_4$, $ThCl_4$) in 2M HCl and a 40mM NTA. This solution was neutralized by successive addition of small amounts of solid $NaHCO_3$.

unphysiologically high salt concentration the solution was dialysed against $20mM$ Tris HCl (pH 8.0) buffer until a negligible amount of ammoinium sulphate was reached. The final AfNiR concentration amounted to 5mg/ml.

The dialysis buffer was used throughout the solution scattering experiments as a cell blank for the different protein samples, respectively.

All proteins as well as buffer materials were filled into a sample cell consisting of a brass holder with two mica windows ($15mm$ in diameter and $25\mu m$ thick) which were tightened by a brass screw. The windows are kept into position on both sides of a teflon ring defining a cylindrical cavity of $1.5mm$ length and a volume of $100\mu l$. The use of teflon is necessary in order to prevent the protein from getting in touch with the metal surface of the sample cell. All samples are injected from the top through a capillary-like opening in the teflon ring.

4.5 Experimental Protocol

After setting up the camera length, i.e. the length of the evacuated pipe between sample cell and detector, and focusing the X-ray beam on the detector position, alignment and calibration of the detector need to be carried out.

Due to the relatively high interaction depth of X-rays ($20mm$) in the gas-filled detector, parallactic effects occur for photons incident at high angles. This can lead to serious resolution broadening, distorting the scattering pattern at high angles. For this reason the tilt angle of the detector had to be chosen correctly. This can be done in combination with the calibration of the detector using the diffraction pattern from a sample of wet rat tail collagen exposed to the beam at the specimen position (see figure 4.6). The

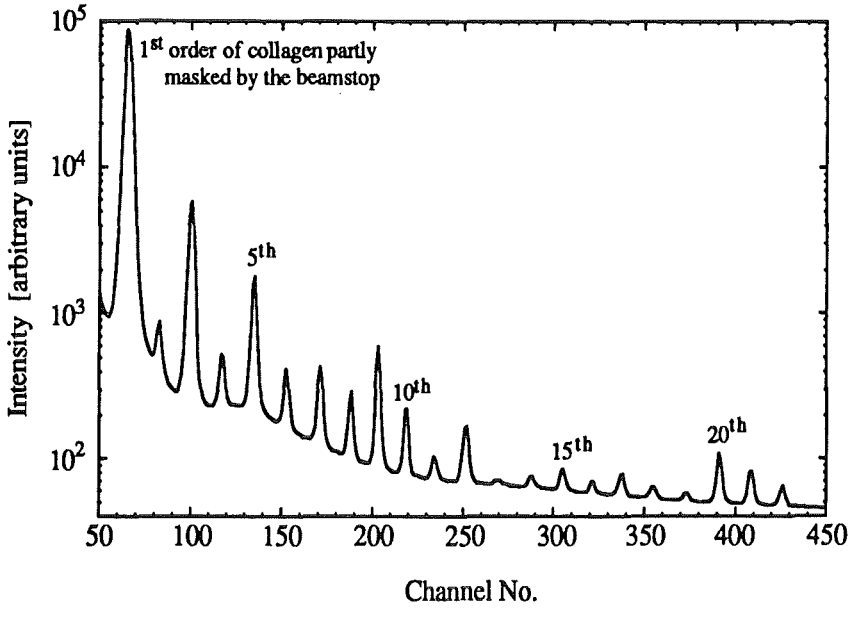


Figure 4.6 : Scattering profile from a wet rat tail collagen sample vertically oriented to the length of the incident X-ray beam. The data were recorded with a camera length of $3m$ and an exposure time of $5min$. The profile is already corrected for detector response and TAC hole.

determination of the average spacing between the first ten diffraction peaks (orders) of collagen allows to calculate the 0^{th} order which corresponds to the centre of the quadrant chamber. A proper detector alignment is given if this order coincides with the position of the direct beam and real detector centre. The use of a special pulse generator circuit connected to the cathode P.C.B. yields the channel number where the centre of the quadrant detector is located (usually channel 50).

The positions of the collagen orders were used to convert detector channels into scattering vectors *s*. The camera length of $3m$ allowed reliable measurements up to a maximum s value of 0.035 $Å^{-1}$ (up to 22 orders of collagen could be clearly identified) with an average sampling length of 0.00012 $Å^{-1}$.

Mylar films in front of the detector are necessary in order to attenuate the flux of scattered photons reaching the quadrant chamber. However, in the case of buffer and protein solutions, which show smoothly varying profiles without any sharp peaks, the acceptable local count rate (photons/mm^2) of $8kHz$ was not exceeded. The overall count rate didn't exceed the critical limit of 150000 counts/*s*. Thus no photon flux attenuators had to be used.

The sample cell was mounted on a horizontally and vertically movable table to allow for accurate positioning. The position of the impinging X-ray beam was checked by colour change of an X-ray sensitive tape sticked on the opening of the cell window. The incident and transmitted photon flux were measured by placing ion chambers on either side of the sample and values were recorded in calibration channels. All experiments were carried out at room temperature (22 to 24°C).

The data collection was performed in a cyclic process. Time frames of usually 100*s* were collected successively so as to check for potential radiation damage or aggregation of material on the cell window, as well as to establish the reproducibility of data obtained in repeated measurements. Each sample was exposed to X-rays for no longer than $15min$. In order to minimize errors due to the subtraction of background and instrument function, measurements of buffer (blank cell) and sample (protein and buffer) were made alternately for equal time intervals.

Chapter 5 : Data Reduction and Analysis

After transmission from the local LSI11 data acquisition system to the central VAX computer, the raw data is available for off-line analysis. Calculations were mainly performed using the software packages OTOKO [101] and GNOM [102]. OTOKO is a programme for data reduction (normalization, background subtraction, corrections for detector response and removal of TAC-hole, and allows the extraction of the scattering pattern $I(s)$ arising from the protein sample) and analysis (calibration, radius of gyration). Subsequently, reading the final one-dimensional scattering curve into GNOM, the evaluation of the particle distance distribution function $p(r)$ can be accomplished. No slit corrections had to be applied since it is unnecessary in view of the point collimation and extremely small divergence of the X-ray beam used in the present experimental camera set up.

5.1 Normalization

Even though the X-ray beam from a storage ring is stable, the beam intensity decays slowly with time. The monitoring of the beam decay allows one to correct the data in such a way that identical experimental conditions can be considered during the measurement of protein sample and blank cell. The normalization is carried out by dividing the collected scattering pattern by the transmitted intensity. The correction value provides the integrated beam intensity over every time frame. This procedure also ensures the automatical compensation for absorption of proteins in a buffer solution as well as for absorption of the buffer itself.

5.2 Check on Radiation Damage

An important requirement for the samples is that they must not be changed by the radiation during exposure time. As the data was collected in a series of time frames (the exposure time of $15min$ and time frames of $100s$ give nine frames per data set for a certain sample), a check on radiation damage could be performed. This was monitored by dividing two frames (in most cases the first and last frame of a data set was sufficient), to see whether a change in the profile occurred. No detectable change was observed that could be due to either radiation damage or deposits on the sample cell windows over a period of 20 minutes. Thus data on a protein sample was collected for only 15 minutes to ensure their integrity. After this time interval a fresh sample was filled in a newly prepared sample cell consisting of unused or cleaned mica windows.

5.3 Background Subtraction

The elimination of the background scattering is a crucial point in the data analysis. In order to obtain the protein scattering curve, the difference between the scattering of the buffer

solution (blank cell), $I_{bs}(s)$, and the protein solution (buffer and protein), $I_{ps}(s)$, must be calculated. The total scattering of the protein solution is the square of the sum of the scattering amplitudes (see equation E3.1) generated by electron density inhomogeneities along the X-ray path through the sample

$$I_{ps}(s) = \left| F_s(s) + F_p(s) + F_{pf}(s) + F_{sf}(s) \right|^2 \quad .$$

$F_s(s)$, $F_p(s)$, $F_{pf}(s)$ and $F_{sf}(s)$ are the scattering amplitudes arising from the solvent, the protein, internal electron density fluctuations in the protein, and fluctuations within the solvent, respectively. For the sake of simplicity, background contributions from the scattering of windows, the sample cell, windows from the camera and air scatter have been omitted as they are the same for the blank cell and the protein sample. Since the blank scattering, $I_{bs}(s)$ is given by

$$I_{bs}(s) = \left| F_s(s) + F_{sf}(s) \right|^2 \quad ,$$

the difference $I_{ps}(s) - I_{bs}(s)$ is actually not identical to the scattering arising fom the protein alone because of the cross or interference terms.

Nevertheless, it is a reliable approximation due to the fact that density fluctuations within protein and solvent cannot be resolved and its contribution to the background is negligibly small. However, the interference term of solvent and protein, $I_{sp}(s)$, as well as protein-protein interactions or aggregations can not be removed using this method of background subtraction. The latter contribution can be significant in the case of high protein concentrations and is usually eliminated by extrapolation to zero concentration (measurements of concentration series need to be carried out). The concentration effect disappears for measurements up to higher angles. It usually becomes negligible beyond values of $s \cdot D > 1$ (where D is the maximum dimension of the particle). Unless otherwise mentioned only low concentrations were used ($5mg/ml$) during the course of this work where concentration effects could be neglected.

Consequently, a good approximation for the true protein curve[8] $I(s)$ in the case of sufficiently low protein concentrations is given by

$$I(s) = I_{ps}(s) - \alpha I_{bs}(s) \quad .$$

The factor α takes account of variations in the concentration for the buffer and protein solutions. In order to determine α the following procedure was developed for the present work. A good empirical description for the scattering intensity of biological macro-

[8] The contribution from solvent-protein interactions is still included (i.e. the effect of hydration) and has to be taken into account during data interpretation.

molecules is given by a s^{-4} dependence for large s values [103, 104]. Along with the approximation that $I_{bs}(s)$ can be taken as a constant in the measured angular range, the scattering profiles for buffer and protein solution should coincide for large s values ($s \rightarrow \infty$). For this reason the ratio $I_{ps}(s)/I_{bs}(s)$ should approach 1, provided the buffer used in blank cell measurement and the buffer for the protein solution are of exactly the same concentration. Thus, a least-squares fit to the ratio $I_{ps}(s)/I_{bs}(s)$ was performed by means of the model function

$$f(s) = \alpha + \beta(channel)^{-4}$$

for the upper end of the spectrum[9] (channels 400-470). Since the discontinuity in the spectra was the same in the blank cell and the protein solution pattern, the effect of the detector response was cancelled out by this quotient.

The fit parameter α determined for each data set according to this procedure (whose values ranged from 0.98 to 1.02) allowed a reliable subtraction of background scattering due to the buffer and other instrumental factors.

5.4 Division by Detector Response

Spatial inhomogeneities in the sensitivity and the sector shaped active area of the quadrant detector requires normalization. This is done by measuring the detector's sensitivity with a Fe^{55} source (see figure 4.5) and dividing the scattering patterns by this response function. This does not only smooth out the spikey profiles but also corrects for the increased detection area at higher scattering angles. In order for the quality of scattering data not to be limited by low statistics in the detector response, it was collected for two to three hours.

5.5 Removal of TAC hole

The gap in the centre of the data is caused by electronical settings in the time to analogue converter (TAC), which was described in section 4.3.2, and does not contain any data. The gap extended from channel 252 to 261 and was removed by shifting the upper block of channels to join the remaining lower channel block.

Finally all data sets belonging to a certain protein sample were added together and averaged. As can be seen from figure 5.1, the signal to noise ratio improved significantly. Despite the favourable synchrotron X-ray intensities, a total data collection time of 90*min*

[9] This procedure was still applied to uncalibrated spectra. The knowledge of s-values is not necessary at this stage of data analysis. In order to complete the picture, the given channel range corresponds to s values between 0.030Å^{-1} and 0.035Å^{-1}. Due to the limited sensitivity of the detector in the outermost part of the spectra (see figure 4.5), the last 40 channels had to be disregarded.

and more turned out to be essential in order to collect sufficient statistics at higher scattering angles. The range of three orders of magnitude is also typical of the strong intensity fall-off in solution scattering experiments (note the logarithmic scale).

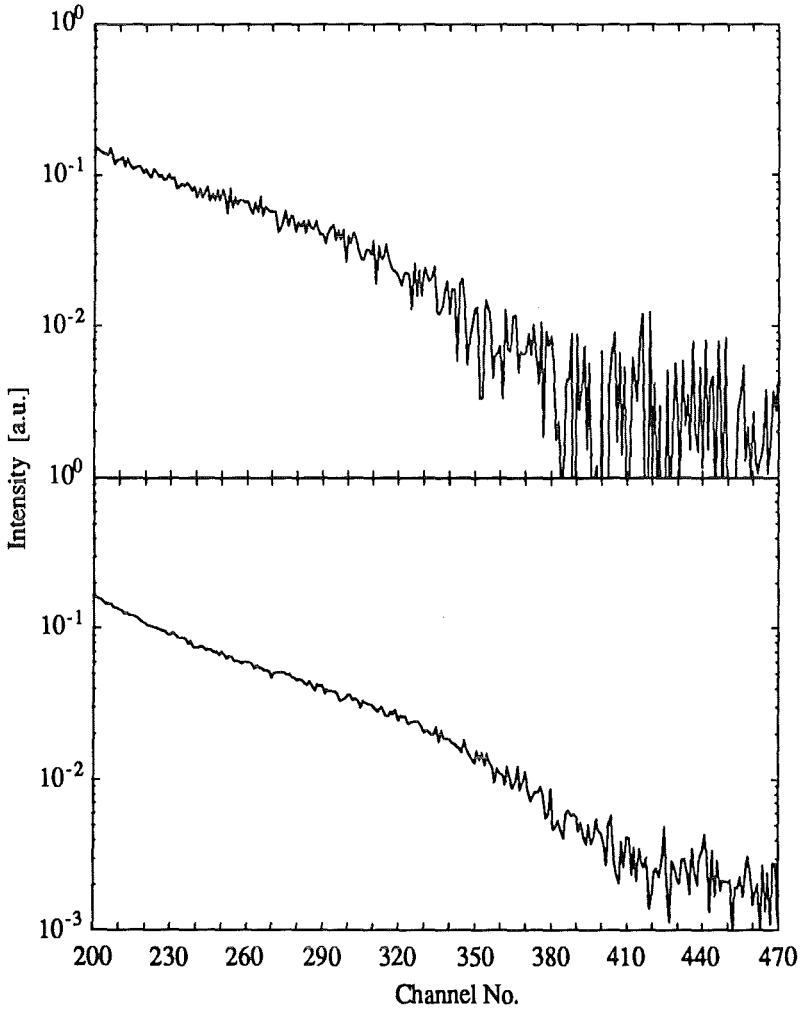


Figure 5.1 : Improvement of the statistics by averaging different data sets of a specific protein sample. A section of the outermost region of the scattering pattern of diferric human serum transferrin is shown after $15min$ (top) and $90min$ (bottom) data collection time.

5.6 Calibration

The position calibration was accomplished by the diffration peaks from a wet rat tail collagen specimen, as shown in figure 4.6. The pattern contained up to 22 orders of diffraction. The recorded refelection orders correspond to a spacing of $(670Å)^{-1}$. To create an x-axis channel numbers had to be converted into the reciprocal space parameter. A plot of the s values (spacing referring to a certain collagen order) as a function of the position of the order (channel number) is shown in figure 5.2 (the position of the detector centre was also included). A polynomial fit of 6^{th} order guaranteed an accurate calibration over the entire measured s-vector range.

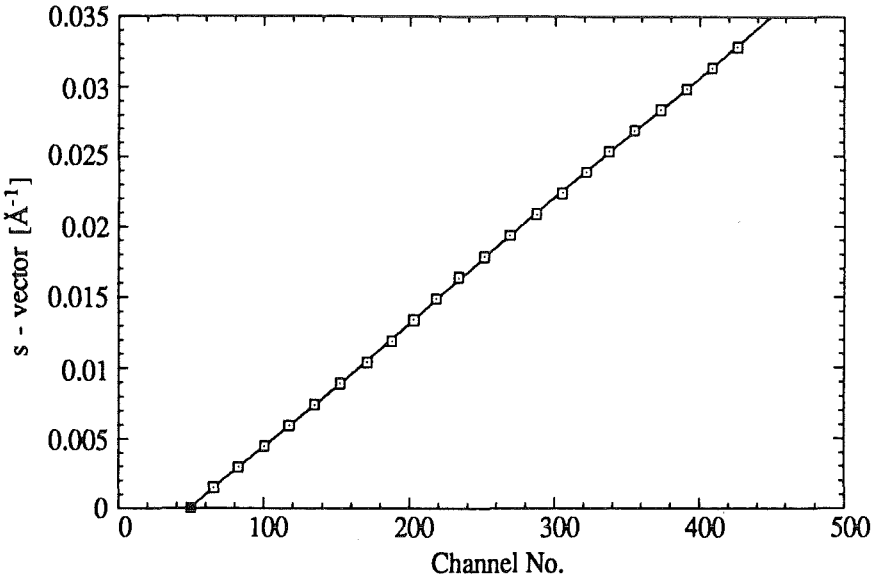


Figure 5.2 :
Plot of the detector channels where the collagen diffraction peaks are observed as a function of the reciprocal space parameter s. The polynomial fit of 6^{th} order included the position of the incident X-ray beam (position of the detector centre).

Detector dimensions and the camera length of $3m$ defined the distance scale probed by the experiments. In view of the indispensable beam stop and limitations in the detector's sensitivity, reliable scattering data were collected in the range $0.005\text{Å}^{-1} < s < 0.035\text{Å}^{-1}$. Thus, according to the sampling theorem [67], the highest spatial resolution r_{res} which can be achieved, is $r_{res} = (2s_{\max})^{-1}$, i.e. 14.3Å. Considering a wavelength of 1.54Å the scattering angles corresponding to $s < 0.035\text{Å}^{-1}$ are smaller than 3.1°.

5.7 Radius of Gyration

The small-angle part of the data was used to scale the scattering curves at zero angle against each other and thus enabled a comparison between spectra resulting from metal-loaded and metal-free transferrins or from the two different species of nitrite reductase. According to the Guinier approximation (see equation E3.4) the intensity distribution in the innermost scattering region can be expressed as

$$\ln(I(s)) = \ln(I(0)) - \frac{4\pi^2}{3} R_g^2 s^2 \qquad \text{(E5.1)}$$

where $I(0)$ denotes the intensity at zero scattering angle or forward scattering and R_g, the radius of gyration, is related to the overall shape of the particle. $I(0)$ is not directly accessable because of the beamstop, which protects the detector against radiation damage. A linear fit to the small angle scattering data in the so-called Guinier plot, $\ln(I(s))$ versus s^2, provided both quantities, $I(0)$ and R_g. Thus, the value of R_g can be extracted without any assumptions regarding the structure of the molecule. The low scattering region is usually defined by the inequality $s \cdot R_g < 1/\pi$. However, the maximum scattering vector, which may be included, depends on the shape of the molecule and can be estimated by the

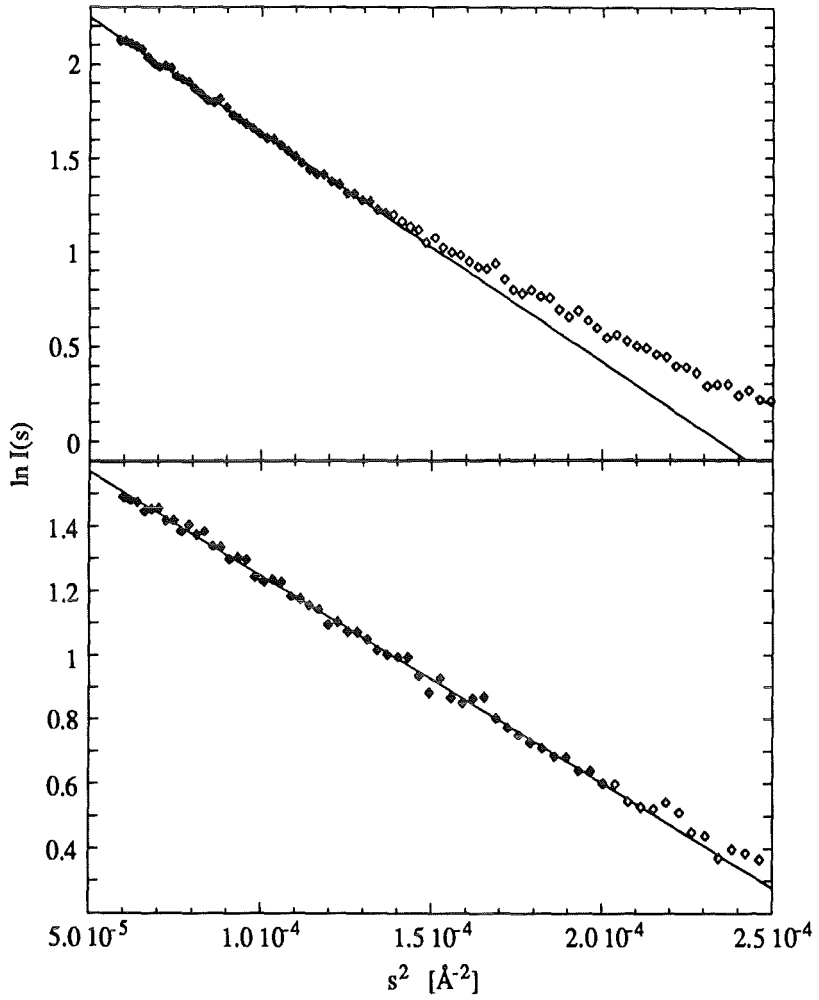


Figure 5.3 :
Typical Guinier plots for (top) the intact protein and (bottom) the N-lobe fragment of iron-free chicken ovotransferrin (protein concentration $5mg/ml$) obtained after an exposure time of $15min$. Only the data points in the linear Guinier range, shown as filled symbols, were used to calculate R_g. It should be noted that the Guinier range extends to higher s-values for the smaller N-lobe fragment ($s \leq 0.014$ Å^{-1}) compared with the whole protein ($s \leq 0.012$ Å^{-1}).

linearity of the Guinier plot. The value of $I(0)$ was used to normalize the scattering curves to unity at zero scattering angle. Figure 5.3 shows two typical Guinier plots, one for the intact protein and the other for the N-lobe fragment of chicken ovotransferrin (COT). The linearity of these plots in the low-angle regime ensured the reliability of the normalization procedure.

The radius of gyration was also calculated from the distance distribution function (see equation E3.7). This method can yield more accurate results than the Guinier approximation as the whole available scattering curve can be included in the calculations.

5.8 Distance Distribution Function

The distance distribution function $p(r)$ was evaluated using the indirect transform method, as implemented in the programme GNOM[10]. The main equations relating the scattering

[10] Due to the limited s-vector range a conventional Fourier transform (see equation E3.5) can result in strong artificial oscillations in the $p(r)$ function that it becomes useless. This termination effect can be minimized with the help of the indirect transform method. However, its main difficulty is that small errors in $I(s)$ may lead to large errors in $p(r)$. For this reason a regularization technique was introduced to stabilize the solution [66] and to construct a criteria for the quality of the solution [105].

intensity to the distance distribution function were already described in section 3.2.2. The data treatment with GNOM allowed a reliable estimation of errors with the help of a polynomial smoothing procedure. The search for the $p(r)$ function of a monodisperse system in real space was performed without taking account of effects such as wavelength smearing or beam divergence.

The following parameters can influence the results decisively: s_{min} and s_{max}, which define the interval in reciprocal space and D, which defines the interval in real space, where the distribution function is assumed to be non-zero. Although the experimental setup determines the scattering range and has to be adapted accordingly, s_{min} was selected so that protein aggregation resulting in elevated intensities close to zero scattering angles was excluded. In view of the low protein concentrations used, calculations could be performed for all samples, apart from lactoferrin, with the entire scattering range between $0.005\mathring{A}^{-1} < s < 0.035\mathring{A}^{-1}$. Lactoferrin tends to aggregate at the measured pH and thus, the lower s-vector limit came to $0.008\mathring{A}^{-1}$.

Since GNOM is based on the indirect transform technique, the maximum particle diameter D is required as an input parameter, though being an unknown structural quantity. For this reason D was chosen in a trial-and-error procedure so as to obtain a plausible solution by determining the optimum value of the regularization parameter [105]. Of course, *a priori* information from crystallographic structures proved helpful to assess the reliability of the solution.

5.9 Scattering Pattern Calculation and Computer Modelling

Theoretical scattering profiles were computed from crystallographic coordinates[11] of HLF [25, 26], RST [28, 29] and AxNiR [50] including all non-hydrogen atoms. According to the explanations given in section 3.4, FORTRAN programmes have been written and run on the CONVEX C220 at the Daresbury computing facility. The more versatile programme DALAI which has been developed at the Daresbury Laboratory [108] could also be used for test purposes. It allows the treatment of structural models in form of closely packed spheres that can represent units as large as individual domains or as small as individual atoms of a molecule.

A pre-requisite for the calculations is the binning of intraparticle distances (a binsize of 0.1Å was used) and results in the maximum particle dimension D. The radius of gyration R_g

---

[11] Crystal structure data from the proteins that have been investigated during the course of the present work were kindly provided by the authors prior to deposition in the Brookhaven Protein Data Bank (PDB) [106, 107]. This computer-based archive contains at present over 800 protein coordinate data sets.

was also calculated from the crystallographic atomic coordinates using equation E3.3 in the form

$$R_g = \sqrt{\frac{\sum_i z_i R_i^2}{\sum_i z_i}} \quad . \tag{E5.2}$$

Here z_i is the atomic number of atom i and R_i is the atomic distance from the centre of the electron charge distribution in the molecule.

In order to assess and interpret the measured scattering spectra computer modelling was performed on a Stardent TITAN and Silicon Graphics INDIGO workstation using the molecular graphics and modelling programmes BIOGRAF (Version 2.0, BioDesign Inc., CA, USA) and INSIGHT (Version 2.2, Biosym Technologies Inc., San Diego, CA, USA). Starting from the crystal structure coordinates various modifications have been explored to study the principal features of the scattering profiles observed for the proteins in solution. The modelling procedure will be reported in detail in the corresponding sections of the following chapter.

Chapter 6 : Results and discussion

6.1 Transferrin

X-ray crystal structures have been reported from a number of members of the transferrin family (see Chapter 2.4). The nature and extent of the large conformational change accompanying iron uptake and release gave ample insights into the mechanism of molecular flexibility. However, the fact that parts of the N-terminal half appear to move substantially whereas no conformational change appears to take place in the C-terminal half, was a most curious feature revealed by the crystal structure of the metal-free state of human lactoferrin [26]. So far, only limited structural information was available on the conformation of these molecules in solution [30-32]. Solution X-ray scattering is a well established technique and, with the availability of a dedicated synchrotron radiation source, the determinination of overall conformations and domain movements of the transferrins with respect to metal binding is feasible under conditions close to physiological environment.

6.1.1 X-ray scattering from intact molecules

The high intensity from the SRS at Daresbury and the use of a quadrant detector have allowed measurements of X-ray scattering data for dilute transferrin solutions with reliable statistics up to scattering vectors of 0.035 $Å^{-1}$. For the first time the experimental scattering profiles of Human Serum Transferrin (HST), Chicken Ovotransferrin (COT) and Human Lactoferrin (HLF), both in the iron-free and iron-loaded state (shown in figure 6.1(a) to (c)), could be compared directly in this scattering range. The scattering data for the three transferrin species show a major change in the intensity curves when iron is bound. While the differences in the two conformations are scarcely resolved in the small-angle range, they are clearly evident in the region of medium s-vector values ($0.020\ Å^{-1} < s < 0.035\ Å^{-1}$). The observed differences between the apo- and holo-proteins are very similar for the three species, suggesting comparable conformational changes when iron is bound. This result is also supported by the distance distribution functions (see insets in figure 6.1).

In view of the close sequence homology among the three proteins (ranging from 50% to 60% identical residues in corresponding positions [16]), the experimental scattering profiles of the transferrins were compared to theoretical profiles based on the crystal structures of apo- and diferric HLT. The calculated scattering curves I(s) as well as their radial distribution functions *p(r)* are shown in figure 6.2. They agree favourably with the overal changes observed in the experimental data. In particular, two characteristic features, the intersection at $s \approx 0.024\ Å^{-1}$ and the noticeable difference between the apo- and holo-forms in the medium s-vector range are consistent with the observed scattering profiles for proteins in solution. Nevertheless, a detailed comparison of the theoretical curves (based on

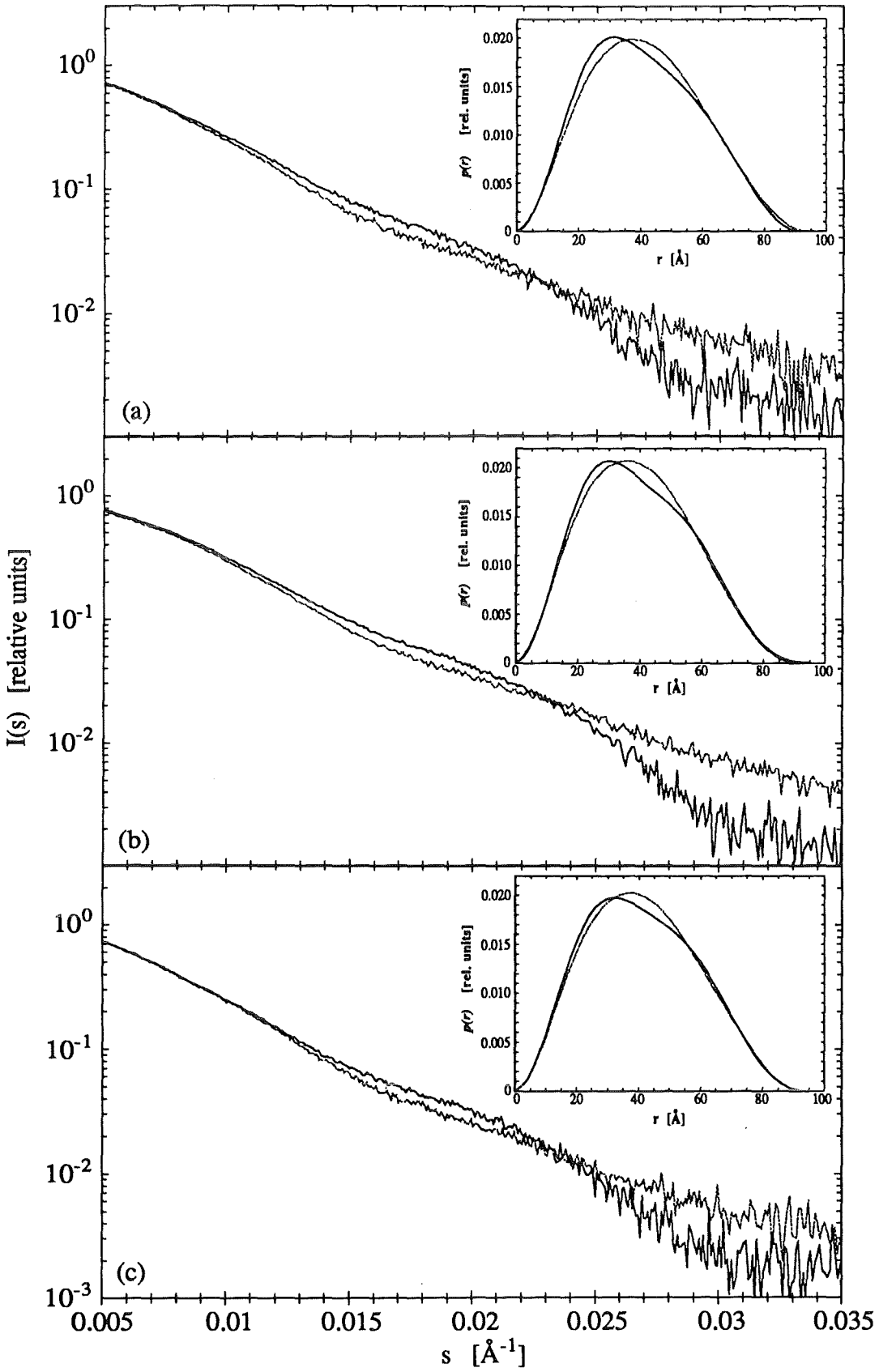


Figure 6.1 :
Experimental results for the iron-free (dotted lines) and iron-loaded (solid lines) samples of intact transferrin molecules. The scattering profiles as well as the distance distribution functions (insets) of (a) HST, (b) COT and (c) HLF are shown.

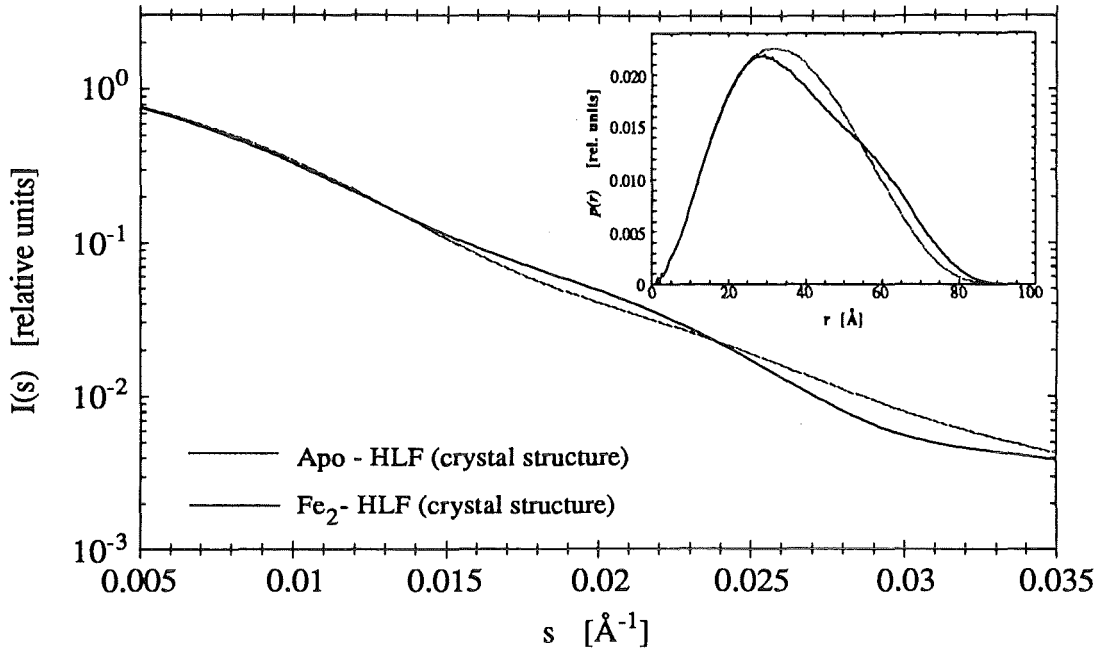


Figure 6.2 :
Calculated scattering spectra and distance distribution functions based on the two crystallographic conformations of HLF. Like in figure 6.1, all scattering profiles have been normalized at s=0 and the *p(r)* functions were plotted relative to each other by normalizing the area under each curve to unity.

crystal structure coordinates) with the experimental data reveals significant differences : the intersection at the small scattering angle corresponding to s=0.013 Å^{-1} is not observed in the experiment. Moreover, the radial distribution function of the crystallographic apo-structure appears to be shifted towards smaller distances. This underlines a subtle difference between the conformational changes due to the metal uptake in the crystalline state and in the aqueous solution, respectively. This difference can also be deduced from the Guinier analysis (see table 6.1). Whereas the holo-form of the transferrins in solution seems to be more compact, characterized by a smaller radius of gyration R_g (as previously observed by Kilár & Simon [31] and Vígh *et al.* [32]), the opposite behaviour was deduced from changes observed in the apo-HLF crystal structure. The R_g values for HST agree rather well with those published for apo- and holo-HST [32], respectively, whereas for apo- and holo-HLF the values are somewhat larger and the difference between the two protein states is more pronounced.

The experimental values for R_g are significantly higher than those calculated with the atomic coordinates of the crystal structures. This difference can be due to several factors e.g. in the theoretical curves no account has been taken of the carbohydrate structure of the glycoprotein, the hydrated volume and possible molecular aggregation. The carbohydrate content of the three transferrins certainly contributes to the differences in the measured R_g

Table 6.1 : Radii of gyration for intact transferrin molecules. The experimental R_g values were obtained by Guinier approximation of the scattering curves observed for a protein concentration of 5 mg/ml. The standard deviations given were derived by the average of different data sets. The theoretical R_g values were computed from the atomic coordinates (according to equation E5.2).

sample	R_g [Å]	ΔR_g [Å]
Experimental		
Apo - HST	32.5 ± 0.2 (33.0 *)	
Fe_2 - HST	31.4 ± 0.2 (31.5 #)	- 1.1 (- 1.5 #)
Apo - COT	30.5 ± 0.2	
Fe_2 - COT	29.7 ± 0.2	- 0.8
Apo - HLF *	33.3 ± 0.2 (36.4 #)	
Fe_2 - HLF *	33.0 ± 0.2 (34.0 #)	- 0.3 (- 2.4 #)
Theoretical		
Apo - HLF	28.4 (29.8 ¶)	
Fe_2 - HLF	29.5	+ 1.1 (- 0.3 ¶)

* Lactoferrin displays a marked tendency to aggregate which is possibly due to its high isoelectric point (table 2.1); # values for zero concentration as reported by Vígh *et al.* [32]; ¶ calculated from the modelled structure of a fully opened apo-HLF molecule, i.e. where the C-lobe is represented by an open conformation analogous to the N-lobe (see section 6.1.3).

values among the species. Both HLF and HST have two sugar side chains (HLF : one in N- and C-terminal half, respectively, HST : both in the C-terminal half), whereas COT only has a single glycan unit in the C-terminal half (see table 2.1). Furthermore, strongly bound water molecules (by two or more hydrogen bonds) form a hydration shell surrounding the protein with properties different from bulk water. In general, these water molecules create an ordered surface layer in order to keep an average orientation [83]. In addition, surface-exposed areas of the three subclasses of the transferrin family show quite different properties (indicated by different isoelectric points, see also table 2.1). Thus, since the carbohydrate, which has not yet been resolved in the crystal structure, presumably due to its high mobility or disorder, and the dynamic solvent shell around the molecules have not been considered in the calculations, experimental and simulated data are not in full agreement (different absolute intensities). However, including this contribution to X-ray scattering will only modify the overall size of the molecule but not the internal subunit arrangement. The effect of hydration on the X-ray scattering profile will be discussed in connection with the N-lobe of HST in section 6.1.5.

6.1.2 X-ray scattering from N- and C-terminal fragments

Individual N- and C-terminal fragments of COT offer an excellent opportunity for investigating the differences in behaviour which may exist for the two lobes when iron is taken up. Figures 6.3(a) and (b) show the scattering patterns for the two lobes in the apo- and holo-state. No additional information was provided by the distance distribution functions, which are therefore not presented. It is obvious that both lobes undergo similar conformational changes. Despite differences in the absolute intensities scale (as discussed above), the observed scattering pattern for the N-lobe can be simulated closely using the crystallographic information for the N-lobe of the intact HLF molecule. In contrast, the data for the C-lobe cannot be reproduced when similar information from the crystallographic structures is used (figure 6.3(c) and (d)).

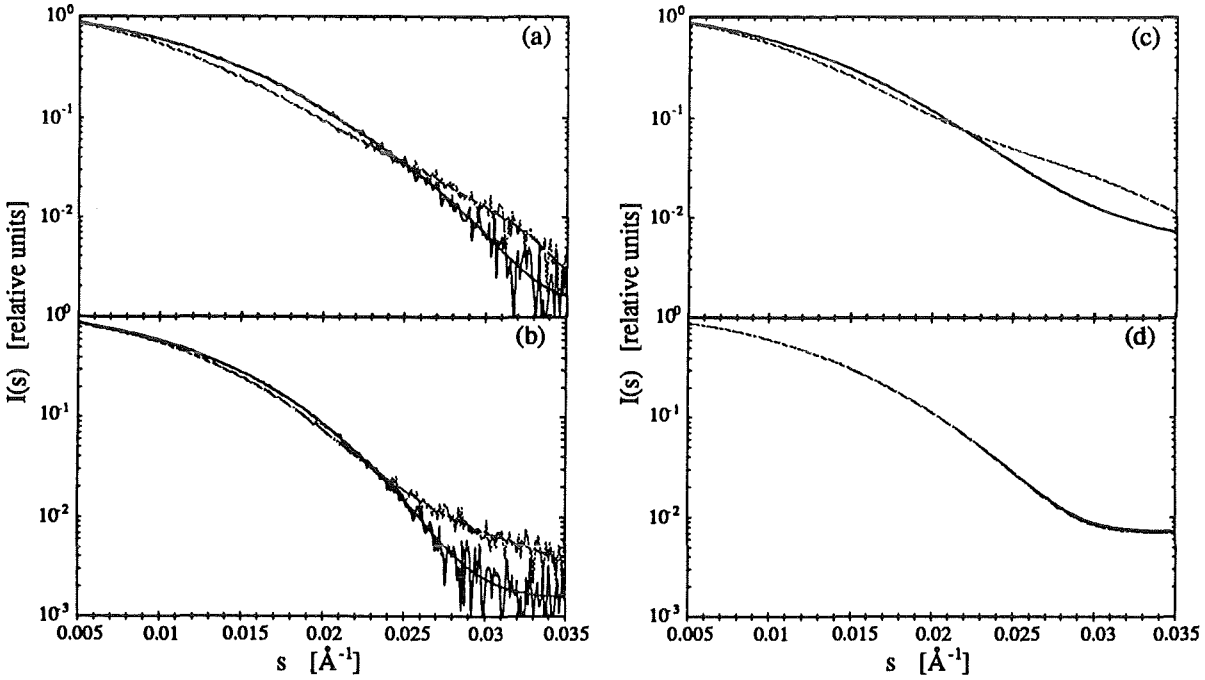


Figure 6.3 :
Comparison of the experimental scattering patterns for the apo (dotted lines) and ferric (solid lines) N-terminal (a) and C-terminal (b) fragments of COT with the results of the simulations obtained from the N-terminal (c) and the C-terminal (d) crystal coordinates of HLF in the iron-free (dotted lines) and iron-loaded (solid lines) configurations. In order to guide the eye, spline fits to the experimental scattering curves are given.

The results of the Guinier analysis in the low angle range were compiled in table 6.2. The experimental and theoretical R_g values for the N-lobe of COT agree within 0.4 Å, whereas they differ by about 1.5 Å in the case of the C-lobe. It is interesting to note that the only glycosylation site of COT is located in the C-lobe and indeed may be responsible for the higher R_g value in the experimental data.

Table 6.2 : Radii of gyration for N- and C-lobe fragments. The experimental R_g values were obtained at a protein concentration 5 mg/ml.

sample	R_g [Å]	ΔR_g [Å]
Experimental		
Apo - N - COT	22.0 ± 0.2	
Fe - N- COT	20.1 ± 0.2	- 1.9
Theoretical	21.6	
Apo - N - HLF		
Fe - N - HLF	19.7	- 2.1
Experimental		
Apo - C - COT	21.8 ± 0.2	
Fe - C- COT	20.8 ± 0.2	- 1.0
Theoretical		
Apo - C - HLF	19.3 (21.3 *, 20.3 #)	
Fe - C - HLF	19.3	± 0.0 (- 2.0 *, - 1.0 #)

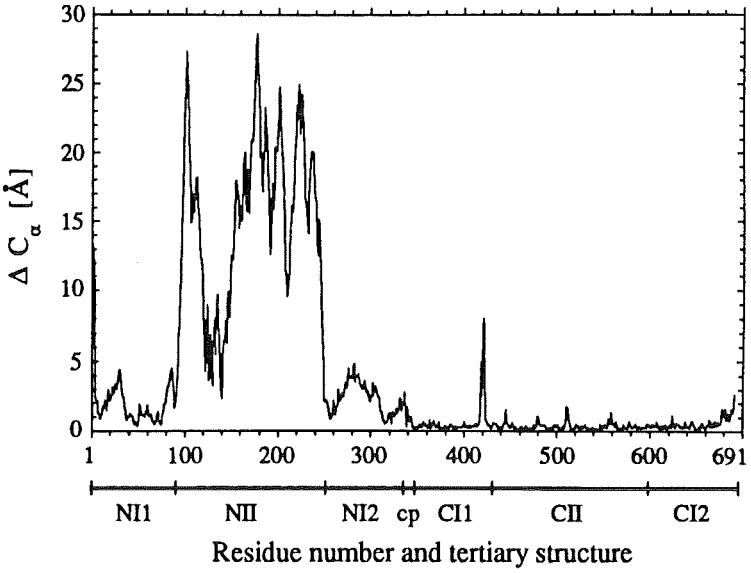
* calculated from a modelled 'open' C-lobe structure with an opening equivalent to that of the N-lobe of HLF observed in the crysrtal structure data; # computed with an opening equivalent to 75% of the N-lobe apo-HLF crystal structure (see section 6.1.3).

The experimental data for the apo- and holo-forms of individual lobes of COT provide clear evidence that very similar conformational changes take place in both forms when iron is bound. In view of the good agreement between the calculated and experimental scattering data for the N-lobe, an attempt was made to achieve a similar opening for the apo-C-lobe.

6.1.3 Computer modelling and simulation

In order to assess the measured differences between iron loaded and iron free transferrins, computer modelling was performed on a TITAN Graphics workstation (Stardent Computer Corporation) using the molecular graphics programme BIOGRAF (Version 2.0, BioDesign Inc., California, USA), starting from the crystal structure coordinates obtained from the 2.8 Å resolution analysis of Human Lactoferrin. HLF is divided into two lobes : the N-lobe consists of amino acids 1-333 and the C-lobe includes 345-691. The remaining eleven residues make up the connecting peptide, in the case of HLF a three-turn α-helix. The N- and C-lobe can be further subdivided into domains (NI, NII, CI, and CII) of approximately

Figure 6.4 :
Difference in C_α positions between the two crystal conformations Fe_2-HLF and apo-HLF after superimposing the main chain atoms of the corresponding C-terminal lobes using a least-squares fit. The fact that the C-terminal half keeps a closed conformation even in the apo-state, guarantees the best superposition. The difference plot shows large displacements of domain NII whereas domain NI as well as the connecting peptide (cp) seem to undergo only minor movements.



160 residues. As a consequence of the high sequence homology (identity about 40% [16]), the two lobes have similar structures. In addition, the close structural similarities of the domains (NI, CI, and NII, CII, respectively) are highly favourable in terms of transferring conformational features from one lobe to the other. This procedure (see below) was used to model a possible opening of the C-terminal half in the apo-protein. A graphical representation of the conformational difference between apo- and diferric HLF in the crystalline state can be seen in figure 6.4, which has been obtained by minimizing the difference of the C_α positions for the CI and CII domains of the apo- and holo-HLF. An 'opened' C-lobe was created essentially by a displacement of domain CII as follows. The main chain atoms of 118 amino acid residues belonging to domain CI were superimposed to domain NI (the root-mean-square (r.m.s.) deviation is 0.75 Å/atom). Keeping the positions of domains NI, CI, and NII of the apo-HLF crystal structure fixed, only domain CII was moved in order to match domain NII. This was done by superposition of the 113 amino acids forming part of domain CII with equivalent residues of domain NII (the r.m.s. for the main chain is 0.59 Å/atom). The newly generated co-ordinates for the C-lobe were used to form an intact ('completely opened') apo-HLF structure in solution by simply replacing the coordinates of the C-lobe in the crystal structure by the modelled coordinates of the open configuration.

It needs a mention that these manipulations are structurally sensible and do not cause steric hindrance between different parts of the molecule. The main constraint imposed upon the modelling seems to be a disulphide bridge (483-677), which is only present in the C-lobe and joins domains CI and CII. The resulting distance between the C_α-atoms of the half-cystines was calculated to be 7.5 Å compared to 5.3 Å for the closed conformation in

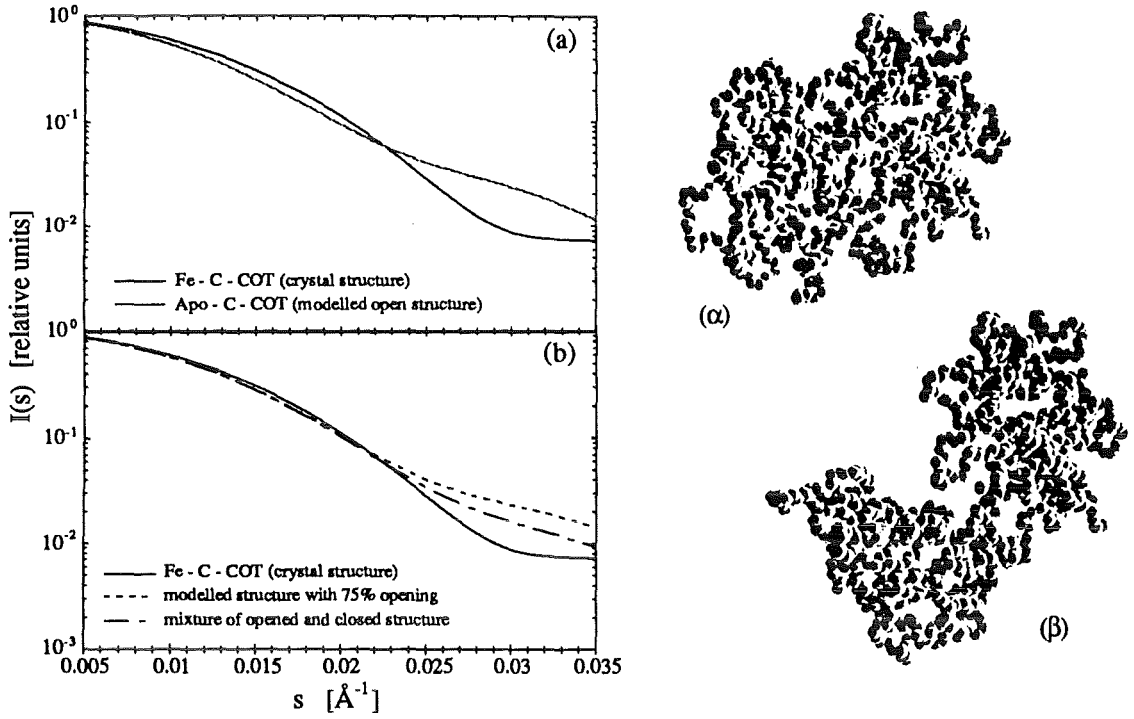


Figure 6.5 :
In order to explain the discrepancy between the experimental solution scattering results and the crystallographic calculations the scattering pattern of an opened C-lobe was simulated (a) and compared to the liganded closed state. On the right hand side the space-filling models (only main chain atoms) of the closed crystal structure (α) and simulated open structure (β) of the metal free C-lobe are shown. The theoretical scattering profiles (b) of a C-lobe intermediate, which is defined by a 75% opening compared to the N-lobe, or a mixture of an opened and closed conformation can also be used to interpret the experimental findings.

the iron loaded state. This could be achieved by increased conformational strain in the disulphide bonds and to some extent by modification in the neighbouring regions.

The agreement between the scattering profile from the modelled open structure and the experimental curve has improved considerably (see figure 6.5(a)). An even better overall correspondence could be achieved by an opening of the apo-C-lobe equivalent to 75% of the opening of the N-lobe (the value of ΔR_g, which is a more reliable quantity than the absolute R_g value, is in excellent agreement with the observed difference, table 6.2). Thus, the structural mechanism responsible for the large-scale conformational change appears to be quite similar in both lobes. However, since the nature of the experiments is such that information is obtained about time and space-averaged molecular configurations, the possibility that the solution X-ray scattering results arise from a mixture of closed and opened conformations cannot be ruled out (figure 6.5(b)).

Nevertheless, there is no doubt about a distinct wide opening of the C-lobe *in solution* similar to that observed for the N-lobe in the crystal structure of apo-HLT when iron is not bound. This is furthermore supported by looking at the intact protein : As a consequence of the modelling studies, differences for the whole protein in the apo- and holo-state are also reproduced accurately when an opened C-lobe conformation (analogous to that of the N-lobe) is assumed for the intact apo-protein of HLF (figure 6.6). In this case, the apo-protein structure represents a less compact conformation compared with the holo-protein (for this model a larger radius of gyration for the apo-protein (29.8Å) compared with the closed iron bound configuration (29.5Å) could be derived, table 6.2).

The question arises, how an intact transferrin molecule where one lobe is in an open and the other in a closed conformation, would affect the experimental X-ray scattering data. An example may be a variant human serum transferrin identified by Evans *et al.* [109]. This variant has been found to differ markedly from the normal type of transferrin in its iron binding properties in the C-terminal half. Keeping in mind the simulation of apo-HLF using the crystal structure and modelled open structure (figure 6.6), the solution X-ray scattering results presented in figure 6.7 indicate a conformation where both lobes cannot be fully closed even in the presence of two bound iron atoms. This is also suggested by the analysis of the R_g values: 32.4 Å for the iron-free variant (which agrees well with the native protein,

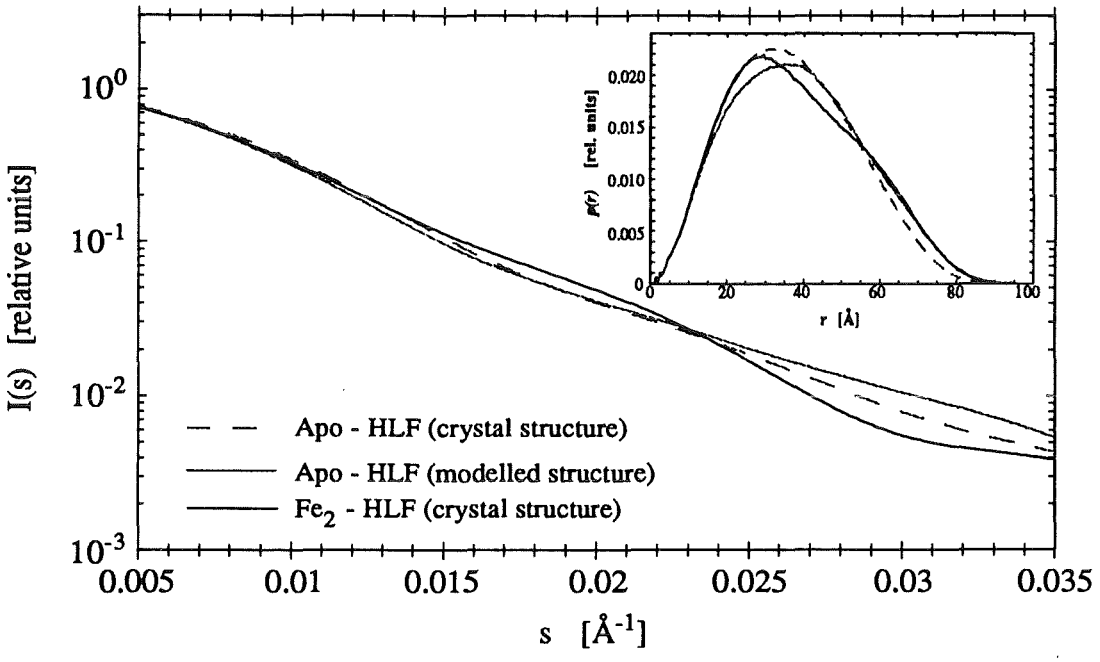


Figure 6.6 :
Comparison of the scattering profiles and radial distribution functions for HLF calculated for the closed iron-loaded conformation (crystal structure) and the iron-free conformation with N- *and* C-lobe opened (modelled structure). From this simple model using the coordinates of all non-hydrogen atoms the essential characteristics of the experimental results (see figure 6.1(c)) could be reproduced. The curves resulting from the apo-HLF crystal structure (N-lobe open and C-lobe closed) are also included.

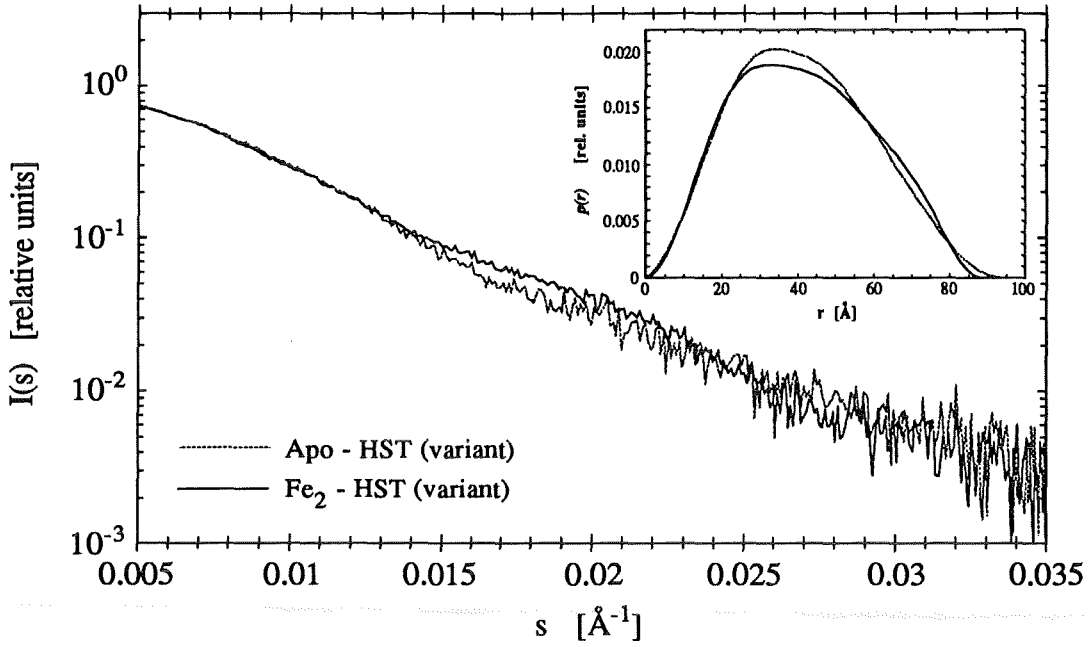


Figure 6.7 :
Scattering profiles and distance distribution functions for the apo- and diferric state of a variant of HST (5 mg/ml). The scattering data of the differic protein appears to result from a closed N-lobe and an open C-lobe (see text).

table 2.1) and 32.7 Å for the diferric variant. As the N-lobe is expected to be normal [110] (i.e. closure of the interdomain cleft in the presence of iron), differences may be due to an open C-lobe in the iron-bound variant protein. However, only for $s \geq 0.023$ Å^{-1} the scattering data for the diferric variant protein resembles the calculated scattering data obtained from the crystal structure of human apo-lactoferrin. The pronounced difference around $s = 0.017$ Å^{-1} points out a different nature of domain and lobe arrangements[12] as known from the conformation with one opened and one closed lobe.

The conformational behaviour of the variant protein can be rationalized at an atomic level. Sequence studies [110] revealed a C-lobe in which a glycine (Gly394), two residues after the aspartic acid (Asp392) involved in iron binding, is replaced by an arginine[13] . It is likely that the positively charged guanidinium moiety of the arginine will interact with the negatively charged carboxylate group of Asp392. Such an interaction will certainly prevent the normal binding of Asp392 to the iron as well as the formation of interdomain hydrogen bonds [111], which are thought to be crucial on closing the interdomain cleft upon binding of iron. Thus, the C-terminal domains are left in an 'open' configuration in the iron-bound

[12] The effects of the connecting peptide joining both lobes and subsequent differences in the arrangement of lobes with respect to each other are still difficult to determine and thus have not been taken into account. By considering only a single lobe of transferrin (e.g. N-lobe), the possibilities of structural rearrangements upon metal binding are reduced and modelling studies are 'easier' to deduce (see section 6.1.5).

[13] See also figure 2.2. Most of the residues located in the iron binding cleft are conserved in both lobes, so that only the numbering of the residues needs to be adapted, e.g. Asp392 in the C-lobe corresponds to Asp63 in the N-lobe.

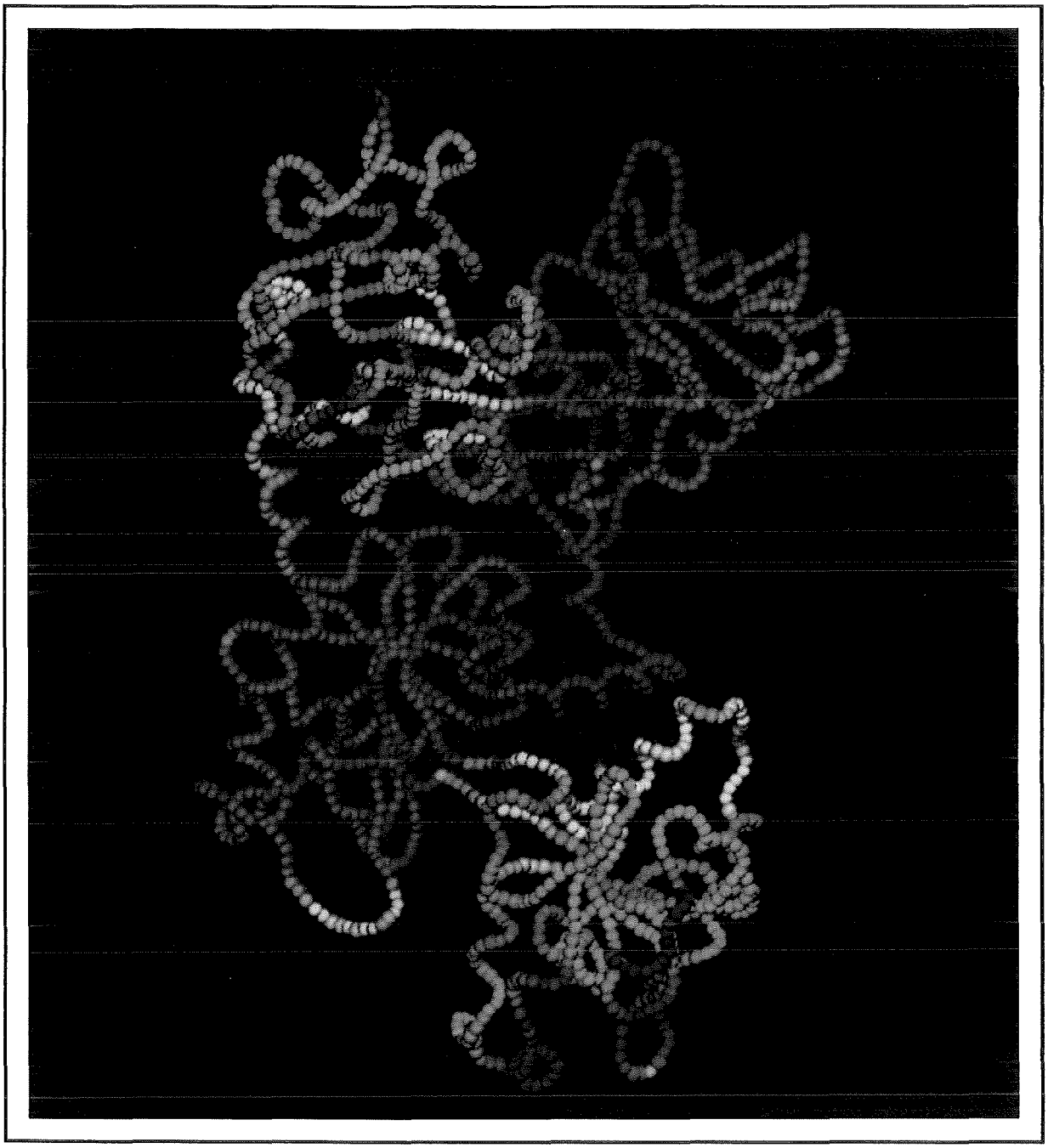


Figure 6.8 :
Smoothed backbone structure of the completely opened apo-HLF molecule as modelled for solution scattering. The movement of residues relative to diferric HLF crystal structure, is represented by colour changes from red to yellow to green to cyan to blue to magenta. Red represents minimum movement. In order to illustrate the conformational change with respect to the diferric structure, a superposition of the main-chain atoms of 236 amino acid residues belonging to the core of the domain NI and CI was performed for the two structures (apo- and Fe_2-HLF). The least square fit resulted in a r.m.s. value of 0.95 Å/atom and a minimum and maximum movement of 0.4 Å and 32.5 Å. The differences between the minimum amd maximum movement were divided into 30 colours and are used for shading the structure of the fully opened apo-structure in solution. Large-scale movements of the domains II in both lobes (N-lobe at the top, C-lobe at the bottom) are clearly illustrated by the green and blue colours.

variant protein due to a break down of the hydrogen bonding network in the metal binding cavity resulting from a single mutation (Gly394 → Arg).

In summary, the use of solution X-ray scattering data in concert with crystal structure information allowed the successful determination of interdomain movement for the two lobes of transferrin. In the presence of iron, both sites are found in a closed configuration (with the exception of a variant of human transferrin), consistent with crystallographic data. However, in the absence of iron, both the N- and C-terminal site open substantially. This behaviour could be deduced from studies of the intact protein and from the individual C- and N-lobes of COT. The large-scale movements of the subdomains in the apo-protein relative to the structurally stable liganded conformation of the diferric protein is demonstrated in figure 6.8.

The conformational changes are likely to be of functional importance and may play a crucial role in receptor recognition. The mechanism of transferrin binding to the receptor as well as the effect of the receptor on iron release from transferrin is yet not well understood. It is known that under physiological conditions (pH 7.4) the affinity of the receptor for diferric transferrin is 25 times higher than for apo-transferrin [112]. Interestingly, a ten times reduced binding constant was found for the complex of the receptor and the iron-loaded variant transferrin discussed above [113]. It is tempting to suggest that the closed conformation of the iron-loaded protein increases the specifity of receptor binding and thus facilitates the incorporation of iron into the iron-requiring cells.

6.1.4 Transferrins loaded with other non-physiological metals

Under physiological conditions *in vivo*, only about 25% to 30% of the iron-binding sites of serum transferrin are occupied [114]. Besides iron, a variety of non-physiological metals (first series transition metals, main group elements, lanthanides and actinides) are capable of binding specifically to the transferrins as demonstrated extensively by a variety of spectroscopic methods (see references in [13]). Two of these metals, Ga and In, are exploited for medical use (tracer and imaging studies), most of the other metals are of toxicological interest, e.g. Al, which is associated with senile plaques in Alzheimer's disease [115] and the actinides which are present in minute quantities in our normal environment as well as potential hazard due to nuclear fall out and their use as nuclear fuel. Although many non-physiological metals, including plutonium, are bound to transferrin in the blood [116, 117] and are often found intracellularly bound to ferritin, they do not follow the metabolic pathways of iron[14]. Despite sharing the same transport protein, i.e. transferrin, non-physiological metals show a significant organ specificity and primarily deposit in the

[14] One of the primary functions of serum transferrin is the transport of ferric iron from sites of absorption and storage to sites of utilization, e.g. from the liver to the bone marrow for incorporation in haem (about 80% of the circulating iron is found in the reticulocytes of the red bone marrow [118]).

liver and on the bone surfaces [119]. Thus, if the closed conformation of iron-loaded transferrin is the key element for receptor recognition, it is of interest to know whether transferrin undergoes similar conformational changes when metals other than iron are bound.

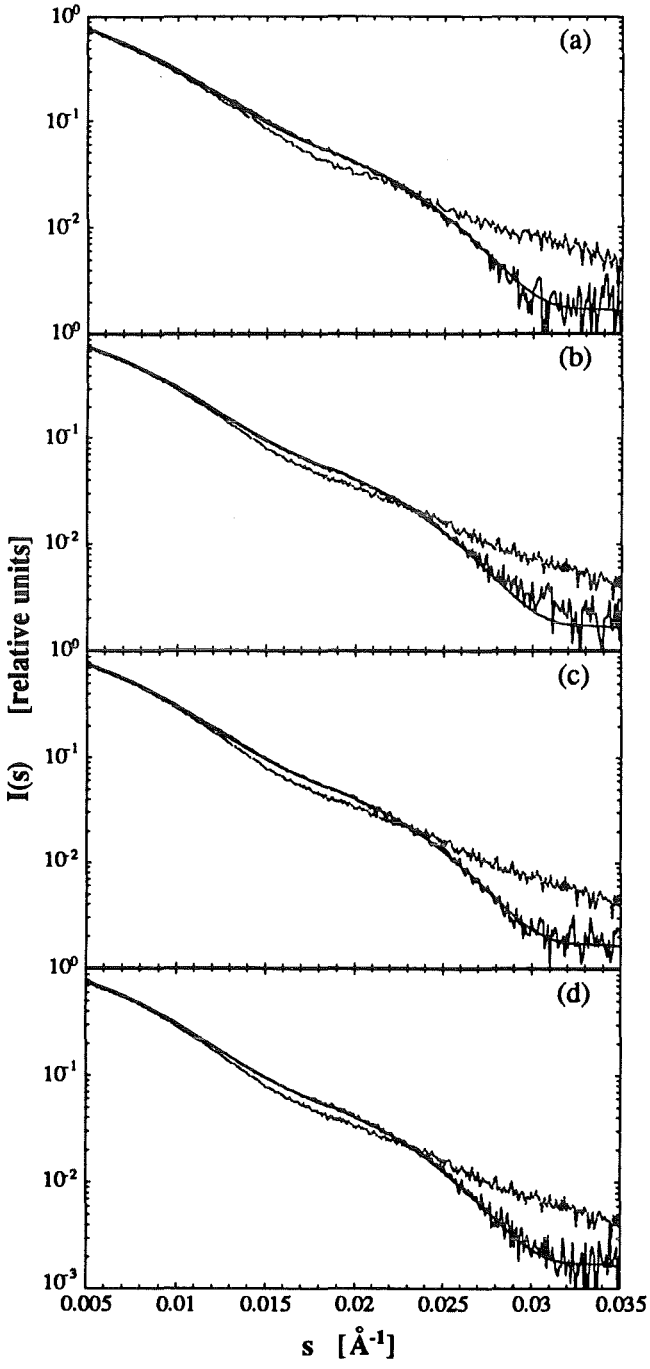


Figure 6.9 :
Solution X-ray scattering profiles of (a) Al_2-, (b) Cu_2-, (c) Ga_2- and (d) In_2-COT (solid lines) compared with Apo-COT (dotted lines) and Fe_2-COT. For the sake of clarity the latter is represented in each plot by the smooth curve resulting from a spline fit to the data shown in figure 6.1(b).

Solution X-ray scattering experiments have been performed to find out whether the uptake of Al, Cu, Ga, Hf, In and Th causes subdomain movements similar to iron. Hf and Th were used as they have been extensively studied as a convenient analogue of Pu because of its similar size, tetravalent character and metabolic behaviour [120, 121]. UV/Vis spectroscopy clearly indicates the binding of two Hf^{4+} ions [122] as well as two Th^{4+} ions [123] per transferrin molecule.

The experimental scattering profiles for Al, Cu, Ga and In labelled COT in comparison to the iron-loaded and metal-free protein are shown in figure 6.9. It is clear that these metals induce conformational changes in COT similar to iron. However, surprisingly, no closure of the interdomain clefts takes place when Hf or Th is taken up by COT, instead an 'open' structure like that of the apo-protein is observed (figure 6.10). These findings are also confirmed by the analysis of the radii of gyration (see table 6.3). Analogous results have been obtained for HST and HLF (not shown).

The specifity of metal binding, e.g. binding one appropriate metal ion much more tightly than others, can be obtained from molar absorptivities derived from ultraviolet difference spectroscopy. In table 6.3 the metal binding affinity relative to iron is given for the metals studied here.

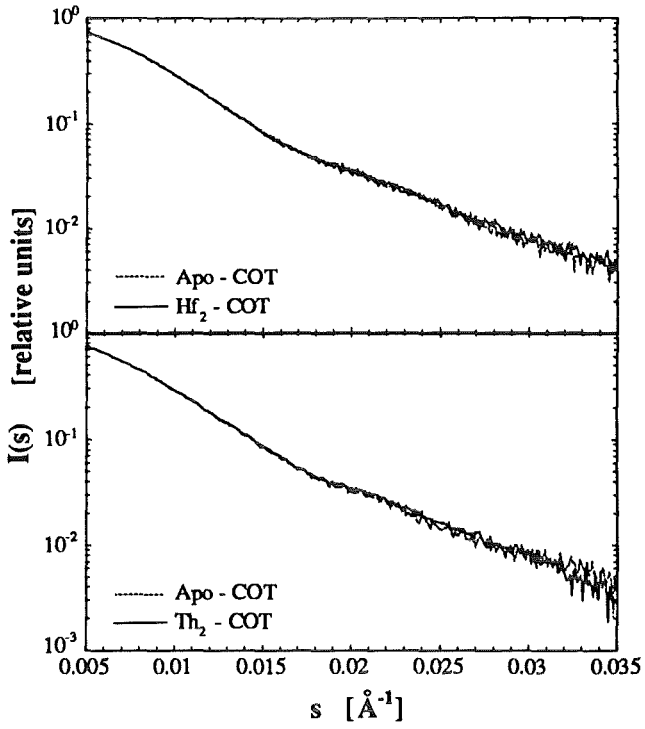


Figure 6.10 :
Comparison of the scattering patterns for Hf_2- and Th_2-COT with respect to the metal-free apo-protein.

It is obvious from the X-ay scattering data that the overall conformation of the metal-loaded protein explains the weak binding of Hf and Th. Thus, it is interesting to ask why the specific incorporation of Hf^{4+} or Th^{4+} does not induce the closure of the molecule. In table 6.3, the ionic radii of Hf^{4+} and Th^{4+} are compared with the radii of the other metals. There is no significant difference in the size of Hf^{4+} compared to In^{3+} which displays the closed conformation. In any case, the structure of dicupric lactoferrin in which the carbonate was substituted by oxalate [124] has demonstrated that the metal binding site has enough flexibility to accommodate larger ions. Thus, it is more likely that the tetravalent nature of Hf and Th and their preference for 8-fold coordination [10, 125] may cause significant changes in the metal binding region, disabling the trigger for the closure of the interdomain cleft.

Assuming that Pu, like Hf or Th, does not induce any conformational change upon binding to transferrin, its metabolic beviour can now be explained at a molecular level. In blood, Pu and many other non-physiological metals are transported by the iron carrier protein transferrin and subsequently deposited in several tissues, specific for the individual metal. This organ specificity has been suggested to be a function of the ionic radius of the metal [119]. For Pu, the target organs are the bone surfaces and the liver, where it is bound to the iron storage protein ferritin. Although this actinide is bound to the proteins of iron metabolism, it does not share the metabolic fate of iron. If transferrin-bound Pu (and many other non-physiological metals) followed the receptor mediated endocytosis process [126], as it is reported for iron transferrin, then red blood cells, which express a very high number

Table 6.3 : Radii of gyration for metal-loaded COT and some metal ion properties

metal ion	R_g* [Å]	metal binding strength # (relative to Fe^{3+})	outer electron configuration	main coordination number ¶	ion radius ‡ [Å]
Fe^{3+}	29.7	10^2	$3d^5$	6	0.65
Cu^{2+}	29.9	$< 10^{-1}$	$3d^9$	5 6	0.65 0.73
Al^{3+}	29.7	10^{-6}	$2p^6$	6	0.54
Ga^{3+}	29.8	10^0	$3d^{10}$	6	0.62
In^{3+}	29.8	10^2	$4d^{10}$	6	0.80
Hf^{4+}	30.4	$< 10^{-10}$	$4f^{14}$	8	0.83
Th^{4+}	30.4	$< 10^{-10}$	$6p^6$	8	1.05
no metal (apo)	30.5				

* R_g values were calculated using the Guinier approximation, the statistical error amounts to 0.2 Å; # only the order of magnitude is given because of missing quantitative considerations or slight disagreements in literature mainly owing to different experimental conditions; ¶ in transferrins Fe^{3+} is coordinated in a distorted octahedral geometry in both lobes, wheres Cu^{2+} coordination is square pyramidal in the N-lobe and distorted octahedral in the C-lobe; ‡ ion radii are taken from Shannon [128] and refer to the given coordination numbers.

of transferrin receptors, would be expected to contain Pu. However, this is not the case [127].

Considering the other metals (Al, Cu, Ga and In) it is difficult to infer the strength of metal binding from the X-ray scattering studies since these metals induce large-scale conformational changes similar to iron. Despite slightly different ionic radii, their most common coordination numbers from the stereochemistry of inorganic structures (table 6.3) clearly suggest the iron-like behaviour but cannot provide a basis for explanations on an atomic level (i.e. local differences in the ligand arrangement). Further structural work including X-ray crystallography and X-ray absorption fine structure (XAFS) is required to address this point. XAFS is a technique which is sensitive to the metal binding environment and like solution scattering has the advantage of allowing the study of the protein under quasi-physiological conditions.

Nevertheless, the question arises whether a protein structure which is the same for different bound metals (such as Al, Cu, Ga and In), implies that these metals could be delivered to cells equally well as iron. In the light of the scattering results, the overall conformation should be indistiguishable towards receptor recognition.

The similarity between the molecular conformation of Fe_2- and Cu_2-HLF has recently been observed crystallographically. The crystal structure of dicupric lactoferrin has been resolved to 2.1 Å resolution [27]. Despite the differences in the vicinity of the binding sites (in contrast to Fe_2-HLF where both metal sites reveal a distorted octahedral geometry, only the C-lobe Cu^{2+} is 6-fold coordinated, whereas the N-lobe Cu^{2+} is 5-fold coordinated;

this was also found by XAFS studies of Cu_2-COT (Garratt *et al.*, 1991)), the dicupric protein structure is the same as for diferric HLF. Thus, if the receptor recognition is provided through the molecular conformation, and if this copper complex was to circulate *in vivo* in the blood stream then it could possibly compete with diferric transferrin for the receptor mediated internalization into the target cells. It is, however, not clear whether the physiological conditions that are appropriate for iron release are equally suitable for copper. In any case, it is known that copper, once ingested, is bound by serum albumin which presumably delivers the metal to the liver where many of the copper oxidases circulating in the blood are synthesized [11]. With this separate 'handling system' for copper the organism is able to direct the two different metals to their target organs.

Both Ga and In are bound to transferrin in plasma. Their radioisotopes (^{67}Ga, ^{68}Ga, ^{111}In, ^{113}In) are used in clinical medicine as a tracer for metabolic processes for the diagnostic scanning of various organs. However the indium transferrin complex differs markedly from other non-ferric metals: In^{3+} appears to bind to serum transferrin and ovotransferrin with an affinity comparable to iron, and unlike other metal ions, it is not displaced from the protein by iron [129]. In addition the comparison of the metabolism of iron and indium bound transferrins in animals [130] revealed that In^{3+} was cleared more slowly from the plasma than iron but very little indium was taken up by the bone marrow or incorporated into red blood cells. Both proteins (di-indium and diferric transferrin) were found to have similar affinity for the receptors [131]. In contrast, Ga^{3+} is not as avidly bound to plasma transferrin as In^{3+} suggesting that the transport of Ga^{3+} may be influenced by transmembrane diffusion of unbound metal ions depending on the transferrin concentration [132]. Due to its similar plasma distribution Ga is often considered as an analogue to Al in metabolic processes [133]. This is of particular interest in view of increased levels of Al in brain which has been implicated as a factor in Alzheimer's disease [134].

Like Ga, the molecular conformation of the aluminium transferrin complex shows a similar closure of the interdomain clefts upon incorporation of Al^{3+} into the molecule. Although transferrin is the major Al carrier protein in the plasma [135], Al^{3+} and Fe^{3+} are metabolized differently. Direct competition between Al^{3+} and Fe^{3+} for binding sites on transferrin have shown an unexpected weaker stability constant for Al^{3+} [136]. This has been interpreted in terms of the smaller size of Al^{3+} (see table 6.3) leading to at least one of the protein metal ligands not being able to bind to the metal [137]. Even though this is a plausible interpretation, it appears unlikely for two reasons: First, the crystallographic structure of Cu_2-HLF indicates that there is sufficient internal flexibility in each binding cleft to allow adaptation in metal coordination [27] and second, Al^{3+} has a high preference for a 6-fold coordination. However, due to its strong preference for 6-fold coordination and a relatively smaller size, it is more likely that one or more of the protein ligands will move significantly to accomodate Al^{3+}. It could alter the hydrogen bonding

network in the metal binding cavity which reduces the stability conditions in the interdomain cleft. This may explain the more labile binding of Ga or its analogue Al by transferrins compared to Fe as well as the increased susceptibility to changes in pH or to competing chelators and electrolyte composition [138]. The release of metals from transferrin depends on the presence of transferrin receptors on the cell surface. Although Al^{3+}-saturated transferrin is assumed to bind to the cell surface of tumors of nervous tissues [139], it is not impossible that Al enters cells (including brain cells) via the formation of low-molecular-weight (e.g. Al-citrate) complexes which could cross the blood brain barrier more easily than protein-bound Al.

Summarizing it can be stated that the solution X-ray scattering results have allowed to define structural changes for transferrins when different metals are bound. Hf^{4+} and Th^{4+} induce conformations that resemble closely the molecular conformation of apo-transferrin. In this case the closed conformation remains the neatest explanation for the high affinity of diferric transferrin for the receptor at the cell surface. Differences in the inorganic chemsitry between Fe and the other metals (Cu^{2+}, Al^{3+}, Ga^{3+}, In^{3+}), including the hydrolysis products and aqueous solubility need to be brought into play in explaining their binding affinities and determining their biodistribution. The nature of the transferrin receptor and the exact mechanism of metal release in the cell are still far from being understood. Interestingly, studies on tumors and inflammatory lesions have shown that metal transport by transferrin may also play a secondary role. A likely modification of the plasma membrane permeability appears to be capable of accomodating metal ions (e.g. Ga^{3+}) bound to a low-molecular weight form and thus allowing metal accumulation in malignant cells [140].

6.1.5 Studies of mutants of the N-terminal fragment

A very efficient and powerful method to study the specifity and function of the metal binding site is the replacement of single amino acids involved in metal binding, interdomain contacts or hinge bending regions. One example with the detrimental effect of iron defficiency was already mentioned: The variant of human serum transferrin in which a naturally occurring single point mutation in the interdomain cleft markedly modifies the overall conformation concomitant abnormal iron binding properties. In this section, variants (mutants) will be discussed in which one amino acid was modified site-specifically in the human transferrin sequence by protein engineering.

Structural studies on transferrins have shown that four of the iron ligands are provided by the protein, only one of these, an aspartic acid, originates strictly from domain I [25, 28]. Examining the crystal structures suggests that the carboxylate group of this amino acid may

play a key role for the following reasons: It has a total charge of -1, divided between its two oxygen atoms. One oxygen atom binds to the metal, thus neutralizing the cation charge by -0.5. The other carboxyl oxygen is held in place in the active site by interdomain hydrogen bonds (see figure 2.2). The feature, i.e. binding the metal ion via one oxygen atom and stabilizing the interdomain stability by hydrogen bonding via the other oxygen atom, not only limits the flexibility of the interdomain cleft but also restrains the metal ion requirements with respect to coordination geometry. The X-ray scattering studies on Hf- and Th-bound tranferrin have led to the suggestion that the 'correct' coordination of metal ion is important for inducing the closed conformation and that the closed conformation is likely to be of functional importance for recognition of receptor by the transferrin.

In order to test directly the role of the aspartic acid (the variant discussed in section 6.1.3 already revealed indirectly the importance of the aspartic acid), it has been replaced by site-directed mutagenesis and subsequently the mutated protein has been investigated by solution X-ray scatterting. Recently, it has become possible to express mutants of the amino-terminal half-molecule of human serum transferrin (HST/2N), a fragment of 337 residues incorporating a single iron binding site [99, 100]. Two mutants, Asp63 → serine (D63S) and Asp63 → cysteine (D63C), both of which show weaker iron binding than the native protein (referred to as wild-type) have been examined by solution X-ray scattering in both the apo- and holo-forms. In the case of D63C, the cysteine residue is blocked by a disulphide bond formation and thus would not be available for ligation to iron. Consequently, the lobe would be expected to remain open. In the case of D63S, the mutation to an uncharged residue would weaken the strength of hydrogen bonds between the two domains and would probably reduce the closure of the interdomain cleft.

Figure 6.11 shows the X-ray scattering results for the native protein and the mutants in the iron-free and iron-loaded state, respectively. Whereas the scattering curves for the wild-type exhibit the typical characteristics of a large-scale conformational change upon iron uptake,

Table 6.4 : Radii of gyration for the wild-type and mutated amino-terminal half-molecules of human serum transferrin.

sample	apo-state (iron-free) R_g [Å]	holo-state (iron-loaded) R_g [Å]
native protein	23.0 ± 0.2 (22.0 *, 22.7 #)	21.4 ± 0.2 (20.1 *, 21.0 #)
Asp63 → Ser	23.5 ± 0.4 ¶	22.3 ± 0.2
Asp63 → Cys	23.0 ± 0.2	22.7 ± 0.2

* computed from the crystal structure and modelled open structure of RST/2N; # calculation including the first hydration shell (see text); ¶ the scattering curve for this sample shows a small aggregation effect in the Guinier region, thus a larger error has been indicated.

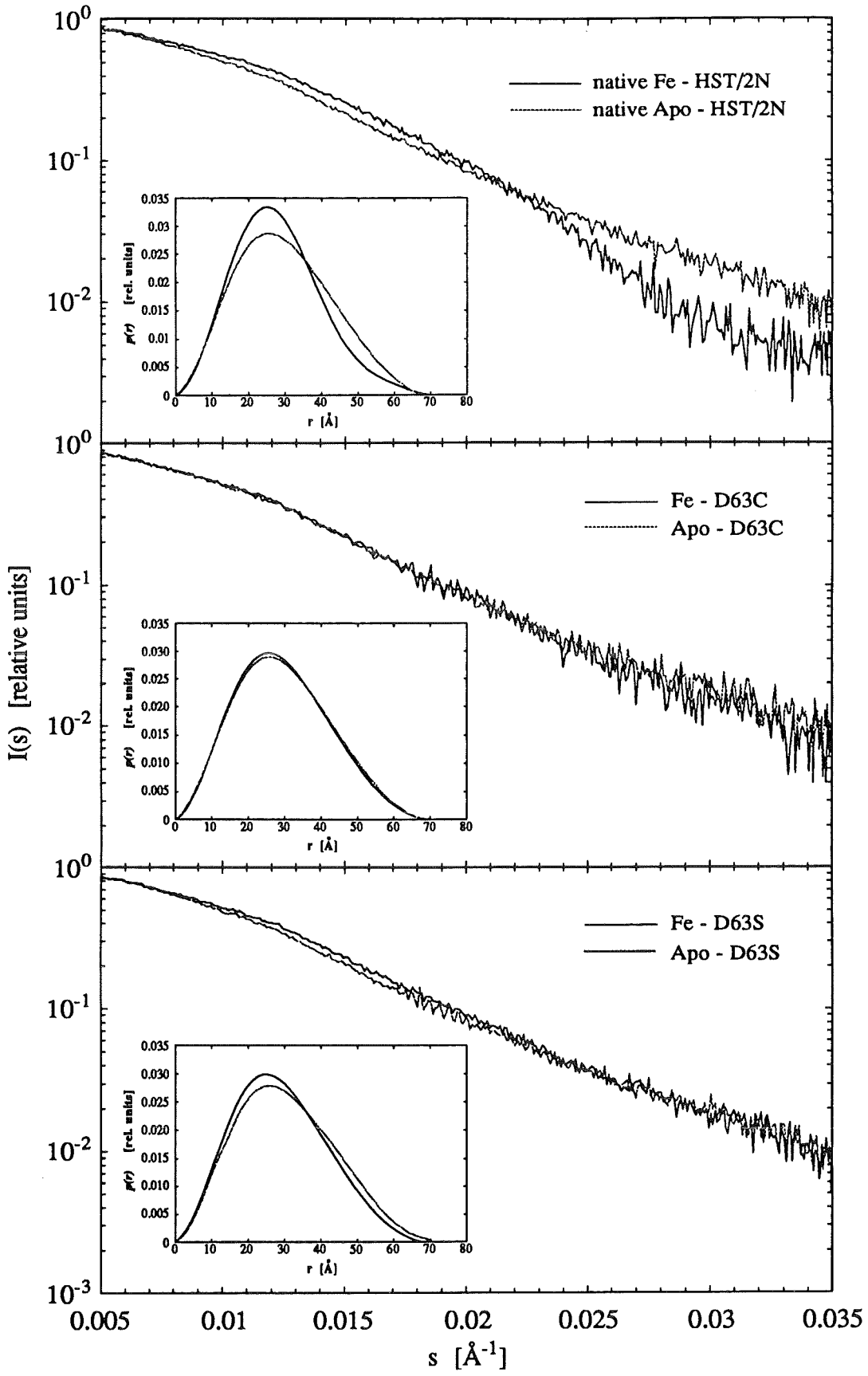


Figure 6.11 :
Scattering profiles of dilute solutions of the wild-type (top) and of the Asp63 → Cys (middle) and Asp63 → Ser half-molecules of human serum transferrin (bottom) recorded in the absence of iron (dotted line) and in the presence of iron (solid line) bound to the active site. Distance distribution functions calculated from the experimental data are shown as insets.

as known from the N-terminal fragment of chicken ovotransferrin (figure 6.3(a)), the curves for the D63C mutant are almost indistinguishable, thus indicating no structural change upon binding of iron as predicted. The scattering curves for the D63S mutant show a small difference only in the low angle scattering range. In contrast, the scattering curves for the expressed native N-lobe differ over the whole range; the differences become very pronounced at $s > 0.02Å^{-1}$, where a distinct intersection point is observed. It is worth noting that all apo-proteins, both wild-type and mutants, do not differ significantly in their scattering behaviour. These findings are described quantitatively in terms of a Guinier analysis using only low-angle data ($s < 0.014Å^{-1}$) and are equivalent to the results calculated using the whole of the experimental scattering curves (applying equation E3.7 by means of the *p(r)* function). Table 6.4 gives the radii of gyration, R_g, for the apo- and holo-states of the wild-type and the two mutants. R_g values for the three apo-proteins are essentially the same. For the D63C mutant, an insignificant reduction in R_g is observed when iron is bound while a small change is observed in the case of D63S. Insets of figure 6.11 provide a comparison of distance distribution functions obtained by Fourier transform of the whole of the scattering patterns for the wild-type and two mutants. Thus, unlike R_g, the distance distribution function does not critically depend on the selection of scattering range and provides a more confident picture. The curves for D63C, show the identity of molecular conformation over the whole radial range. The distance distribution functions for D63S show a behaviour somewhat similar to those for the wild-type, except that the magnitude of change is significantly smaller.

It is clear from these data that iron does not induce a structural rearrangement in the D63C mutant but the lobe remains in an open, 'apo-like', conformation. In contrast, the iron binding to the D63S mutant induces a small structural change which corresponds to a 'slightly closed' conformation. These mutational studies represent an impressive test of the abrogation of normal iron binding[15]. In both mutations an uncharged polar group replaced a negatively charged aspartic acid side chain. Although the serine side chain is of only a marginally different length, the substitution results in a major perturbation at the iron binding site, resulting in an inhibition of the domain closure and a lower affinity for iron. In this case, the negatively charged carboxylate group of the aspartic acid is replaced by an uncharged polar group. The energy contribution for a hydrogen bond to a charged residue is three fold more than the contribution due to an uncharged residue [142]. It is, therefore, most probable that the hydrogen bonds in the mutant are not sufficiently strong to tighten the metal binding cavity and bring the two domains closer together when iron is bound. The D63C mutant showed that with the non-availability of this ligand for iron binding, the trigger mechanism for domain closure is completely switched off. This demonstrates the

[15] Mutational experiments [100] were initiated because of the amino acid modification (serine instead of aspartic acid) in the C-terminal lobe of melanotransferrin, a membrane-bound transferrin-like molecule which is present in most human melanomas [141]. Its role in cellular iron metabolism is not yet known but seems to differ plainly from that provided by serum transferrin.

delicate balance which the molecule possesses in order not only to be specific to a receptor but also for a metal.

The explanation of the 'slightly closed' conformation revealed by the scattering data of the D63S mutant requires an extensive process of computer modelling not only on the basis of structural information but also owing to functional characteristics (e.g. the mechanism of iron uptake).

As a first step, it is important to reproduce reliably the experimental scattering features of the native protein in the iron-free and iron-loaded state. The simulation for the closed, iron-loaded conformation of the N-lobe of human serum transferrin (HST/2N) can be based on the crystal structure coordinates of the N-terminal half molecule of rabbit serum transferrin [29]. In view of the sequence homolgy between lactoferrin and serum transferrin, it is favourable to 'copy' the open configuration of the N-lobe of apo-lactoferrin into the N-terminal half of rabbit serum transferrin (RST/2N). This structural transfer was achieved similar to the procedure described in section 6.1.3: The superposition of the main chain atoms of 113 residues from the core of domain NII (r.m.s. deviation of 0.58 Å/atom),

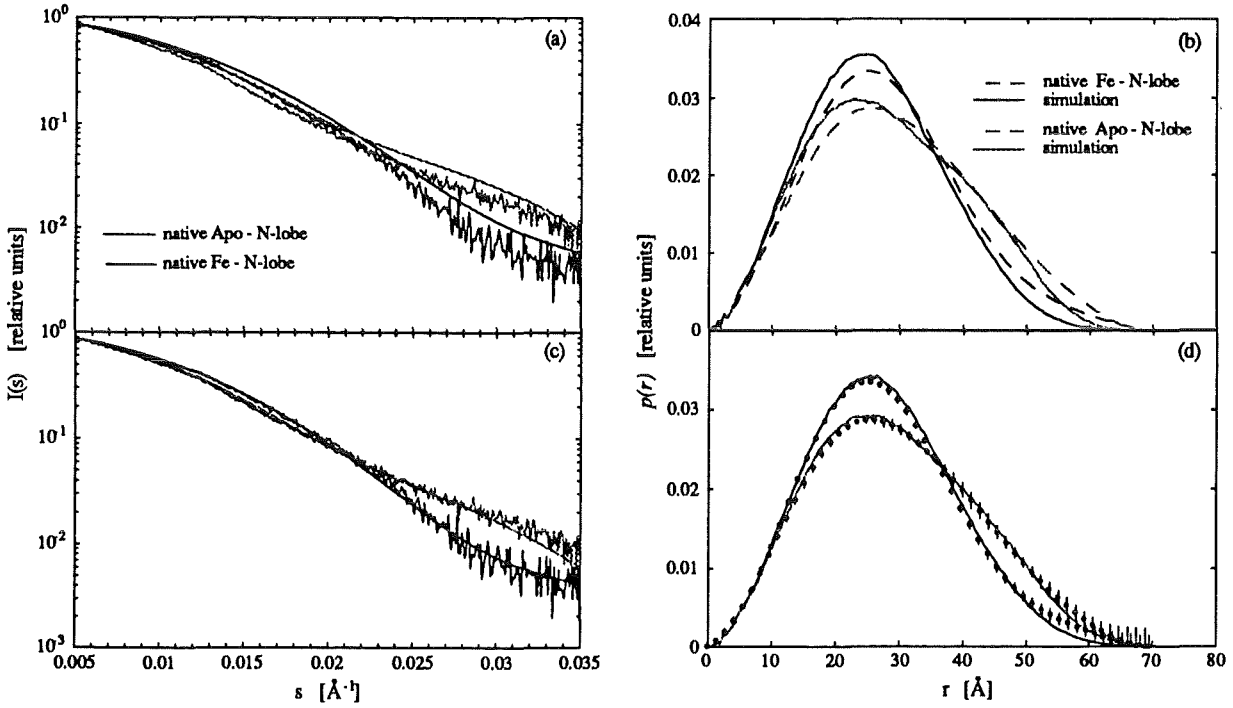


Figure 6.12 :
Comparison of experimental and theoretical scattering profiles and distance distribution functions for the iron-free and iron-loaded N-lobe fragment of serum transferrin. The simulated curves in (a) and (b) are based alone upon non-hydrogen atoms of the molecular models for the open and closed N-lobe. By including water oxygens of the first hydration layer a reliable fit to the experimental data could be obtained (c) and (d). In (d) only circles with error bars represent the experimental p(r) functions in order to emphasize the good agreement with the model calculations.

defines domain NII being the basis of both the lactoferrin and rabbit serum transferrin structure. Breaking up the peptide chain of RST/2N at residues 93 and 247, enables the opening of domain NI. This was done by matching its counterpart, domain NI of apo-lactoferrin (superposition of the main chain atoms from 81 core residues yielded a r.m.s. value of 0.61 Å/atom).

The calculations using the models for the apo- and holo-state of RST/2N were directly compared with the experimental results of native HST/2N in figure 6.12(a,b). Although the essential features representing the characteristic domain arrangement were achieved, the discrepancies between theory and experiment are obvious in the absolute scale of the scattering profiles as well as in the noticeable shift of *p(r)* functions. It was already mentioned in section 6.1.1 that hydration as well as carbohydrate content do affect the scattering intensity so as to modify the overall size of the molecule. Due to the absence of any carbohydrate in the N-lobe of human transferrin, only the hydration shell around the protein will contribute. This has been accounted for by incorporating the first water layer (up to 3 Å from the protein's surface) into the scattering pattern simulations (see Appendix B). The used approach also includes the optimization of the opened configuration at positions where computer modelling introduced strains owing to the rupture of the polypeptide chain. The results presented in figure 6.12(c,d) clearly yielded a much better fit to the experimental data supporting the overall correspondence of the applied models.

In order to deal in more detail with the strongly reduced interdomain closure of the iron-loaded D63S mutant it will be essential to take functional models for iron uptake by

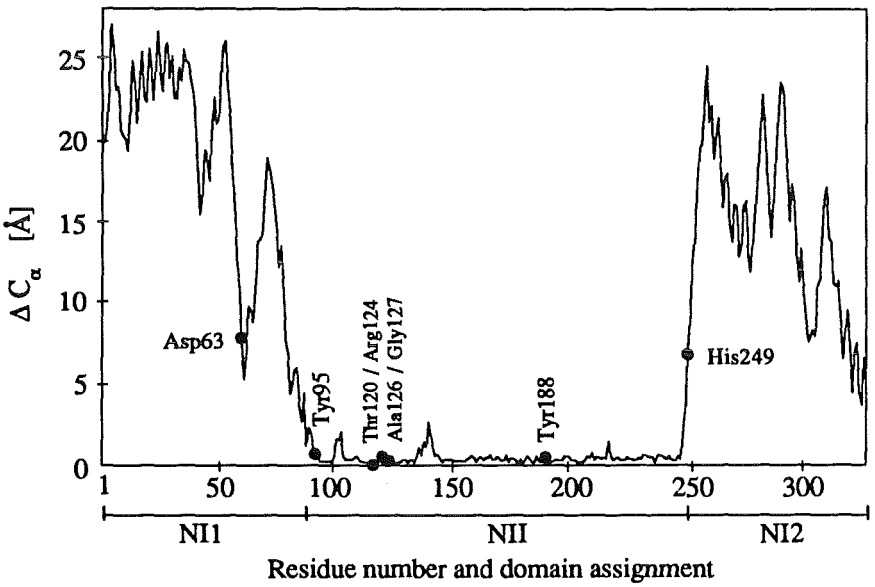


Figure 6.13 :
Difference plot of C_α positions between the closed and open structure of the N-lobe of transferrin. The positions of the iron binding and carbonate anchoring ligands are indicated (residue numbering according to HST).

transferrins into consideration. As was suggested by Baker [24] and most recently by Lindley [111], the iron along with the synergistic carbonate anion binds to domain II including the ligands belonging to the two backbone strands (see again figure 2.2) while the protein is in the open conformation. Iron binding will be complete when the remaining ligand of domain I, the aspartic acid, closes over the metal site to fulfil optimal iron coordination. This metal uptake mechanism may be used to develop a model for the D63S mutant in iron-loaded state: Figure 6.13 illustrates the C_α difference plot between the apo- and holo-state of the N-terminal half molecule after superimposing the main chain atoms of domain NII of both protein configurations. Assuming iron and carbonate binding at domain NII has already taken place, this graph can also be interpreted as the C_α difference of the iron-loaded, closed structure and an intermediate, iron-loaded but still open structure: Apart from Asp63 and His249 the other ligands involved in iron binding (Tyr95 and Tyr188) and carbonate anchoring (Thr120, Arg124, Ala126 and Gly127) are already in place owing to their almost unchanged positions in both structures (figure 6.13). The His249 is located at a position with minor constraints by the surrounding protein structure (no hydrogen bonds to any protein atom), so that no large-scale peptide movement would be necessary to approach the metal ion. In contrast, the binding of Asp63 would precede a more pronounced peptide movement which is indicated by the increased C_α differences in the vicinity of Asp63. Thus, the weak iron binding of the D63S mutant and its almost 'apo-like' conformation strongly suggest that Ser63 is not in contact with the metal ion, whereas His249 is. The adaptability of His249 may induce slight structural rearrangements from residue 250 onwards (up to about 280) leading to the difference in overall conformation which has been observed in the scattering studies.

Computer modelling of such an intermediate structure is still in progress. However, further experiments with a series of other mutants (where other iron binding ligands were exchanged site-specifically, e.g. Tyr95 → histidine) appear to support the suggested model.

A vast amount of structural information on the transferrin family could be collected during the past few years. Although solution X-ray scattering is only a low resolution technique, the results presented here are significant and underline the importance of flexibility of transferrins in solution. However, many questions are still not or not completely answered. In the concluding chapter 7 some future prospects will be given and the role of solution X-ray scattering in providing possible answers will be discussed.

6.2 Nitrite Reductase

As was pointed out in Chapter 2.5, reports on the copper-containing nitrite reductases (NiR) from the denitrifying bacteria of *Alcaligenes faecalis* (AfNiR), *Achromobacter cycloclastes* (AcNiR) and *Achromobacter xylosoxidans* (AxNiR) vary in their molecular weights as well as their subunit arrangements in solution. Furthermore, deducing the oligomeric structure from studies of gel filtration and sedimentation equilibrium centrifugation revealed conflicting results within the same species (in the case of AcNiR [43, 50] and AxNiR [47, 49]).

However, the crystal structure determination of AcNiR [50] yields that the molecule must be a trimer which is in accord with sedimentation equilibrium centrifugation studies. The similarity in the amino acid composition between AcNiR and AxNiR and the physicochemical properties (specific activities and copper content), which they have in common [49], strongly suggest the close structural relatedness of the two proteins. The amino acid composition and spectral properties of AcNiR are also similar to AfNiR, and, in addition, both NiRs receive electrons from closely related blue copper proteins [45]. On these assumptions one might expect similar molecular conformations in solution for the three species of copper-containing NiRs.

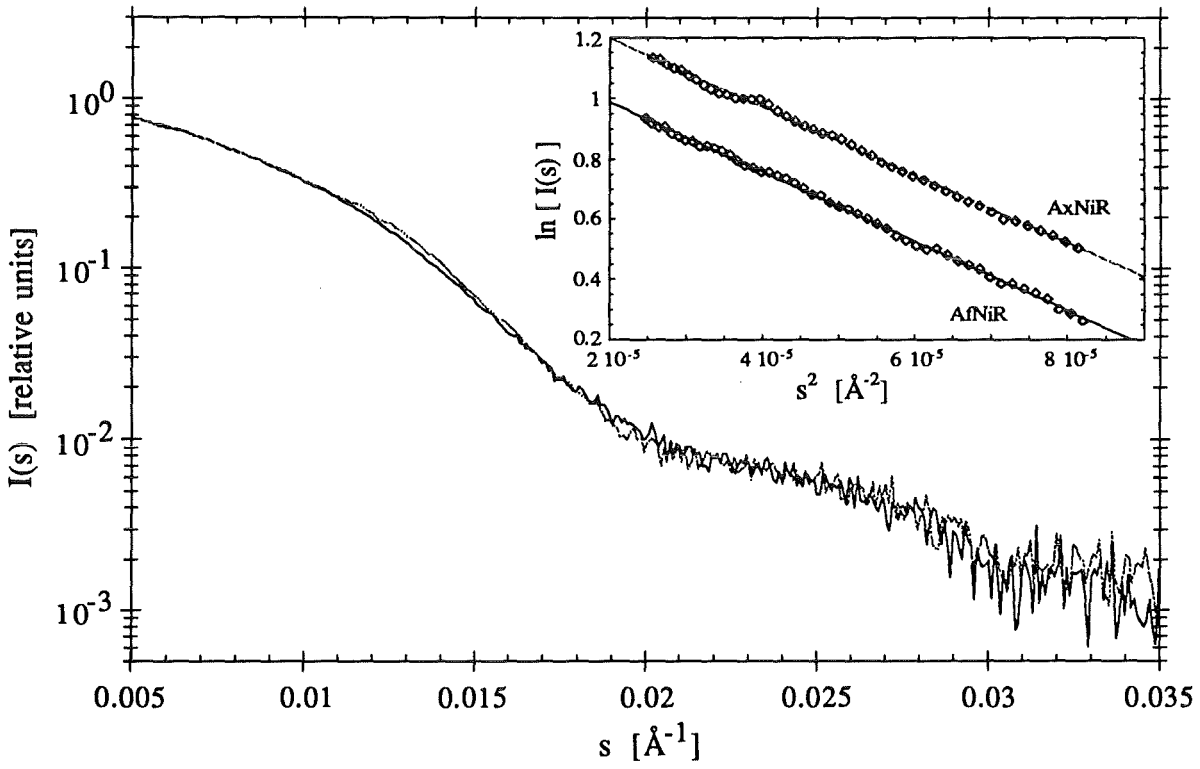


Figure 6.14 :
Experimental scattering profiles for AxNiR (dashed line) and AfNiR (full line) at a sample concentration of 5 mg/ml. The curves were normalized to unity at zero scattering angle. The inset shows the Guinier plots for both proteins from which the R_g values were derived.

6.2.1 Experimental results

Solution X-ray scattering experiments on AxNiR (sample concentrations of 5 mg/ml and 10 mg/ml) and AfNiR (sample concentration of 5 mg/ml) have been carried out in order to probe their overall conformation. Figure 6.14 presents the averaged and complete experimental scattering profiles for 5 mg/ml (the profile for 10 mg/ml of AxNiR was equivalent and is therefore not shown, thus indicating no obvious concentration dependence). The monodispersity of both species in solution is also given according to the linearity in the low angle range as seen in the Guinier plots (insets of figure 6.14). Both species (AxNiR and AfNiR) appear to have very similar conformations in solution which is also expressed in the comparison of the distance distribution functions (given in figure 6.15) and the derived distance parameters for AxNiR and AfNiR (compiled in table 6.5).

Values of the radius of gyration R_g were calculated over the angular range s_{min} = 0.005 Å^{-1} to s_{max} = 0.009 Å^{-1}. The use of the Guinier approximation was further examined by computing R_g using equation E3.7. This enables a more accurate determination of R_g by

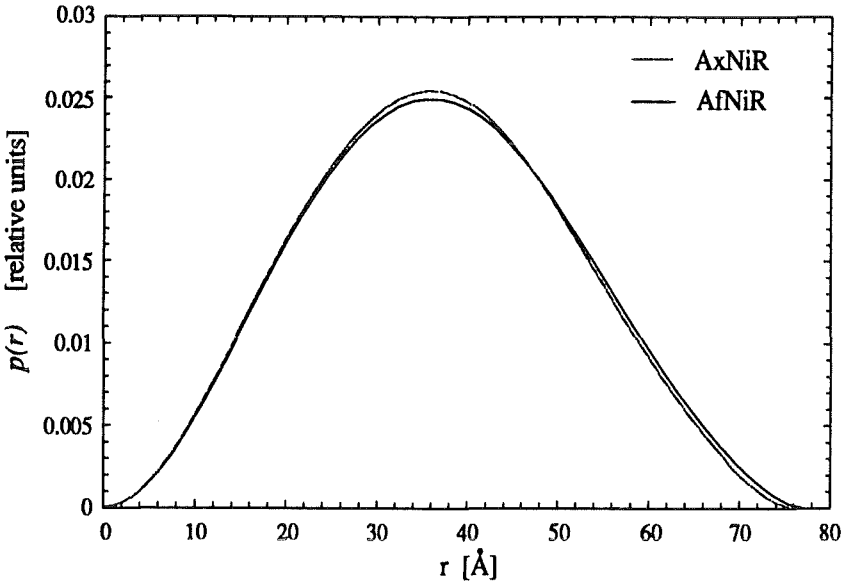


Figure 6.15 : Distance distribution functions *p(r)* calculated from the experimental scattering data of AxNiR and AfNiR at a sample concentration of 5 mg/ml. The area under each curve is normalized to unity.

sample	concentration	R_g from Guiner *[Å]*	R_g from *p(r)* *[Å]*	*D [Å]*
AxNiR	5 mg/ml	28.7 (0.5)	28.0 (0.2)	76 (3)
	10 mg/ml	28.3 (0.5)	27.7 (0.2)	76 (3)
AfNiR	5 mg/ml	28.9 (0.5)	28.3 (0.2)	77 (3)

Table 6.5 : Experimental values of the radius of gyration (R_g) and the maximum particle size (D) of the two species of NiR (estimated standard deviations in parentheses).

means of the whole information content of the scattering profiles ($s \leq 0.035$ Å^{-1}). The Guinier approximation is restricted to a small range (the innermost part of the scattering curve) which is not completely available due to the beamstop. Values of the maximum particle size *D* have been derived from the calculation of the *p(r)* function.

6.2.2 Model building

Calculation of the one-dimensional *p(r)* function and distance parameters of a molecule in solution is clearly not sufficient for the unambiguous determination of molecular structure. They provide constraints on the structure that can be used to construct models. Thus, structural information from other sources is essential and can be examined for consistency with the solution scattering results. A crucial source of information on the molecular conformation of AfNiR and AxNiR is the known high-resolution crystal structure of the related AcNiR [50]. In order to assess the effect of different oligomeric conformations on the scattering behaviour, molecular modelling was performed, using as starting parameters the atomic coordinates from the 2.3Å resolution analysis [50]. AcNiR is associated as a trimer (not as a dimer as previous studies suggested, see table 2.2). The monomeric subunit contains 340 amino acids and has a molecular mass of 36kD. Modelling studies included residues 8 to 340 as the actual position of the amino-terminal peptide of each monomer couldn't be resolved in the crystal structure (probably as a result of proteolysis [143]). This fact will not limit conformational considerations since the amino-terminal end (i.e. residue 8) lies close to the main body of the monomer in contrast to the carboxy-terminal end which forms a protruding arm and interacts with another monomer in the crystal structure.

Apart from the trimeric assembly, several models of a dimer and tetramer of NiR were build in different ways (see figure 6.16). The structural basis for the model building was mainly geometric and included symmetry considerations as well as distance constraints derived from the experimental scattering data. Moreover, the docking in the case of dimeric association was comprising the formation of a large interface which will help to exclude water and thus stabilize such a dimeric configuration in solution. The chemical nature of the molecular surface and, in particular, the maintenance of the type II copper site (at the monomer-monomer interface) have not been taken into account in all models. No energy minimization has been applied to the models since it only helps to find hydrogen bonds, correct side chain positions along the monomer-monomer interface and to perfect artificial complexes up to atomic resolution. However, the difference between the original and energy minimized structure will be hardly noticeable in the simulated, low to medium resolution scattering profiles. In addition, the energy values themselves are not meaningful. Finally, it has to be pointed out that this modelling study assumes that the subunit association occurs without major conformational change in the monomers.

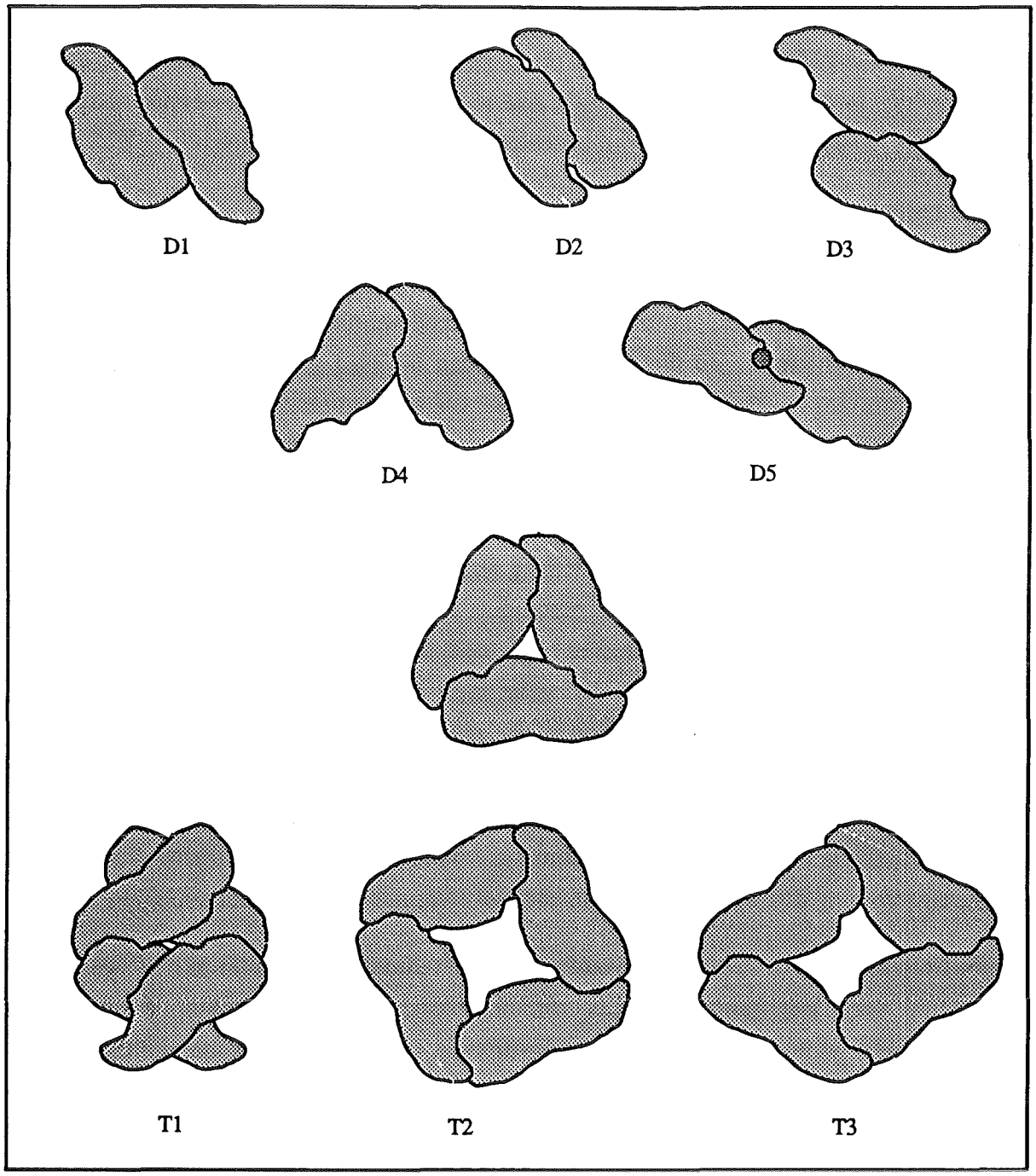


Figure 6.16 :
Schematic representations of subunit arrangements for nitrite reductase. Models of dimeric (D1-D5) and tetrameric (T1-T3) molecules have been based on the trimer (figure centre) of AcNiR revealed by the 2.3Å X-ray structure analysis [50]. The dimer D4 is obtained by taking a monomer away from the trimer and thus the association of the monomers in D4 is the same as in the trimer. Dimer D4 has been used as a template for tetramers T1 and T3. In the tetramer T1 the two dimers behave like two pairs of pincers that interlock or engage with each other without clashing. Apart from the models where the dimeric structure known from the trimer was maintained, only dimer D5 turned out to form a type II copper site at its monomer-monomer interface (see dark circle). Four histidine residues (2x His 255 and 2x His 306) can provide a tetrahedral ligand arrangement around a copper atom. Thus, on the basis of 6 copper atoms in the trimeric structure, the copper content for the respective models would be 2 (D1-D3), 3 (D4 and D5), 4 (T2) and 6 (T1 and T3).

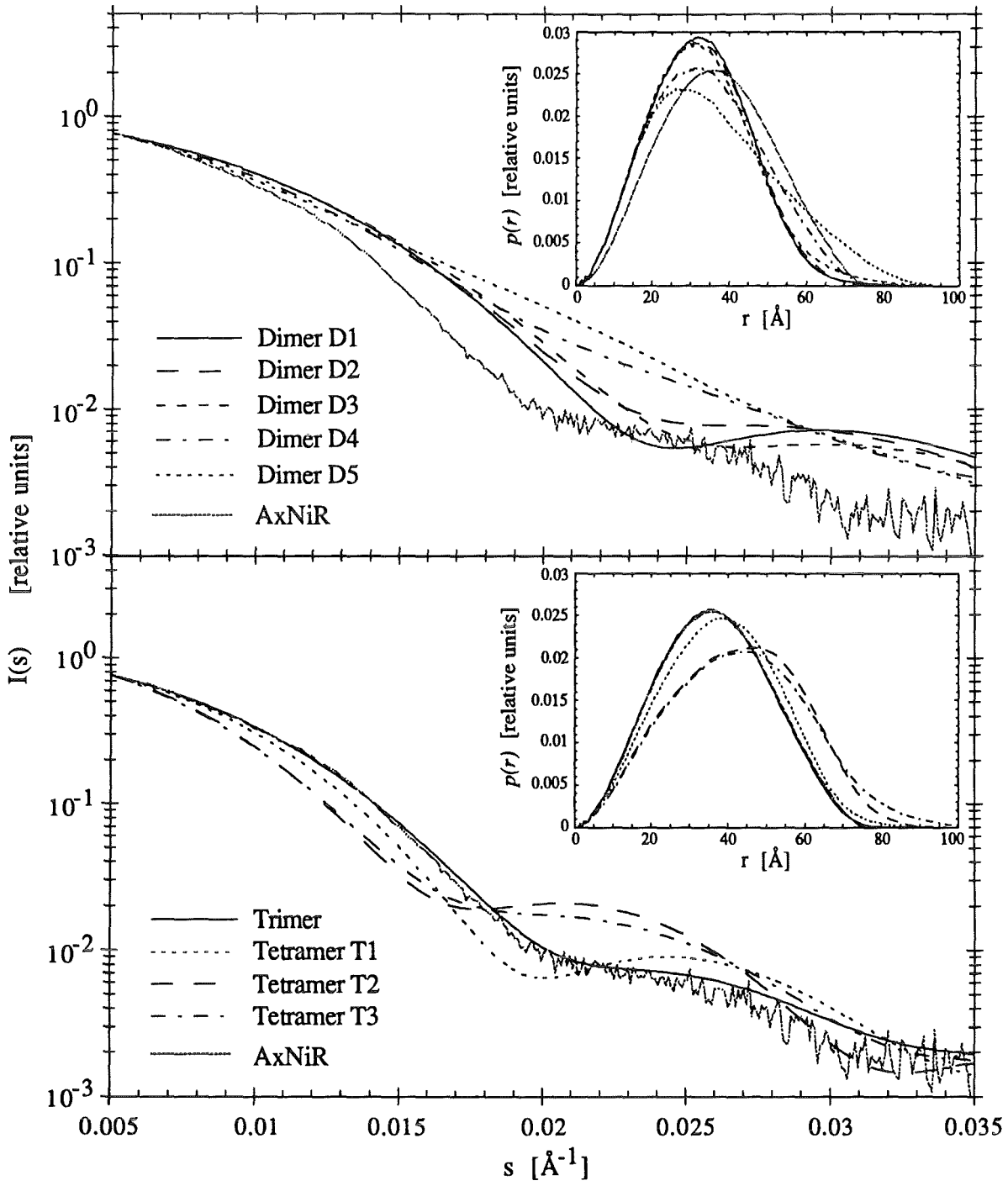


Figure 6.17 :
Comparison of the scattering curve I(s) and the distance distribution *p(r)* of AxNiR with those of the models shown in figure 6.16. The I(s) profiles are normalized at s = 0.005Å^{-1} with respect to AxNiR, and *p(r)* functions are plotted by normalizing the area under each curve to unity.

6.2.3 Which model is consistent with the scattering results ?

Simulated scattering profiles and the corresponding distance distribution functions of the models are presented along with the experimental results for AfNiR (figure 6.17). Only

AxNiR is presented since the difference between the experimental data of AxNiR and AfNiR is not significant. The calculations (performed as described in Chapter 3.3) are based upon the crystal structure data of AxNiR [50], using 5148, 7722 and 10296 non-hydrogen atoms for the dimer, trimer and tetramer, respectively (copper atoms not included). The computational results for the structural parameters of the dimeric, trimeric and tetrameric models for NiR are compiled in table 6.6. In order to check the reliability of the oligomeric models, their solvent-accessible surface areas have been calculated with the Lee & Richards [144] algorithm implemented in the programme X-PLOR (Version 3.0, Yale University, New Haven, CT, USA). A probe radius R_{H_2O} of 1.4 Å was used. It has been shown that the solvent-accessible surface area A_S of an oligomeric protein, which is the sum of the areas of its component atoms, is directly related to its molecular mass [145, 146]. Values of A_S and the predicted accessible surface area A_P according to Miller *et al.* [145] are also given in table 6.6. The models have accessible surface areas that lie within the limits of the predicted value (the power law by Miller *et al.* [145] predicts values to within 5% on average).

Model		R_g [Å]	D [Å]	A_S [Å^2]	A_P [Å^2]	$\Delta A / A_S$
Dimer	D1	24.7	103	27650		5.8 %
	D2	24.7	96	28000		7.0 %
	D3	25.3	122	27350	26050	4.8 %
	D4	26.2	87	27110		3.9 %
	D5	28.5	101	26420		1.4 %
Trimer		27.8	87	34760	35460	-2.0 %
Tetramer	T1	29.2	105	40120		-9.9 %
	T2	32.2	104	47250	44120	6.6 %
	T3	33.2	121	47270		6.7 %

Table 6.6 : Structural parameters (R_g and D) of NiR models in figure 6.16 and their solvent-accessible surface areas (A_S and A_P). Theoretical values were calculated according to $A_P = 5.3\, M^{0.76}$ derived from 23 oligomeric proteins [145]. The mass of 36kD has been assumed for the monomer of NiR. The isolated monomer was calculated to have 15530 Å^2 of solvent-accessible area. $\Delta A/A_S$ is the deviation of A_S from the predicted value A_P.

It is apparent that the trimeric model is able to closely approximate the observed scattering data for the NiR molecule in solution at the concentrations used. This result is clearly shown by the data in reciprocal and in real space (figure 6.17). The alternative dimeric and tetrameric associations do not match the striking characteristics of the experimental scattering curve (the amplitude and periodicity of oscillation). The necessity of the scattering pattern simulation is obvious as the structural parameters as well as the solvent-accessible surface area do not determine the oligomeric conformation (dimer D5 and tetramer T1 have R_g values very similar to the experimental values).

It needs a mention that interatomic distances greater than 77 Å only contribute to 0.03% of the total amount of interatomic distances in the trimeric association. However, this small contribution gives rise to a discrepancy (of more than 10 Å) between experimental

and calculated *D* values for the trimer. This difference arises from small errors in intensity measurements in the low angle region.

The simualtion of the scattering profile could be improved further by incorporating a hydration layer surrounding the trimeric structure (see figures 6.18 and 6.19). Although these water molecules are not yet accessible crystallographically, an energy minimization procedure was used to estimate the amount of hydration (see Appendix B). A good fit to the experimental data can be obtained by including the first hydration shell up to a distance of about 3 Å from the protein surface (here approximately 700 water molecules were included). This result can be associated with a hydration of approximately 0.12 g H_2O per g protein and gives an estimate of the water content of NiR. However, this rather small solvent contribution compared to 0.38 g H_2O / g protein corresponding to a full hydration [83] can only be considered as the amount of tightly bound water representing a 'permanent' part of the trimer framework. A relatively conservative estimate of the total number of waters expected to be found in the crystal structure is one per residue, which here would be 340 x 3 or 1020. However, a comparatively lower solvent contribution is indicated by the fact that 3950 $Å^2$, or 25.4%, of the monomer solvent-accessible surface area is buried upon

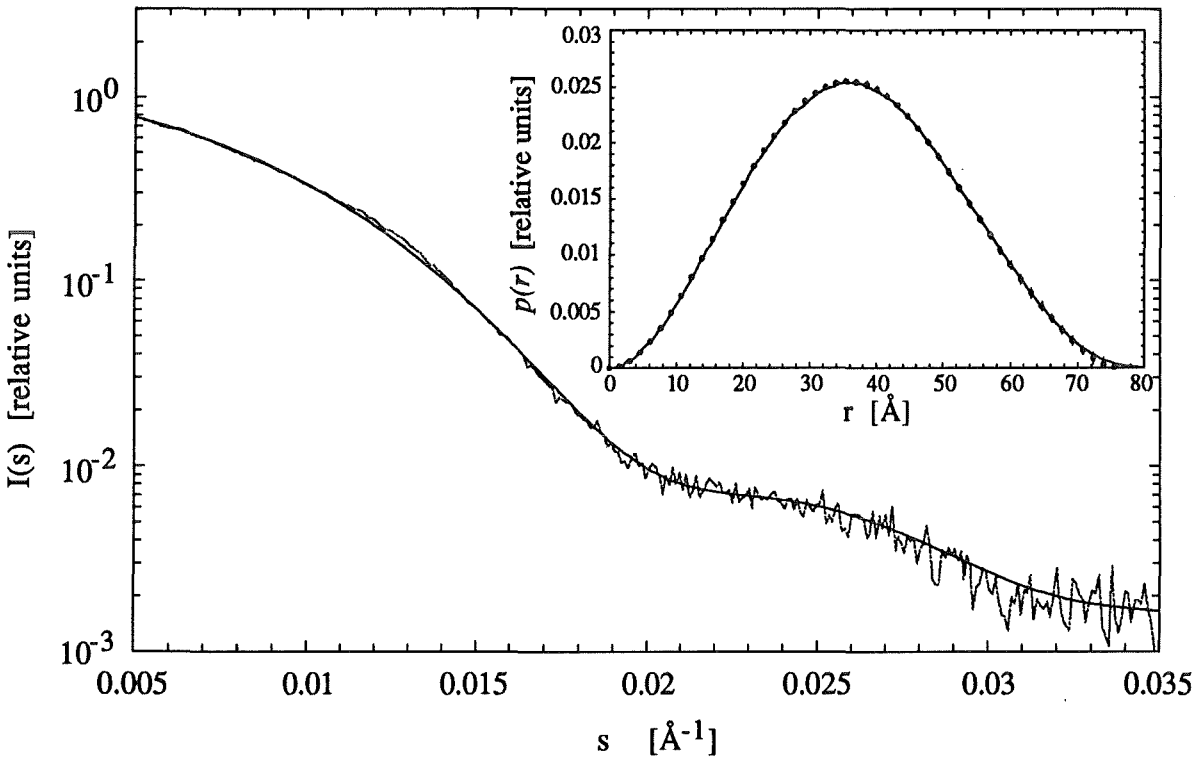


Figure 6.18 :
Comparison of the experimental data of AxNiR (5 mg/ml) with the calculated scattering pattern and distance distribution function (inset) of the NiR trimer including the first hydration shell (about 700 water oxygens were added to the 7722 non-hydrogen atoms of the trimer). Circles with error bars represent the experimental *p(r)* data. The radius of gyration of the hydrated structure is 28.2Å. The intensity profiles are normalized at s = 0.015$Å^{-1}$.

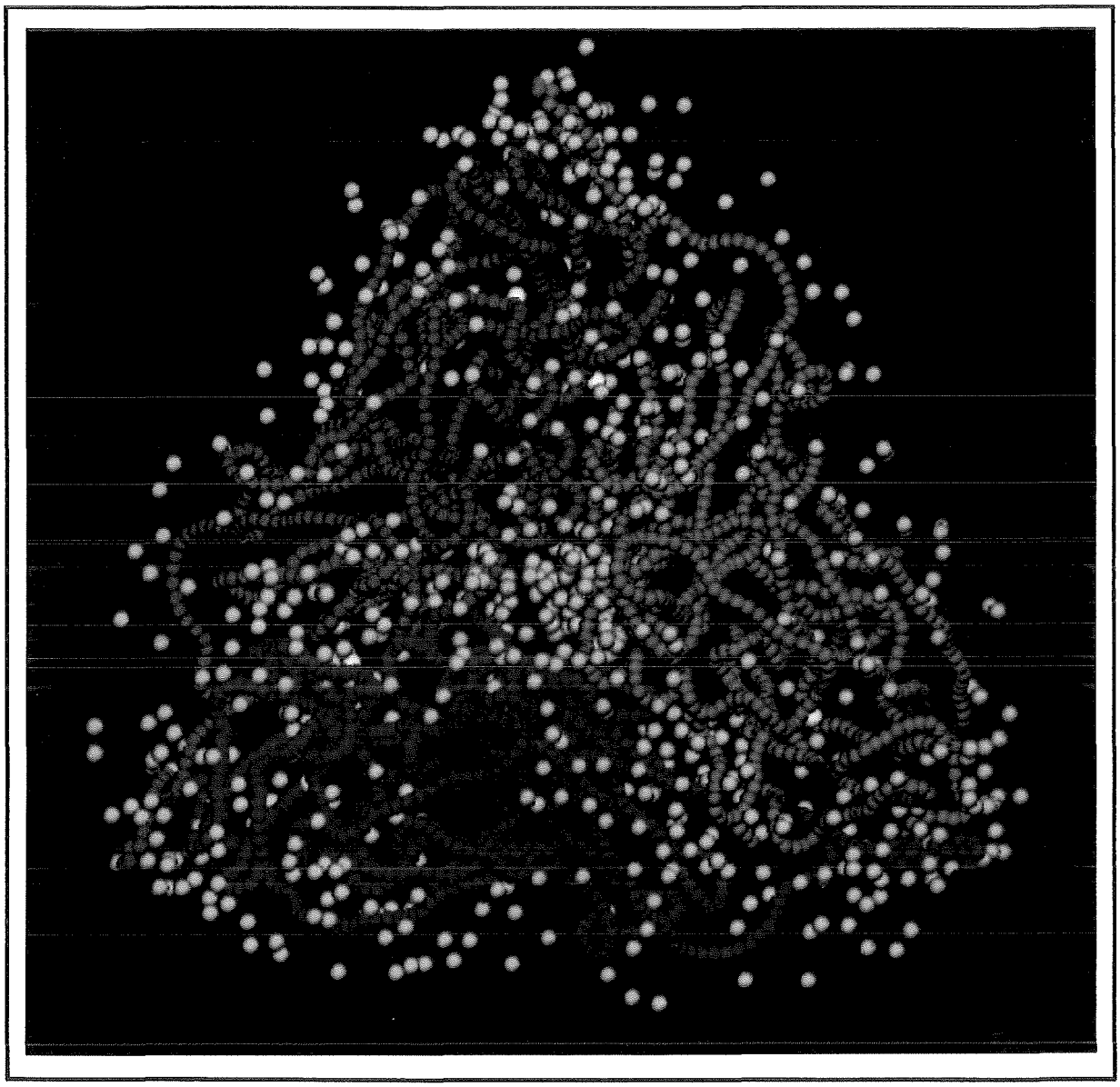


Figure 6.19 :
The model of the hydrated trimer of NiR used for the scattering pattern simulation shown in figure 6.18. In order to illustrate the trimeric association, the monomers are represented by their smoothed backbone structure using different colours (red, green and blue). Water oxygens are shown as yellow spheres. The copper atoms can be located as white spheres.

trimerization. Conversely, the total accessible surface area of the trimer is 34% smaller than the area of an isolated, monomeric protein of the same mass (calculated according to [145]). Hence only about two thirds of the protein surface is accessible to the solvent compared to a monomeric protein of 1020 amino acid residues. This rather large solvent-inaccessible area is a clear sign of extensive intersubunit interaction and thus it also demonstrates the high stability of the trimeric geometry.

A study of the hydropathy[16] reveals that the solvent-accessible surface area of the trimer is 52.3% non-polar, 34.5% polar and 13.2% charged (for comparison, the accessible

[16] The hydropathy of a protein surface is a measure for the affinity for attracting or repelling water. It can be evaluated by dividing a protein into non-polar, polar and charged components. The relative contribution to

surface area of the isolated monomer is 54.5% non-polar, 34.5% polar and 11.0% charged). The accessible surface area for monomeric and oligomeric proteins are equally non-polar (57% to 58%) [146]. Thus, this balance between hydrophobic and hydrophilic contributions of the surface area as well as of the buried interface area cannot be *the* driving force for trimerization. It is clear that each of the three copper atoms at the interfaces between two monomers do not only play a crucial role in enzyme activity but also in structure stabilization.

A striking feature of the trimer of NiR is a large channel or pore (~ 50 Å long and between 1.5 Å and 6 Å wide) along the threefold symmetry axis [50]. An interesting result of the performed energy minimization is that the pore is filled with water molecules building a solvent channel (see also figure 6.19). On top of it, each type II copper atom (bound between two monomers) was found to be ligated by one water molecule at a distance between 1.8 Å and 1.9 Å, though no water molecule has been located in the vicinity of the active site before starting the minimization procedure. This suggests clearly that a small molecule or substrate (such as NO_2^-) is easily capable of reaching the interior of the pore, i.e. the active site of the enzyme.

The trimeric structure as revealed by crystallographic studies is fully consistent with the solution scattering data, at least for the concentrations used here. On this basis, the copper-containing nitrite reductases from the three species (*Achromobacter cycloclastes*, *Achromobacter xylosoxidans* and *Alcaligenes faecalis*) appear to have more similarities in terms of their solution properties and redox centres than previously thought. The crystal structure of AcNiR has revealed that the type II copper is bound at the interface of the monomeric subunits within the trimeric association [50]. Clearly, retention of the trimeric structure (or minimally a dimeric structure), is necessary to maintain the relative orientation of the two domains which bind the copper atom. The finding here, that AfNiR and AxNiR are also trimers, and contain both type I and type II copper [41, 49] strongly suggests that a similar interface will be found in these two molecules. In addition,the model-building studies revealed that no dimeric association can be build that retains two type II copper centres at the monomer-monomer interface (assuming the monomers not to undergo a substantial conformational change). The difference in the colour of the two proteins (see table 2.2) may reflect a difference in the type I copper sites which dominates the absorption spectrum for visible light. Different ligands and/or a different coordination geometry are likely to be responsible for this absorption property. Further structural work will be necessary to understand these differences.

---

the accessible surface area of each component determines the hydropathy of the protein. All carbon atoms are assumed to be non-polar. Nitrogen, oxygen and sulphur atoms are taken to be polar when they carry no electric charge. Charged nitrogen and oxygen atoms appear in amino and guanidium groups (Lys and Arg) and carboxylate groups (Asp, Glu) as well as in the N- and C-terminal end of the polypeptide chain)

Chapter 7 : Conclusion and Future Objectives

Solution X-ray scattering has been applied to study two families of metalloproteins, a class of proteins where one or a few metal ions are located at or near the active site. Two important aspects of protein conformation have been observed in the solution scattering studies on the family of *transferrins* and the copper-containing *nitrite reductases*:

The transferrins change their conformation substantially after binding of particular metal ions and thus provide a prerequisite for receptor recognition and the subsequent metal transport across the cellular membrane.

The nitrite reductase must adopt a certain oligomeric structure in order to facilitate its enzymatic function of nitrite reduction. The copper atoms do not only act as active centres of the protein but also stabilize its conformation.

The determination of the solution structure of both metalloproteins brought further insights in the understanding of structural factors which control their function. However, we are still left with more questions than answers. The technique of solution X-ray scattering in concert with model-building studies based on the available crystallographic information is certainly able to contribute answers to special problems related with protein conformations in solution. For considered continuing experiments, some perspectives and possibilities of solution X-ray scattering will be given. These comments are not restricted to the proteins studied in this work.

The multi-domain composition of the *transferrins* confers their selectivity to act as a specific iron binding protein in company with a characteristic, so-called synergistic anion (carbonate). X-ray crystallography has defined their tertiary structures and has markedly influenced the interpretation of functional aspects in the transferrin family. The crystal structure analysis of the metal-free protein has been possible so far only in one case [26] but it impressively revealed a large-scale conformational change which takes place with respect to the metal-loaded transferrin molecule. However, an inconsistent single-sided domain flexibilty in the crystal structures of human lactoferrin was observed (only in the N-lobe pronounced domain movements seemed to accompany the metal binding [25, 26]). The solution X-ray scattering studies in combination with molecular modelling reported here, clearly show that in solution both lobes undergo a similar conformational change which is consistent with the transition from an opened to a closed interdomain cleft when iron is taken up by the protein. Thus, the "Venus' flytrap" model as suggested for another bilobal protein [147], is a very appropriate picture to describe this phenomenon. The structural studies in solution were further extended to the uptake of non-physiological metals which also bind to transferrin with lower binding affinities compared to iron. These experiments

have led to the suggestion that the 'correct' coordination of metal ion is crucial for inducing the closed conformation and that the closed conformation is likely to be of functional importance for receptor recognition by the transferrin. Moreover, the first direct experimental evidence for the existence of a trigger mechanism for the closure of the interdomain cleft could be observed. This trigger mechanism is disrupted by mutation of the structurally important aspartic acid residue, which is the only ligand of iron from the main body of domain I.

The role of basic residues in the iron binding site and in the interdomain hinge region is an important factor to elucidate. It is not only the mechanism of uptake but also release of iron which bears far-reaching consequences: The natural occurring variant of human serum transferrin gives an impressive example that mutations of residues in the iron binding cavity can have large effects on stability and functionality. A fact that is clearly proven by the metabolic abnormality of iron deficiency [109]. However, it is still difficult to predict effects of any substitution on protein structure and/or stability. Structural studies and examination of multiple mutations at the same positions can help to reveal the different factors that contribute to the stability of metal binding and subunit association. This, of course, is valid for any metalloprotein. Thus solution X-ray scattering experiments with mutant and wild-type proteins remain a challenge for the future, however, the validity of the deductions depends heavily on the quality of the models used. Fortunately, transferrins are a highly homologous family of proteins. All structures known to date have the same overall fold, similar active sites and presumably similar structural mechanism of metal uptake.

Large conformational changes involving rotation between domains induced by the binding of ligands is a recurring functionally important feature of a variety of proteins [7]. Little is known about the mechanism and dynamics associated with this type of change. In general, the transferrins offer outstanding possibilities to learn more about the conformational flexibility of interdomain or hinge regions. For example, the coupling of salt bridges can be of general importance to the stability and function of proteins [148]. There is a significant difference in the iron binding strength among the transferrins (see table 2.1). The remarkable structural similarities of the binding sites cannot account for this functional difference. Consequently, interactions between the two domains within a lobe have to be considered. It will be interesting to see whether an interdomain salt bridge in lactoferrin which is not present in serum transferrin can destabilize the domain arrangement and thus alter the iron binding properties. A salt bridge can be blocked e.g. pH-induced, by high salt concentrations (counter ions are expected to neutralize the charges on the protein surface) or replaced by site-directed mutagenesis.

The two complementary functions of transferrin (chelating ferric iron in a stable complex to prevent its hydrolysis by maintaining its solubility at physiological pH and binding to cell surface receptors to facilitate the uptake of iron) are likely to apply to other metals. There has been great interest in the role of transferrin in transporting metals not

normally present in the body, in particular those with possible toxicological consequences. The effect of metal uptake on transferrin conformation was studied in this work only for a few metal ions. A more comprehensive examination would include a systematic investigation of the wide variety of metal ions. Despite the fact that the role of transferrin as a mediator of metal transfer may not be equivalent to iron, protein engineering brings exciting opportunities within the realms of possibility: For example, the replacement of residues at the iron binding site and interdomain cleft so that the protein is 'designed' to take up metal ions which are harmful for the body and excrete them again through specific organs. Interestingly, by virtue of their high affinity for iron, transferrins can retard the growth of microorgansims, which essentially require iron, by making this element available on a reduced level. In addition, there is a striking resemblance of transferrins to other "Venus' flytrap" proteins (see below). This feature might even initiate the modification of transferrin as a chelator not only for metal ions but also for small molecules (e.g. drugs) which could be delivered to specific tissues.

Future progress in understanding metal ion transport requires detailed information of co-operative interactions with the transferrin receptor. Although structural studies on the soluble portion of the human transferrin receptor are already in progress [149], the work on forming a stable transferrin-receptor complex is at an early stage. This is essential before reliable solution scattering studies will be feasible.

The crystallographic structures of two copper-containing *nitrite reductases* from the denitrifying organisms *Alcaligenes faecalis* and *Achromobacter xylosixidans* are still not determined. Solution X-ray scattering provided an immediate way of establishing its similarity to the NiR from *Achromobacter cycloclastes*. Both species have been shown to be trimers in solution at concentrations between 5 and 10 mg/ml, which implies that all three NiRs have very similar solution structures. The trimeric subunit arrangement was based upon the crystallographic studies of the related nitrite reductase from *Achromobacter cycloclastes*, where, unexpectedly, the enzyme was found to associate as a trimer [50]. Several questions arise in view of the role of the bound copper ions, in particular the type II coppers, which are coordinated by three histidines and an aquo or hydroxo moiety in an unusual pseudotetrahedral array at the interface between two subunits [50]. These copper atoms may help to stabilize the trimeric association of the protein. Metal-induced protein stability and association is a well-known phenomenon in molecular interactions of biological molecules. For instance, the bacteriophage gene 32 protein (a single-stranded DNA protein) contains one zinc atom per molecule. In the case of zinc bound to the protein a pronounced effect on the cooperativity of binding to nucleic acids can be detected [150]. In addition, results from NMR and X-ray scattering studies suggest that the apo-protein is monomeric whereas the zinc-containing protein is dimeric [151]. NMR experiments on a

synthetic peptide (representing one domain of the calcium-binding protein troponin-C) gave evidence for dimeric assembly in the presence of calcium [152].

Although the roles of the different types of copper in NiR are not yet clear, the type II copper site may be of functional importance in view of the fact that it is easily accessible by solvent (this was also found in energy minimization studies reported here), which is in marked contrast to the type I copper site being buried inside the protein and protected from water and small ions. Interestingly, a copper-nitrosyl complex which has been identified as a possible intermediate in the enzymatic nitrite reduction [153], is likely to be correlated with the type II copper centre.

As in the case of mutational studies of the iron binding site in transferrin, the understanding of the nature of the copper centres and their influence on the spectrochemical properties, on NiR activity as well as on the protein conformation will definitely increase with the involvement of protein engineering. The replacement of key residues involved in copper-binding or in the subunit interface of the trimeric enzyme (even in a region remote from the active site) are expected to reveal substantial insights in specific molecular interactions, a question which is central to molecular biology.

The self-organization of living matter in the form of complex macromolecular assemblies is unique in the physical world. Like the metal transfer mechanism correlated with the transferrin-receptor interaction, electron transfer reactions are ubiquitous in biological processes and are therefore subject of intense interest. The reduction of nitrite by the copper-containing NiRs requires small blue copper proteins as electron donors [39]. The structure of several of these donors have been solved crystallographically [34]. Thus, structural information on the complex formation will be necessary to understand the role of protein-protein interaction in the elementary electron transfer pathways.

Interestingly, the tight trimeric complex, the existence of ß-barrel domains and the specific solvent channel which contains the active sites, evidently imply remarkable similarities between NiR and the porins, proteins in the bacterial outer membranes [53]. In parallel, structural analogues are established for the transferrins: Despite little sequence homology to the transferrins, the bacterial periplasmatic binding proteins (including proteins specific for phosphate and sulphate binding) reveal a striking similarity regarding domain structure, polypeptide chain folding and conformational flexibility due to substrate binding [154, 155]. Furthermore the recent crystal structure determination of a key enzyme in the heme pathway also yielded a close structural resemblance with a single lobe of transferrin [156]. These examples suggest that similar structural motifs are exploited by nature in carrying out very different biological functions. It is the main target of present-day molecular biology to fit together the pieces of this tremendous puzzle. Certainly, it needs the combination of different interdisciplinary techniques and cooperative efforts. The impacts on biotechnology and medicine won't be long to wait for.

It should be obvious from the work presented in this thesis, that the technique of solution X-ray scattering can certainly provide valuable and immediate information on the overal size and conformation of single molecules or complexes under 'physiological' conditions. Model-building studies are a necessity for interpreting solution scattering data. Crystallographic structure information of protein subunits and related, homologous proteins, are a prerequisite for unambigous structure determination in solution. Fortunately, such an information has been available for the proteins studied in this work. On the other hand, the biological functions are evident only with proteins in solution, outside the crystal lattice. In solution they have enhanced degrees of flexibility not apparent in the crystal structures and applies especially to multi-domain proteins. Therefore the identity of crystal and solution conformation must not be taken for granted. For instance, the metal-free conformation of apo-lactoferrin in the crystalline state and in solution revealed that the C-lobe is held in a specific orientation owing to interactions with neighbouring molecules in the crystal. Specific aspects of the crystallization conditions (e.g. pH, sample and salt concentration) may also affect local features of the conformation.

Nevertheless, the final optimal answer to structures of biological systems on atomic scale cannot be provided by solution scattering only but has to await results from high-resolution techniques such as X-ray crystallography or NMR. Moreover, electron microscopy is another technique capable of diagnosing overal molecular conformations. The chances of finding the right shape of a biological molecule are usually higher than for solution scattering since real space information is directly obtainable. Thus, one may ask, why do we need solution X-ray scattering? The following brief remarks will deal with this question and present some promising as well as critical aspects regarding solution X-ray scattering applied to biological systems:

- All liquid related problems (disordered systems) can be studied, i.e. experiments can be performed under approximate physiological conditions. It is important that the sample is highly purified and a biochemical characterization has been performed prior to the scattering studies.
- Water and its interactions with the protein plays a crucial role in the functioning of living systems, thus an understanding of protein hydration is of particular interest. The hydration layer has features which are quite different from bulk solvent. Only a technique that can be applied in hydrous solution will be able to provide reliable answers with respect to structural changes dependent on e.g. temperature, salt concentration or pH value.
- Unfolding and refolding of proteins are processes which can only be studied in solution. By destroying the native state of a molecule (e.g. by extreme temperature or pH conditions), one hopes to obtain insights into the self-reorganization and possibly into

the folding mechanism of proteins. It is true that denaturation does not force the polypeptide into a random coil, on the contrary, there is some residual conformation. The information about such a system is still meagre.

- Solution scattering offers clear advantages for systems where time resolution is essential (in particular the transient processes of biological systems such as enzyme reactions or the growth and differentiation of cellular components). This favourable factor has become very useful with the availability of intense synchrotron radiation sources: Time-resolved scattering experiments with millisecond resolution are possible. In view of advances in technology (the third generation of synchrotron radiation sources with high flux and brilliance is being built), the time required for measurements will be even shortened significantly and processes in the sub msec region may be observable.
- Crystallographic structures can be tested and used as propositions for solution structures ('structural checking'). In addition, in the absence of crystal structures, a computer model on the basis of homologous proteins can be used to interpret experimental results which will test it or differentiate between a number of competing models. However, the determination of the radius of gyration and maximum dimension of a molecule is not sufficient for an unambigous interpretation of experimental scattering profiles. The recording of scattering data up to medium resolution (10-20Å) is a crucial condition.
- The theoretical approach of molecular modelling based upon empirical force fields still needs further work before applying to more complicated systems. An optimal parameter set for general use in biomolecular simulations certainly requires the analysis of an increasing rate of experimental information in order to test predicted models which e.g. deal with overall conformations, hinge-bending mechanisms or intermolecular processes.

An unwritten law of solution X-ray scattering says: "*Never do anything that crystallographers or electron microscopists can do better*". Solution X-ray scattering results will certainly become obsolete one time when high-resolution structures will be available. Nevertheless, the results presented point out clearly that this method is a useful and unique tool for providing detailed insight into the conformation of proteins *in solution*. It offers a variety of opportunities, which can contribute to the understanding of the structure-function relationship of biological molecules. An improvement on the experimental side will be the availability of highly intense and collimated X-rays from a new generation of synchrotron radiation sources along with elaborate data acquisition systems. Of course, being a low to medium resolution technique, its usefulness can only bear fruit on the basis of atomic coordinates as 'starting' data taken from X-ray crystallography and sophisticated molecular modelling procedures.

Appendix A : Theoretical Concept of Molecular Model-Building

Proteins play a crucial role in all living systems. Knowledge of the three-dimensional structure of a protein is a pre-requisite to understanding its physical, chemical and biological properties. The number of different proteins in a human being is estimated to be of the order of 10^5 to 10^6. Although it is expected that by the year 2000 about 10 000 macromolecular models will have been experimentally determined, the ability to predict models based on known structures is gaining recognition and is likely to become an integral part of molecular biology.

The building units of proteins are amino acids, 20 in number, joined together to form the long molecular chains, known as polypeptides (see figure A.1). The polypeptide backbone consists of a repeated sequence of three atoms: the amide N, the alpha carbon C_α and the carbonyl C. The numbering of the amino acid residues starts fom the amino end (the N-terminal). Moreover, globular proteins frequently contain so-called prosthetic groups such as metal ions or heme moities (iron-protoporphyrin complexes) and their biological functions include interactions with one or more small molecules (ligands, substrates, hormones) or another macromolecule.

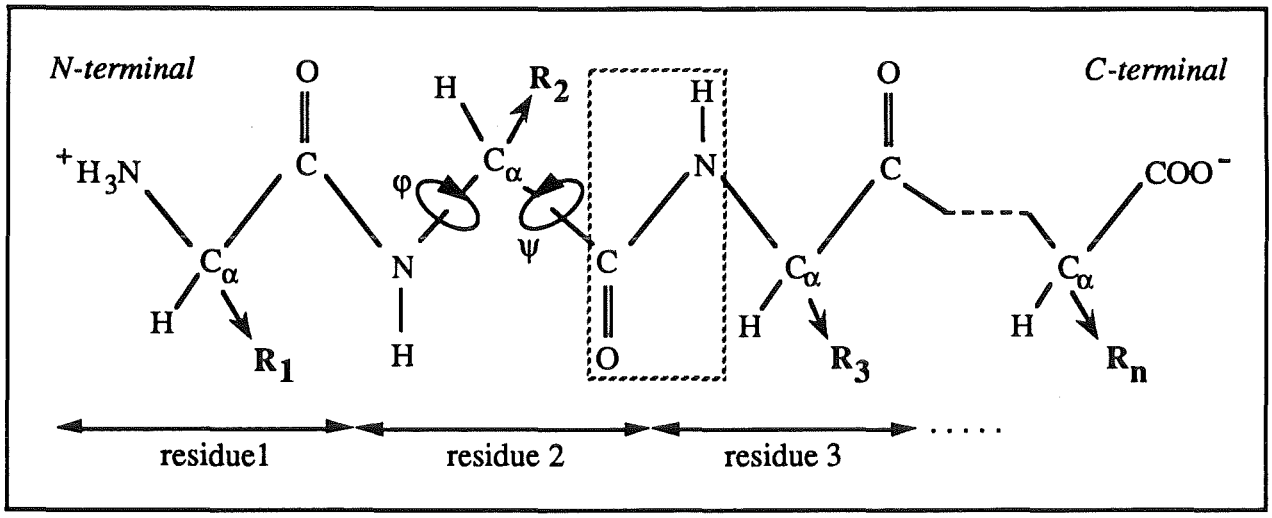


Figure A.1 :
Primary structure of a protein illustrating a portion of a polypeptide chain which begins with the α-amino group or N-terminal end and proceeds to the C-terminal end carrying a carboxyl group. At pH 7 the α-amino group is protonated and the carboxyl is deprotonated. Each amino acid residue contributes a group to the main polypeptide chain (the N-C_α-C backbone) and possesses a characteristic side chain **R** that differs for each of the 20 amino acids. In principle, rotation may occur about any of the three bonds of the polypeptide backbone. However, since almost no rotation is possible about the peptide bond (all atoms in the marked box lie in the same plane) the polypeptide configuration is determined by the torsional angles ψ and φ (indicated for amino acid residue 2) which are significantly affected by the nonbonded interactions of the side chains **R**.

The basic concept of molecular model-building is that a molecule or a group of molecules assumes a given structure due to the interaction of forces upon it. If these forces are allowed to assume their lowest possible values the most stable and lowest energy structure will ensue. In view of the large number of atoms constituting biological molecules only numerical methods can help to solve quantitative structural models. Due to the n^2-dependence of the degrees of freedom[17], which have to be optimized, the computation or CPU time increases strongly with larger molecules. However, the planar configuration of the peptide bond allows an enormous reduction in the number of variables necessary for a complete description of a protein molecule. Even though the force fields for molecular mechanics of extremely large molecules such as proteins and nucleic acids have to be simple as well as consistent with known experimental observations.

Statical (e.g. energy minimization) as well as dynamical calculations (e.g. motions or interactions) for a macromolecule are performed by means of force field and potential energy functions. Since it is impracticable to obtain such functions by the methods of quantum chemistry, the empirical energy function to model macromolecular systems is typically of the form (an extensive treatment is given in the descriptions of macromolecular modelling programmes such as AMBER [157], CHARMM [158] and GROMOS [159])

$$
\begin{aligned}
E = & \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 \\
& + \sum_{dihedrals} k_\phi[1+\cos(n\phi-\delta)] + \sum_{impropers} k_\omega(\omega-\omega_0)^2 \\
& + \sum_{nonbonded\ pairs} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \sum_{partial\ charges} \frac{q_i q_j}{4\pi\varepsilon_0 D r} \quad .
\end{aligned}
$$

The first two terms represent the contribution from bond and angle deformations. The use of harmonic potentials is justified in most cases owing to ordinary temperatures and the absence of chemical reactions. The natural bond lengths and bond angles are given by b_0 and θ_0, respectively, and k_b and k_θ are the force constants for bond stretching and angle bending. Along with the first two terms the third and fourth term belong to the internal energy terms (resulting from bonded interactions) and describe energy contributions from torsion (dihedral angle) and chirality or planarity (improper torsion angle). These four-atom based terms are defined according to figure A.2, their associated force constants (energy barriers) are k_ϕ and k_ω. The quantity ω_0 is the reference value for improper torsions, whereas δ is the dihedral reference angle (phase angle) and n is a dimensionless rotational

[17] Each non-linearily arranged molecule consisting of m atoms possesses 3m-6 internal degrees of freedom.

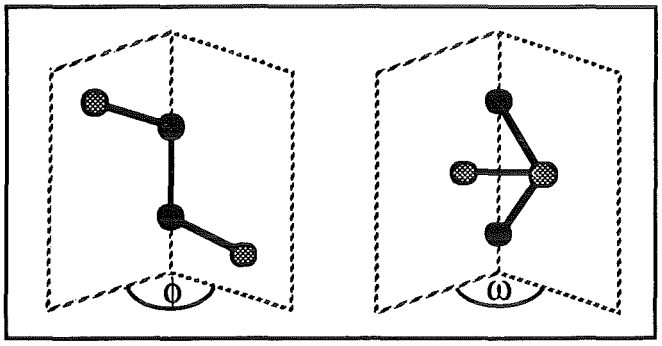


Figure A.2 :
Definition of the dihedral angle ϕ and the improper torsion angle ω.

symmetry factor. Typical values for n are 2 and 3; for instance, two energy minima occur for $n = 2$ (corresponding to a sp^2 bond), and there are three energy minima in the case $n = 3$ (sp^3 bond).

The terms five and six make up the nonbonded interactions, hence the summations exclude interactions between bonded atoms and also between atoms included in angle bending and in torsional terms. A potential function, that is frequently used in computer simulation studies, is the Lennard-Jones potential (fifth term) representing an approximation for the van der Waals energy. In this function the attractive force is proportional to r^{-6}, while the repulsive force varies as r^{-12}. ε gives the depth of the potential well and σ is the value of r for which the potential is zero. The van der Waals interaction is very small at long range and a saving in computational time can be achieved by neglecting the interaction for values of r greater than a cut-off value (about 9Å). Coulombic interactions between partial charges on atoms in the molecule are given by the sixth term. Because of the need to limit the number of pair interactions, different schemes for truncating the electrostatic potential are used. These are primarily linked with the use of the dielectric constant D, which actually describes a bulk phenomenon. Although it has no real meaning when applied to the scale of individual atoms, the nature of buried amino acids in proteins suggest that proteins have an internal dielectric constant D of the order of 1 to 5 [160]. However, a value different from unity is often used to allow for the dielectric behaviour of a system in which the solvent is not explicitly included. A distance-dependent D (e.g. $1/r$ is in wide-spead use) approximates solvent screening and polarization effects, however, there is no physical justification for this.

Contributions due to hydrogen bonding have been omitted in the above potential function. The effect of hydrogen bonds can be taken implicitly into account by appropriate parametrization of the partial charges and the van der Waals parameters. Nevertheless, a representation of an explicit hydrogen bonding term is advantageous for studies of H-bond network formation and protein-solvent interactions. Furthermore, in modelling studies the contribution from disulphide bridges plays an important role. Representing a crucial

conformational constraint, the connection between the sulphurs of two cysteine residues known to form a covalent bond, is usually incorporated in advance and treated as bonded interaction.

The parameters used in this potential energy function have been first determined by the consistent force field method of Lifson and coworkers [161-163] which includes all relevant and available experimental data (e.g. crystal structure parameters, vibrational and infrared spectra) of whole families of simple organic molecules. Extensive optimization [164-167] has been carried out in order to find the best agreement between calculated and experimental data. Nevertheless, the nonbonded parameters are the least well defined and the most empirical in nature. These values have been derived mainly from crystal packing properties including sublimation energy as well as from computer simulations of liquids and fit to quantum mechanical calculations to derive protein-atom partial charges. Thus an optimal parameter set for general use and advanced modelling approaches in biomolecular simulations requires the analysis of more experimental information. In addition, one has to keep in mind, the more accurate a parameter set the longer the calculation takes and the smaller the number of atoms that can be accomodated by the programme. For this reason the development of powerful computing machines contributes substantially to the progress of such theoretical studies.

Appendix B : Energy Minimization and Hydration

Most of the functional properties of proteins are only evident in solution, outside the crystal lattice. It is necessary to describe their properties under these conditions. Thus, interactions between the protein and the hydration shell contribute considerably to the energy of such a system. The protein determines the structure of the water up to a certain distance from the protein 'surface' and the water influences the protein dynamics. A common approach to protein energetics uses theoretically based potential functions with terms such as bond and angle distortion energies, electrostatic energy and van der Waals energy (see Appendix A). However, the explicit simulation of water presents additional challenges, in part because water interacts with proteins through electrostatic and van der Waals as well as hydrophobic[18] forces and in part because the structure of water is less well understood than the structure of proteins.

For modelling studies on biological macromolecules, it is of importance to consider also the solvent contribution. Unfortunately, X-ray diffraction experiments do not always reveal all the positions of protein-bound water molecules. Here, the solvent structure must be generated, completed or approximated by computational techniques. Due to the limited resolution of solution X-ray scattering, approaches of protein solvation have been proposed either by putting the dry molecule into a grid of water molecules [72] or by rescaling the crystal coordinates with a factor which approximates a hydrated model [169]. The method reported here, presents an approach based upon energy minimization to generate a hydration shell around a protein structure. Parts of this water shell will be included in the scattering pattern simulation in order to interpret and improve the fit to the solution X-ray scattering data. For a macromolecular system, however, it is frequently impossible to determine the global energy minimum. The minimization of the total energy yields a local minimum of the energy surface, i.e. generally a minimum in the neighbourhood of the model structure is examined. Thus, the purpose of the minimization was to provide an energetically reasonable structure including a sufficient amount of water molecules, which represent the first hydration layer[19] around the protein's surface. Furthermore, the minimization does not allow for large-scale or global changes, but rather optimizes the local structure where the protein had been modelled. The absolute value of the total energy is not meaningful. Here, only the models of the N-lobe of rabbit serum transferrin (RST/2N) in the crystallographic closed and in a modelled open state as well as the trimeric crystal structure of nitrite reductase (AcNiR) have been investigated.

---

[18] The hydrophobic interaction results from the ordering of water molecules around non-polar atoms of a protein (e.g. pentagonal rings of water have been found in various high-resolution crystal structures of small proteins [168]).

[19] Although being labile the water molecules belonging to this first hydration shell can occupy the same sites on a protein surface repeatedly [170]. These would be the sites expected to be seen in a X-ray experiment.

B.1 Computational procedure

The programme used in the present calculation was X-PLOR by A.T. Brünger (Version 3.0, Yale University, New Haven, CT, USA), a comprehensive programme for simulations and structural refinement of macromolecules. It has been implemented on a CONVEX computer system. X-PLOR has evolved from the CHARMM programme [158] and has therefore similar parameter and topology data structures as well as potential energy functions. An explicit hydrogen bond term completes the energy function given in Appendix A. Prior to any energy minimization, the positions of all hydrogen atoms which are needed for the force fields were generated. Finally, the structural models of transferrin and nitrite reductase were surrounded by solvent water using the programme BIOGRAF (Version 2.0, Biodesign Inc., CA, USA) running on a Stardent TITAN graphics workstation. Since the location and orientation of these water molecules (they have been placed on a diamond shaped grid) were still arbitrary, the determination of an optimal structure was attained by the following energy minimization procedure.

All calculations using X-PLOR were performed with the parameter set 'parmallh3x', extended by the parameters for the three site water model TIP3P [171]. For this water-protein system, a dielectric constant D of 1 was used. The method of conjugate gradients according to the algorithm of Powell [172] was employed for all energy minimization calculations. The most time consuming portion of the minimization is the calculation of the long range non-bonded and hydrogen bond interactions. Therefore, distance cutoffs are used to reduce the number of terms. However, simply cutting the non-bonded interactions off at a given distance leads to discontinuities in the energy and its derivatives. Thus, a so-called switching function (starting at 8 Å and reaching zero at 10 Å) was introduced in order to smoothly turn off the interactions and guarantees the mitigation of the discontinuities.

Enough water molecules were included to approximate solution. The water/vacuum interface was shifted up to a reasonable distance from the protein surface (7 Å to 10 Å) in order to eliminate energy discontinuities with respect to the first hydration layer. Thus, instead of using a convergence limit for the gradient of the total energy, the minimization calculation was considered to be converged when the number of water molecules within the first hydration shell remained stable (see figure B.1). The convergence was checked after every 100^{th} minimization step.

Throughout the energy minimization the atoms of the metal binding sites (i.e. metal ion and atoms of the protein ligands) were kept fixed[20]. In addition, as soon as the protein started to shrink (noticeable by a decrease of the radius of gyration) the backbone atoms of the protein were also constrained to their positions. Thus, despite the large number of atoms

[20] No standard force fields for modelling metalloproteins are yet available. Even though several force fields for metal ions have been developed, there is the fundamental problem that they can be used only for certain sets of ligands and for a well specified stereochemistry of the binding sites.

Table B.1 : Parameters of the energy minimization procedure for the three protein molecules

molecule	thickness of complete water shell	No. of water molecules	total No. of atoms	No. of water molecules in first hydration shell (start/end)	end of mini-mizatio n (steps)	used CPU time
RST/2N (closed)	10.5 Å	3485	15626	90/300	2300	10 h
RST/2N (open)	7.0 Å	1925	10943	110/350	1800	8 h
AcNiR	7.5 Å	2807	23670	450/700	1500	14 h

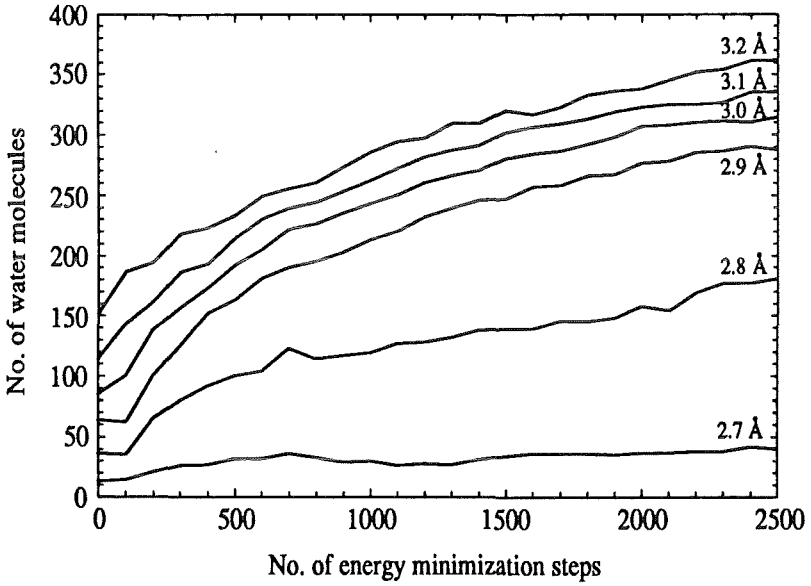


Figure B.1 : Integrated fractions of water molecules for distances between 2.7Å and 3.2Å from the protein surface of the iron-loaded N-lobe of rabbit serum transferrin evolving during the minimization procedure. The used convergence criterium required that the amount of water within about 3.0Å from the protein surface came to a saturation.

the computational efforts could be reduced. Some properties of the investigated structures are compiled in table B.1.

B.2 Effect of protein on water

The formation of hydration shells is clearly shown in figure B.2. Only the results for the closed, iron-loaded N-lobe of rabbit serum transferrin are presented. Similar results have been obtained for the other two structures, which are therefore not shown. The dramatic effect of the protein on the clustering of water molecules within 3 Å from the protein surface (i.e. the first hydration shell) indicates that the simulation is realistic. It is also supported by the fact that the inclusion of this solvent contribution for calculating the theoretical X-ray scattering profiles improves significantly the fit to experimental data (see Chapter 6). Moreover, the amount of water within the first hydration shell agrees well with the number expected (as a rule of thumb, about one water molecule per amino acid residue make up the closely bound solvent molecules).

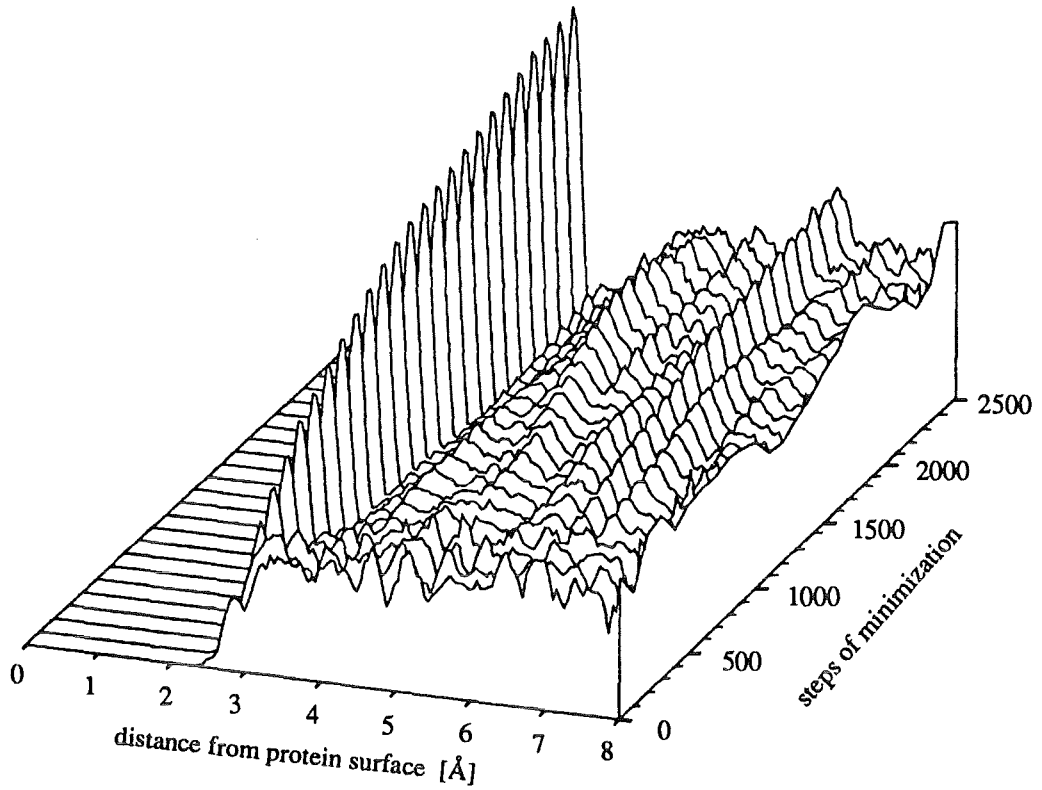


Figure B.2 :
Distance distributions of water molecules from the surface of the iron-loaded N-lobe of rabbit serum transferrin dependent on the course of energy minimization. The distance is defined as the distance from a water oxygen to the closest non-hydrogen atom of the protein. The amount of water molecules within the pronounced first hydration shell are a decisive factor in the structural composition and functional properties of a protein.

Although the inclusion of water molecules is able to improve the simulation of solution X-ray scattering profiles, this energy minimization procedure for approximating a solvent shell around a protein cannot yet be considered accurate. On one hand, there is no optimal parameter set for general use in molecular mechanics calculations and, on the other hand, not only water but also salt molecules and other ions (as well as carbohydrates in the case of glycoproteins) will contribute to the close environment of biological macromolecules in solution. Nevertheless, the limited resolution of solution X-ray scattering experiments does not require such an accuracy and the results presented here, indicate the importance of the first hydration layer as an additional contribution needed for the interpretation of X-ray scattering data.

A suggestion for further studies would be the comparison of the location of water molecules which have been simulated with those which might be found in a crystal structure refined to an even higher resolution. This method has been already used in the case of small proteins in order to improve simulations and also to refine crystallographic structures.

References

[1] Blundell, T. L. & Johnson, L. N. (1976)
Protein Crystallograpy.
Academic Press, New York.

[2] Helliwell, J. R. (1992)
Macromolecular crystallography with synchrotron radiation.
University Press, Cambridge.

[3] Wüthrich, K. (1986)
NMR of Proteins and Nucleic Acids.
Wiley, New York.

[4] Pilz, I. (1982)
Proteins. In *Small Angle X-ray Scattering* (Glatter, O. & Kratky, O., eds), pp. 239-293,
Academic Press, London & New York.

[5] Stuhrmann, H. B. (1990)
The role of small angle scattering in structure determination. In *Synchotron Radiation and Biophysics* (Hasnain, S. S., ed), pp. 223-242,
Ellis Horwood, Chichester.

[6] Watterson, J. G. (1991)
Role of water in cell function
Biophysics **36**, 1-26

[7] Bennett, W. S. & Huber, R. (1984)
Structural and functional aspects of domain motions in proteins.
CRC Crit. Rev. Biochem. **15**, 291-384.

[8] Huber, R. (1987)
Flexibility and rigidity, requirements for the function of proteins and protein pigment complexes.
Biochem. Soc. Trans. **15**, 1009-1020.

[9] Shortle, D. (1992)
Mutational studies of protein structures and their stabilities.
Quart. Rev. Biophys. **25**, 205-250.

[10] Cotton, F. A., Wilkinson, G. & Gaus, P. A. (1987)
Basic Inorganic Chemistry.
John Wiley & Sons, New York.

[11] Frausto da Silva, J. J. R. & Williams, R. J. P. (1991)
The Biological Chemistry of the Elements.
Claredon Press, Oxford.

[12] Brock, J. (1985)
Transferrins. In *Metalloproteins* (Harrison, P. M., eds), pp. 183-262,
Verlag Chemie, Weinheim.

[13] Harris, D. C. & Aisen, P. (1989)
Physical Biochemistry of the Transferrins. In *Iron Carriers and Iron Proteins* (Loehr, T. M., Gray, H. B. &Lever, A. B. P., eds), pp. 239-351,
VCH Publishers, Weinheim.

[14] Aisen, P. (1989)
Physical Biochemistry of the Transferrins: Update, 1984-1988. In *Iron Carriers and Iron Proteins* (Loehr, T. M., Gray, H. B. &Lever, A. B. P., eds), pp. 353-371,
VCH Publishers, Weinheim.

[15] Crichton, R. R. (1991)
Inorganic biochemistry of iron metabolism.
Ellis Horwood, Chichester.

[16] Metz-Boutigue, M.-H., Jolles, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J. & Jolles, P. (1984)
Human lactotransferrin : amino acid sequence and structural comparisons with other transferrins.
Eur. J. Biochem. **145**, 659-676.

[17] Spik, G., Coddeville, B. & Montreuil, J. (1988)
Comparative study of the primary structures of sero-, lacto- and ovotransferrin glycans from different species.
Biochimie **70**, 1459-1469.

[18] Aasa, R., Malmström, B. G., Saltman, P. & T.Vänngård (1963)
The specific binding of iron(III) and copper(II) to transferrin and conalbumin.
Biochim. Biophys. Acta **75**, 203-222.

[19] Aisen, P. & Leibman, A. (1968)
The stability constants of the Fe^{3+} conalbumin complexces.
Biochem. Biophys. Res. Commun. **30**, 407-413.

[20] Aisen, P. & Leibman, A. (1972)
Lactoferrin and transferrin: a comparative study.
Biochim. Biophys. Acta **257**, 314-323.

[21] Legendre, J. M., Moineau, M. P., Menez, J. F. & Turzo, A. (1985)
Analyse rapide des complexes transferrine-fer par isoélectrofocalisation et densitométrie radioactive sur le gel.
Path. Biol. **33**, 741-745.

[22] Roberts, T. K., Masson, P. L. & Heremans, J. F. (1973)
Heterogeneity of human lactoferrin. In *Immunology of Reproduction* (Bratanov, K., ed), pp. 706-709, Bulgarian Academy of Sciences Press, Sofia.

[23] Wenn, R. V. & Williams, J. (1968)
The Isoelectric Fractionation of Hen's-Egg Ovotransferrin.
Biochem. J. **108**, 69-74.

[24] Anderson, B. F., Baker, H. M., Dodson, E. J., Norris, G. E., Rumball, S. V., Waters, J. M. & Baker, E. N. (1987)
Structure of human lactoferrin at 3.2 Å resolution.
Proc. Nat. Acad. Sci, U.S.A. **84**, 1769-1773.

[25] Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W. & Baker, E. N. (1989)
Structure of human lactoferrin: crystallographic structure analysis and refinement at 2.8 Å resolution.
J. Mol. Biol. **209**, 711-734.

[26] Anderson, B. F., Baker, H. M., Norris, G. E., Rumball, S. V. & Baker, E. N. (1990)
Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins.
Nature (London) **344**, 784-787.

[27] Smith, C. A., Anderson, B. F., Baker, H. M. & Baker, E. N. (1992)
Metal Substitution in Transferrins: The Crystal Structure of Human Copper Lactoferrin at 2.1 Å Resolution.
Biochemistry **31**, 4527-4533.

[28] Bailey, S., Evans, R. W., Garratt, R. C., Gorinsky, B., Hasnain, S. S., Horsburgh, C., Jhoti, H., Lindley, P. F., Mydin, A., Sarra, R. & Watson, J. L. (1988)
Molecular Structure of Serum Transferrin at 3.3 Å Resolution.
Biochemistry **27**, 5804-5812.

[29] Sarra, R., Garratt, R., Gorinsky, B., Jhoti, H. & Lindley, P. (1990)
High-Resolution X-ray Studies on Rabbit Serum Transferrin: Preliminary Structure Analysis of the N-Terminal Half-Molecule at 2.3 Å Resolution.
Acta Cryst. **46**, 763-771.

[30] Martel, P., Kim, S. M. & Powell, B. M. (1980)
Physical Characteristics of Human Transferrin from Small Angle Neutron Scattering.
Biophys. J. **31**, 371-380.

[31] Kilár, F. & Simon, I. (1985)
The Effect of Iron Binding on the Conformation of Transferrin. A Small Angle X-Ray Scattering Study.
Biophys. J. **48**, 799-802.

[32] Vígh, R., Cser, L., Kilár, F. & Simon, I. (1989)
Different Segmental Flexibility of Human Serum Transferrin and Lactoferrin.
Arch. Biochem. Biophys. **275**, 181-184.

[33] Baker, E. N. & Lindley, P. F. (1992)
New Perspectives on the Structure and Function of Transferrins.
J. Inorg. Biochem. **47**, 147-160.

[34] Adman, E. T. (1991)
Copper Protein Structures.
Advan. Protein Struct. **42**, 145-197.

[35] Heizmann, C. W. & Hunziker, W. (1991)
Intracellular Calcium-binding Proteins: More Sights than Insights.
Trends Biochem. Sci. **16**, 98-103.

[36] Vallee, B. L. & Auld, D. S. (1990)
Zinc Coordination, Function, and Structure of Zinc Enzymes and Other Proteins.
Biochemistry **29**, 5647-5659.

[37] Garratt, R. C., Ph.D. thesis (1989),
X-ray Structural Studies on Transferrins,
University of London.

[38] Mellor, R. B., Ronnenberg, J., Campbell, W. H. & Diekmann, S. (1992)
Reduction of nitrate and nitrite in water by immobilized enzymes.
Nature (London) **355**, 717-719.

[39] Hochstein, L. I. & Tomlinson, G. A. (1988)
The enzymes associated with denitrification.
Ann. Rev. Microbiol. **42**, 231-261.

[40] Brittain, T., Blackmore, R., Greenwood, C. & Thomson, A. J. (1992)
Bacterial nitrite-reducing enzymes.
Eur. J. Biochem. **209**, 793-802.

[41] Kakutani, T., Watanabe, H., Arima, K. & Beppu, T. (1981)
Purification and Properties of a Copper-Containing Nitrite Reductase from a Denitrifying Bacterium, Alcaligenes faecalis Strain S-6.
J. Biochem. **89**, 453-461.

[42] Petratos, K., Beppu, T., Banner, D. W. & Tsernoglou, D. (1986)
Preliminary Characterization of Crystals of Nitrite Reductase Isolated from Alcaligenes Faecalis Strain S-6.
J. Mol. Biol. **190**, 135.

[43] Iwasaki, H. & Matsubara, T. (1972)
A Nitrite Reductase from Achromobacter cycloclastes.
J. Biochem. **71**, 645-652.

[44] Iwasaki, H., Noji, S. & Shidara, S. (1975)
Achromobacter cycloclastes Nitrite Reductase. The Function of Copper, Amino Acid Composition, and ESR Spectra.
J. Biochem. **78**, 355-361.

[45] Liu, M.-Y., Liu, M.-C., Payne, W. J. & Le Gall, J. (1986)
Properties and Electron Transfer Specifity of Copper Proteins from the Denitrifier "Achromobacter cycloclastes".
J. Bacteriol. **166**, 604-608.

[46] Iwasaki, H., Shidara, S., Suzuki, H. & Mori, T. (1963)
Studies on Denitrification. VII. Further Purification and Properties of Denitrifying Enzymes.
J. Biochem. **53**, 299-303.

[47] Masuko, M., Iwasaki, H., Sakurai, T., Suzuki, S. & Nakahara, A. (1984)
Characterization of Nitrite Reductase from a Denitrifier, Alcaligenes Sp. NCIB 11015. A Novel Copper Protein.
J. Biochem. **96**, 447-454.

[48] Sykes, A. G. (1991)
Active-site properties of the blue copper proteins.
Advan. Inorg. Chem. **36**, 377-408.

[49] Abraham, Z. H. L., Lowe, D. J. & Smith, B. E. (1993)
Purification and characterization of the dissimilatory nitrite reductase from *Alcaligenes xylosoxidans* subsp. *xylosoxidans* (NCIMB 11015): Evidence for the presence of both type 1 and type 2 copper centres.
Submitted to *Biochem. J.*.

[50] Godden, J. W., Turley, S., Teller, D. C., Adman, E. T., Liu, M. Y., Payne, W. J. & Le Gall, J. (1991)
The 2.3 Angstrom X-ray Structure of Nitrite Reductase from Achromobacter cycloclastes.
Science **253**, 438-442.

[51] Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992)
Crystal structure of a dUTPase.
Nature (London) **355**, 740-743.

[52] Mattevi, A., Obmolova, G., Schulze, E., Kalk, K. H., Westphal, A. H., DeKork, A. & Hol, W. G. J. (1992)
Atomic Structure of the Cubic Core of the Pyruvate Dehydrogenase Multienzyme Complex.
Science **255**, 1544-1550.

[53] Weiss, M. S., Abele, U., Weckesser, J., Welte, W., Schiltz, E. & Schulz, G. E. (1991)
Molecular Architecture and Electrostatic Properties of a Bacterial Porin.
Science **254**, 1627-1630.

[54] Glatter, O. & Kratky, O. (1982)
Small Angle X-ray Scattering.
Academic Press, London.

[55] Feigin, L. A. & Svergun, D. I. (1987)
Structure Analysis by Small-Angle X-Ray and Neutron Scattering.
Plenum Press, New York & London.

[56] Luzzati, V. & Tardieu, A. (1980)
Recent developments in solution X-ray scttering.
Ann. Rev. Biophys. Bioeng. **9**, 1-29.

[57] Doniach, S. (1985)
Small-Angle X-ray Scattering and Structural Biology.
Physics Today **4**, 68-76.

[58] Moore, P. B. (1988)
On the Role of Small-Angle Scattering in Biological Research.
J. Appl. Cryst. **21**, 675-680.

[59] Ueki, T. (1991)
Biopolymers and solution X-ray scattering. Is it useful in structural research in biology?
Nucl. Instr. Meth. Phys. Res. **A303**, 464-475.

[60] Debye, P. (1915)
Zerstreuung von Röntgenstrahlen.
Ann. Phys. (Leipzig) **46**, 809-823.

[61] Guinier, A. & Fournet, G. (1955)
Small Angle Scattering of X-rays.
John Wiley & Sons, Inc., New York.

[62] Debye, P. & Bueche, A. M. (1949)
Scattering by an Inhomogeneous Solid.
J. Appl. Phys. **20**, 518-525.

[63] Glatter, O. (1977)
A New Method for the Evaluation of Small-Angle Scattering Data.
J. Appl. Cryst. **10**, 415-421.

[64] Moore, P. B. (1980)
Small-Angle Scattering. Information Content and Error Analysis.
J. Appl. Cryst. **13**, 168-175.

[65] Provencher, S. W. (1982)
A constrained regularization method for inverting data represented by linear algebraic or integral equations.
Computer Phys. Commun. **27**, 213-227 & 229-242.

[66] Svergun, D. I., Semenyuk, A. V. & Feigin, L. A. (1988)
Small-Angle-Scattering-Data Treatment by the Regularization Method.
Acta Cryst. **A44**, 244-250.

[67] Shannon, C. E. (1949)
Communication in the Presence of Noise.
Proc. I.R.E. **37**, 10-21.

[68] Stuhrmann, H. B. (1970)
Interpretation of Small-Angle Scattering Functions of Dilute Solutions and Gases. A Representation of the Structures Related to a One-Particle-Scattering Function.
Acta Cryst. **A26**, 297-306.

[69] Stuhrmann, H. B. (1970)
Ein neues Verfahren zur Bestimmung der Oberflächenform und der inneren Struktur von gelösten globulären Proteinen aus Röntgenkleinwinkelmessungen.
Z. Phys. Chem. **72**, 177-184.

[70] Svergun, D. I. & Stuhrmann, H. B. (1991)
New Developments in Direct Shape Determination from Small-Angle Scattering. 1. Theory and Model Calculations.
Acta Cryst. **A47**, 736-744.

[71] Ninio, J., Luzzati, V. & Yaniv, M. (1972)
Comparative Small-angle X-ray Scattering Studies on Unacylated, Acylated and Cross-linked *Escheria coli* Transfer RNA_I^{Val}.
J. Mol. Biol. **71**, 217-229.

[72] Fedorov, B. A., Ptitsyn, O. B. & Voronin, L. A. (1972)
X-ray diffuse scattering of globular protein solutions: consideration of the solvent influence.
FEBS Lett. **28**, 188-190.

[73] Fedorov, B. A. & Denesyuk, A. I. (1978)
Large-Angle X-ray Diffuse Scattering, a New Method for Investigating Changes in the Conformation of Globular Proteins in Solutions.
J. Appl. Cryst. **11**, 473-477.

[74] Müller, J. J. (1983)
Calculation of Scattering Curves for Macromolecules in Solution and Comparison with Results of Methods using Effective Atomic Scattering Factors.
J. Appl. Cryst. **16**, 74-82.

[75] Pickover, C. A. & Engelman, D. M. (1982)
On the Interpretation and Prediction of X-Ray Scattering Profiles of Biomolecules in Solution.
Biopolymers **21**, 817-831.

[76] Lattman, E. E. (1989)
Rapid Calculation of the Solution Scattering Profile From a Macromolecule of Known Structure.
Proteins: Struct. Funct. Genet. **5**, 149-155.

[77] Glatter, O. (1980)
Computation of Distance Distribution Functions and Scattering Functions of Models for Small Angle Scattering Experiments.
Acta Phys. Austr. **52**, 243-256.

[78] Rol'bin, Y. A., Kayushina, R. L., Feigin, L. A. & Shchedrin, B. M. (1974)
Calculation of the small-angle X-ray scattering intensity on a computer using a macromolecule model.
Sov. Phys. Crystallogr. **18**, 442-444.

[79] Olson, A. J. & Goodsell, D. S. (1992)
Macromolecular graphics.
Curr. Opin. Struct. Biol. **2**, 193-201.

[80] Brooks, C. L., Karplus, M. & Pettitt, B. M. (1988)
Proteins: a theoretical perspective of dynamics, structure, and thermodnamics.
John Wiley, New York.

[81] Peticolas, W. (1979)
Low frequency vibrations and the dynamics of proteins and peptides.
Meth. Enzymol. **61**, 425-458.

[82] Smith, J. C. (1991)
Protein dynamics: comparison of simulations with inelastic neutron scattering experiments.
Quart. Rev. Biophys. **23**, 227-291.

[83] Rupley, J. A. & Careri, G. (1991)
Protein Hydration and Function.
Advan. Protein Chem. **41**, 37-172.

[84] Richardson, J. S., Richardson, D. C., Tweedy, N. B., Gernert, K. M., Quinn, T. P., Hecht, M. H., Erickson, B. W., Yan, Y., McClain, R. D., Donlan, M. E. & Surles, M. C. (1992)
Looking at proteins: representations, folding, packing, and design.
Biophys. J. **63**, 1186-1209.

[85] Thornton, J. M. (1992)
Lessons from analyzing protein structures.
Curr. Opin. Struct. Biol. **2**, 888-894.

[86] Müller, J. J., Pankow, H., Poppe, B. & Damaschun, G. (1992)
The database BIOSCAT: a tool for structure research by scattering and hydrodynamic methods.
J. Appl. Cryst. **25**, 803-806.

[87] *SERC Daresbury Laboratory*, Annual Report 1991/1992.

[88] Wille, K. (1991)
Synchrotron radiation sources.
Rep. Prog. Phys. **54**, 1005-1068.

[89] Margaritondo, G. (1988)
Introduction to Synchrotron Radiation.
Oxford University Press, New York - Oxford.

[90] Krinsky, S., Perlman, M. L. & Watson, R. E. (1983)
Characteristics of Synchrotron Radiation and of its Sources. In *Handbook on Synchrotron Radiation (Vol.1A)* (Koch, E.-E., ed), pp. 65-171,
North-Holland Publishing Company, Amsterdam-New York-Oxford.

[91] Luke, Y. L. (1975)
Mathematical Functions and their Approximations.
Academic Press, New York.

[92] Schwinger, J. (1949)
On the Classical Radiation of Accelerated Electrons.
Phys. Rev. **75**, 1912-1925.

[93] Bordas, J. (1985)
Synchrotron Radiation X-Ray Scattering and Diffraction : Applications to Structural Research in Biology.
Z. Phys. B - Condensed Matter 389-395.

[94] Towns-Andrews, E., Berry, A., Bordas, J., Mant, G. R., Murray, P. K., Roberts, K., Sumner, I., Worgan, J. S. & Lewis, R. (1989)
Time-resolved X-ray diffraction station: X-ray optics, detectors, and data acquisition.
Rev. Sci. Instrum. **60**, 2346-2349.

[95] Lewis, R., Sumner, I., Berry, A., Bordas, J., Gabriel, A., Mant, G., Parker, B., Roberts, K. & Worgan, J. (1988)
Multiwire X-ray detector system at the Daresbury SRS.
Nucl. Instr. Meth. Phys. Res. **A273**, 773-777.

[96] Worgan, J. S., Lewis, R., Fore, N. S., Sumner, I. L., Berry, A., Parker, B., D'Annunzio, F., Martin-Fernandez, M. L., Towns-Andrews, E., Harries, J. E., Mant, G. R., Diakun, G. P. & Bordas, J. (1990)
The Application of Multiwire X-ray Detectors to Experiments using Synchrotron Radiation.
Nucl. Instr. Meth. Phys. Res. **A291**, 447-454.

[97] Charpak, G. & Sauli, F. (1984)
High-resolution electronic particle detectors.
Ann. Rev. Nucl. Part. Sci. **34**, 285-349.

[98] Oe, H., Doi, E. & Hirose, M. (1988)
Amino-terminal half-molecules of ovotransferrin: preparation by a novel procedure and their interactions.
J. Biochem. **103**, 1066-1072.

[99] Funk, W. D., MacGillivray, R. T. A., Mason, A. B., Brown, S. A. & Woodworth, R. C. (1990)
Expression of the Amino-Terminal Half-Molecule of Human Serum Transferrin in Cultured Cells and Characterization of Recombinant Protein.
Biochemistry **29**, 1654-1660.

[100] Woodworth, R. C., Mason, A. B., Funk, W. D. & MacGillivray, R. T. A. (1991)
Expression and Initial Characterization of Five Site-Directed Mutants of the N-Terminal Half-Molecule of Human Transferrin.
Biochemistry **30**, 10824-10829.

[101] Boulin, C., Kempf, R., Koch, M. H. J. & McLaughlin, S. M. (1986)
Data Appraisal, Evaluation and Display for Synchrotron Radiation Experiments: Hardware and Software.
Nucl. Instr. Meth. Phys. Res. **A249**, 399-407.

[102] Semenyuk, A. V. & Svergun, D. I. (1991)
GNOM - a Program Package for Small-Angle Scattering Data Processing.
J. Appl. Cryst. **24**, 537-540.

[103] Porod, G. (1951)
Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen (1. Teil).
Kolloid-Zeitschrift **124**, 83-114.

[104] Luzzati, V. (1960)
Interpretation des Mesures Absolues de Diffusion Centrale des Rayons X en Collimation Ponctuelle ou Lineaire: Solutions de Particules Globulaire et de Bâtonnets.
Acta Cryst. **13**, 939-945.

[105] Svergun, D. I. (1992)
Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria.
J. Appl. Cryst. **25**, 495-503.

[106] Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., F. M., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977)
The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures.
J. Mol. Biol. **112**, 535-542.

[107] Abola, A., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987)
Protein Data Bank. In *Crystallographic Databases - Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G. &Sievers, R., eds), pp. 107-132,
Data Commision of the International Union of Crystallography, Bonn/Cambridge/Chester.

[108] Pantos, E., SERC Daresbury Laboratory, unpublished.

[109] Evans, R. W., Williams, J. & Moreton, K. (1982)
A variant of human transferrin with abnormal properties.
Biochem. J. **201**, 19-26.

[110] Evans, R. W., Meilak, A., Aitken, A., Patel, K. J., Wong, C., Garratt, R. C. & Chitnavis, B. (1988)
Characterization of the amino acid change in a transferrin variant.
Biochem. Soc. Trans. **16**, 834-835.

[111] Lindley, P. F., Bajaj, M., Evans, R. W., Garratt, R. C., Hasnain, S. S., Jhoti, H., Kuser, P., Neu, M., Patel, K., Sarra, R., Strange, R. & Walton, A. (1993)
The Mechanism of Iron Uptake by Transferrins: the Structure of an 18kD NII-Domain Fragment from Duck Ovotransferrin at 2.3 Å Resolution.
Acta Cryst. **D49**, 292-304.

[112] Young, S. P., Bomford, A. & Williams, R. (1984)
The effect of the iron saturation of transferrin on its binding and uptake by rabbit reticulocytes.
Biochem. J. **219**, 505-510.

[113] Young, S. P., Bomford, A., Madden, A. D., Garratt, R. C., Williams, R. & Evans, R. W. (1984)
Abnormal *in-vitro* function of a variant human transferrin.
Br. J. Haematol. **56**, 581-588.

[114] Aisen, P. & Brown, E. B. (1975)
Transferrin structure and function.
Progr. Haematol. **9**, 25-56.

[115] Candy, J. M., Oakley, A. E., Klinowski, J., Carpenter, T. A., Perry, R. H., Atack, J. R., Perry, E. K., Blessed, G., Fairbairn, A. & Edwardson, J. A. (1986)
Aluminosilicates and senile plaque formation in Alzheimer's disease.
Lancet **1**, 354-357.

[116] Taylor, D. M. (1967)
The effects of desferrioxamine on the retention of actinide elements in the rat.
Health Phys. **13**, 135-140.

[117] Taylor, D. M. (1972)
Interactions between transuranium elements and the component of cells and tissues.
Health Phys. **22**, 575-581.

[118] Jandl, J. H. & Katz, J. H. (1963)
The plasma to cell cycle of transferrin.
J. Clin. Invest. **41**, 314-320.

[119] Durbin, P. W. (1962)
Distribution of the transuranic elements in mammals.
Health Phys. **8**, 665-671.

[120] Then, G. M., Appel, H., Duffield, J., Taylor, D. M. & Thies, W.-G. (1986)
In vivo and *in vitro* studies of Hafnium-binding to rat serum transferrin.
J. Inorg. Chem. **27**, 255-270.

[121] Duffield, J. R. & Taylor, D. M. (1987)
A Spectroscopic Study on the Binding of Plutonium(IV) and its Chemical Analogues to Transferrin.
Inorg. Chim. Acta **140**, 365-367.

[122] Taylor, D. M., Thies, W.-G., Appel, H., Neu-Müller, M., Schwab, F. J. & Weber, H. W. (1988)
Spectroscopic studies of hafnium-transferrin-complexes: differences between the two binding sites.
In *Metal Ion Homeostasis: Molecular Biology and Chemistry, UCLA Symposia on Molecular and Cellular Biology* (Hamer, D. H. & Winge, D. R., eds), pp. 179-188, Alan R. Liss Inc., New York.

[123] Harris, W. R., Carrano, C. J., Pecoraro, V. L. & Raymond, K. N. (1981)
Siderophilin Metal Coordination. 1. Complexation of Thorium by Transferrin: Structure-Function Implications.
J. Am. Chem. Soc. **103**, 2231-2237.

[124] Smith, C. A., Baker, H. M. & Baker, E. N. (1991)
Preliminary Crystallographic Studies of Copper(II)- and Oxalate-substituted Human Lactoferrin.
J. Mol. Biol. **219**, 155-159.

[125] Cotton, S. (1991)
Lanthanides and actinides.
Macmillan Education Ltd., London.

[126] Dautry-Varsat, A. & Lodish, H. F. (1984)
How Receptors Bring Proteins and Particles into Cells
Scientific American **251** (5), 48-54.

[127] Durbin, P. W. (1975)
Plutonium in mammals, influence of plutonium chemistry route of administration and physiological status of the animal on initial distribution and long-term metabolism.
Health Phys. **29**, 495-510.

[128] Shannon, R. D. (1976)
Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides.
Acta Cryst. **A32**, 751-767.

[129] Evans, R. W. & Ogwang, W. (1988)
Interaction of indium with tranferrin.
Biochem. Soc. Trans. **16**, 833-834.

[130] McIntyre, P. A., Larson, S. M., Eikman, E. A., Colman, M. & Scheffel, U. (1974)
Comparison of the metabolism of iron labeled tranferrin and indium labeled transferrin by the erythropoietic marrow.
J. Nucl. Med. **15**, 856-862.

[131] Evans, R. W., Ogwang, R. W., Patel, K. J., Marsden, P. J., Smith, F. A. & Bomford, A. (1989)
Studies on the interaction of indium with transferrin (Abstract).
9th International Conference on Iron and Iron Proteins, 1989. Brisbane, Australia.

[132] Raijmakers, P. G. H. M., Groeneveld, A. B. J., Den Hollande, W. & Teule, G. J. J. (1992)
Transport of Gallium-67 and Indium-111 across a membrane, role of plasma binding and concentration gradients.
Nucl. Med. Commun. **13**, 349-356.

[133] Farrar, G., Morton, A. P. & Blair, J. A. (1987)
The intestinal absorption and tissue distribution of aluminium, gallium and scandium: a comparative study.
Biochem. Soc. Trans. **15**, 1164-1165.

[134] Perl, D. P. (1985)
Relationship of aluminium to Alzheimer's disease.
Environ. Health Perspect. **63**, 149-153.

[135] Fatemi, S. J. A., Kadir, F. H. A. & Moore, G. R. (1991)
Aluminium transport in blood serum, binding of aluminium by human transferrin in the presence of albumin and citrate.
Biochem. J. **280**, 527-532.

[136] Martin, R. B. (1986)
The chemistry of aluminium as related to biology and medicine.
Clin. Chem. **32**, 1797-1806.

[137] Cochran, M., Coates, J. & Neoh, S. (1984)
The competitive equilibrium between aluminium and ferric ions for the binding sites of transferrin.
FEBS Lett. **176**, 129-132.

[138] McGregor, S. J., Brown, D. & Brock, J. H. (1991)
Transferrin-gallium binding in Alzheimer's disease.
Lancet **338**, 1394-1395.

[139] Morris, C., Candy, J. M., Court, J. A., Witford, C. A. & Edwardson, J. A. (1987)
The role of transferrin in the uptake of aluminium and manganese by the IMR 32 neuroblastoma cell line.
Biochem. Soc. Trans. **15**, 498.

[140] Anghileri, L. J. (1988)
Accumulation mechanism of gallium-67 by tumors and inflammatory lesions.
J. Med. Nucl. Biophys. **12**, 81-88.

[141] Rose, T. M., Plowman, G. D., Teplow, D. B., Dreyer, W. J., Hellstrom, K. E. & Brown, J. P. (1986)
Primary structure of the human melanoma-associated antigen p97 (melanotransferrin) deduced from the mRNA sequence.
Proc. Nat. Acad. Sci, U.S.A. **83**, 1261-1265.

[142] Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Wayne, M. M. Y. & Winter, G. (1985)
Hydrogen bonding and biological specifity analyzed by protein engineering.
Nature (London) **314**, 235-238.

[143] Fenderson, F. F., Kumar, S., Adman, E. T., Liu, M.-Y., Payne, W. J. & Le Gall, J. (1991)
Amino Acid Sequence of Nitrite Reductase: A Copper Protein from *Achromobacter cycloclastes*.
Biochemistry **30**, 7180-7185.

[144] Lee, B. & Richards, F. M. (1971)
The Interpretation of Protein Structures: Estimation of Static Accessibility.
J. Mol. Biol. **55**, 379-400.

[145] Miller, S., Lesk, A. M., Janin, J. & Chothia, C. (1987)
The accessible surface area and stability of oligomeric proteins.
Nature (London) **328**, 834-836.

[146] Janin, J., Miller, S. & Chothia, C. (1988)
Surface, Subunit Interfaces and Interior of Oligomeric Proteins.
J. Mol. Biol. **204**, 155-164.

[147] Mao, B., Pear, M. R., McCammon, J. A. & Quiocho, F. A. (1982)
Hinge-bending in L-Arabinose-binding Protein.
J. Biol. Chem. **257**, 1131-1133.

[148] Horovitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A. R. (1990)
Strength and Co-operativity of Contributions of Surface Salt Bridges to Protein Stability.
J. Mol. Biol. **216**, 1031-1044.

[149] Borhani, D. W. & Harrison, S. C. (1991)
Crystallization and X-ray Diffraction Studies of a Soluble Form of the Human Transferrin Receptor.
J. Mol. Biol. **218**, 685-689.

[150] Giedroc, D. P., Keating, K. M., Williams, K. R. & Coleman, J. E. (1987)
The function of Zinc in Gene 32 Protein from T4.
Biochemistry **26**, 5251-5259.

[151] Pan, T., Giedroc, D. P. & Coleman, J. E. (1989)
^{1}H NMR Studies of T4 Gene 32 Protein: Effects of Zinc Removal and Reconstitution.
Biochemistry **28**, 8828-8832.

[152] Shaw, G. S., Hodges, R. S. & Sykes, B. D. (1990)
Calcium-Indiuced Peptide Association to Form an Intact Protein Domain: ^{1}H NMR Structural Evidence.
Science **249**, 280-283.

[153] Suzuki, S., Yoshimura, T., Kohzuma, T., Shidara, S., Masuko, M., Sakurai, T. & Iwasaki, H. (1989)
Spectroscopic evidence for a copper-nitrosyl intermediate in nitrite reduction by blue copper-containing nitrite reductase.
Biochem. Biophys. Res. Commun. **164**, 1366-1372.

[154] Baker, E. N., Rumball, S. V. & Anderson, B. F. (1987)
Transferrins: insights into structure and function from studies on lactoferrin.
Trends Biochem. Sci. **12**, 350-353.

[155] Pflugrath, J. W. & Quiocho, F. A. (1988)
The 2 Å Resolution Structure of the Sulfate-binding Protein Involved in Active Transport in *Salmonella typhimurium*.
J. Mol. Biol. **200**, 163-180.

[156] Brownlie, P., Louie, G. V., Lambert, R., Cooper, J. B., Blundell, T. L., Wood, S. P., Warren, M. J., Woodcock, S. C. & Jordan, P. M. (1992)
Porphobilinogen Deaminase - a flexible multidomain polymerase structure.
Pittsburgh Diffraction Conference, University of Pittsburgh.

[157] Weiner, P. K. & Kollman, P. A. (1981)
AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions.
J. Comp. Chem. **2**, 287-303.

[158] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983)
CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations.
J. Comp. Chem. **4**, 187-217.

[159] Van Gunsteren, W. F. & Berensden, H. J. C. (1987)
Groningen Molecular Simulation (GROMOS) Library Manual, Biomos, Groningen.

[160] Pethig, R. (1979)
Dielectric and Electronic Properties of Biological Materials.
John Wiley & Sons, Chichęster.

[161] Lifson, S. & Warshel, A. (1968)
Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkene and n-Alkane Molecules.
J. Chem. Phys. **49**, 5116-5129.

[162] Warshel, A. & Lifson, S. (1970)
Consistent Force Field Calculations. II. Crystal Structures, Sublimation Energies, Molecular and Lattice Vibrations, Molecular Conformations, and Enthalpies of Alkanes.
J. Chem. Phys. **53**, 582-594.

[163] Lifson, S., Hagler, A. T. & Dauber, P. (1979)
Consistent Force Field Studies of Intermolecular Forces in Hydrogen-Bonded Crystals. 1. Carboxylic Acids, Amides, and the C=O···H- Hydrogen Bonds.
J. Am. Chem. Soc. **101**, 5111-5121.

[164] Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975)
Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids.
J. Phys. Chem. **79**, 2361-2381.

[165] Némethy, G., Pottle, M. S. & Scheraga, H. A. (1983)
Energy Parameters in Polypeptides. 9. Updating of Geometrical Parameters, Nonbonded Interactions, and Hydrogen Bond Interactions for the Naturally Occuring Amino Acids.
J. Phys. Chem. **87**, 1883-1887.

[166] Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta Jr., S. & Weiner, P. (1984)
A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins.
J. Am. Chem. Soc. **106**, 765-784.

[167] Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986)
An All Atom Force Field for Simulations of Proteins and Nucleic Acids.
J. Comp. Chem. **7**, 230-252.

[168] Teeter, M. M. (1991)
Water-Protein Interactions: Theory and Experiment.
Annu. Rev. Biophys. Biophys. Chem. **20**, 577-600.

[169] Perkins, S. J., Nealis, A. S., Sutton, B. J. & Feinstein, A. (1991)
Solution Structure of Human and Mouse Immunoglobulin M by Synchrotron X-ray Scattering and Molecular Graphics Modelling.
J. Mol. Biol. **221**, 1345-1366.

[170] Otting, G. & Wüthrich, K. (1989)
Studies of Protein Hydration in Aqueous Solution by Direct NMR Observation of Individual Protein-Bound Water Molecules.
J. Am. Chem. Soc. **111**, 1871-1875.

[171] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983)
Comparison of simple potential functions for simulating liquid water.
J. Chem. Phys. **79**, 926-935.

[172] Powell, M. J. D. (1977)
Restart Procedures for the Conjugate Gradient Method.
Math. Prog. **12**, 241-254.