



SAPIENZA
UNIVERSITÀ DI ROMA

Structural and spectroscopic studies of proteins in stress conditions

Coordinator of the PhD Course: Prof. Osvaldo Lanzalunga
Dottorato di Ricerca in Scienze Chimiche – XXIX Ciclo

Candidate

Alessandra Del Giudice
ID number 1209079

Thesis Advisor

Prof. Nicolae Viorel Pavel

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemical Sciences
October 2016

Thesis defended on 19th December 2016
in front of a Board of Examiners composed by:
Prof. Massimo Curini (chairman)
Prof. Alessandra Gentili
Prof. Fabio Ramondo

Structural and spectroscopic studies of proteins in stress conditions

Ph.D. thesis. Sapienza – University of Rome

© 2016 Alessandra Del Giudice. All rights reserved

This thesis has been typeset by L^AT_EX and the Sapthesis class.

Version: January 18, 2017

Author's email: alessandra.delgiudice@uniroma1.it

Dedicated to Nicola, Giovanna and Fortunato

Abstract

This thesis incorporates studies about the structural characterization of proteins in solution mainly by means of small angle X-ray scattering (SAXS).

The major part of the thesis has as main focus the study of the effect of stress conditions on the conformation of Human Serum Albumin (HSA), the most abundant plasma protein. The consequences of the hypochlorite-induced oxidation at increasing doses were explored with SAXS in combination with UV-visible spectroscopy and fluorescence, together with additional characterization. Despite the chemical modification, the native shape was preserved up to oxidant/HSA molar ratio < 80 , above which a structural transition occurred in the critical oxidant/HSA molar ratio range between 80-120. This conformational variation involved the drifting of one of the end-domains from the rest of the protein and corresponded to the loss of one third of the α -helix and a net increase of the protein negative charge. The high reproducibility and well-defined nature of this transition suggested that it represents a structural response characteristic of this multi-domain protein that was never observed before.

The unfolding in acid conditions is another typical conformational variation of serum albumin, the existence of which is well known in this case. We managed to implement an experimental setup in which the process could be followed *in-situ* as a function of a continuous decrease of pH in time. This allowed for the correlation between the structural and spectroscopic data as a function of pH together with an attempt to resolve the structural conformers involved in the transition.

In the framework of a collaboration with a structural biology - molecular plant physiology group of the University of Bologna, this thesis also includes a structural study in solution of redox-sensitive proteins of photosynthetic organisms and of their regulatory complexes. The SAXS analysis aided by 3D-modeling methods provided an insight into the shape of the individual enzymes and the multi-protein constructs, and into the way these proteins assembly and disassembly in response to redox regulation.

Acknowledgments

The results presented in this thesis are not only the outcome of my own work in the last three years: I want to acknowledge the passionate dedication of the people who contributed.

I would like to express my gratitude to my supervisor Prof. Pavel and to Prof. Luciano Galantini for the patience they had in working together with me and for always having given their scientific approval and sincere encouragement.

I want to say thanks to Cedric Dicko who friendly welcomed me at the Division for Pure and Applied Biochemistry in the Chemistry Department in Lund: he represented an inspiring mentor and thanks to him I have learned a lot.

I want to warmly acknowledge the team at the University of Bologna with whom I had the opportunity to collaborate: Prof. Simona Fermani, Prof. Francesca Sparla, Prof. Paolo Trost and Libero Guerrieri. Doing research together has really been a wonderful experience from both the scientific and the personal point of view for me.

I would like to thank Ornella Di Stefano for the help with the spectroscopic experiments of sections 3.8.1 and 3.9.

I am grateful to the beamline staff for the technical support and tutoring at the large scale facilities in Paris (Javier Pérez and Aurélien Thureau), Grenoble (Adam Round and Martha Brennich) and in Lund (Sebastian Lages, Christopher Söderberg and Thomás Plivelic). Special thanks go to Sebastian and Thomás for sharing with me very interesting discussions.

I want to say thanks to Prof. Karin Schillén and Solmaz Bayati for their support during beamtime in Lund and thesis writing in Rome.

I would like to acknowledge Prof. P.O. Larsson for his highly experienced advice about SEC.

I would also like to acknowledge Prof. Ulf Olsson and Prof. Leandro Barbosa for having accepted to be the referees of the thesis. Their comments were highly gratifying and contributed to give me the motivation to go on.

I am thankful for the economic support received by Sapienza University of Rome and for the opportunity given by the Unipharma Graduates Erasmus + project.

Contents

Overview	1
1 Main techniques of investigation	3
1.1 Small angle X-ray scattering	3
1.1.1 Theory	3
1.1.2 Information achievable by SAXS	8
1.1.3 SAXS of proteins	14
1.2 Component analysis	22
1.3 2D correlation spectroscopy	24
2 Albumin	27
2.1 Overview of albumin functions: the “tramp steamer”	27
2.2 Albumin structure	28
2.3 Ligand binding	30
2.4 Some physical-chemical properties	32
2.4.1 Ionic properties	32
2.4.2 Spectroscopic properties	32
2.5 Conformational variations	35
2.5.1 pH-induced rearrangements	35
2.5.2 Denaturation	36
2.5.3 Allostery hypotheses	37
2.6 Albumin as an antioxidant	38
3 Hypochlorite-induced oxidation of Human Serum Albumin	41
3.1 Introduction	41
3.2 Experimental procedures	43
3.2.1 Sample preparation	43
3.2.2 SUrF platform experiments	43
3.2.3 Circular dichroism	44
3.2.4 Zeta potential	45
3.2.5 Complementary UV-visible and fluorescence measurements	45
3.2.6 SEC-SUrF at the SWING beamline	46
3.2.7 Data analysis	46
3.3 Combined spectroscopic and SAXS characterization of oxidized HSA	49
3.4 Time evolution of the oxidized protein spectra	54
3.5 SEC-SAXS experiments: structural models	59

3.5.1	Oxidation of HSA oligomers	63
3.6	Zeta potential and circular dichroism results	63
3.7	Application of component analysis to the observed oxidation-dependent structural transition	66
3.8	Investigation of the effect of ligands	70
3.8.1	Spectroscopic characterization at low oxidant doses	73
3.8.2	Absence of effects on the structural transition	73
3.9	About the reversibility of the structural transition	77
3.10	A note about mixing conditions	81
3.11	Summary and discussion	84
3.12	Conclusions	87
4	Acid unfolding of Human Serum Albumin	89
4.1	Introduction	89
4.2	Experimental procedures	93
4.2.1	Sample preparation	93
4.2.2	Time-resolved SURF experiments	93
4.2.3	Complementary measurements	94
4.2.4	Data analysis	95
4.3	Monitoring structural and spectral features of HSA as a function of pH	97
4.4	Looking for structural components in the acid unfolding	104
4.5	Summary and discussion	109
5	SAXS analysis of Calvin cycle enzymes and their regulatory complexes	113
5.1	Introduction	113
5.2	Experimental procedures	115
5.2.1	Sample preparation	115
5.2.2	SAXS data acquisition and analysis	118
5.2.3	DLS measurements	122
5.2.4	Hydrodynamic calculations	122
5.3	PRK, A ₄ -GAPDH/CP12 and their ternary complex	123
5.3.1	Oxidized and reduced PRK	123
5.3.2	The binary complex A ₄ -GAPDH/CP12	125
5.3.3	The ternary complex A ₄ -GAPDH/CP12/PRK	128
5.4	Oligomerization of AB-GAPDH	135
5.5	Discussion and conclusions	149
	Concluding comments	151
	List of abbreviations	153
	Bibliography	154

Overview

The study of molecular structure under varying conditions can help elucidating the underlying mechanism of complex processes and linking structure to function. In the case of biological macromolecules the interesting length-scales of phenomena like conformational changes or formation of supramolecular complexes can be monitored with the technique of small angle X-ray scattering (SAXS, section 1.1), which offers the advantage of probing the structure directly in solution conditions close to the natural functioning state or modified in a controlled way to discover structural responses upon “stress”-inducing environments [1,2]. In addition, the power to resolve the mechanism of convoluted effects can be enhanced by the combination of techniques that can report on both the structure and chemistry of the samples [3]. As an example investigation methods based on SAXS analysis aided by optical spectroscopy can give a useful structural insight into complex processes involving proteins under changing solution environment or application of external perturbations [4].

This idea was applied to the case of Human Serum Albumin, the most abundant protein of body fluids which has a “modular” three-domain structure potentially responding to stress by means of changes of conformation (Chapter 2).

Besides acting as the main carrier protein of the circulation, HSA is also an obvious target of extracellular reactive oxidant species due to its high abundance in plasma. It is for this reason considered the main anti-oxidant defense in blood. It is foreseeable that repeated exposure of HSA to oxidative environments could significantly affect its biological activity and a correlation between the structure and the oxidation state of HSA is expected to be crucial for rationalizing the effect of oxidative modifications on the functional ability of this protein. In the first part of this thesis (Chapter 3) a study addressed at probing the structural changes of albumin induced by the chemical damage caused by the oxidant hypochlorite is presented. The pivotal experiments of this study benefited from the use of a multi-technique instrumental platform which combined the simultaneous collection of SAXS, UV-vis absorbance spectra and fluorescence emission on the same sample volume and it is in the following referred to as SURF (SAXS, UV-vis and Fluorescence).

HSA is also known to undergo conformational variations as a function of pH and in particular the rearrangements observed when going from neutral to acid pH are defined as the N-F transition (pH 4.3) and the acid expansion (below pH 3.5) [5]. In Chapter 4 the results of a SAXS investigation combined with the monitoring of the protein intrinsic fluorescence are described. We managed to implement an experimental setup in which the process of acid unfolding could be followed *in-situ*

as a function of a continuous decrease of pH in time.

The final Chapter of the thesis (Chapter 5) deals with the outcome of a collaboration with a structural biology - molecular plant physiology group of the University of Bologna. The mechanisms which regulate the sugar metabolism in photosynthetic organisms under fluctuating illumination conditions are a fundamental object of study in plant physiology. One of these mechanisms involves the formation of supramolecular multi-enzyme complexes, which accumulate in the dark when the synthetic pathways have to slow down [6]. The potential of the SAXS structural investigation applicable directly in solution conditions was exploited to the study of the assembly of regulatory complexes involving the photosynthetic enzyme of the Calvin cycle glyceraldehyde phosphate dehydrogenase (GAPDH) in both isoforms A₄ and AB. In both cases the SAXS experiments gave for the first time a structural insight into the shape of the constructs in solution and suggested the dynamic nature of the multi-protein complexes.

Chapter 1

Main techniques of investigation

In this chapter the theoretical bases of small angle X-ray scattering (SAXS), the main technique of investigation employed in the studies of this thesis, are briefly presented. The treatment is not aimed to be fully comprehensive but is rather intended to clarify the data analysis approaches used, with also reference to specific software commonly employed. The experimental details of the measurements performed with both SAXS and other techniques will be described later on in dedicated sections of each chapter dealing with one particular topic. Then, some background about mathematical treatments applicable in the evaluation of series of spectra or scattering profiles collected as a function of a varying parameter (e.g. chemical modification, pH, protein concentration) is briefly outlined.

1.1 Small angle X-ray scattering

Small angle X-ray scattering probes the size and shape of electron density dishomogeneities ranging in the nanometer to micrometer scale. The theoretical basis enabling the extraction of structural information from a small angle scattering experiment were initially developed by Guinier in the 1930s [7]. The advances in instrumentation capability and computation efficiency lead to a further spreading of this method of investigation in the subsequent years which become one of the main tools in the field of soft matter science [8]. In the last years the application of the technique to the study of biological macromolecules in solution [9] underwent an enormous development mainly thanks to the accessibility of synchrotron radiation (SR) facilities with experimental setups and data analysis pipelines optimized for high-throughput characterizations [1,2].

1.1.1 Theory

Scattering from a single electron The basic physical principle of an X-ray scattering event can be thought in terms of the interaction between the electromagnetic wave carried by an X-ray photon with the electrons of the atomic shells. This phenomenon can be described in the framework of the classical electromagnetic theory. When an electromagnetic wave is incident on the electrons in the atom, the oscillating electric and magnetic components of the wave exert a Lorenz force on

the charged particles, accelerating them into an oscillatory motion. The accelerated electrons behave in turn as oscillating dipoles which become source of an oscillating electromagnetic field propagating in all directions, the *scattered field*. The solution of Maxwell's equations for the electromagnetic field from an oscillating electric dipole at large distances from the dipole (far-field-spherical wave limit) provides the following expression for the intensity of the electromagnetic wave scattered from a single electron, in the case the incident wave is not polarized:

$$I_e(\theta) = I_p r_e^2 \frac{1}{a^2} \frac{1 + \cos^2 2\theta}{2} \quad (1.1)$$

in which

$$r_e = \frac{1}{4\pi\epsilon_0} \frac{e^2}{m_e c^2} = 2.817940325(28) \cdot 10^{-13} \text{cm} \quad (1.2)$$

is the Thomson scattering length of the electron often called also the *classical electron radius* and represents the entity of its intrinsic power to scatter radiation; in addition I_p is the intensity of the incident radiation, a is the distance of the point of observation from the electron and 2θ (which appears in the so called polarization factor) is the angle formed between the direction of propagation of the incident electromagnetic wave and the direction of observation of the scattered wave.

We note that in the case of the X-ray scattering at small angles, the angular dependence of the intensity due to the polarization factor is negligible, being practically 1 for the angles typically probed. This expression describes the so called *Thomson scattering* and is valid under the assumption of a “free” electron so that the electric field it experiences coincides with that of the incident radiation, without perturbation from the other electrons of the sample. This approximation is valid when the photon energy is far higher than the binding energies of the electrons in atoms and is so fulfilled for the scattering of hard X-rays (with $0.5 < \lambda < 2 \text{ \AA}$). In addition the Thomson expression neglects the inelastic scattering due to the Compton effect, which is negligibly weak at very small angles.

Scattering from a distribution of electrons The intensity of X-ray radiation scattered by a distribution of diffusing elements, i.e. of electrons, if the probability of multiple-scattering events is low and so negligible, is the result of the interference between the waves scattered by the single elements.

Each of the secondary waves resulting from a single scattering event can be represented as:

$$E_s = E_{0s}^{e-} \exp(i\frac{2\pi}{\lambda}x) \exp(i\frac{2\pi}{\lambda}t) \quad (1.3)$$

The wave amplitude is $E_{0s}^{e-} \exp(i\frac{2\pi}{\lambda}x)$ in which $E_{0s}^{e-} = \sqrt{I_{Th}}$. I_{Th} is $r_e^2 \approx 7.94 \cdot 10^{-26} \text{cm}^2$ for unit of incident intensity when the distance from the detector is 1 cm; this scaling factor is usually omitted when describing the angular dependence of the scattered intensity and only introduced when calculating it in absolute units (section 1.1.2).

The factor $\exp(i\frac{2\pi}{\lambda}x)$ expresses the wave phase difference relatively to a reference for which the wave path is $x = 0$. Since all contributions have the same wavelength, the amplitude of the total scattered electric field can be computed as the sum over

the amplitudes scattered by all the electrons. They differ for the phase term which depends on the position occupied by the diffusing electron relatively to a reference origin. The derivation of an expression of the phase factor as a function of the geometrical positions of the diffusing centers is depicted in Fig. 1.1. The direction of the incident beam is indicated by the unit vector $\mathbf{s}_0$ whereas that of the scattered beam by $\mathbf{s}$; they form an angle 2θ .

The difference between the path traveled by the wave scattered from an electron in the origin O and another located in point P, whose position is represented by vector $\mathbf{R}$ results from the geometrical construction as:

$$\Delta x = |\mathbf{s} \cdot \mathbf{R} - \mathbf{s}_0 \cdot \mathbf{R}| \quad (1.4)$$

and the corresponding phase difference:

$$\Delta\varphi = 2\pi \frac{|\mathbf{s} \cdot \mathbf{r} - \mathbf{s}_0 \cdot \mathbf{r}|}{\lambda} = 2\pi \frac{(\mathbf{s} - \mathbf{s}_0) \cdot \mathbf{r}}{\lambda} = \mathbf{q} \cdot \mathbf{r} \quad (1.5)$$

where the *scattering vector*

$$\mathbf{q} = \frac{2\pi}{\lambda} (\mathbf{s} - \mathbf{s}_0) \quad (1.6)$$

was defined, having absolute value:

$$q = |\mathbf{q}| = \frac{4\pi \sin \theta}{\lambda} \quad (1.7)$$

The equation 1.5 therefore shows that every j -th scatterer located in $\mathbf{r}_j$ will contribute to the total scattering amplitude with a $\exp i(\mathbf{q} \cdot \mathbf{r}_j)$ term of the sum. We can consider an electron density distribution $\rho(\mathbf{r})$ to describe the positions of the scatterers in space, which can be defined as the number of electrons in a unit volume, so that a volume dV located at position $\mathbf{r}$ will contain $\rho(\mathbf{r})dV$ electrons. The total amplitude scattered from the distribution of electrons can be calculated (neglecting the constant factor E_{0s}^{e-}) by integration over the volume V in which the scatterers are placed:

$$F(\mathbf{q}) = \int_V \rho(\mathbf{r}) \exp i(\mathbf{q} \cdot \mathbf{r}) dV \quad (1.8)$$

From a mathematical point of view, the equation 1.8 shows that the scattered amplitude $F(\mathbf{q})$ as a function of the vector $\mathbf{q}$ is the *Fourier transform* of the distribution of electron density in real space $\rho(\mathbf{r})$.

What is experimentally measured is the intensity of the scattered radiation $I(\mathbf{q})$ which is the product of the total scattered amplitude and its complex conjugate:

$$I(\mathbf{q}) = I_{Th} F(\mathbf{q}) F^*(\mathbf{q}) = I_{Th} |F(\mathbf{q})|^2 = I_{Th} \int_V \int_V \rho(\mathbf{r}_1) \rho(\mathbf{r}_2) \exp i[\mathbf{q} \cdot (\mathbf{r}_1 - \mathbf{r}_2)] dV_1 dV_2 \quad (1.9)$$

This expression still represents a Fourier integral, containing the relative distance $(\mathbf{r}_1 - \mathbf{r}_2)$ between each pair of scatterers, which we can call $\mathbf{R} = \mathbf{r}_1 - \mathbf{r}_2$. The double integral can be seen as a first integration of $\rho(\mathbf{r}_1) \rho(\mathbf{r}_1 - \mathbf{R})$ over all the pairs with the same relative distance $\mathbf{R}$, and then as a Fourier integral of this quantity over

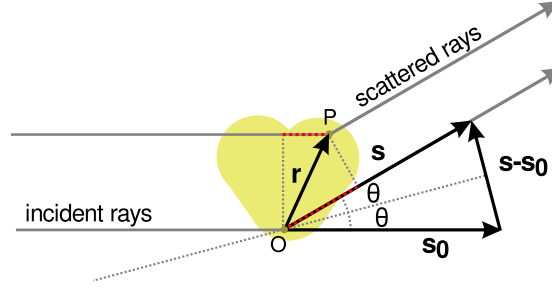


Figure 1.1. Geometry for the determination of the phase difference between two waves scattered from different positions.

all the possible distances. The first integration is the same as the operation of autocorrelation of the electron density function:

$$\tilde{\rho}^2(\mathbf{R}) = \int_V \rho(\mathbf{r}_1) \rho(\mathbf{r}_1 - \mathbf{R}) dV_1 \quad (1.10)$$

Introducing the autocorrelation function of the electron density, the expression for the total scattered intensity becomes:

$$I(\mathbf{q}) = I_{Th} \int_V \tilde{\rho}^2(\mathbf{R}) \exp i(\mathbf{q} \cdot \mathbf{R}) dV \quad (1.11)$$

which shows that the scattered intensity as a function of q , i.e. in reciprocal space, is the Fourier transform of the autocorrelation function of the distribution of scatterers $\tilde{\rho}^2(\mathbf{R})$ and is so uniquely determined by the structure of the system.

Scattering from particles In the case of a system constituted by a particle in solution, the scattering from the electrons of the solvent having a uniform electron density ρ_0 also contributes:

$$F(\mathbf{q}) = \int_{V_s} \rho(\mathbf{r}) \exp i(\mathbf{q} \cdot \mathbf{r}) dV + \rho_0 \int_{V-V_s} \exp i(\mathbf{q} \cdot \mathbf{r}) dV = F_S(\mathbf{q}) + F_M(\mathbf{q}) \quad (1.12)$$

where V is the overall volume of the sample, V_s is the volume occupied by the scattering particle so that the medium has volume $V - V_s$. The total scattered amplitude is therefore the sum of two contributions:

$$F_S(\mathbf{q}) = \int_{V_s} (\rho(\mathbf{r}) - \rho_0) \exp i(\mathbf{q} \cdot \mathbf{r}) dV \quad (1.13)$$

$$F_M(\mathbf{q}) = \rho_0 \int_V \exp i(\mathbf{q} \cdot \mathbf{r}) dV \quad (1.14)$$

The contribution of a homogeneous medium $F_M(\mathbf{q})$ remains constant for all observable values of $\mathbf{q}$.

If the internal density fluctuations within the particle are small with respect to the difference between the mean electron density of the particle and that of the solvent, one can treat the particle as a uniform scattering density object with a mean contrast $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_0$ which would replace the electron density $\rho(\mathbf{r})$ in the

equations above. In the limit of a particle with the same average electron density as the medium, in practical terms a negligible small angle scattering would be observed, as the contributions to the small angle scattering arising from the internal density fluctuations are generally very weak.

In order to experimentally obtain the scattering contribution of the particles as stated above, the intensity scattered by the medium has to be subtracted from the intensity scattered by the particles in the medium. In both cases the scattering contribution of the sample holder also illuminated by the incident beam is included in the intensity experimentally measurable, and cancels out in the subtraction.

Very relevant is the case of isotropic systems in which the angular pattern of the scattered intensity depends only on the modulus of the scattering vector $q = |\mathbf{q}|$ and not on its direction. This is the case for a scattering particle which can freely rotate in solution, for which the measured $|F(\mathbf{q})|^2$ is an average over all possible orientations $\langle |F(\mathbf{q})|^2 \rangle_\Omega$ with respect to the direction of the incident beam. If all orientations are equally probable, using the fundamental Debye formula:

$$\langle \exp i(\mathbf{q} \cdot \mathbf{r}) \rangle_\Omega = \frac{\sin(qr)}{qr} \quad (1.15)$$

and by considering the orientationally-averaged autocorrelation of the electron density contrast:

$$\langle \tilde{\Delta\rho}^2(\mathbf{R}) \rangle_\Omega = \langle \gamma(\mathbf{R}) \rangle_\Omega = \gamma(R) \quad (1.16)$$

the spherically averaged scattered intensity (from eq. 1.11) results:

$$I(q) = 4\pi \int_0^\infty \gamma(R) R^2 \frac{\sin(qR)}{qR} dR \quad (1.17)$$

The extension of what shown so far for a single particle to a solution of N particles is straightforward only in the double assumptions of *ideality* and *monodispersity* of the solution.

The first condition requires a solution dilute enough to neglect the particle-particle interactions. In this case the intensity scattered by each particle ($I_i(q)$) independently contributes to the total $I(q)$ as a term of a sum and the interference effects due to correlation between the relative positions of distinct particles do not have to be taken into account.

In the further assumption of a monodisperse solution comprising identical particles ($I_i = I_j = I_1 \forall i, j$), the total intensity can be obtained as $I(q) = NI_1(q)$.

According to the basic inverse relationship of each scattering technique, if inter-particle correlations are present they would affect the measured scattered intensity at the lowest q values, where the contribution of the biggest inter-electron distances probed by the scattering experiment can be mainly seen, whereas would have little effect at higher q where the scattered intensity is sensitive to local structure. The interactions are usually approximately described by factorizing the total scattered intensity into a contribution depending on the electron distribution of the single particle as described before and a *structure factor* representing the Fourier transform of the radial distribution function of the particles [10], which approaches unity at higher values of q .

In the context of this thesis, the dilute solution conditions allowed for an analysis in which the structure factor could be neglected; in cases in which a certain contribution was apparent at low q we limited ourselves to assessing the slight deviation from ideality but without adopting any further analytical approach to take it into account.

The pair distance distribution function The relationship between the measured SAXS intensity and the particle shape expressed by eq. 1.17 can be stated again after defining the pair distance distribution function $P(R)$:

$$P(R) = R^2 \gamma(r) \quad (1.18)$$

It was introduced by Kratky and Porod in 1948 [11] and, for a particle with homogeneous electron density, defines the frequency with which a pair of scatterers are placed at a distance between R and $R + dR$ inside the particle, representing a sort of histogram of the internal distances. According to its definition, the $P(R)$ becomes zero for distances larger than the maximum dimension of the particle D_{max} . By introducing the $P(R)$ the equation 1.17 becomes:

$$I(q) = 4\pi \int_0^\infty P(R) \frac{\sin(qR)}{qR} dR \quad (1.19)$$

and, since a Fourier transform relationship stands:

$$P(R) = \frac{1}{2\pi^2} \int_0^\infty I(q) q R \sin(qR) dq \quad (1.20)$$

The function $P(R)$ provides a mathematical relationship between the scattered intensity profile, of which it is essentially the inverse Fourier transform (eq. 1.19), and the electron density distribution of the particle, to which is related by means of the autocorrelation (eqs. 1.16, 1.18). An important property of $P(R)$ can be obtained from equation 1.19; for $q = 0$ the term $\frac{\sin(qR)}{qR}$ becomes 1 and it results in:

$$I(0) = 4\pi \int_0^\infty P(R) dR \quad (1.21)$$

showing how the “zero-angle” scattered intensity ($I(0)$) is proportional to the area subtended by the $P(R)$ function.

1.1.2 Information achievable by SAXS

Guinier approximation

The radius of gyration (R_g) of an electron density distribution $\rho(r)$ is the square root of the mean square distance of the electrons from their center of gravity (the second moment of the distribution), in analogy to the definition of the radius of inertia of a mass distribution:

$$R_g^2 = \frac{\int_V \rho(r) r^2 dV}{\int_V \rho(r) dV} \quad (1.22)$$

Guinier showed that the first part of the scattering profiles at small values of q could be approximated as a Gaussian function whose decay was dependent on the R_g of the scattering object:

$$\begin{aligned} I(q)_{0 < q < R_g} &\approx I(0) \exp\left(-\frac{q^2 R_g^2}{3}\right) \\ \ln I(q) &\approx \ln I(0) - \frac{q^2 R_g^2}{3} \end{aligned} \quad (1.23)$$

According to the Guinier approximation, by plotting the natural logarithm of the scattered intensity as a function of q^2 (Guinier plot) a linear behavior should be observed at low q , with slope dependent on only the R_g and intercept extrapolating the zero-angle scattered intensity $I(0)$.

The range of q in which linearity of the Guinier plot is expected is usually considered to be $0 < q < 1.3/R_g$. Since the data points at very low q are generally affected by large experimental errors with 2D detectors or by contribution of interactions, the accessible data range for R_g evaluation is usually poorer for bigger particles. In addition, the presence of a mixture of particles of different sizes or of particle-particle interaction could both determine the impossibility to recognize a reliable linear range of the Guinier plot.

The R_g is also related to the $P(R)$ function through the following relation:

$$R_g^2 = \frac{\int_0^\infty P(R) R^2 dR}{2 \int_0^\infty P(R) dR} \quad (1.24)$$

By means of this expression an alternative estimate of the R_g can be evaluated by integration of the obtained $P(R)$ which is dependent on the whole SAXS profile rather than on only the low- q portion, and the agreement with the Guinier estimate represents a check for the reliability of the measure.

The validity of the Guinier approximation could be justified in the following way. Let's consider the Mc Laurin's expansion of the term $\sin(qr)$ which appears in eq. 1.19:

$$\sin(qR) \approx qR - \frac{1}{3!}(qR)^3 + \frac{1}{5!}(qR)^5 - \dots \quad (1.25)$$

If we truncate the approximation at the second term and replace the obtained expression in the formula of the scattered intensity (eq. 1.19) we obtain:

$$I(q) \approx 4\pi \int_0^\infty P(R) dR - \frac{4\pi q^2}{6} \int_0^\infty P(R) R^2 dR \approx I(0) \left(1 - \frac{q^2 R_g^2}{3}\right) \quad (1.26)$$

in which substitutions according to the identities $I(0) = 4\pi \int_0^\infty P(R) dR$ and $R_g^2 = \frac{\int_0^\infty P(R) R^2 dR}{2 \int_0^\infty P(R) dR}$ have been performed.

The equation 1.26 is in agreement with the series expansion of the exponential approximation (eq. 1.23) up to the second term ($e^{-x} \approx 1 - x$), valid for $q \rightarrow 0$:

$$I(q) \approx I(0) \exp\left(-\frac{q^2 R_g^2}{3}\right) \approx I(0) \left(1 - \frac{q^2 R_g^2}{3}\right) \quad (1.27)$$

In the case of elongated particles it was demonstrated that a similar reasoning lead to find an approximate formula for the $I(q)$ in the “intermediate” q range which

allows to extract the radius of gyration of the cross-section of the elongated object R_c [8]:

$$I(q) \approx \frac{I(0)}{q} \exp\left(-\frac{q^2 R_c^2}{2}\right) \quad (1.28)$$

Indirect Fourier Transform

As deduced from eq. 1.19 the $P(R)$ function of the particle could be obtained as the Fourier transform of the experimental intensity profile. Even when instrumental factors determining “smearing” of the angular profile of the scattered intensity have not to be considered, as in the case of the highly monochromatic and collimated X-ray beams produced at a synchrotron beamline, a direct Fourier transform operation to obtain the $P(R)$ cannot be performed since the $I(q)$ profile is available only in a limited q -range sampled at discrete points rather than in an infinite and continuous interval, and large truncation errors would result otherwise. In addition, the experimental data are affected by noise.

The mathematical approaches to obtain the $P(R)$ therefore are rather based on an Indirect Fourier Transform (IFT), i.e. the regularized optimization of an estimated $P(R)$, so that the $I(q)$ back-calculated from it (eq. 1.19) would fit the experimental data. The method originally introduced by Glatter [12] consists in a constrained weighted least-squares minimization. The $P(R)$ is approximated as a linear combination of N cubic B-spline functions ($\phi_v(R)$) within a range of distances in real space $0 \leq R \leq D$ determined by a chosen maximum D , above which one expects the $P(R)$ to be zero:

$$P(R) = \sum_{v=1}^N c_v \phi_v(R) \quad (1.29)$$

The optimal coefficients c_v are determined by means of the minimization of the weighted least squares discrepancies between the experimental data $I(q)$ and the estimate determined by application of the proper linear transform $\hat{F}$ (Fourier transform and eventually convolution with smearing effects) to the $P(R)$ expressed as in eq. 1.29:

$$L = \frac{\sum_{i=1}^M [I(q_i) - \sum_{v=1}^N c_v \hat{F} \phi_v(R)]^2}{\sigma^2(q_i)} = \frac{\sum_{i=1}^M [I(q_i) - \sum_{v=1}^N c_v \chi_v(q_i)]^2}{\sigma^2(q_i)} = \min \quad (1.30)$$

where M is the number of experimental points, $\sigma^2(q_i)$ is the variance of the i^{th} point and $\chi_v(q)$ are the results of the Fourier transformation of the $P(R)$ real space components $\phi_v(R)$. A regularization condition is necessary in order to solve this ill-posed problem [13]. The function of the regularization condition is to constrain the solution to be artifact-free, i.e. without unphysical oscillations which would also be beyond the low real space resolution of a SAXS experiment, but at the same time allowing for a model flexible enough to reproduce the experimental data. Therefore in order to ensure smoothness of the $P(R)$, the c_v coefficients are constrained according to the equality:

$$N_{c'} = \sum_{v=1}^{N-1} (c_{v+1} - c_v)^2 \quad (1.31)$$

Overall, according to the method of the Lagrange multiplier, the constrained optimization consists in the minimization of:

$$(L - \lambda N_c) = \min \quad (1.32)$$

Multiple solutions according to different values of the multiplier λ , which represents the “weight” of the regularization constraint, are obtained and the multiplier value must be adjusted appropriately. This has been accomplished using a “stability plot” to choose the regularization parameter in a range of stability of the calculated solutions [12]. The GNOM method [14] which is included now in the ATSAS package [15] uses “perceptual criteria” for the estimate of the best regularization multiplier. These criteria include requirements of smoothness, positivity and validity of the chosen real space interval for the calculated $P(R)$ and of absence of systematic deviations in the fit of the scattering profile.

These approaches have the disadvantage that the criteria used are reasonable but not built on fundamental principles. On the contrary, the work of Hansen [16], available as the program BayesApp [17], considered the IFT in the Bayesian statistical inference approach, hence it is based on fundamental principles, which allows one to determine a solution objectively. This method relies on the implementation of equations based on the Bayesian treatment of the probability of a model. The procedure is able to determine a probability distribution of the solutions given by different smoothness parameters, which in practice correspond to the noise level of the data, and different estimates of the maximum D . The most probable $P(R)$ is calculated without any necessity of input by the user, if needed.

Both the GNOM and the Bayesian IFT methods were employed to estimate $P(R)$ functions from the $I(q)$ data in the studies presented in this thesis. BayesApp was particularly useful when analyzing series of tens of scattering profiles, since reliable results were obtained without the necessity to input an *ad hoc* estimate of the maximum limit of the real space range D for each experimental curve. In addition, the BayesApp software includes an estimate of the particle elongation expressed as the axial ratio of the equivalent ellipsoid [18]. The evaluation of this parameter is based on an empirical relation which links the axial ratio of an ellipsoid of revolution to the ratio between the third moment of the $P(R)$ and the cubic power of its first moment:

$$t = \frac{\int_0^{D_{\max}} P(R) R^3 dR}{(\int_0^{D_{\max}} P(R) dR)^3} \quad (1.33)$$

The empirical relation was obtained in BayesApp by evaluating the t parameter for ellipsoids of revolution with given axial ratio, by integration of the corresponding $P(R)$ calculated analytically [19], and by fitting then the trend of t vs. axial ratio. The obtained relation is then used in the software to estimate an axial ratio value from the t parameter calculated with the $P(R)$ provided by the IFT method. The estimate is considered reliable in the range of values between 0.1 and 6.

Information content In order to estimate the information content of a small angle scattering experiment, the Shannon sampling theorem [20] had been applied to the problem by describing it as a procedure to retrieve the $P(R)$ signal in real space between 0 and D , from the $I(q)$ experimental data in reciprocal space between

$q_{\min}$ and $q_{\max}$ [21, 22]. A first important consequence of this treatment is that the experimental data in the q range should be sampled with a spacing at least as small as $\frac{\pi}{D}$ in order to contain information about distances as large as D . Since the spacing between datapoints available experimentally is usually not a limitation in the SAXS experiments ($\Delta q \ll \frac{\pi}{D}$), this prescription has rather an impact on the requirement of a minimum scattering vector value $q_{\min} < \frac{\pi}{D}$ to be included in the analysis in order to retrieve information on a size D , and in the case procedures of rebinning are considered [23]. Indeed the sampling theorem predicts that the number of Shannon channels, interpretable as the minimum number of datapoints spaced $\frac{\pi}{D}$ required to retrieve the information or, on the other hand, the maximum number of independent observable which could be retrieved and so a measure of the information content of the scattering data, is determined by:

$$N_s = \frac{D_{\max}(q_{\max} - q_{\min})}{\pi} \quad (1.34)$$

Estimates based on this expression were used to evaluate the number of splines to be used in the IFT procedures [21]. However, the signal-to-noise ratio is not taken into account in this estimate and furthermore the noise is partially compensated by the fact that data are usually measured with an angular increment Δq much smaller than the distance between the Shannon channels $\frac{\pi}{D}$. The amount of information in the data must be related to both the level of experimental error and the degree of oversampling. These aspects have been recently re-discussed in the small angle scattering community [23–26].

Vestergaard and Hansen suggested a method to account for the signal-to-noise ratio in the estimate of the information content, by using the “number of good parameters” whose evaluation depends on the regularization parameter of the IFT as an index of the level of noise of the analyzed data [24].

Also, an evaluation of the useful data range (especially at high q) of an experimental scattering profile was introduced [26]. The useful number of Shannon channels is empirically evaluated by considering the quality of the Shannon fit to the data and to the $P(R)$ with increasing number of Shannon components, and then the maximum useful $q_{\max}$ calculated back from this estimate (eq. 1.34).

Rambo & Tainer [25] suggested that an estimate of the N_s was also needed to more correctly evaluate the statistical significance of a model fitting the data by means of the reduced χ^2 :

$$\chi_{red}^2 = \frac{1}{\nu} \sum_{i=1}^N \frac{[I_{obs}(q_i) - I_{calc}(q_i)]^2}{\sigma_i^2} \quad (1.35)$$

involving the knowledge of the number of degrees of freedom ν , which is not a trivial matter. While the χ_{red}^2 is usually calculated using all the available data points N and by using $\nu = N - 1$, they pointed out that the correct way should involve a rebinning of the data according to the estimated number of Shannon channels in order to satisfy the requirement of “independent observations”. Furthermore a correct assessment of the number of parameters of the model would be needed in order to properly estimate the number of degrees of freedom.

The importance of rebinning the oversampled data in order to estimate the information content relatively to a chosen model has been pointed out also in

Ref. [23]. In practice it is not easy to ensure that all the critical aspects of a correct evaluation are fulfilled, so that the goodness of fit of a model to small angle scattering data has more commonly a relative meaning when performing comparisons rather than a statistically valid interpretation. An alternative method based on the estimate of the probability of systematic deviations assessed by correlation of a series of data points has been suggested to overcome these requirements and also the concern for a correct estimate of the standard deviation of the data points [27].

Overall the problem of the content of information is still under debate, especially in the view that a standardized way to present the results of SAXS data analysis is needed especially in the field of structural biology where they usually involve three-dimensional structural models whose limitations have to be correctly stated (section 1.1.3) [28].

Porod invariant, volume and approximation

From equations 1.17 and 1.16 it can be deduced that the total amount of scattered intensity at zero angle ($q=0$) depends on the square of the total amount of contrast electrons in the scattering sample:

$$I(0) \propto (\Delta\rho)^2 V^2 \quad (1.36)$$

In addition, Porod showed that the integral of $I(q)q^2$ also only depends (in addition to instrumental factors affecting absolute intensity) on the absolute amount of contrast electron density in the scattering sample and not on the shape in which it is distributed, so that this quantity was defined “invariant” Q [29]:

$$Q = \int_0^\infty I(q)q^2 dq = 2\pi^2(\Delta\rho)^2 V \quad (1.37)$$

This integral quantity proved to be useful to deduce the excluded scattering volume of one particle V without the need for absolute scaling of the intensity, since it can be obtained as:

$$V = \frac{2\pi^2 I(0)}{Q} \quad (1.38)$$

However a reliable integration in the full range of q to calculate Q is not possible without extrapolation of the experimental data according to limiting laws. Extrapolations based on the Guinier approximation at $q \rightarrow 0$ (eq. 1.23) and on the Porod’s law are possible. This second approximation was demonstrated to hold in the case of systems with a defined boundary delimiting the homogeneous particle volume from the medium volume and describes the asymptote of the small-angle scattered intensity $I(q)$ for large q values:

$$I(q)_{q \rightarrow \infty} = \frac{K_P}{q^4} \quad (1.39)$$

An implementation of this approximation optimized to estimate the Porod excluded volume of biological macromolecules in solution from the regularized scattering profile is present in the ATSAS package and involves the subtraction of a constant to enforce the q^{-4} decay at large q , estimated in the $q < 7.5/R_g$ range to avoid accounting for the internal particle structure [30].

Absolute intensity and molecular mass

The value of the intensity scattered by a solution of particles in absolute units (cm^{-1}), defined as the *macroscopic differential scattering cross section* $\frac{d\Sigma}{d\Omega}$ is the quantity which can be directly related to the number of electrons in the particle, with the purpose of estimating its molecular weight.

In order to obtain the absolute intensity $\frac{d\Sigma(q)}{d\Omega}$ from the experimentally available intensity, calibration procedures are needed. They are based on the measurement of a reference sample for which the $\frac{d\Sigma}{d\Omega}$ is known. The flat scattering of pure water in the same conditions as the samples is usually used in the case a sufficient photon flux is available, since its $\frac{d\Sigma}{d\Omega}$ at a specific temperature can be calculated knowing the water isothermal compressibility ($\chi_T [\text{Pa}^{-1}]$) and scattering length density ($\rho [\text{cm}^{-2}] = \frac{r_e [\text{cm}/\text{electron}] (z_{H_2O} N_A) [\text{electrons}/\text{mol}] d_T [\text{g}/\text{cm}^3]}{\text{MW}_{H_2O} [\text{g}/\text{mol}]}$) [31]:

$$\frac{d\Sigma}{d\Omega} = \rho^2 K_B T \chi_T \quad (1.40)$$

The following relation expresses the intensity in absolute units in terms of the characteristics of the sample of particles in solution:

$$\frac{d\Sigma(0)}{d\Omega} [\text{cm}^{-1}] = I(0) = \frac{c [\text{g}/\text{cm}^3] \Delta\rho_{\text{M,part}}^2 [\text{cm}^2/\text{g}^2] \text{MW} [\text{g}/\text{mol}]}{N_A [\text{mol}^{-1}]} \quad (1.41)$$

where c is the mass concentration of the sample, MW is the mass of the particle, N_A is the Avogadro number and $\Delta\rho_{\text{M,part}}$ is the scattering contrast per mass. Therefore the MW can be estimated if the $I(0)$ in absolute units (cm^{-1}) and the scattering contrast per mass are known. $\Delta\rho_{\text{M,part}}$ (in cm/g) can be calculated as:

$$\Delta\rho_{\text{M,part}} = (\rho_{\text{M,part}} [\text{electrons}/\text{g}] - \tilde{\nu} [\text{g}/\text{cm}^3] \rho_s [\text{electrons}/\text{cm}^3]) r_e [\text{cm}/\text{electron}] \quad (1.42)$$

in which $\rho_{\text{M,part}}$ is the number of electrons per mass of the particle and can be estimated on the basis of the chemical composition, $\tilde{\nu}$ is the partial specific volume of the particle and can be obtained from measures of density, ρ_s is the electron density of the solvent and r_e is the Thomson scattering length.

The electron density for pure water at room temperature is about $0.334 \text{ electrons}/\text{\AA}^3$ and this value is commonly used to estimate the contrast also when the medium is an aqueous solvent with dilute salts, as in the case of biological molecules. For ρ_{M} , in the case of proteins the value $3.22 \cdot 10^{23} \text{ electrons}/\text{g}$ is used [32]. As an example we can estimate the expected $\Delta\rho_{\text{M}}^2$ for a globular protein and $\frac{d\Sigma(0)}{d\Omega}$ for a $1 \text{ mg}/\text{cm}^3 = 0.001 \text{ g}/\text{cm}^3$ solution of protein having mass 66.5 kDa (Human Serum Albumin): assuming a partial specific volume $\tilde{\nu} = 0.735 \text{ g}/\text{cm}^3$, a scattering contrast per mass $\Delta\rho_{\text{M}}^2 = 4.648 \cdot 10^{20} \text{ cm}^2/\text{g}^2$ is calculated and a zero-angle scattered intensity of 0.0513 cm^{-1} is obtained.

1.1.3 SAXS of proteins

General issues Generally speaking, all the biomolecules to be analyzed in solution by SAXS have similar requirements in terms of sample quality, experimental setup

and data-analysis methods needed, so that the enormous expansion of the “BioSAXS” field in the context of structural biology has led to develop dedicated SAXS beamlines at synchrotron facilities and specific data analysis tools for this kind of experiments [1, 9, 33–39].

One common characteristic of samples of proteins in solution, differently from other soft matter systems usually investigated by SAXS, is that they give a relatively low signal compared with the solvent background that has to be subtracted. This originates from the fact that experiments in dilute conditions are usually performed and the average electron density of the particles is only slightly higher compared to that of the aqueous solvent, giving a relatively weak contrast ($\rho_{\text{protein}} \approx 0.42$ electrons/ $\text{\AA}^3$ vs. $\rho_{H_2O} \approx 0.334$ electrons/ $\text{\AA}^3$).

From this aspect the criticality of the buffer chosen to collect the signal for background subtraction results. In order to have usable data the solution measured as background has to match as much as possible the real solvent in which the macromolecules are dispersed in the sample. Due to low concentration, the volume fraction of the particles is usually neglected in performing the subtraction. In addition, in several fitting procedures the possibility of an adjustable constant background is often allowed to account for small errors; the subtraction is more critical in the higher q range where the signal due to the protein becomes of the same order as the background signal and sometimes negative points could arise in the subtracted data.

Another issue in performing SAXS on protein samples is the damage induced by X-rays that arises after a critical radiation dose. During X-ray exposure radicals are generated from the water photolysis whose visible consequence in the SAXS spectrum is protein aggregation [40]. In the case of high photon flux the experimental setups therefore involve either continuous sample flowing across the X-ray illuminated portion or oscillation of a certain sample volume in order to avoid exposing the same volume of protein solution for too long, while acquiring enough signal [41, 42].

The derivation of sound information about the protein shape relies on the assumptions of ideal and monodisperse solution. In order to check the first assumption SAXS measurements at different protein concentrations in given conditions are usually needed. A dilute solution of pure protein should in principle satisfy the definition of monodisperse solution due to the intrinsic uniformity of molecular weight but in some conditions the main population of macromolecules could be accompanied by larger aggregates due to reversible or irreversible protein-protein attractive interactions. Therefore the SAXS measurements were combined with online size-exclusion purification as a standard set-up nowadays widespread at synchrotron facilities (SEC-SAXS). A fine SAXS sampling of the elution peak(s) can then be performed providing an independent warranty that the resulting experimental data arise from a freshly produced monodisperse sample. In general, a structural investigation of more complex systems involving multiple oligomerization states can also be attempted with SEC-SAXS experiments.

The Kratky plot An additional way of presenting SAXS data which has proved to be of semiquantitative usefulness for partially folded or denatured states of proteins or other macromolecules is a measure of compactness obtained from examination of

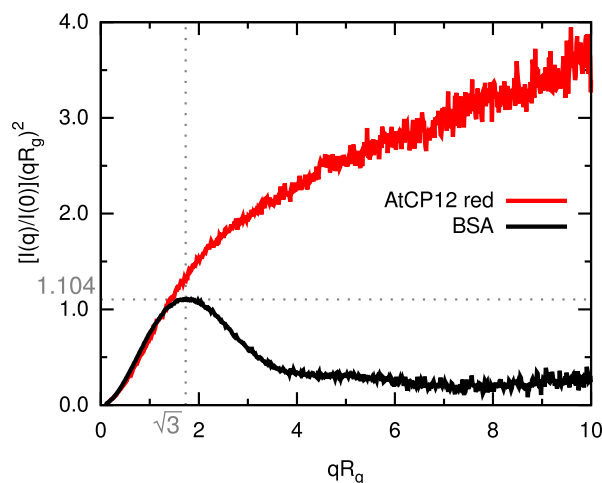


Figure 1.2. *Dimensionless Kratky plots which exemplify the behaviors of a globular folded protein and an unfolded protein.* Dimensionless Kratky plots calculated from the experimental data of the intrinsically disordered protein AtCP12 with reduced disulfide bridges (red line, $R_g = 29 \text{ \AA}$ and $MW = 9 \text{ kDa}$; protein involved in the study of Chapter 5, section 5.3) and of Bovine Serum Albumin (BSA) at pH 7.5 (black line, $R_g = 30 \text{ \AA}$ and $MW = 66 \text{ kDa}$; standard control for BioSAXS) which exemplifies the behavior of a typical globular folded protein.

the Kratky plot [43,44]. This representation of the scattered intensity $I(q)$ multiplied by the squared scattering vector q^2 as a function of q ($I(q)q^2$ vs. q) was introduced as a tool to probe the conformational state of polymer molecules [8]. As shown by Kratky, the scattering profile for a random coil state of a polymer asymptotically falls as $1/q$ at large q . Hence, in the Kratky plot the curve rises linearly at large angles and so may be very clearly distinguished from that of a compact protein where, following the work of Porod, the clearly defined surface leads to a scattering profile which falls as $1/q^4$. The result is that native states or partially folded states with a relatively small amount of random coil disorder show a pronounced peak in the Kratky plot while fully denatured states of proteins, which are not constrained by disulfide bonds, show an upward rising feature in the Kratky plot (Fig. 1.2).

Flexible multi-domain proteins can also potentially be identified from the Kratky plot, displaying a mixture of characteristic features of both folded and unfolded proteins similar to that observed for a partially unfolded state. However simulated data demonstrated the potential ambiguity in the assessment of the level of flexibility in the case of highly flexible modular proteins; an ensemble average of different conformations made by folded domains interconnected by flexible linkers can lead indeed to Kratky plots in which the absence of a clear random coil signature would have suggested relatively rigid compact globular structures [45].

In order to compare the folding state of macromolecules with different masses and dimensions, the normalized or dimensionless Kratky plot was introduced [46,47]. If the zero-angle scattered intensity $I(0)$ and the radius of gyration R_g are estimated, by plotting $[I(q)/I(0)] \times (qR_g)^2$ as a function of qR_g , for a globular particle following the Guinier approximation in an extended q range, a peak at $qR_g = \sqrt{3}$ and

ordinate=1.104 is predicted. A shift of the maximum to higher qR_g values is considered an index of elongated conformations.

Estimates of molecular mass In order to estimate the protein molecular weight (MW) from the absolute intensity scattered at zero angle (eq. 1.41), the mass concentration of the sample has to be known with reasonable accuracy, at least within 5-10% error in order to assess the protein oligomerization state [32]. This requirement could be sometimes difficult to achieve in practice; for example this can be the case if the spectrophotometric determination of the protein content is hampered by the presence of ligands or cosolvents with UV absorbance which interferes with the protein characteristic peak, or if a SEC-SAXS experiment is performed which does not involve the monitoring of the protein concentration directly at the X-ray exposure site. Both these conditions were indeed found in the experiments presented in Chapter 5, section 5.4. In these contexts, and also as an additional general check even when the protein concentration is known, methods of estimating a molecular weight from the SAXS data independently from the absolute scale of the intensity can be useful.

They rely on empirical correlations between the MW and scale-independent metrics estimated from the scattering profile. These correlations were found by statistically examining a large number of proteins either with known mass and for which experimental SAXS data were collected or with known structure for which simulated scattering profiles were obtained [25,30,48].

Fischer & Craievich demonstrated a method based on a linear relationship between the dry protein volume and an “apparent” Porod volume (eq.1.38) obtained by calculation of the Porod invariant (eq.1.37) on a truncated q range which is experimentally available. The empirical relation includes the effects of truncation at different $q_{\max}$ and assumes a general estimate of the protein density of $\rho = 1.37 \text{ g/cm}^3$ [48] to convert the dry volume into protein mass. Another empirical correlation with the molecular mass of proteins was then shown with a different concentration- and conformation-independent quantity whose calculation requires, in addition to the R_g and $I(0)$, the estimate of the $\int_0^{q_{\max}} I(q)q dq$ integral, which unlike the Porod invariant converges also for flexible systems with a divergent Kratky plot at large q [25]. In this case it was demonstrated to hold in a broader range of protein masses.

In addition, empirical relations based on the statistical evaluation of simulated SAXS profiles of several proteins with known structure were suggested to make estimates based on the already calculated Porod volume V_P (page 13) as $MW \approx 0.625V_P$ or based on the the average volume V_d of *ab-initio* dummy-atom reconstructions (page 18) as $MW \approx 0.5V_d$ [30]. These two approximate estimates are not supposed to work well outside the requirement of globular folded proteins.

3D-modeling approaches

The very big limitation of small angle scattering in solution as a structural technique is the fact that the monodimensional $I(q)$ scattering signal corresponds to an orientational average of the object (eq. 1.17), leading to a consistent loss of information compared to the original three-dimensional shape. The retrieval of a 3D model of the scatterer from the intensity profile is so an intrinsically ambiguous

problem with potentially multiple solutions. Even though, the addition of constraints deriving from independent knowledge about the sample can help in guiding towards a defined meaningful solution. In the case of proteins the fact that the requirement for monodispersity can be fulfilled, the fact that the average electron density contrast is known and can be considered constant inside the particle with good approximation at low resolution, the knowledge of the primary sequence and also the possibility that partial structural information from other techniques or even atomic-resolution models are available, allow for the implementation of methods which can often provide credible low-resolution models fitting the data, whose reliability has to be however carefully evaluated [35, 49]. The most used implementations in this thesis were those included in the ATSAS package [30].

***Ab-initio* shape reconstruction** *Ab-initio* approaches are aimed to find a 3D shape whose spherically averaged computed scattered intensity fits the experimental scattering profile of a dilute monodisperse solution of the particle at low resolution, i.e. accounting only for the overall particle shape and not for internal structure [35]. In practice many of these programs use additional information external to the scattering profiles to constrain the search for a model and the term *ab-initio* solely refers to lack of a pre-defined input structure [34]. The assumption of a homogeneous electron density contrast and also the requirements for continuity of the shape, with additional imposition of symmetry if known *a priori* for example from the oligomeric state of the protein, are usually employed to constrain the search for a model. The search can be performed with different approaches including multipole expansion in spherical harmonics to define envelopes [50, 51], and beads models [52, 53] optimized with genetic algorithm [54] or simulated annealing approaches [55].

The DAMMIN/DAMMIF method is based on the representation of the 3D shape as an assembly of densely packed spheres (dummy-atoms) disposed on a starting search grid, which is modified in order to optimize a target function including both the goodness of fit to the given scattering profile and the additional constraints of connectivity and compactness given by means of penalties for loose arrangements [55, 56]. The theoretical scattered intensity is computed by means of the spherical harmonics expansion and the search is performed by an iterative Monte-Carlo approach with gradual decrease of the threshold for acceptance in a simulated annealing procedure. The size and the number of the beads and the shape of the search grid are defined according to the maximum size of the particle to be modeled, resulted from an IFT previously performed with GNOM. The regularized scattering profile obtained by GNOM at a reduced number of knots is the target for computing the goodness of fit. An adjustable flat background is also included in the fit.

The *ab-initio* approach implemented in GASBOR is thought for better reproducing the small angle scattering of a protein with an approximation which can be considered reliable up to a higher q compared to the dummy-atom approach [57]. The particle is represented by dummy-residues of 3.8 Å diameter that are constrained to adopt a spatial arrangement with connectivity requirements which reproduce the C α chain of proteins. They are also constrained to be in a fixed number defined by the number of amino acids expected for the protein to be modeled. The average form factor of an amino acid in water is used to represent the scattered amplitude of

each dummy-residue and in addition dummy-water beads are placed on the surface to reproduce the hydration layer. The overall scattering is computed by the Debye formula ($\sum_{i=1}^N \sum_{j=1}^N f_i(q) f_j(q) \frac{\sin qR_{ij}}{qR_{ij}}$, where $f_i(q)$ are the scattering factors from individual dummy-residues or dummy-water molecules) and the optimization of the 3D representation is therefore performed with simulated annealing [57].

Calculation of theoretical intensities from high-resolution structures The approach to compute a theoretical scattering from a hypothetical model in order to compare it to the experimental scattering profile can be considered always informative and allows for discarding structural hypothesis [1]. This is possible thanks to methods to evaluate the small angle scattering profile from a defined arrangement of atoms (e.g. a protein structure from crystallography or NMR, or a simulated high resolution model). They are founded on the basic equation of scattering (eq. 1.9) for computing the amplitude scattered by the target molecule with its ordered solvation layer *in vacuo*, but have also to include subtraction of the excluded volume contribution that takes into account the missing scattering of bulk solvent due to the existence of the solute. The excluded volume term can be determined by defining the shape of the molecule and calculating the scattering from it as if it were filled by an electron density equivalent to bulk solvent [34]. The methods implemented in the several programs which have been released throughout years basically differ in the way the orientational average is performed and in which the hydration layer of the macromolecule, which gives a non-negligible contribution to the scattering [58], and the excluded volume are described [59].

In the case of the software CRY SOL [60] the spherically averaged scattered intensity is computed using the spherical harmonics formalism of the multipole expansion which speeds up the computing time, but sacrificing the possibility to accurately describe the scattering from the particle internal structure ($q > 0.5 \text{ \AA}^{-1}$) [59]. The number of spherical harmonics to be used (L) should be higher for bigger particles and it has been advised to estimate it as $L = 5 + \frac{q_{max} D_{max}}{2}$ [61]. In CRY SOL the protein hydration is taken into account by surrounding the particle with a continuous envelope representing the solvation shell; it has a thickness of 3 Å and a default density about 10% more than the bulk water, which can be slightly optimized when fitting experimental data. An alternative version in which an atomic-level model of water molecules surrounding the protein is generated has been implemented and apparently should fix the issues in representing the hydration of cavities. The excluded volume treatment in CRY SOL is based on the representation of each atomic group as a dummy-atom with a given average atomic excluded volume which in fitting mode could be adjusted around the default value given by the average over the high resolution structure. On request, an additional constant background can be included when fitting experimental data to account for incorrect blank subtraction. Overall, a maximum of three parameters (background, average atomic excluded volume, contrast of the hydration shell) are adjusted when fitting an experimental profile with the scattered intensity of a high resolution structure.

The advantage of this method is mainly in the way it can transform the partial amplitudes calculated for single subunits upon application of rotations, so that the computing time for performing a rigid-body search aimed at modeling an assembly of

multiple high-resolution domains can be reasonable [62,63]. In the software SASREF a simulated annealing algorithm is employed to optimize the relative positions of multiple subunits which form together the multi-protein assembly that has to be modeled. The search is performed by optimizing a target function involving both the agreement between theoretical and experimental scattering profile and penalties aimed at ensuring absence of clashes, interconnectivity of the model and fulfillment of user-defined contact conditions if known *a priori*. In addition the imposition of symmetry constraints allows for speeding up both the computing time of the scattered intensity and the search for subunits' positions, and reduces the degrees of freedom. The disadvantage is that the hydration layer is implicitly included for each of the subunits and is not recalculated for the full construct. This represents a limitation especially for models in which the subunits are in close interaction.

A situation often encountered is the possibility that the known high resolution structure of the full-length protein construct presents unknown and/or flexible portions which cannot be modeled as rigid atomic coordinates. In these cases the application of hybrid methods is possible in which the optimization involves both the rigid-body modeling based on the scattered amplitude of known high resolution domains and the *ab-initio* representation of the missing portions which could have the role of flexible linkers.

This is the case of the application CORAL, in which the method implemented in SASREF and briefly outlined above is added with the possibility to reproduce the missing parts by dummy-residues whose possible conformations are taken from a library of linkers [30].

In the BUNCH code, instead, a single peptide chain can be modeled using high-resolution domains interconnected by portions with unknown structure. They are represented as dummy-residue chains constrained by penalties to avoid clashes with the rigid moieties, unphysical extension above a critical R_g value and to ensure a proper distribution of bond and dihedral angles compatible with a $C\alpha$ chain [63].

For all these optimization algorithms based on simulated annealing, it is very important to assess the variability of the best-fit models obtained by multiple runs. A way to compare the 3D models is given by the pairwise superimposition algorithm SUPCOMB [64], which works both with atomic models and beads distributions. The orientation and position of a 3D object is optimized to match the reference structure by minimizing a normalized spatial discrepancy parameter (NSD), which is a measure of the degree of overlap between the two spatial distributions stable against outlying points and normalized for being independent of the overall object's size. The search is performed by starting from an initial orientation in which the inertia axes of the two objects have been matched.

Procedures for the automatic comparison among a group of structures are implemented in the applications DAMAVER [65] and DAMCLUST [30]. After pairwise superimposition among all the structures, the one showing the lowest average $\langle \text{NSD} \rangle$ is selected as reference for superimposing all others, while the models whose $\langle \text{NSD} \rangle$ exceeds 2 standard deviations from the mean are considered outliers and discarded. In DAMCLUST the possibility to ameliorate the overall superimposition by clustering the structures in different families whose internal similarity is distinctly

checked is evaluated.

When a number of spatial distributions are superimposed an average grid of all possibly occupied positions among the models is created and the positions are placed in decreasing order of occupancy probability. From that ranking a filtered envelope made of the subset of positions with highest occupancy and volume equal to the average volume of one single model is calculated.

Ensemble approaches There are cases in which the requirement of monodispersity is not strictly fulfilled but still *a priori* information can help for a description of the system in terms of models. If a number k of different structural models are thought to be present in solution and contributing to the total scattering, the overall $I(q)$ of the mixture can be approximated, assuming absence of interactions, as:

$$I(q) = \sum_{i=1}^k I_i(q) \nu_i \quad (1.43)$$

in which I_i is the intensity scattered by the i^{th} component if it was the only one present, and ν_i is the volume fraction of the i^{th} component in the mixture. Assuming a certain fixed model for the I_i profiles, the coefficients ν_i can be optimized in order to fit the experimental scattering curve.

This kind of hard-modeling is implemented in the OLIGOMER tool of the ATSAS package [66] or in the GENFIT suite from Spinozzi and colleagues [67]. This approach is applicable for example if the study of equilibrium systems is attempted, in which a small number of defined oligomeric species can coexist.

If the protein is flexible in solution, as indicated by its scattering profile (paragraph 1.1.3), a reasonable description of the intensity scattered is an ensemble average of the diffusion profiles originating from the multiple conformations accessible.

Differently from the case of oligomeric mixtures, for the modeling of flexible systems the number of possible components is enormous and the ensemble size is not known *a priori*. Ensemble methods are based on a common strategy containing two consecutive steps: first the computational generation of a pool of static models for possible conformations of a given protein (thousands of models), and then the selection of a sub-ensemble of conformations that together describe the experimental profile utilizing eq. 1.43. The development of realistic models to generate pools of conformers which adequately represent unstructured states of proteins and of optimization methods to select the ensemble are both challenging [68].

For the ensemble optimization method (EOM) [69, 70] an algorithm to generate pools of conformations of amino acids' chains is implemented which relies on the distributions of bond angles vs. dihedral angles in a Ramachandran plot for structured and unstructured regions of known proteins [71], with additional checks to avoid steric clashes. The initial pool may also include domains with known structure represented by high resolution coordinates. Unfortunately an automatic check for steric clashes between the dummy-residues of the flexible regions and the coordinates of the rigid portions has still to be implemented and an additional screening is needed before obtaining a meaningful pool in case large structured portions known *a priori* are present in the flexible model. The theoretical scattered intensity is then calculated from all the structures of the pool by means of CRY SOL and a genetic

algorithm is applied to perform the selection of an ensemble of conformers whose averaged scattering profiles fit the experimental data, repeating then the process multiple times. The meaningful observations from these approaches come from the comparison between the distribution of structural parameters (R_g , maximum C α -C α distance) of the starting pool which should cover the whole conformational space available and of the selected ensembles which fit the experimental data.

1.2 Component analysis

When a large number of spectra or scattering profiles are measured as a function of a varying parameter (e.g. time, pH, temperature, concentration of reactant,...) a possible way to describe the process undergone by the system is in terms of a linear combination of structural/spectral components whose fraction varies as a function of the perturbation.

The mathematical method of the singular value decomposition (SVD) offers a tool to estimate the number of independent components which contribute, as a linear combination, to the observed variation of the data due to the process which has to be described, but in the presence of experimental noise.

The series of data can be indeed arranged to form a matrix $\mathbf{A} = [\mathbf{I}_j(\nu)]$ of dimensions $M \times N$ where each column $\mathbf{I}_j(\nu)$ is one of the spectra/scattering profiles (with M data points ν_i) corresponding to the different values of the perturbation at which the observations were made ($j = 1, \dots, N$).

According to a theorem of linear algebra [72] every matrix $\mathbf{A}$ of dimensions $M \times N$ with $M \geq N$ can be decomposed as the product of: a $M \times N$ matrix with orthogonal columns $\mathbf{U}$, a $N \times N$ diagonal matrix $\mathbf{S}$ with non-negative elements s_j called singular values, and the transpose of a $N \times N$ orthogonal matrix $\mathbf{V}$:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (1.44)$$

A numerical algorithm can perform the decomposition and calculate the three matrices $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}^T$ [73]. The result is that the full series of experimental curves can be exactly expressed by the product of a set of N orthogonal components $u_j(q)$ which constitute the columns of $\mathbf{U}$ multiplied by the corresponding weights, given by the elements of the matrix $\mathbf{S}\mathbf{V}^T$, which are naturally placed in decreasing order of importance when calculated by the algorithm. An equivalent formulation in terms of decomposition in eigenvectors-eigenvalues is described by the identities:

$$\mathbf{A}^T\mathbf{A} = (\mathbf{U}\mathbf{S}\mathbf{V}^T)^T(\mathbf{U}\mathbf{S}\mathbf{V}^T) = \mathbf{V}\mathbf{S}\mathbf{U}^T\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{V}\mathbf{S}^2\mathbf{V}^T \quad (1.45)$$

in which the diagonal elements of $\mathbf{S}^2$ (i.e., the squares of the singular values of $\mathbf{A}$) are the eigenvalues, and the columns of $\mathbf{V}$ are the corresponding eigenvectors, of the matrix $\mathbf{A}^T\mathbf{A}$.

The benefit of the linear decomposition is that one can find the *minimum number of independent components* $\ll N$ which are necessary to describe all the variation of the experimental data, neglecting those that just contribute to the noise. Indeed a propriety of the algorithm is that the first L columns of $\mathbf{U}$ multiplied by the corresponding columns of $\mathbf{V}$ and diagonal elements of $\mathbf{S}$ produce the best

approximation of rank L for the $\mathbf{A}$ matrix, in the least-square sense [74]:

$$\mathbf{A} \approx I_{calc}(\nu, t) = \sum_{j=1}^L u_j(\nu) w_j v_j^T(t) \quad (1.46)$$

Unfortunately the evaluation of the number of significant components L is not a standardized task; it can be made according to the following aspects: the relative magnitude of the singular values; the change of the goodness of fit of the rank L linear approximation at varying L , estimated by $\chi^2 = \frac{1}{M(N-L)} \sum_{i=1}^M \sum_{j=1}^N \frac{(I_{calc}(i,j) - I(i,j))^2}{\sigma(i,j)^2}$ and by visual inspection of the fit; the level of signal-to-noise ratio in the independent components, which can be evaluated by calculating their autocorrelation [74]. In the context of SAXS, the SVD method had been applied to study the kinetics of folding or the equilibrium of unfolding of proteins and RNA [43, 75–78] and used to demonstrate the necessity of assuming intermediate states to describe the processes, or, on the other hand, that only a native and an unfolded conformer were sufficient.

However the orthogonal components obtained by SVD do not represent spectra or scattering profiles of molecular species. An interpretation of the data in terms of physically meaningful components can be possible if assumptions are made on the identity of these basis functions deriving from *a priori* knowledge [79].

Alternatively, a soft-modeling procedure to decompose the experimental data series in a weighted linear combination of pure components, with also the imposition of constraints aimed at giving a physical meaning to the model, is the multivariate curve resolution method using an alternating least square (MCR-ALS) algorithm [80, 81].

The purpose of the optimization is to find both the pure spectra of the species in the mixture and their concentration profiles which provide the best least-square approximation to the experimental data, but by fulfilling at the same time some reasonable constraints derived from the physical nature of the system and from prior knowledge of the problem under study. The imposition of constraints reduces the so-called “rotation” and “scale” freedom of the factorization problem. Usual examples for constraints are the condition of non-negativity for both the basis components and the coefficients of the linear combination, and a closure condition for the concentration profiles; these soft requirements already allow for looking at the optimized components as spectra or scattering profiles of the underlying molecular species and to the coefficients of the linear combination as their volume fractions. Possibly more stringent requirements on the coefficients can be imposed, for example deriving from kinetic or thermodynamic models for the process. The components can also be forced to coincide with some known spectra. The procedure starts from an initial guess and continues in an iterative optimization [82]. The quality and reliability of the MCR-ALS solution may be assessed using parameters that quantify the dissimilarity among the experimental data matrix and the data modeled by MCR-ALS. The visualization of the solution space in terms of quality of the reconstruction as a function of the optimized components and concentration profiles can help assessing the ambiguity of the modeling procedure [83].

In the context of SAXS, MCR-ALS had been applied to isolate the scattering profiles of the monomeric and oligomeric protein species in a series of data of

equilibrium mixtures at different total protein concentrations, providing an optimized description that was in agreement with both the known thermodynamic constant of the association equilibrium and the available structural models [84].

1.3 2D correlation spectroscopy

Two-dimensional correlation analysis (2DCOS) is a mathematical technique to extract information from a series of spectral data measured on a system undergoing some external perturbation. The spectral intensities are function at the same time of two independent variables, $I(\nu, t)$: ν is the spectral variable of the experimental method used to probe the samples (e.g the wavelength for UV-vis and fluorescence spectra or the scattering vector for SAXS data) and t is an external perturbation variable which describes the variation of the samples across the experiment (e.g. time, pH, temperature, concentration of reactant). By spreading the original data over the second dimension, the spectral resolution is enhanced, and the features not readily observable in the conventional spectra are emphasized. Moreover, it can probe the specific sequential order of the spectral intensity variations under the external perturbations. Particularly informative could be indeed the correlation between two series of spectra deriving from different types of techniques applied to the same system under similar external perturbation, which is called 2D hetero-spectral correlation analysis (2DHCOS). This situation can be implemented by means of multi-technique instrumental platform like the SURF which was used in the experiments for this thesis [4].

Typically, the 2DCOS provides a pair of synchronous and asynchronous correlation spectra, which represent the overall coincidental trends and the out-of-phase character of the spectral intensity variations, respectively.

The work of Noda provided both a generalized mathematical description of the two-dimensional correlation and a practical method to implement the calculation of the 2D synchronous and asynchronous correlation matrices of discrete datasets [85].

The original experimental data are converted in a dynamic spectrum by subtracting a reference spectrum:

$$\tilde{y}(\nu, t) = \begin{cases} y(\nu, t) - \bar{y}(\nu), & \text{if } t_{\min} \leq t \leq t_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (1.47)$$

The average of all the spectra over $t_{\min} \leq t \leq t_{\max}$ is usually picked as a reference. The correlation spectrum is then defined as:

$$\chi(\nu_1, \nu_2) = \langle \tilde{y}(\nu_1, t) \cdot \tilde{y}(\nu_2, t) \rangle \quad (1.48)$$

where the symbol $\langle \cdot \rangle$ represents the operation of cross-correlation. The intensity of the 2D correlation spectrum $\chi(\nu_1, \nu_2)$ represents a quantitative comparison of the patterns of spectral intensity variations occurring along the external variable t which are observed at two different values of the spectral variable, ν_1 and ν_2 .

Using a complex number notation, the cross-correlation can be simplified and divided into synchronous, Φ , and asynchronous, Ψ , parts as follows:

$$\chi(\nu_1, \nu_2) = \Phi(\nu_1, \nu_2) + i\Psi(\nu_1, \nu_2) \quad (1.49)$$

In order to compute the 2D correlation spectra in practice, Noda showed that some simplification can be made when the data are equidistantly spaced, yielding for the synchronous spectra:

$$\Phi(\nu_1, \nu_2) = \frac{1}{m-1} \tilde{y}(\nu_1)^T \cdot \tilde{y}(\nu_2) \quad (1.50)$$

where m is the number of spectra equidistantly measured between $t_{\min}$ and $t_{\max}$. For computing the asynchronous spectra, the Hilbert-Noda matrix is used:

$$N_{jk} = \begin{cases} 0, & \text{for } j = k \\ \frac{1}{\pi(k-j)}, & \text{otherwise} \end{cases} \quad (1.51)$$

$$\Psi(\nu_1, \nu_2) = \frac{1}{m-1} \tilde{y}(\nu_1)^T N \tilde{y}(\nu_2)$$

The hetero-correlation, 2DHCOS, is simply an extension of the above homo-correlation, 2DCOS, by treating ν_1 and ν_2 from different sources, e.g. SAXS and UV-vis.

The synchronous and asynchronous correlation matrices are usually represented as contour plots in which magnitudes with positive and negative sign are indicated in different colors (see e.g. Fig. 4.8 in Chapter 4).

The intensity of a synchronous correlation map represents the simultaneous occurrence of coincidental changes of two spectral intensity variations. The synchronous correlation matrix is symmetric, and the diagonal section is called power spectrum, with peaks referred to as auto-peaks. The magnitude of the power spectrum reflects the extent by which a spectral region changes across the experimental data series at varying t . The off-diagonal peaks or cross-peaks represent the simultaneous or coincidental changes observed in two spectral region of interest. While the sign of the auto-peak is always positive, the cross-peaks can be either positive or negative: the cross-peak has a positive intensity if the correlated spectral features increase or decrease together, whereas a negative sign indicates that one of the spectral features increases while the other one decreases.

On the other hand, the changes in the asynchronous map are indicative of sequential or successive but not coincidental changes in spectral intensities. The asynchronous map does not have any auto-peak and consists exclusively of cross-peaks. The asynchronous cross-peak develops only if the intensities of the two spectral features change out of phase with each other (e.g., delayed or accelerated). If the intensity at the two spectral coordinates changes in the same direction (both increase or both decrease), the sign of an asynchronous peak is positive if the change at ν_1 happens before that at ν_2 and negative if *vice versa*.

Chapter 2

Albumin

Albumin is the most abundant soluble protein in the body of all vertebrates, and due to its availability in purified form inherited from the plasma fractionation program of the World War II, it became the favorite model system for protein chemists. Albumin is probably the most studied of all proteins, having attracted the interest of both physicians and basic research scientists for generations [5]. In this chapter some information about the protein functions and its structural and physical-chemical features are briefly presented as introductory material for Chapters 3 and 4, in which two studies about conformational changes of this protein in stress conditions are reported.

2.1 Overview of albumin functions: the “tramp steamer”

Human Serum Albumin (HSA) is produced in the liver cells and then released in the extracellular circulation where each molecule resides on average for 27 days before being degraded and replaced. It is the most abundant protein of the blood stream: albumin represents about 60% of the total protein mass and reaches a typical concentration of 42 g/l in the serum of an adult, being its blood levels examined as index of good nutrition and health. In addition it can diffuse in extravascular compartments so that it is also present to some degree depending on the tissue permeability in every fluid of the body [5].

Due to its abundance it largely contributes to the total colloidal osmotic pressure of plasma and to the maintenance of the blood pH. However the most recognized role of albumin is the ability to bind and carry in the circulation a large number of compounds, especially of hydrophobic nature, which would be otherwise scarcely soluble in the serum. Therefore great part of its physiological relevance and potential biotechnological applications derive from the extraordinary ligand-binding capacity, providing a depot and carrier for many endogenous and exogenous species including for example: fatty acids, heme group, a broad range of drugs, hormones, bilirubin, other potentially toxic metabolic products, and also metal cations. Indeed, HSA represents the main carrier for fatty acids, affects pharmacokinetics of many drugs, provides the metabolic modification of some ligands (e.g. de-esterification of aspirin), renders potential toxins harmless, accounts for most of the anti-oxidant capacity of human plasma [86].

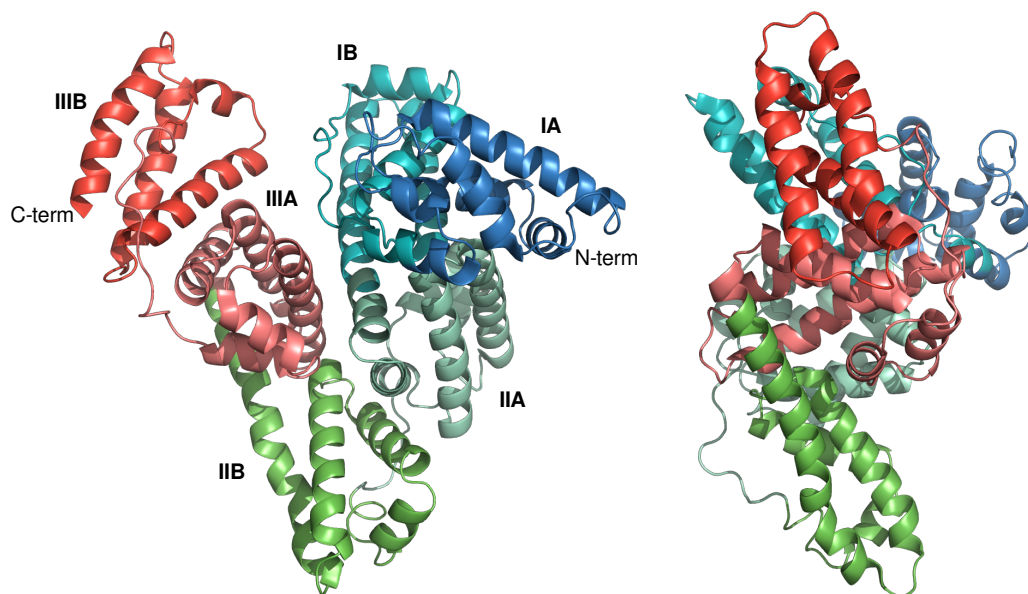


Figure 2.1. *Crystal structure of HSA.* In the protein structure (pdb entry 1AO6 [88]) shown in cartoon representation the subdomains are highlighted by different colors: IA in blue, IB in cyan, IIA in blue-green, IIB in green, IIIA in pink, IIIB in red. On the right the structure is also shown rotated of 90° about the vertical axis.

This willingness to take on a varied cargo causes albumin to be likened to a sponge or to a “tramp steamer” of the circulation [5]. The flexibility of the albumin structure adapts it readily to ligands, and its three-domain design (section 2.2) provides a variety of sites [5].

2.2 Albumin structure

HSA is a monomeric globular protein of 585 amino acids with a molecular mass of 66.5 kDa without neither prosthetic groups or glycosidic moieties. Its crystal structure, which was solved by He and Carter in 1992 [87] and also later again at higher resolution [88], shows an asymmetric protein with a hearth shape, or in other words an almost triangular shape with edges of 80 Å and thickness of 30 Å (Fig. 2.1).

The crystal structure shows also the high content of α -helix as secondary structure (67%), in agreement with the circular dichroism signal and with prediction methods based on the sequence. The helices are spaced out by turns (10%) and regions of unstructured polypeptide chain (23%).

The albumin structure is characterized by a sequence of three homologous domains numbered starting from the N-terminus as I, II and III which probably evolutionarily originated by a gene duplication mechanism. It has been noticed that the general architecture resembles that of other less abundant plasma proteins as transferrin and transthyretin.

A peculiar feature of albumin is the presence of 17 disulfide bridges involving 34 of the 35 cysteine residues (Cys34 is free, the first in the sequence). These

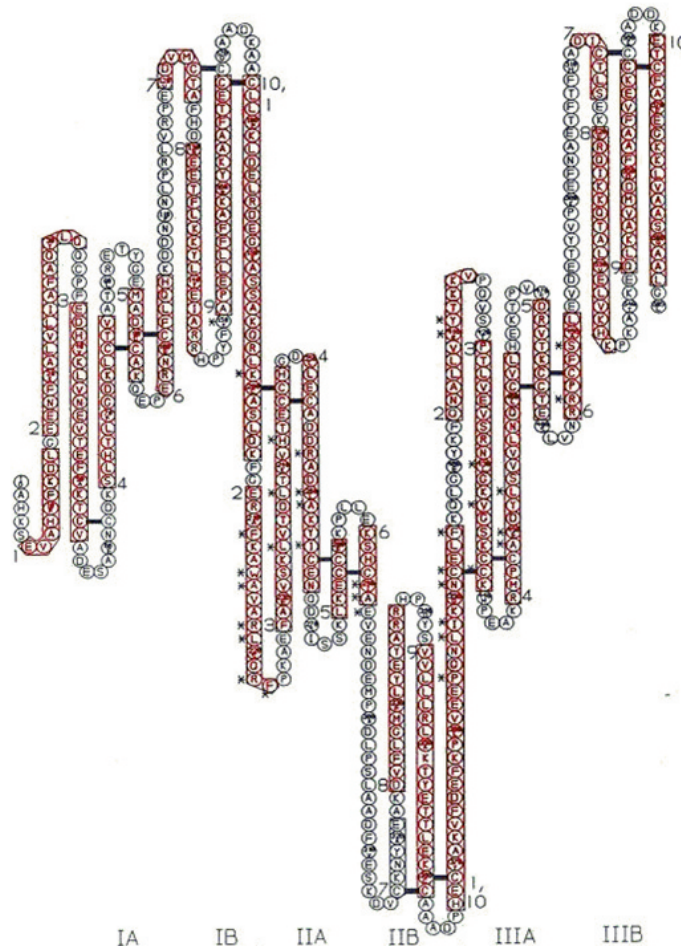


Figure 2.2. Amino acid sequence of HSA shown in an arrangement resembling its heart-shaped structure. The disulfide bridges are shown as black segments; the regions corresponding to α -helices are surrounded by red rectangles and numbered from 1 to 10, for each domain (the long helices at the end of subdomains IB and IIB are inter-domain). The amino acids involved in ligand binding are highlighted by asterisks. The scheme was reproduced from Ref. [5].

covalent bonds are disposed in adjacent pairs along the polypeptide chain (Fig. 2.2), determining a packing of the helical segments in nine rigid “double loops”. Each homologous domain involves three consecutive loops, of which the first two are grouped as subdomain A and the third represents subdomain B. The A and B subdomains contain six and four α -helices respectively.

The size and shape of albumin in solution at physiological pH had been debated in the past. The first works described albumin in solution as a prolate ellipsoid with approximate dimensions $140 \text{ \AA} \times 40 \text{ \AA} \times 40 \text{ \AA}$. This model was initially proposed on the basis of sedimentation equilibrium experiments on Bovin Serum Albumin (BSA) [89] and then considered coherent with a variety of hydrodynamics properties and also small angle neutron scattering experiments [90]. Later, when the crystal structure became available, a more compact shape was suggested by Ferrer and colleagues [91], who ruled out a “cigar-shaped” hydrodynamic model and suggested a

triangular prism shape with 80 Å edge and 30 Å thickness, which more resembled the tertiary structure found in crystal form. Then, more recently, SAXS measurements confirmed that the hearth-shaped structure derived from the crystal coordinates was also found in solution [39, 92].

2.3 Ligand binding

The fascinating binding properties of albumin have been studied for more than 50 years [93, 94]. These several investigations revealed the complexity of the multiple binding sites located in the protein structure and also a certain degree of flexibility which allows albumin to partially adapt for a better interaction with the diverse molecular species which can host. The case of the albumin as extremely versatile carrier represents one of the most interesting structure-function correlations ever reported for a monomeric protein [95]. Long and medium-chain fatty acids are the primary physiological ligands for albumin, so that about two fatty acid molecules can be found bound to each protein molecule on average *in vivo* [96]. Even though the existence of seven sites for fatty acid interaction and two main binding clefts for drug molecules had been hypothesized thanks to binding experiments in solution, the high resolution crystallographic analysis of several complexes between HSA and fatty acids, endogenous ligands, drugs and other small molecules revealed the precise architecture of the albumin sites (Fig. 2.3).

Seven common binding sites for fatty acids were identified, named as FA1 to FA7. Out of them three have the highest affinity and are located at the interface between subdomains IA and IIA (FA2) and in domain III (FA4 and FA5), offering the most solvent-protected binding environments inside the protein. These three main sites allow for the binding of the fatty acid chains in an almost fully linear conformation and provide specific interactions between the carboxylic head group and a basic residue of the protein.

The albumin clefts can also host other ligands than fatty acids. In particular FA7 and FA3-FA4 largely coincide with the two main pockets for drug binding, which are called site I and site II respectively, according to the nomenclature introduced by Sudlow [97, 98]. The site I is a big cavity located in subdomain IIA where dicarboxylic acids and/or big heterocyclic compounds with a negative charge in the center of the molecule preferentially bind. The site II is preferred by aromatic carboxylic acids with hydrophobic groups in an extended conformation. Warfarin (a coumarin derivative used as anti coagulant drug) and ibuprofen (a non-steroidal anti inflammatory drug) represent the stereotypical ligands for site I and site II respectively.

In addition the “D-shaped” cavity of FA1 in subdomain IB is also the specific binding site of the heme group [99].

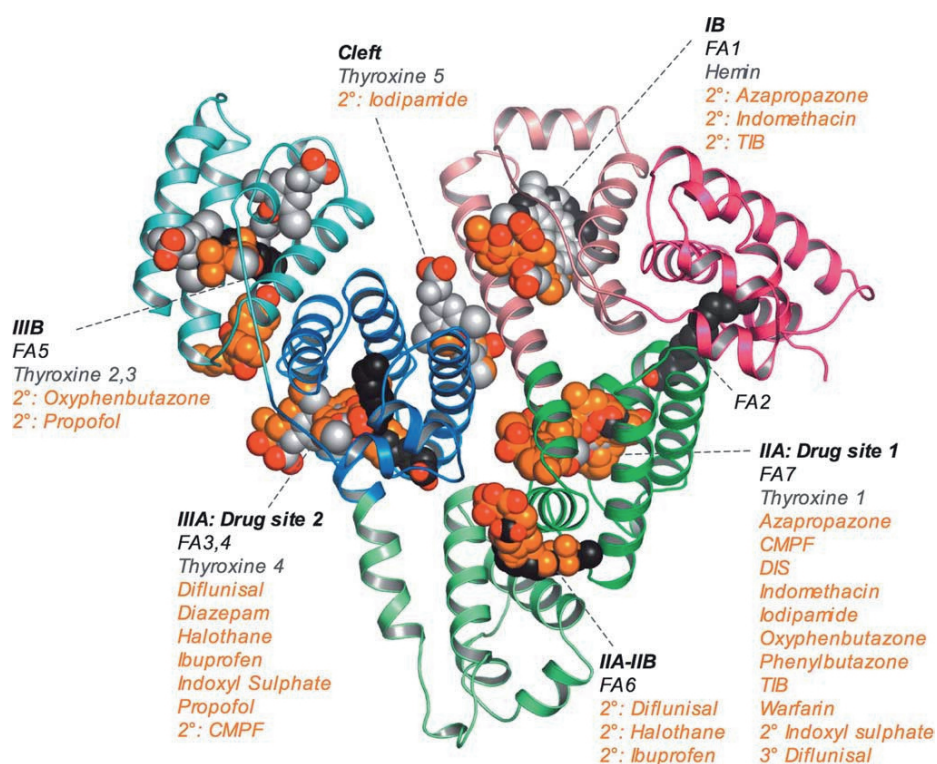


Figure 2.3. Visual summary of the ligand binding possibilities of HSA as assessed by crystallography. The albumin structure is shown in cartoon representation. The ligands are represented as spheres in which the oxygen atoms are colored in red; all the other atoms of the fatty acid molecules (myristic acid), of the other endogenous ligands (heme, thyroxin) and of drug molecules are represented in dark grey, light grey and black respectively. The figure was reproduced from Ref. [96].

2.4 Some physical-chemical properties

2.4.1 Ionic properties

Generally two pH values are defined to characterize the ionic properties of proteins. The *isoionic point* is defined as the pH of a thoroughly deionized solution of the protein. For albumin a value of 5.2 is reported and at this pH essentially all carboxylic groups should be deprotonated and the amino, guanidino and imidazole group should be protonated, giving the pH of highest total number of charged residues.

The *isoelectric point* on the other hand is the pH at which the total net charge of the molecule, including any bound ion, is zero. In this conditions the molecule would not migrate in an electric field. For the albumin not deprived of fatty acids naturally bound and in a solution with NaCl 0.15 M the isoelectric point is reported to be 4.7. At the isoelectric point the effect of electrostatic repulsion between different protein molecules with equal electric charge is absent: the solubility of albumin was observed to markedly decrease at this pH, more than that of other proteins [5].

The net charge at physiological pH (7.4) calculated from the HSA amino acid sequence is overall -15, but the charges are not uniformly distributed in the three domains: -9 for domain I, -8 for domain II and +2 for domain III. In principle the true ionization state of the acid/basic residues can differ from the one calculated on the basis of the reference pK_a , due to local effects of the protein structure, if the amino acids are embedded in pockets and engaged in specific hydrogen bond interactions. However the outstanding early work of Tanford [100] and Foster [101] about the HSA titration lead to retrieve the number of potentially ionizable groups and their average pK_a ; these predictions turned then to be in excellent agreement with the true amino acid sequence of the protein which was still unknown at their time. This fact essentially suggests that all the potentially ionizable groups would be accessible to the surrounding water solvent in solution. Some of the pK_a reported by Foster resulted anomalous compared to the values found in other proteins: half of the carboxylic groups were predicted to have a pK_a lower than 3.7, about 0.5 units less than usual. These slightly altered pK_a values, the origin of which could be based on the precise local structure, were suggested by Foster [102] to have a role in the conformational transitions observed at acid pH (section 2.5.1).

2.4.2 Spectroscopic properties

UV spectrum In the near UV region the absorbance spectrum of HSA is similar to that of many proteins without prosthetic groups, showing a peak centered approximately at 280 nm, the common wavelength used for estimating protein concentration; more precisely the maximum is found at 278.5 nm (Fig. 2.4b). The overall signal is mainly due to the aromatic amino acids: tryptophan (Trp), tyrosine (Tyr) and phenylalanine (Phe), in decreasing order of activity. HSA possesses only one tryptophan residue in position 214 which is located in subdomain IIA, in correspondence of the Sudlow's site I. The 18 tyrosine residues are mainly concentrated in subdomains IB and IIB. Also 31 phenylalanine residues are present, with negligible absorbance at 280 nm but which contribute with vibrational features of the benzene ring absorbance in the 250-270 nm region, observable in good quality spectra. In addition also the several S-S bonds should contribute to some extent [103]. Compared to the average

value for proteins, the molar absorptivity of HSA is lower due to the presence of only one tryptophan residue out of 585 amino acids. This reflects the predominant contribution of this amino acid compared to the others, being the molar absorptivities for Trp, Tyr and Phe at 280 nm equal to 5540, 1480 and $\approx 0 \text{ M}^{-1}\text{cm}^{-1}$, respectively.

In the far UV region the absorbance of the peptide bonds is found, which has a peak around 187 nm. Therefore below 240 nm the absorbance starts to greatly increase, and is also largely influenced by the sample turbidity.

Fluorescence The HSA fluorescence emission is mostly due to the only tryptophan, with additional minor contributions of the several tyrosine residues, depending on the excitation wavelength λ_{exc} (Fig. 2.4c). If $\lambda_{exc} \geq 295 \text{ nm}$ the tryptophan can be selectively excited to obtain its emission spectrum; it shows a maximum around 345 nm but it is sensitive to the local environment and in general can show blue shift if embedded in a more hydrophobic surrounding or a red shift if more exposed to a polar solvent [104]. When $\lambda_{exc} = 280 \text{ nm}$ also tyrosine residues can be excited and the emission spectrum will show contributions of both aromatic amino acids, with tyrosines having their intensity peak around 305 nm. According to the crystal structure, the Trp214 is located in subdomain IIA close to the Sudlow's site I, relatively exposed to the solvent due to the presence of a 10 Å access cavity, but not on the protein surface.

BSA possesses two tryptophan residues: Trp213 is homologous to the Trp214 of HSA, whereas Trp134 is located in subdomain IB.

Circular dichroism The circular dichroism (CD) spectrum of albumin in the far UV region (190-250 nm) is very typical for a protein whose secondary structure is mostly α -helical. Characteristic features are the two minima at 208 nm and 222 nm (Fig. 2.4d). The percentages of α -helix structure reported in the literature can range between 55 and 75 % according to the algorithm used to estimate it from the spectrum and to the inclusion or not of partially distorted regions. In the near-UV a weak CD signal can be observed due to the fact that the aromatic amino acids and the S-S bonds are part of a chiral framework. It is therefore considered a “fingerprint” of the protein tertiary structure and used to monitor conformational variations; two minima with negative ellipticity at 262 and 268 nm are recognizable (Fig. 2.4d, inset).

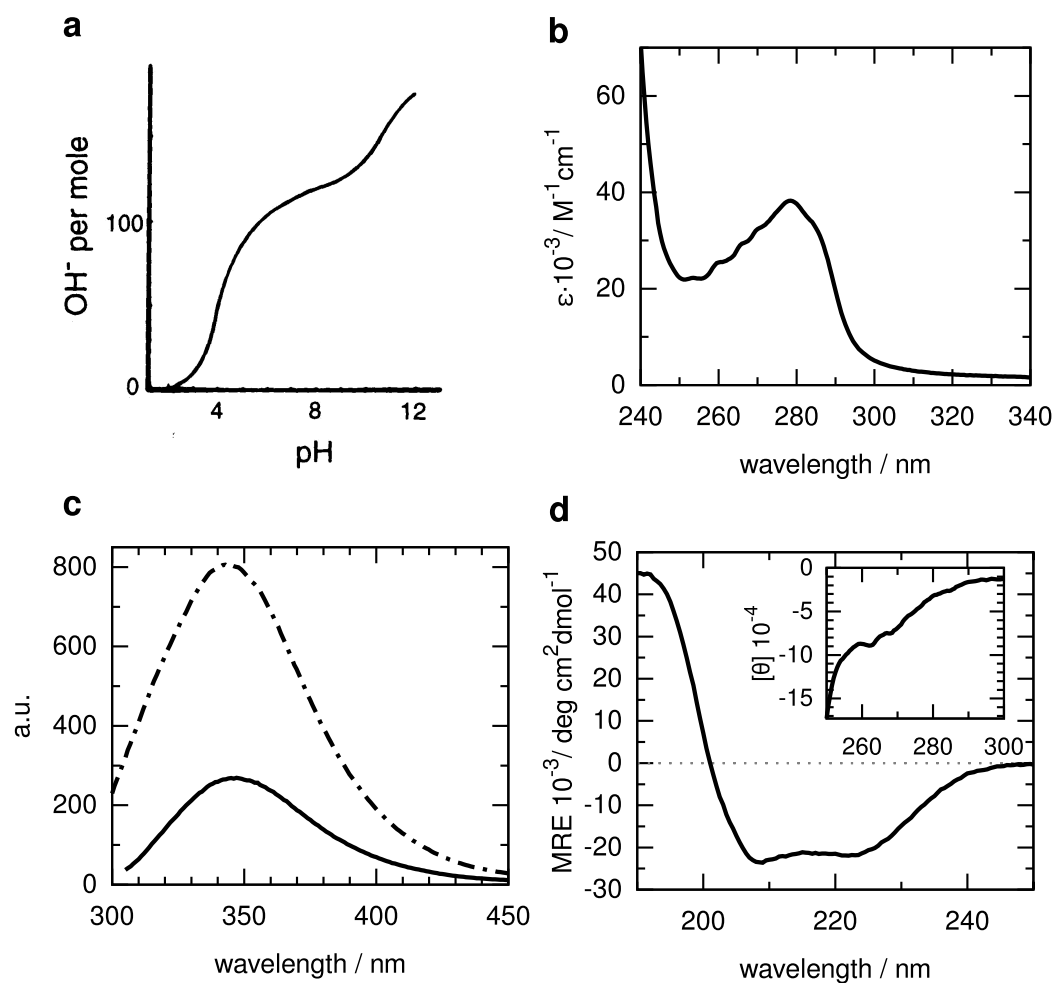


Figure 2.4. *Some physical-chemical properties of HSA. (a) Titration curve from Ref. [5]. (b) UV absorption spectrum. (c) Fluorescence emission excited at 280 nm (dash-dotted line) and at 295 nm (solid line). (d) Circular dichroism spectrum in the far-UV. In the inset, CD signal in the near-UV region.*

2.5 Conformational variations

The modular structure of albumin which involves a succession of loops kept together by disulfide bridges and spaced out by more flexible regions, confers to the protein both a notable degree of flexibility and a remarkable resilience in extreme conditions. Conformational variations were indeed recognized in the protein behavior and they are observed usually in non-physiological conditions.

2.5.1 pH-induced rearrangements

The property of albumin to undergo reversible conformational variation as a function of pH was first demonstrated with BSA by Luetscher [105], who observed different isomers in the electrophoretic migration of the protein. This behavior was reexamined later by Foster who suggested a classification of the different forms [101, 102, 106]:

- “N” form or normal, prevailing at neutral pH;
- “F” form or fast-migrating at $\text{pH} \approx 4$;
- “E” form or extended, at $\text{pH} < 3.5$;
- “B” form or basic, observed at $\text{pH} \approx 8$;
- “A” form or aged, observed after prolonged treatment in alkaline solution ($\text{pH} > 10$).

The structural modifications accompanying these transitions have been mainly predicted on the basis of the several physical-chemical data obtained, but are not fully characterized by assigning a 3D structure to the non-native isomers. Most of the information about the albumin isomerizations were obtained by studying BSA, but can be largely extended to HSA. The existing knowledge about the transitions at acid pH is further reviewed in the introduction of Chapter 4.

The N-F transition was interpreted as an opening of the molecule, implying an increase of the viscosity, a decrease of the solubility and a loss of less than 10% of the helical structure [106]. At these pH values (4.5-4) a blue shift of the Trp214 fluorescence can be observed, explainable by admitting that the F conformation determines a more rigid packing of the subdomains IIA and IIB around the fluorophore.

Below pH 3 albumin undergoes a further opening, with significant reduction of α -helical structure. This is thought to involve mainly the unfolding of the long inter-domain helices between domains I and II and between II and III. The fluorescence emission of Trp was observed to further decrease up to pH 2, and a SAXS analysis at pH 1.7 suggested that HSA is present in a conformation expanded up to the maximum extension allowed by the S-S bonds [92].

At moderately alkaline pH HSA seems to undergo another rearrangement, less substantial than those previously cited, and with a more gradual nature, which was suggested to have physiological significance [86]. In the literature it is indeed largely assumed that at pH 8-9 HSA exists in the so-called B form. The experimental investigations lead to consider the N-B isomerization as a structural fluctuation, a relaxation of the molecule with loss of rigidity, which should mainly involve the N-terminus and is enhanced by the presence of some ligands [5].

The A form was recognized by Sogami and colleagues in 1969 [107] as a new band in the electrophoretic migration of a BSA sample which had been deprived of the

bound fatty acids and was left for 3-4 days in a highly alkaline solution. This band was migrating slower than the N form and amounted to around half of the total albumin of the preparation. Later attempts showed that the A isomer formation was inhibited by alkylation of the free sulfhydryl group of Cys34 and was accelerated by thiol compounds in solution: these observations suggested that the isomerization was the result of disulfide interchange reactions originating from a free S-H group. The necessity for highly basic conditions indicated that the deprotonated thiolate was involved in the reactions. Both CD and fluorescence signals consistently decrease at $\text{pH} > 12$, suggesting a largely unfolded state.

2.5.2 Denaturation

Due to the extensive use of albumin as a model protein, the studies of its denaturation processes induced by both heat and chaotropic agents, traditionally employed to probe protein stability, have been several and performed with plenty of techniques [108–113]. Being the binding properties the most important features of albumin, the way in which the interactions with ligands could enhance or reduce the protein stability has also been object of interest [79, 114–117].

The loss of the native structure was observed to be almost always reversible, except for high temperature and high alkalinity conditions that can cause aggregation. Even the reduction of the disulfide bridges which can happen only if specific reducing agents are applied, resulted to be reversible [5].

Albumin is rather resistant to high temperatures: the clinical preparations undergo a pasteurization process at 60 °C for 10 hours to remove potential pathogens and it results that no appreciable changes occur. At temperature higher than 62 °C however a decrease of α -helical content can be seen and the “defatted” albumin starts aggregating beyond a temperature of 63 °C [5].

The behavior of HSA undergoing thermal unfolding and then refolding was investigated and a multi-step mechanism was pointed out: below 50 °C only a reversible separation of domains II and III was deduced; above 70 °C an irreversible unfolding of domains II and III was evidenced [118].

However the observation of a fully unfolded albumin molecule due to thermal treatment would be difficult since aggregation occurs at the same time during heating [119]. It is probable that the thermal aggregation process starts from a partially unfolded conformation.

The study of the equilibrium unfolding processes of albumin induced by denaturing agents (mainly urea and guanidinium chloride, GdmCl) strongly suggested multi-step mechanisms as a function of increasing concentrations of the co-solvents, in which partially folded conformers can be populated and probably are ascribable to the multi-domain nature of the protein [5].

The maximum loss of structure was reached in GdmCl solution at a concentration of 2.3 M, while in the presence of urea at a molarity as high as 9 M. It has to be considered that these conditions are not able to impair the covalent S-S bonds, however. In both cases most of the literature assumes an unfolding mechanism with at least one intermediate form between the native and the fully unfolded

conformations. Solution scattering and spectroscopic studies lead to suggest a step-wise unfolding of the three protein domains at increasing denaturant concentration, involving two intermediate states sequentially populated, in which one and then two domains unfold, respectively.

The “opening” order of the three domains had been debated; it was then proposed that the most labile domain and so the first to unfold would be the C-terminal domain III in the presence of GdmCl as denaturing agent, whereas it could be the N-terminal domain I when the unfolding is induced by urea [79, 112, 117, 120]. In addition the presence of hydrophobic ligands (fatty acids, drugs) was observed to clearly favor more compact conformations in the urea denaturation process, whereas the fatty acids had a negligible effect on the overall protein dimensions in the GdmCl-induced unfolding [79, 116, 117].

It should be noticed that several studies of the unfolding process were based on the use of fluorescent probes with affinity for specific albumin binding sites to also infer an order of stability for the three domains. However it is likely and also suggested by the findings cited before that the interaction with ligands can affect the energy landscape of the protein conformations in denaturing conditions; for example these molecules could induce a stabilizing effect on the folded state of the domain in which they are bound.

One could think about the albumin unfolding as a path which potentially admits a number of partially folded conformations based on the protein modular structure, whose relative stability can be modified by the specific conditions, included the interaction with ligands.

2.5.3 Allostery hypotheses

Some experimental evidences suggested that the conformational adaptability of albumin upon ligand binding can involve rearrangements which extend beyond the immediate surrounding of the interaction site and therefore can determine a reciprocal influence between the occupation of different sites. A hypothesis of albumin as an allosteric protein has been therefore put forward envisaging that the multi-domain architecture could provide for behaviors resembling those of multimeric proteins [121].

The first observation was the conformational transition which occurs upon binding of the fatty acids: from the comparison of the crystal structures of “defatted” HSA and in the presence of several fatty acids, a 16° rotation of the domains I and III relative to domain II is observed, which is induced by the tilt of the long interdomain helix connecting subdomains IB and IIA. Crystallographic evidences and also solution experiments showed that the transition occurs cooperatively when the protein is half-saturated with fatty acids [94]. Since the FA2 and FA3 are located at the interface between two domains (Fig. 2.3), their occupancy was suggested as a possible trigger for the transition. It was also proposed that this transition could physiologically work as a switch mechanism to guide the receptor-mediated endocytosis in the process of release of the fatty acids to liver cells; this hypothesis would imply the existence of a receptor able to selectively recognize the “loaded” albumin conformation [86]. The observed changes seem to have an impact on the affinity of ligands for some binding sites, in particular the Sudlow’s site I (subdomain

IIA).

Indeed the crystal coordinates showed that the binding of a fatty acid molecule in FA2 determines a change of position of Tyr150, which results attracted towards the polar head group of the fatty acid. In defatted HSA Tyr150 protrudes instead towards the cavity of Sudlow's site I where it can interact by means of hydrogen bond with the polar groups of the drug molecules that can occupy that site. Regarding how the affinity of site I for its ligands is affected by the presence of fatty acid molecules, solution studies provided apparently contrasting results: Ascenzi and colleagues reported a higher affinity of warfarin for "defatted" HSA compared to "fatted" HSA [121], whereas Ghuman and colleagues reported that the addition of fatty acids up to a molar ratio 1:4 could increase the affinity of warfarin for site I, but a competition mechanism took over at higher fatty acid concentrations [96]. The structural detail of the re-positioning of Tyr150 do not rule out either of the hypotheses: the first one is explainable with the loss of a specific hydrogen bond interaction, and the second one is justifiable due to the increase of the available volume in the cavity, with reduction of the steric hindrance.

Other observations regarded the relationship between the Sudlow's site I and the heme binding site [122]. Ascenzi and colleagues observed that saturating concentrations of drugs with affinity for site I decreased of one order of magnitude both equilibrium and kinetic constants for the formation of the HSA/heme-Fe(III) complex. A negative effect on the HSA/heme-Fe(III) affinity was also reported for ibuprofen, supposedly when interacting at its secondary site in FA6. Effects on the reactivity of the HSA/heme-Fe(II) complex were also observed, connected to the coordination state of the Fe atom. HSA/heme-Fe(II) is indeed able to reversibly bind CO and the radical NO[•] [123], facilitating the detoxification from the nitrogen reactive species. Saturating concentrations of drugs binding in site I seemed to accelerate the dissociation process, with the movement of the Tyr150 possibly involved.

These hypotheses of allosteric behavior are so far based on crystallographic data which gave a detailed but static picture of the albumin structure. However dynamic features of the protein structure could play a significant role in the understanding of its physiological functions.

2.6 Albumin as an antioxidant

Several lines of research evidenced the importance of albumin as extracellular antioxidant resource [124–127].

The ways in which albumin could contribute to a reduced incidence of the oxidative damage in the extracellular environment could be grouped according to these main mechanisms [126]:

i) the albumin amino acids can sacrificially react with reactive oxygen species (ROS) and reactive nitrogen species (RNS), preventing in this way the damage to more "important" targets which cannot be easily replaced; the high abundance of the protein would statistically provide a plenty of potential reactive sites by means of its amino acids susceptible of modification (free cysteine, methionines, lysines above all);

ii) metal ions like Cu^{2+} and Fe^{2+} can be sequestered in the albumin binding sites therefore decreasing their availability for the Fenton reactions with H_2O_2 which produce the detrimental hydroxyl radical $\cdot\text{OH}$;

iii) lipid molecules potentially susceptible of free-radical damage like polyunsaturated fatty acids could find protection by binding to the protein hydrophobic sites;

iv) the complex of albumin with bilirubin, which can bind specifically to the protein, was observed to inhibit lipid peroxidation;

v) by binding to albumin, some products of oxidative reactions like oxysterols and homocysteine would be less available in free form so that the incidence of their oxidation- and inflammation-promoting effects could be potentially reduced.

The existence of these mechanisms has been largely proved in *in vitro* experiments.

In the biomedical literature it was suggested that the epidemiological correlations between reduced albumin levels and increased mortality risk could be ascribed to the incidence of oxidative stress in many pathological conditions [125]. More recently the occurrence of oxidatively modified albumin forms in the circulation of patients affected by chronic states of renal and pancreatic pathologies has been indeed observed, providing direct evidence of the scavenging activity of albumin *in vivo* [128].

Chapter 3

Hypochlorite-induced oxidation of Human Serum Albumin

3.1 Introduction

In the introductory Chapter 2 it was mentioned that the carrier protein albumin is also considered the main anti-oxidant defense in blood [125–127, 129]. A simple reason for this role is that due to its high abundance in plasma, albumin is an obvious target of extracellular reactive oxidant species.

The oxidant hypochlorite One of the main oxidants released in the extracellular environment is *hypochlorite*, a key component of the inflammatory response [130]. Upon stimulation of the blood cells in charge of the mammalian immune system (i. e. the leukocytes) a cascade of reactions based on a respiratory burst involving high oxygen consumption leads to the production of a number of reactive and oxidant species with the role of damaging the pathogens at a molecular level in order to kill them. In the specific case of the neutrophils (one type of leukocytes) it was recognized that when stimulated they release the heme-enzyme myeloperoxidase which is able to catalyze the conversion of hydrogen peroxide (H_2O_2) and chloride ion (Cl^-) to hypochlorous acid (HOCl) [131, 132]. The pK_a of HOCl is around 7.4 [133] so in physiological conditions “hypochlorite” exists as an equilibrium mixture of HOCl and ClO^- . Its pivotal role in the immune defense due to its strong bactericidal activity had been soon recognized but it has also become apparent that collateral damage to host tissues is a key event in a number of human pathologies linked with inflammation [134–137].

Due to the possible role in human disease, there has been increasing interest in understanding the reactions of hypochlorite with biological substrates. Rate constants for reaction of hypochlorite with individual biological components [138] have shown that proteins are by far the major kinetic targets in plasma, being amines and sulfur-containing compounds very reactive.

Typically, facile oxidation of methionine and cysteine residues was observed, together with widespread formation of chloramines ($\text{RR}'\text{N-Cl}$) at higher hypochlorite concentrations. Modifications of the aromatic side chains (tryptophan and tyrosine) were also observed. [139, 140]

These reactivity scales were typically built on the basis of the rate of the oxidant consumption but the mechanisms and the products of the underlying reactions are in most cases not completely understood and characterized [141].

Albumin oxidation in the literature HSA, as the most abundant protein in body fluids, is thought to be the most probable target of extracellular oxidation in general. Oxidatively modified forms of HSA have indeed been detected in several pathological conditions characterized by oxidative stress (e.g., nephropathies, uremia) [128,142,143]. Regarding specifically the hypochlorite-induced modifications, the hypothesis that a “chlorinated” form of HSA could be implicated in specific receptor recognitions with potential roles in physiological processes like clearance of the damaged protein or inflammation stimulation, has also been formulated [144–148]. At a molecular level, the characterization of oxidized HSA has mainly been addressed by means of model systems and a quite wide literature exists [127] including very early works [149]. Mainly chemical essays and also mass spectrometry methods were employed to quantify and identify chemical modifications induced by hypochlorite [150–153]. Spectral variations in circular dichroism (CD), UV-vis absorption and fluorescence spectroscopy experiments were also addressed in order to deduce the consequences of the chemical damage, including evaluation of possible changes in the binding properties [154–159]. The molecular response to oxidation constitutes a complex picture as the extent of protein modification induced could widely change depending on the oxidant-to-protein molar ratio and the time elapsed from the first fast reactions. Conversely, most studies focused on one specific condition and identified the treated protein as “oxidized albumin”, sometimes finding contrasting results. Some interesting systematic studies have been reported by Pattison and coworkers [153] on the homologue Bovine Serum Albumin (BSA) clarifying the effect of increasing equivalents of the oxidant hypochlorite on the amino acidic composition. It is therefore difficult to create a complete image of the hypochlorite-induced damage of HSA from the existing literature, also because the most used approaches can only indirectly address the protein overall structure.

It is foreseeable that repeated exposure of HSA to oxidative environments at high oxidant/protein molar ratios could significantly affect its biological activity. With knowledge of the well-established strict dependence of protein functionality on protein structure, a correlation between the structure and the oxidation state of HSA is expected to be crucial for rationalizing the effect of oxidation on the functional ability of this protein. In this chapter of the thesis the results of a biophysical characterization of the damaged protein at increasing doses of the oxidant hypochlorite are reported.

We correlated chemical damage and protein structural stability over a broad range of oxidant/protein ratios comprising both relevant physiological oxidant concentrations (high-micromolar or low-millimolar levels in inflamed tissues [160]) and higher oxidant/protein molar excesses explored in *in-vitro* experiments. We used a new multi-technique instrumental platform which combined the simultaneous collection of SAXS, UV-vis absorbance spectra and fluorescence emission on the same sample volume and it is in the following referred to as SURF (SAXS - UV-vis

and Fluorescence) [4]. The reliability of the structural information on the oxidized protein monomer in solution was enhanced by the in-line separation with size exclusion chromatography (SEC). Further characterization was also performed by means of CD and zeta potential measurements.

3.2 Experimental procedures

3.2.1 Sample preparation

Human serum albumin (HSA) lyophilized powder, fatty acid- and globulin-free ($\geq 99\%$, type A3782), obtained from Sigma, was dissolved in 10-mM sodium phosphate buffer at pH 7.4 and centrifuged at $18 \cdot 10^3$ g for 10 minutes at 4 °C. The protein concentration was measured by collecting a UV spectrum of the stock solution with a 1-mm path length, using the absorbance at 280 nm and the extinction coefficient $37500 \text{ M}^{-1}\text{cm}^{-1}$ [103]. The hypochlorite-oxidized HSA samples were prepared by mixing equal volumes of HSA and hypochlorite (NaOCl) stock solutions. A series of fresh NaOCl working solutions with the concentration required to reach the desired oxidant/protein ratio (OPR, $[\text{NaOCl}]/[\text{HSA}]$) in the final sample were used in the experiments. They were prepared by diluting a stock solution of commercial NaOCl (from Sigma, type 425055, reagent grade, available chlorine 10-15%) diluted in buffer. The concentration of the NaOCl stock solution (about 80 mM) was determined by measuring the UV absorbance at 293 nm after dilution in a NaOH solution with pH 12, in a 1-cm quartz cuvette, using the molar extinction coefficient value of $350 \text{ M}^{-1}\text{cm}^{-1}$ [133]. The working solutions were stored on ice or in the refrigerator at 4 °C during the day of the experiments. Samples with oxidant/protein ratios from 0 up to 170 were prepared. Also stock solutions of albumin complexed with ligands were prepared (HSAPalm, HSAQuerc, HSACurc). The ligands (palmitic acid $\geq 99\%$ type P0500, quercetin $\geq 95\%$ type Q4951 and curcumin $\geq 99.5\%$ type 78246, from Sigma) were solubilized in ethanol of spectroscopic grade and small volumes were added to the protein solution, in order to have a protein:ligand ratio of 1:1 for quercetin and curcumin and 1:10 for palmitic acid. The HSA-palmitic acid solution was left to equilibrate for at least 12 hours before the experiment. The final ethanol volume was less than 2%. To estimate the quercetin concentration in the fresh prepared stock ethanolic solution the absorbance value at 374 nm and the extinction coefficient $25,000 \text{ M}^{-1}\text{cm}^{-1}$ [161] were used. Similarly, the absorbance value at 429 nm and the extinction coefficient $55,000 \text{ M}^{-1}\text{cm}^{-1}$ [162] were used for curcumin. All the experiments, unless otherwise specified, were performed in degassed 10 mM phosphate buffer with pH 7.4, at room temperature (24 ± 1 °C).

3.2.2 SUrF platform experiments

The SUrF multi-probe set-up is a sample environment for SAXS experiments at a synchrotron beamline which allows users to collect optical spectroscopy data (here UV-vis absorbance and fluorescence emission) by means of fiber optics on the same capillary volume which is exposed to X-rays. The outline of this platform has been presented in Ref. [4]. We used the second generation of the SUrF which is described in *Dicko C., Haas S., Plivelic T., SUrF: combined SAXS/UVvis/Fluorescence a*

new tool to resolve complex interactions in biological samples (to be submitted). One of the main upgrades of the new setup was the implementation of the fluorescence detection as an alternative to Raman scattering and the substitution of a scanning-mode collection of the UV-vis spectra based on a moving grating with diode-array detection, allowing for a faster time sampling. The spectroscopic techniques result sub-optimal in the multi-probe platform based on fiber optics but this drawback is overcome by the possibility to correlate data from the different probes without the intrinsic variability of the sample preparation.

The experiments were performed at the I-911 SAXS beamline at MAXlab for which the technical specifications are reported in Ref. [163]. The sample-to-detector distance was 2400 mm and the X-ray wavelength 0.91 Å. The accessible angular range in term of the scattering vector $q = 4\pi \sin \theta / \lambda$ (where 2θ is the angle between incident and scattered beam and λ is the radiation wavelength) was 0.01-0.42 Å⁻¹. For the fluorescence emission collection, the excitation wavelength was set at 293 nm with an estimated full width at half height of 15 nm. The measurements were performed both directly on the oxidized samples loaded into the capillary (experiments referred to as “single shot”) and in combination with SEC separation. In the SURF “single shot” experiments, the final protein concentration was 5 g/l (75 μM). For each sample loaded into the capillary the first SAXS, UV-vis, and fluorescence data were collected 5 minutes after mixing HSA and NaOCl. In total four SAXS exposures of 1 minute were collected, spaced out by a waiting time of 2 minutes. The overall duration of a “single shot” experiment was about 10 minutes. During each X-ray exposure, two UV-vis transmission spectra (one at the beginning and one at the end of the exposure) and one fluorescence emission were collected. The acquisition time was 2 seconds for each UV-vis spectra and 18 seconds for fluorescence. The UV-vis transmission spectra (I) were converted to absorbance using buffer transmission (I_0) and the absorbance definition $A = \log_{10}(I_0/I)$. The buffer contribution was subtracted from the fluorescence emission spectra.

Size exclusion chromatography

In the SEC experiments the protein concentration in the samples injected into the column was about 10 g/l (150 μM). The separation run was started 5 minutes after mixing of HSA and NaOCl solutions. A 150-μl injection loop was loaded, injecting 400 μl of the samples, which were then eluted with one column volume. 1-minute SAXS exposures and 9 UV-visible spectra for each SAXS exposure were continuously collected during the elution. SEC was performed using the Fast Protein Liquid Chromatography system ÄKTA pure with a Superdex 200 Increase 10/300 GL column, in-line with the SURF platform. The SEC was run at variable flow rate, reducing it from 0.5 to 0.1 ml/min during the elution of the protein species (elution volumes between 7.5 and 13.5 ml) in order to collect a larger number of SAXS data in a single elution peak.

3.2.3 Circular dichroism

The CD spectra were recorded between 260 and 190 nm in a JASCO J-750 spectropolarimeter at 25 °C using a bandwidth of 0.5 nm, a 50 nm/min scan rate and

averaging 3 acquisitions. The demountable quartz cuvette path length was $b = 0.01$ cm and the protein concentration was $c = 22.75 \mu\text{M}$. The raw data in millidegrees θ were subtracted by the buffer contribution and converted into Medium Residue Ellipticity ($\text{MRE} = \frac{\theta \cdot 10^6}{585 \cdot c \cdot b}$, units $[\text{deg cm}^2 \text{ dmol}^{-1}]$) considering that HSA has 585 amino acids.

3.2.4 Zeta potential

In order to measure the zeta potential of oxidized HSA, the samples were prepared as previously described and then desalted 10 minutes after oxidation with 10 mM sodium phosphate buffer at pH 7.4, using a GE-healthcare PD-G25 Mditrap buffer-exchange column. The HSA concentration in the oxidized and then desalted samples was around $30 \mu\text{M}$. The measurements were performed at 25°C , using a MALVERN NanoZetaSizer apparatus (Malvern Instruments LTD, UK) equipped with a 5-mW HeNe laser. This instrument's technique combines laser Doppler velocimetry and phase analysis light scattering to determine the average mobility and the mobility distribution. The zeta potential ζ of the particles (in mV) was obtained from the electrophoretic mobility u measured using the Smoluchowski relation $\zeta = u\eta/\epsilon$, where η and ϵ are the viscosity and the permittivity of the solvent phase (water), respectively. The measurements were performed using a folded capillary cell (Malvern, UK), and the runs were carried out in triplicate for each sample, each consisting of 12 to 24 sub-runs.

3.2.5 Complementary UV-visible and fluorescence measurements

More accurate spectral characterization of the oxidized HSA samples was performed using bench-top instruments.

Fluorescence spectra were collected with a Cary Eclipse Fluorescence spectrophotometer. The intrinsic fluorescence was measured by exciting the protein solution at $\lambda_{exc} = 280$ nm, where both tyrosines and tryptophan are excited, and at 295 nm, selective for tryptophan. The emission spectra were collected with a band-pass of 5 nm for both excitation and emission monochromators, with a step of 1 nm and 0.2 s acquisition time. Filters were used both in excitation (250-395 nm) and in emission (295-1100 nm). Additional experiments with different settings were also performed. Quartz cuvettes with a path length of 1 cm in both excitation and emission directions were used. Absorption spectra were collected just before the fluorescence spectra in order to be able to correct the emission intensities for the inner filter effect according to the approximate expression $I_{corr} = I_{meas} 10^{(Abs(\lambda_{exc}) + Abs(\lambda_{em}))/2}$ [104], where I_{meas} is the fluorescence intensity subtracted for the contribution of the buffer, $Abs(\lambda_{ex})$ and $Abs(\lambda_{em})$ are the absorbance values of the same sample at the excitation and emission wavelength, respectively and I_{corr} is the corrected fluorescence signal. The samples were obtained as reported in section 3.2.1 using also magnetic stirring in the mixing since bigger volumes were prepared directly into the cuvettes.

The UV-visible absorption spectra mentioned above were acquired with a Varian Cary 1E double beam spectrophotometer with a spectral bandwidth of 0.5 nm, a wavelength step of 0.5 nm and a 0.2 s sampling time for each wavelength point. For kinetics experiments with extensive time sampling an Agilent Cary 60 single

beam spectrophotometer with a flash Xenon lamp was employed. For samples with a protein concentration around 5 g/l, 1 mm quartz cuvettes were used. When measuring a large number of samples at the same time disposable PMMA cells with 1-cm path length were employed and the sample was prepared directly mixing the stock solution into the cuvette. In this case the spectra were considered unreliable for a wavelength lower than 260 nm due to poor transmission.

3.2.6 SEC-SUrF at the SWING beamline

Additional experiments with the SuRF platform in combination with SEC were performed also at the SWING beamline, for which some technicals details are reported for example in Ref. [41]. The sample-to-detector distance was 2750 mm and the X-ray wavelength 1.03 Å. The accessible q range was 0.004-0.4 Å⁻¹.

SEC was performed using the HPLC system Agilent series 1200 available at the beamline endowed with an automatic injection robot. A volume of 50- μ l of the protein samples (concentration 155 μ l, about 10 g/l) was injected and separated through an Agilent Bio SEC-3 column with 5-ml volume, whose output was connected with the SuRF platform. A 30-mM Tris buffer at pH 7.7 without or with the addition of 0.15 M NaCl was used for the elution, which was run at a flow rate of 0.2 ml/min.

SAXS exposures of 750 ms, spaced out by a waiting time of 2250 ms, were continuously collected for 3 minutes before the elution of the protein sample in order to acquire frames to be used for background subtraction, and then during the elution of the protein for 12.75 minutes. The full elution was also monitored with the UV-vis probe of the SuRF platform collecting 2-s exposures continuously.

The samples were prepared as explained in section 3.2.1; an additional series of samples was prepared using a 10 mM sodium phosphate buffer added with 0.15 M NaCl. As a test described in section 3.9, a sample of oxidized HSA with hypochlorite/protein ratio of 120 was added with methionine (DL-Methionine $\geq 99\%$, type M9500 from Sigma) 10 minutes after the preparation. A volume of 238 μ l of a 0.1 M solution was mixed with the 400- μ l protein sample in order to reach a 3:1 molar excess of methionine relative to the nominal hypochlorite molarity in the final sample.

3.2.7 Data analysis

SAXS data analysis

SAXS data reduction (radial averaging, buffer subtraction) was performed with the software RAW [164], using silver behenate to calibrate the range of the scattering vector (q) [165] and water to convert the detector data in absolute intensity [31].

A first estimate of the radius of gyration (R_g) according to the Guinier approximation [7] at low q was made by fitting the data with PRIMUS [30], considering a q range between $q_{\min}=0.012$ Å⁻¹ and $q_{\max}=1.3/R_g$.

Model-free analysis through Indirect Fourier Transformation, providing the pair distance distribution function ($P(R)$), from which the maximum particle distance ($D_{\max}$), and the direct space R_g and scattering intensity at zero angle ($I(0)$) could be calculated, was performed with GNOM [14] and IFT-BayesApp [17]. The same q range (0.012-0.35 Å⁻¹) was selected for all data.

IFT-BayesApp also gave the equivalent ellipsoid axis ratio [18] that is very diagnostic of a transition from a globular to elongated shape.

Data treatment in the SEC-SAXS experiments

The analysis was aimed at the comparison of the average scattering patterns and UV-vis spectra among the increasingly oxidized HSA samples. For this purpose, the frames collected in correspondence with the maxima of the two elution peaks (the main peak of the monomer and the minor peak of the protein oligomers, mostly dimers) were averaged (generally 2-3 SAXS frames and 20 UV-vis transmission data). Buffer subtraction of the SAXS data and transmission-to-absorbance conversion of the UV-vis data were performed using frames corresponding to the buffer eluted after protein peaks (larger elution volumes).

In the case of the SEC-SAXS experiments performed at SWING (section 3.2.6) the data reduction was performed using FoxTrot software, developed in-house at Soleil. An automatic analysis of the subtracted protein profiles was performed with a Matlab script. The script used the tools of the ATSAS package [30] to automatically evaluate $I(0)$ and R_g via the Guinier approximation and the $P(R)$ function via the Indirect Fourier Transform implemented in GNOM [14]. Estimates of the molecular weight (MW) were also determined both from the Porod invariant [166] as 0.625 times the Porod volume for globular proteins [30] and by the invariant volume-of-correlation length (V_c), through a power-law relationship between V_c , R_g and MW that has been parametrized [25].

3D-modeling based on SAXS data

The *ab-initio* program GASBOR [57] was used to fit the SAXS data up to $0.35 \text{ \AA}^{-1}$. This software employs the average scattering factor of amino acids and makes it possible to model the particle as a protein chain-like assembly of dummy-residues, whose number is constrained according to the known sequence of the protein. The calculations were performed with P1 symmetry. A series of 20 models was generated, and the similarity of the obtained structures was verified by DAMAVER [65] and DAMCLUST [30], in which the superposition was performed by the SUPCOMB code [64]. A normalized spatial discrepancy (NSD) value was calculated and used as a parameter to determine the difference between two three-dimensional models. The model most similar to the others was chosen as representative. The combined rigid-body/*ab-initio* approach implemented in BUNCH [63] was also used to model possible partially unfolded conformations of HSA fitting the scattering data of the oxidized protein. High-resolution coordinates found in the pdb entry 1AO6 [88] were employed to create the starting model to be optimized, which was composed of rigid moieties interconnected by flexible linkers.

Secondary structure calculation from far-UV CD spectra

The mean residue ellipticity (MRE) at 222 nm (units $[\text{deg cm}^2 \text{ dmol}^{-1}]$) was used to estimate the α -helix secondary structure content, according to the method suggested by Chen and colleagues. [167] ($\alpha\text{-helix} = 100 \times (-\text{MRE}_{222\text{nm}} - 2,340)/30,300$). In this case error bars were computed estimating the standard deviation of the MRE

values dependent on a 1% uncertainty of the protein concentration value, a 0.5% uncertainty of the demountable cell pathlength and a 1% uncertainty due to noise in the spectra. In addition, the full CD spectra between 190 and 250 nm were analyzed by means of both the BeStSel [168] web-server and the CONTIN/LL algorithm of the software package CDpro [169]. Both methods provide a least squares fit of the far-UV CD spectra in terms of secondary structure components (e.g., α -helix, β -strands, turns, disordered/other). In the case of HSA, the α -helical (both regular and distorted at the ends of the segments) and disordered/turn structures accounted for most (>93%) of the secondary structure predicted in all conditions.

2D-correlation spectroscopy

Two-dimensional correlation analysis was introduced in section 1.3 as a method to highlight and correlate the most varying features in experiments in which any kind of spectral data are collected as a function of a perturbing parameter. The spectral intensities used for data correlation were the scattering data in the form of Kratky plots ($I(q) \cdot q^2$ VS q) and the UV-vis spectra ($Abs(\lambda)$); they were correlated using the oxidant/protein ratio or time, in the case of kinetics experiments, as the perturbation variable. The computation of the synchronous and asynchronous 2D-correlation spectra was carried out with the Matlab code using the method introduced by Noda, slightly modified to take into account the fact that the oxidant/protein molar ratio variable was not evenly spaced [85]. The average spectrum was used to compute the dynamic spectra. The diagonal section of the 2D-correlation matrix of a set of spectra is called power spectrum and shows for each data point (wavelength or q value) the total squared variation from the average, observed across the experimental data. It allowed us to quickly visualize at which wavelengths or q values the signal varied at most across the experimental data series (increasingly oxidized samples or time-series). In the 2D contour maps, the white-filled regions are defined as the positive correlation intensities, while the yellow-colored regions represent negative correlation intensities.

Component analysis

The Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) method (section 1.2) was applied to the data matrices built by the series of the UV-vis spectra and of the SAXS scattering data, in order to estimate the minimum number of linearly independent spectral components needed to adequately reproduce all the experimental variation observed in the HSA oxidation, and so give a hint on the minimum number of coexisting species required to account for the data. The analysis was performed by means of a Matlab script including the SVD algorithm. The SAXS subtracted data between 0.02 and 0.2 $\text{\AA}^{-1}$, normalized dividing the intensity by $I(0)$, and the UV-vis spectra between 250 and 500 nm rebinned with a 1 nm step, were used to build the data matrices. In order to estimate the most probable number of significant components underlying the data, we looked at: i) the percentage of variance explained by including each additional component, ii) the presence of systematic deviation from zero in the computed eigenvectors, evaluated looking at their autocorrelation function iii) the improvement of the agreement

between the experimental data and the PCA estimate including each additional component, evaluated by means of the value of χ^2 , the presence of significant systematic variation of the residuals and visual inspection. After PCA analysis, the multivariate curve resolution - alternating least squares (MCR-ALS) [170] method was also applied in order to find spectral components satisfying some physical constraints (non-negativity and closure of the molar fractions) so that they could be interpreted in terms of coexisting species. The optimization was performed by using the MCR-ALS toolbox implemented as a Matlab graphical interface [83]. In the case of the SAXS data we also tried to model the experimental data by constraining the possible components to be the experimental scattering pattern for chosen samples, typically the native protein and the fully oxidized protein within the experimentally investigated samples (oxidant/protein ratio=0 and 170, respectively). The data collected for intermediately oxidized samples (oxidant/protein ratio =80, 90 or 105), were assumed as pure components in the attempts to improve the fit trying to reproduce all the experimental data by simple linear combination of three instead of only two scattering patterns. The least-square fit of the experimental data as a linear combination of defined components was performed using the program OLIGOMER [66], which gives the best weight coefficients for the components, that can be interpreted as molar ratios.

3.3 Combined spectroscopic and SAXS characterization of oxidized HSA

The SAXS analysis of the series of hypochlorite-treated HSA samples collected after 5 minutes from the oxidant addition (black dots in Figure 3.1a,b,c) clearly showed a critical behavior as a function of the increasing oxidant/protein molar ratio. A steep increase of the total particle dimensions (R_g and D_{max} , 3.1a,b) and axial ratio (Fig. 3.1c) was observed at critical oxidant/protein ratio values around 80, and a plateau was seen for values higher than 120. It is interesting to note that the variation of these parameters as a function of the increasing oxidant dose was approximately sigmoidal. The $P(R)$ profiles suggested that a transition to a more elongated particle occurred (Fig. 3.1d). The almost constant value (within 7% from the reference sample) of the zero-angle scattering intensity suggested that this elongated form still belongs to protein monomers. The loss of the native tertiary structure was evidenced by a characteristic evolution of the $P(R)$ profiles from the typical bell shape of the native form to a bimodal distribution.

In addition, the visualization of the scattering data in the form of a Kratky plot highlighted that this structural variation did not imply a significant increase of the protein chain flexibility that is typical of unfolded states (section 1.1.3). In the case of oxidized HSA, the initial bell shape of the Kratky plot demonstrated an increasing positive skewness when the OPR values increased, indicating a more elongated structure [46], but was far from reaching a plateau at higher q , typical for a protein behaving as a Gaussian-like coil (Fig. 3.2).

The use of the SUrF platform enabled us to collect UV-vis absorption and fluorescence emission spectra on the same sample volume as the SAXS data (Fig. 3.3), and both the spectral data series showed how the chemical damage of HSA had to be

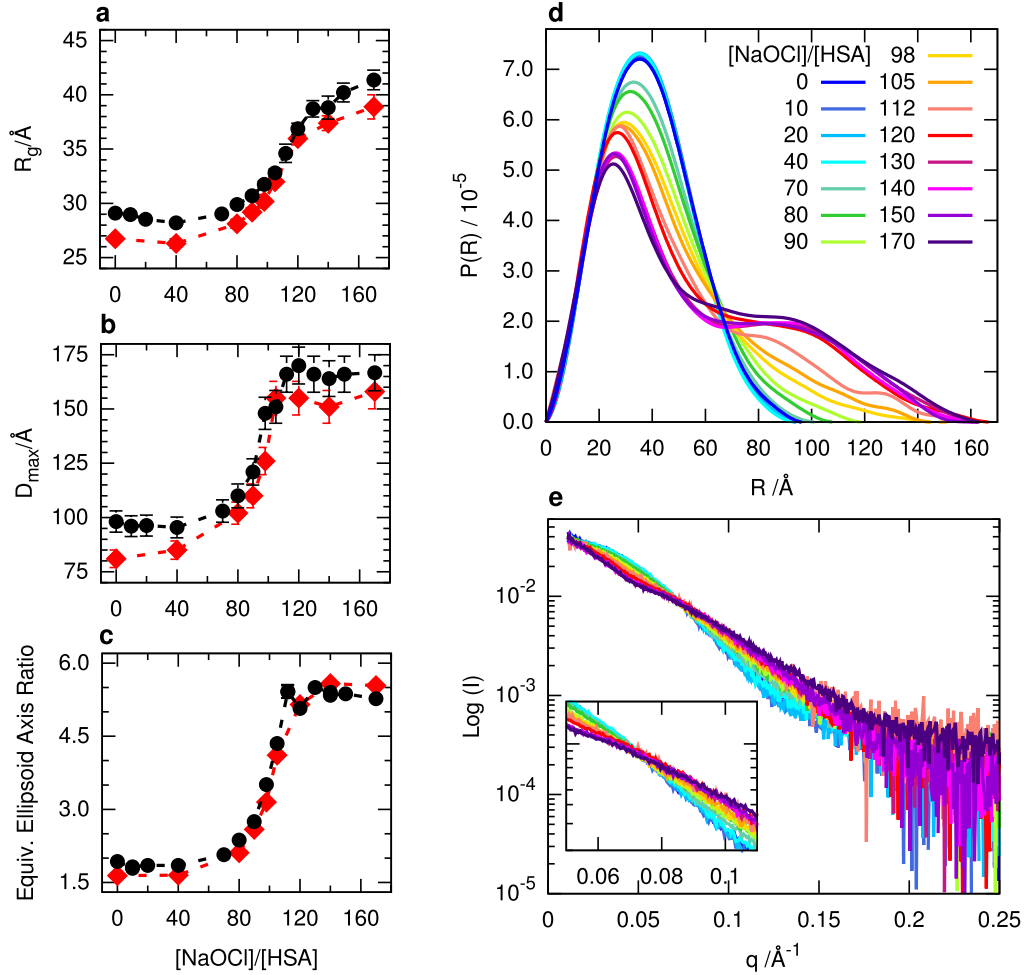


Figure 3.1. The variation of structural features derived from the SAXS data as a function of the oxidant/protein ratio ($[\text{NaOCl}]/[\text{HSA}]$) suggests that the oxidized protein monomer assumes a progressively elongated shape. The parameters derived from the “single shot” experiment without SEC (black) and from the SEC-SAXS data collected at the monomer peak maximum (red) are compared. (a) Radius of gyration (R_g); the error bars show the standard deviation estimated by the fitting procedure. (b) Maximum intra-particle distance (D_{max}); error bars corresponding to 5% of the value are displayed. (c) Axis ratio of the equivalent ellipsoid of revolution. (d) Pair distance distribution functions ($P(R)$) derived from the “single shot” experiment for increasingly oxidized HSA. (e) SAXS profiles showing an isosbestic point highlighted in the inset. The sequence of colors is the same as in (d).

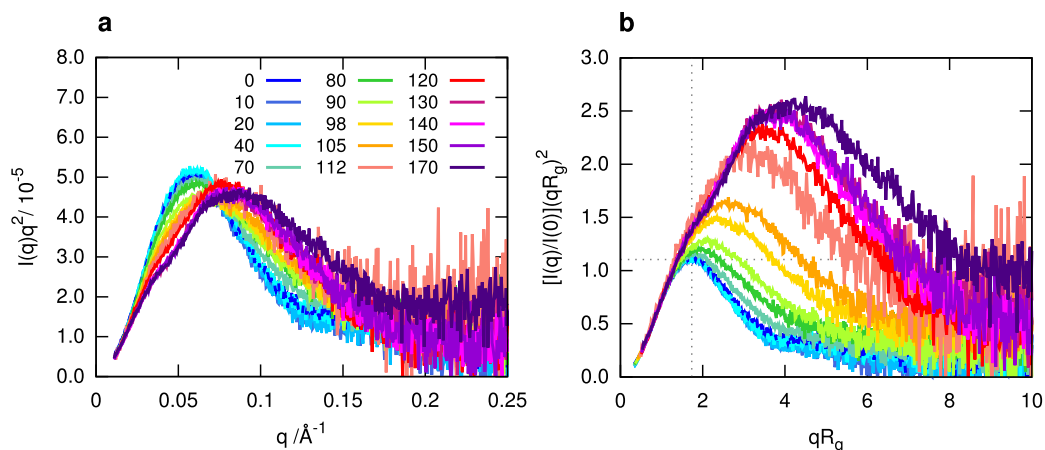


Figure 3.2. Representation of the scattered intensity $I(q)$ profiles collected for increasingly oxidized HSA samples in the form of (a) Kratky plot ($I(q)q^2$ vs q) and (b) dimensionless Kratky plot (section 1.1.3).

quite severe in order to observe a change in the protein conformation probed by SAXS. This could be seen from a comparison between the trend of the SAXS dimensional parameters (Fig. 3.1a,b,c) and the most varying spectral features (Fig. 3.3a).

The maximum fluorescence emission very quickly decayed, becoming already less than 60% of the initial value for the lowest oxidant/protein ratio value of the experiment (OPR=10) and almost undetectable in our conditions for an oxidant/protein ratio value equal or greater than 80. We considered this effect as very selectively probing the oxidation of the sole tryptophan residue which mainly contributes to the fluorescence in the native protein (Fig. 3.3c). So, in the range of oxidant/protein ratio in which the fluorescence signal quickly decayed, the SAXS data still did not show any appreciable deviation from the native protein.

The UV-vis spectra of the oxidized samples confirmed that significant chemical modification of the protein occurred for oxidant/protein ratios as low as 20. The spectra obtained as a difference with respect to the curve of the non-oxidized sample showed that the main variations in the absorbance occurred at about 250 and 304 nm (Fig. 3.3b). The absorbance at these wavelengths was observed to almost linearly increase as a function of the oxidant ratio up to a value of 120. For higher ratios, a second linear pattern with a larger slope was detected (Fig. 3.3a). Interestingly, the changes observed in the spectra as a function of the oxidant/protein ratio did not match the sigmoidal pattern of the SAXS parameters. The absorbance values maintained a linear increase in the range of oxidant doses which determined the abrupt structural expansion (oxidant/protein ratio between 80 and 120). At higher doses, when the increased protein dimensions reached the plateau (oxidant/protein ratio > 120), a steeper increase of the absorbance values at these wavelengths was observed.

2D-correlation analysis The 2D-correlation analysis (section 3.2.7) of SAXS data and UV-vis spectra using the hypochlorite/albumin molar ratio as perturbation variable gave some more details about the relationship between distinct features

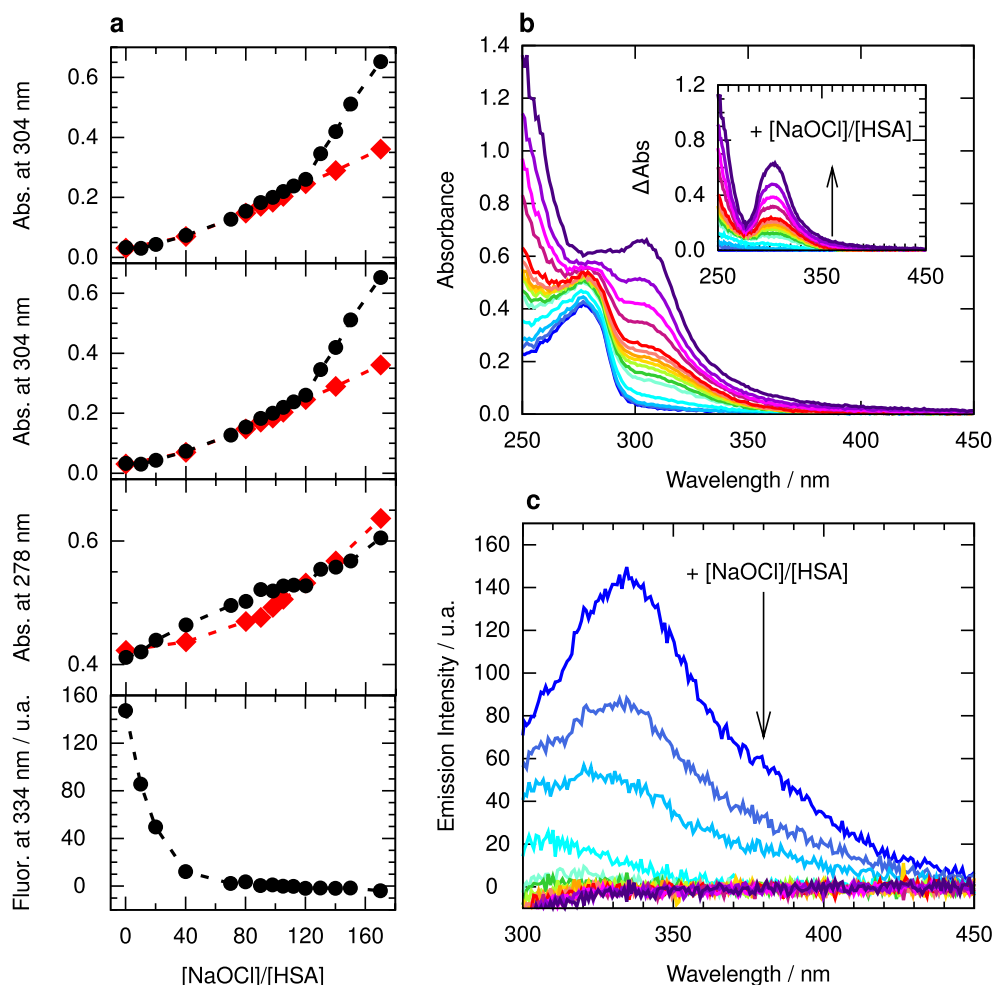


Figure 3.3. Variation of spectral features collected simultaneously with the SAXS data as a function of the oxidant/protein ratio ($[\text{NaOCl}]/[\text{HSA}]$). (a) From the top to the bottom, the absorbance values at 304 nm, 250 nm and 278 nm, and the maximum fluorescence intensity are shown. In the first three plots the absorbance data from the “single shot” experiments without SEC (black) are compared to those from the main elution peak of the SEC experiment, (red). The values from the SEC experiment are scaled according to the concentration estimated by the absorbance at 280 nm of the reference sample ($[\text{NaOCl}]/[\text{HSA}] = 0$). (b) UV-vis absorbance spectra of increasingly oxidized HSA. In the inset, the difference between UV-vis absorbance spectra of oxidized samples and the non-oxidized reference ($[\text{NaOCl}]/[\text{HSA}] = 0$) is shown. (c) Fluorescence emission spectra, obtained with an excitation wavelength set at 293 nm (section 3.2.2). In (d) and (e) the arrows represent a guide in order to follow the series of spectra in increasing order of $[\text{NaOCl}]/[\text{HSA}]$, with values: 0, 10, 20, 40, 70, 80, 90, 98, 105, 112, 120, 130, 140, 150, 170 (color code from blue to purple).

observed to vary with oxidation.

In the SAXS spectra, the main intensity variations at low (about $0.04 \text{ \AA}^{-1}$) and intermediate (about $0.11 \text{ \AA}^{-1}$) scattering vector values appeared to happen synchronously as a function of the oxidant ratio, so both the q regions can be interpreted as sensitive to the same unique process affecting the protein low resolution structure (Fig. 3.4a).

The 2D-correlation analysis involving the UV-vis spectra was predominantly affected by the greatest increase in absorbance observed for oxidant/protein ratios higher than 120. Indeed a strong synchronous positive correlation was found between the broad increases of absorbance around 250 and 304 nm (Fig. 3.4b). On the other hand, the asynchronous correlation highlighted some minor features changing with oxidation that were not easily observable in the synchronous analysis since masked by the very high peaks. Small cross-peaks centered at 278 nm revealed that an increase in absorbance around this wavelength happened predominantly earlier on the doses scale compared with the most intense bands. The increasing trend of the absorbance value at 278 nm did not show a second larger slope at higher oxidant doses as the 250 and 304 bands, indeed (Fig. 3.3a).

The 2D correlation analysis between SAXS, fluorescence and UV-vis data confirmed the strong asynchronicity among the main variations observed through the different techniques. Apart from the obvious early depletion of fluorescence emission compared to the SAXS transition, the hetero-correlation showed also how the strongest increase in UV absorbance observed in the highly oxidized samples took place across the hypochlorite/protein ratio range after the major change in the HSA scattering pattern had already happened (Fig. 3.4c). This evidenced that the hypochlorite-induced oxidation above a critical threshold (oxidant/protein ratio > 120) triggered the formation of further UV-absorbing species, even if no further significant evolution of the scattering was observed.

In order to assess potential correlation in the earlier oxidation-induced variation of the UV-vis spectra and SAXS profiles, the analysis was applied to the restricted subset of data obtained excluding the oxidant/protein ratios ≥ 130 and the results were compared to the case of the full set.

The UV-vis/UV-vis correlation (Fig. 3.4e) showed roughly the same bands in the synchronous spectrum, but with overall lower values (since less absolute variation occurred in the restricted range); in the asynchronous spectrum, bands with the same sign pattern were observed, but with notable broadening. In addition, the wavelength range in which the main increase in absorbance happened before the major changes at 250 and 304 nm was located around 268 nm and not 278 nm as when considering the full oxidant ratio range.

This analysis suggests a complex underlying population of possible UV-absorbing species formed upon oxidation, and the existence of different oxidation “regimes” in terms of UV-absorbing species produced by increasing hypochlorite molar equivalents. This scenario, although foreseeable, was not immediately perceived by the inspection and comparison of the measured spectra.

Considering the restricted data set, also the SAXS/UV-vis hetero-correlation showed roughly the same peaks in the synchronous spectrum, but presented noticeably inverted sign in the asynchronous (Fig. 3.4f). So, while including the most oxidized samples the conclusion was that the main UV-absorbance variation hap-

pened after the main SAXS variations, restricting to the range of oxidant/protein ratio between 0 and 120 the conclusions would be: i) the increase in absorbance at 304 nm determined no visible peak in the asynchronous spectrum, probably being too low; ii) the main spectral variation in this oxidation range (around 250 nm) happened mainly before the SAXS transition. So, around 250 nm species formed by oxidation mainly absorbed, which had no immediate consequences on the protein conformation.

3.4 Time evolution of the oxidized protein spectra

The SAXS data collected on the same sample at short times (between 5 and 15 minutes) from the oxidant addition did not show any variation, (Fig. 3.5b) demonstrating that the structural transformation induced by oxidation was fast and that, once in the highly oxidized state, albumin retained the open conformation and was not subject to aggregation or further degradation in this timeframe. Measurements performed at longer times (>24 hours) showed a slight decrease of the scattered intensity for the highly oxidized samples demonstrating a possible partial fragmentation of the protein [171]. The phenomenon was modest and did not significantly affect the SAXS curve and $P(R)$ shapes (Fig. 3.6).

On the other hand, the UV-vis spectra of the samples oxidized with an oxidant/protein ratio higher than 120 showed a clear and relatively fast evolution in the time of the SURF experiment (Fig. 3.5a). The main features observed to vary in time were a decrease in absorbance at 304 and 250 nm, together with an increase of the signal at 278 nm and in a broader spectral range around 360 nm. The spectra of the less oxidized samples showed only minor variations on a short time-scale (Fig. 3.5c), but enough to appreciate how the direction was inverted compared to the higher ratios: for $[\text{NaOCl}]/[\text{HSA}] \leq 120$ a slight increase instead of a decrease was observed in the absorbance at 250 and 304 nm. These observations were confirmed by kinetics measurements of the UV-vis absorbance performed offline for longer times (Fig. 3.7), which allowed also for a better assessment of the spectral variations slowly undergoing in the less oxidized protein. Since the time evolution of the UV-spectra of an oxidized sample was observed even after exchanging the buffer in order to remove any residual of the added oxidant after the initial reaction, these spectral variations suggest the presence of secondary reactions involving the protein species themselves (Fig. 3.7d).

The most relevant observation from the UV-vis kinetics was the presence of a critical behavior across the oxidant/protein ratio range (Fig. 3.7e and Fig. 3.5c), with a faster and characteristic time-evolution of the spectra for the highly oxidized samples. The boundary between the most recognizable regimes was located at the same hypochlorite doses as the change of slope in the absolute absorbance increasing trend (Fig. 3.3) and, importantly, as the plateau in the structural transition revealed by SAXS (Fig. 3.1).

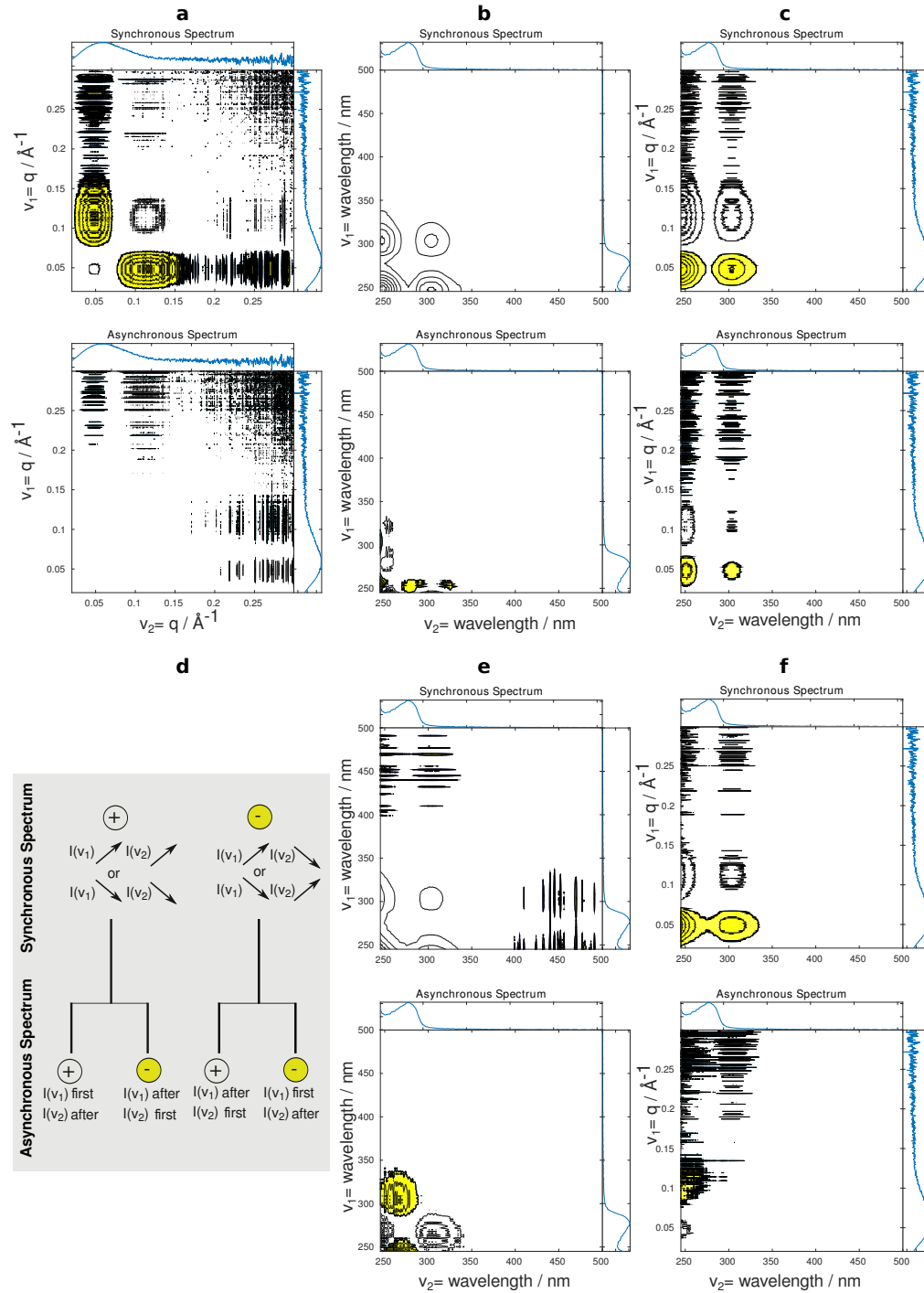


Figure 3.4. 2D-correlation spectra from the analysis of SAXS and UV-vis data on increasingly oxidized HSA. (a) homo-correlation of the SAXS data in the form of Kratky plots (b) homo-correlation of UV-vis spectra (c) hetero-correlation of SAXS data (Kratky plots) and UV-vis spectra. The analysis shown in (b) and (c) is compared to that obtained from a subset of the data restricted to oxidant/molar ratios lower than 120, in (e) and (f) respectively. (d) A guide to the interpretation of the relative signs of peaks found in synchronous and asynchronous correlation spectra is also presented. More explanations are given in section 1.3.

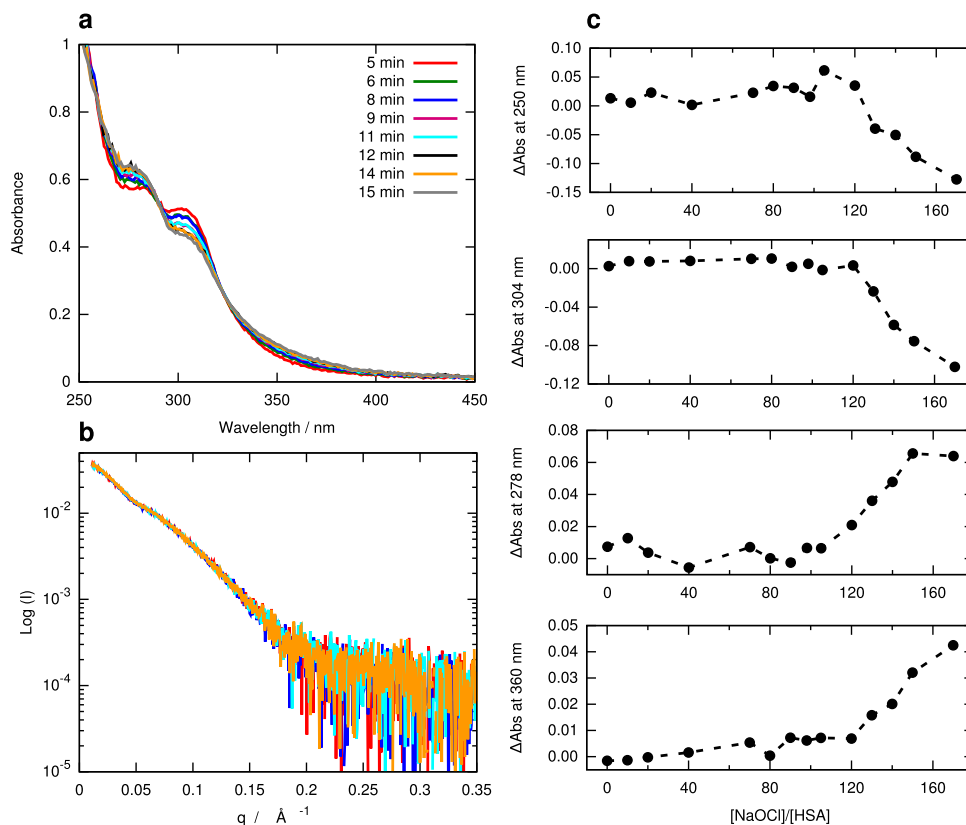


Figure 3.5. A time evolution of the UV spectra collected with the SURF platform between 5 and 15 minutes after addition of the oxidant to HSA is observed for the highly oxidized samples. The experimental data for the sample with oxidant/protein ratio = 150 are shown: (a) UV-vis absorbance, (b) SAXS intensity. The color code for the time elapsed from the oxidant addition is: 5 minutes (red), 6 minutes (green), 8 minutes (blue), 9 minutes (purple), 11 minutes (cyan), 12 minutes (black), 14 minutes (orange), 15 minutes (gray). No variation in the SAXS data is observed, while a clear time evolution of UV absorbance is detected in the eight subsequent exposures. In (c) the absorbance difference at four relevant wavelength values (250 nm, 304 nm, 278 nm and 360 nm) between the last and the first UV-vis spectra collected across the 15-minute timeframe (Δ Abs) is shown as a function of the oxidant/protein ratio ($[\text{NaOCl}]/[\text{HSA}]$).

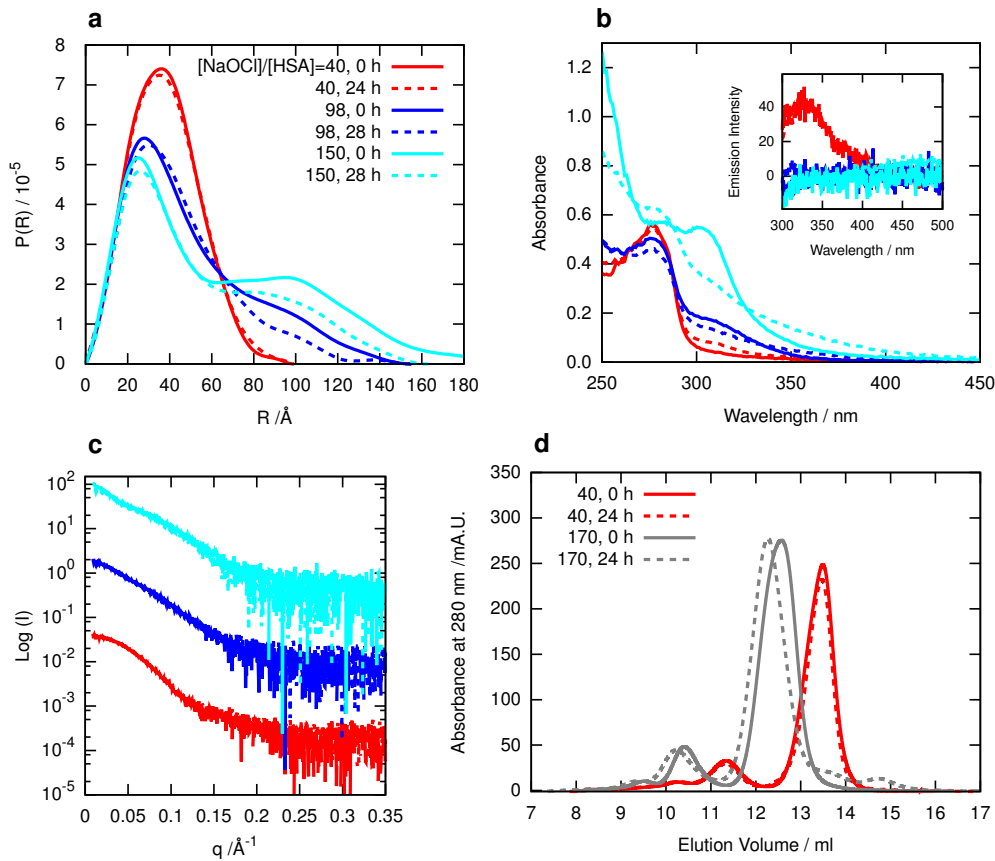


Figure 3.6. Long-time evolution of oxidized HSA. The SURF data collected after 5 minutes (solid lines) and after > 24 hours (dashed lines) since the addition of the oxidant for the samples with oxidant/protein ratio = 40 (red), 98 (blue) and 150 (cyan) are shown: (a) P(R) functions, (b) UV-vis absorption spectra and fluorescence emission in the inset, (c) SAXS profiles. In (d) the SEC traces recorded by the UV-280 nm detector of the SEC apparatus from an independent experiment are shown. The elution profile of the oxidized samples injected 5 minutes (solid lines) and 24 hours (dashed lines) since the starting of the oxidation are compared for two samples with oxidant/protein ratio = 40 (red) and 170 (gray). The elution buffer contained 0.15 M NaCl.

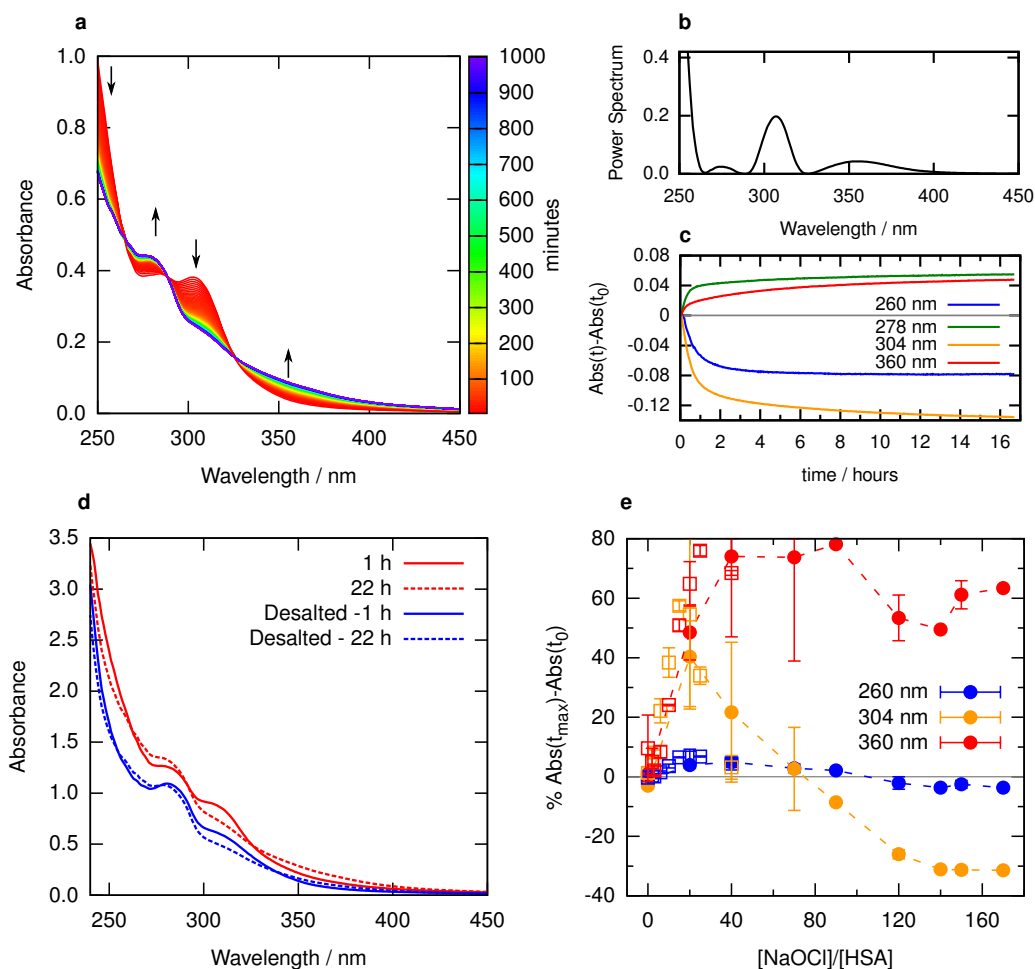


Figure 3.7. *UV-vis kinetics analysis of oxidized HSA.* (a) Evolution of the absorption spectra of a sample with $[\text{NaOCl}]/[\text{HSA}] = 170$ between 4 minutes and 16.5 hours since the addition of the oxidant. (b) Power spectrum of the kinetics series shown in (a) which highlights the most varying features. (c) Absorbance values at selected wavelengths as a function of time, from the experiment in (a). (d) Of two samples oxidized with $[\text{NaOCl}]/[\text{HSA}] = 170$, only one underwent a buffer-exchange as mentioned in section 3.2.4. The spectra collected 1 hour and 22 hours after the addition of the oxidant are compared. A similar time evolution occurred in both cases irrespective of the desalting procedure. (e) Summary of the time evolution of the UV-vis spectra observed in HSA samples at increasing oxidant doses. In the kinetics experiments the first spectrum was collected 20 minutes after the addition of the oxidant and the last 17 hours after. The percentage of absorbance variation between the last and the first spectra at selected wavelengths ($\%[\text{Abs}(t_{\text{max}}) - \text{Abs}(t_0)] = [(\text{Abs}(t_{\text{max}}) - \text{Abs}(t_0))/\text{Abs}(t_0)] \cdot 100$) is shown as a function of $[\text{NaOCl}]/[\text{HSA}]$. The dots with error bars represent the average of two replicates. The squares represent additional experiments in which only the low-doses range was explored.

3.5 SEC-SAXS experiments: structural models

The SEC analysis of the reference non-oxidized sample showed as expected [39,172] a highly monomeric protein solution with a small contribution of oligomers. We observed that all the oxidized samples preserved the same monomer/oligomer relative populations as the native protein solution (Fig. 3.8a). The oxidized samples showed a peak broadening and a systematic shift of the elution profile towards lower elution volumes (Fig. 3.8b). The drift towards shorter retention times could be due to a combination of effects: both the increase of the hydrodynamic radius and the decrease of some affinity or increase of repulsive interactions of the protein towards the stationary-phase in the oxidized protein forms. When the ionic strength of the elution buffer was increased through the addition of NaCl 150 mM, the shift was significantly reduced (gray squares in Fig. 3.8b), suggesting that the interactions were largely dictated by an electrostatic contribution. Some changes in the hydrophobicity of the protein surface due to oxidation possibly affecting protein-column interactions could also not be excluded to contribute [125]. A variation of the monomer shape in the highly oxidized states shown by the scattering data should also play a role.

The structural parameters obtained by the SAXS analysis of the monomer fraction had the same critical behavior as a function of the oxidant doses as in the experiment without separation, showing just a slight downward shift (Fig. 3.1). This suggested that the contribution of oligomers was rather negligible in the process. For the highest hypochlorite doses (OPR > 120), the UV-vis spectra of the main peak of the SEC-eluted oxidized HSA showed less intense absorbance at 304 nm and 250 nm than the unseparated samples (Fig. 3.3). Considering that the UV-vis absorbance of these bands decreased rapidly for the protein at the high oxidation level, the evolution of the sample in the time elapsed between the injection and the elution of the main peak (45-70 minutes depending on the oxidant/protein ratio) could explain this lower absorbance.

Since the scattering data came from a relatively homogeneous population of protein species, at least in the sense they behave as a defined though broad SEC peak, they could be employed to recover low-resolution dummy-residue models representative of the average conformation assumed by the oxidized monomer species of HSA in solution.

The oxidation-induced conformational change could be well depicted by comparing the most representative models for each oxidant/protein ratio. In contrast with the heart shape typical of the native form at low doses (oxidant/protein ratio <40), at high oxidant/protein ratio values (≥ 120) the models presented a pronounced protuberance (Fig. 3.9). In the latter case, the average number of residues found in the extended lobe allowed us to roughly estimate that it was equivalent in size to one of the end-domains of HSA. An alternative modeling approach which assumes the known domain-structure of HSA and was used in the past to model the HSA unfolding in urea [79,117], was then tried to fit the same scattering data. The intact HSA domains II and III and the three loops of domain I were used as structural units interconnected by flexible linkers. Repeated optimizations converged to the same spatial distribution of domains, and supported the fact that a detachment of one of the end-domains of HSA induced by the increasing amount of chemical damage was compatible with the experimental data (Fig. 3.10).

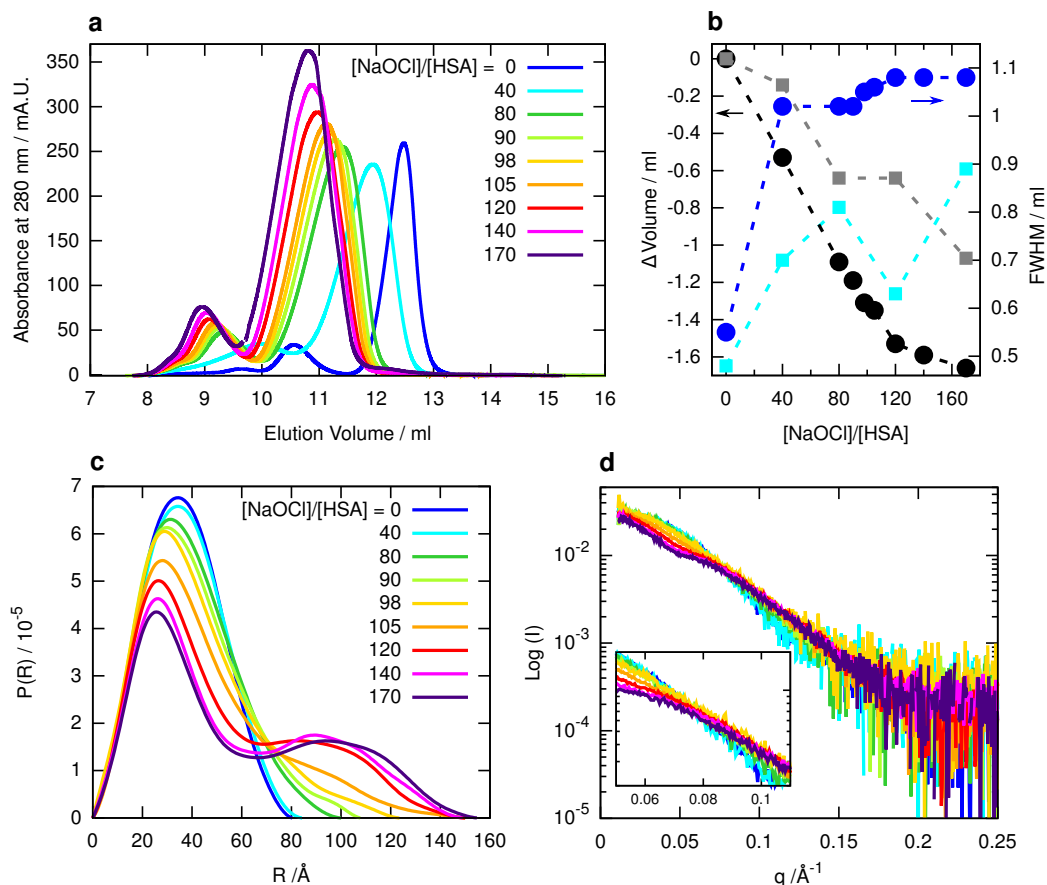


Figure 3.8. SEC-SAXS analysis of increasingly oxidized HSA samples. (a) Chromatogram traces recorded by the UV-280 nm detector of the SEC apparatus during the elution of oxidized HSA samples with increasing oxidant/protein ratios. (b) Relative decrease of the elution volume compared with the reference sample (ΔVolume , ordinate axis to the left, black dots) and increase of full width at half maximum (FWHM, to the right, blue dots) of the main elution peak. The increase of the ionic strength of the elution buffer (containing 0.15-M NaCl) determined smaller volume shifts (gray squares) and peak broadening (cyan squares). (c) Pair distance distribution functions derived from the SEC-SAXS data collected at the monomer peak maximum for increasingly oxidized HSA. (d) SAXS profiles collected at the monomer peak maximum.

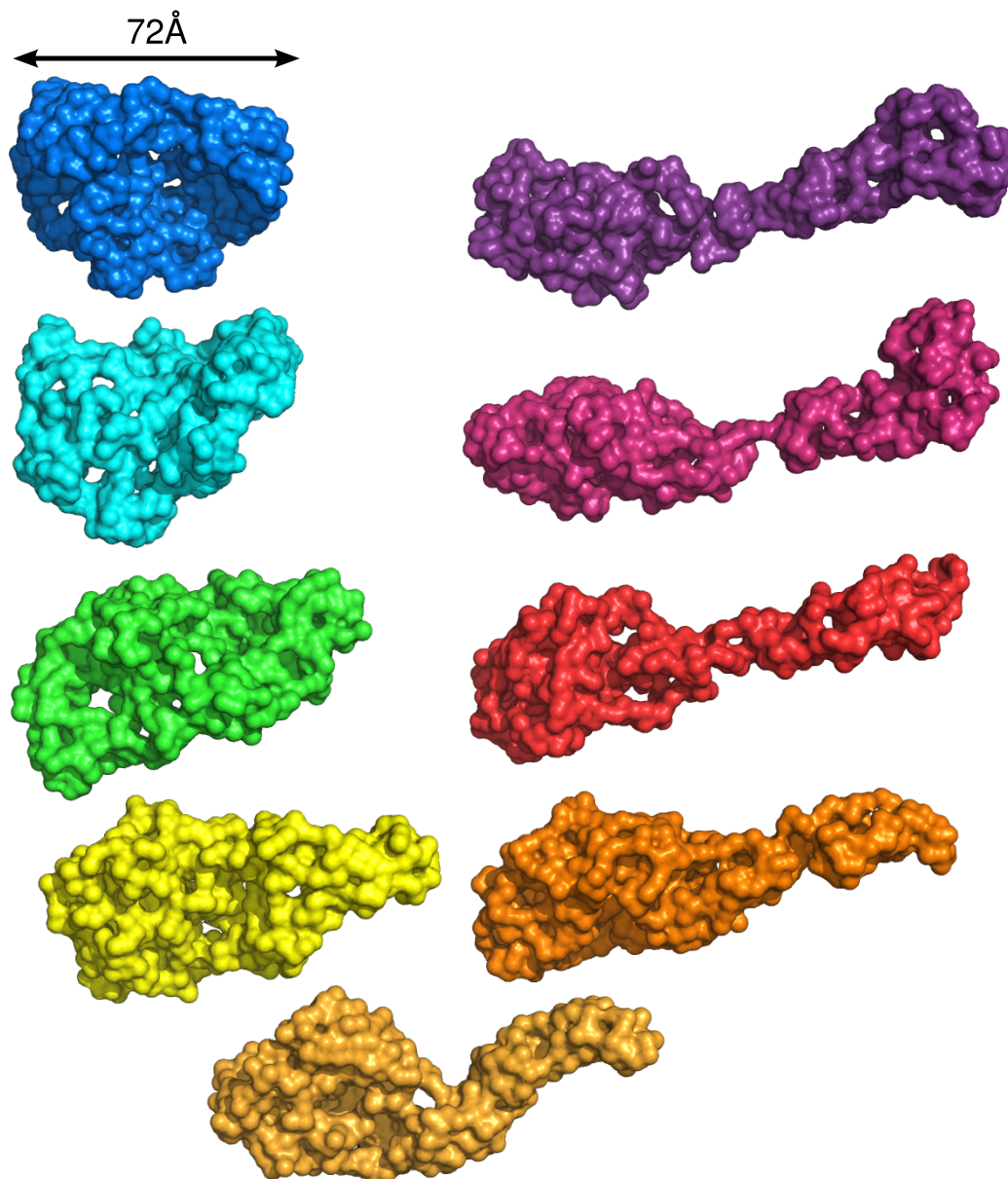


Figure 3.9. *Dummy-residue models of the increasingly oxidized HSA monomer derived from SAXS data fitting.* Representative dummy-atom models fitting SEC-SAXS data (from the maximum of the main elution peak) of oxidized HSA samples up to 0.35 Å^{-1} . Going from the top left to the top right the experimental $[\text{NaOCl}]/[\text{HSA}]$ ratios were: 0 (blue), 40 (cyan), 80 (green), 90 (yellow), 98 (light orange), 105 (orange), 120 (red), 140 (purple), 170 (violet). A dimensional bar corresponding to 72 Å is shown as a reference.

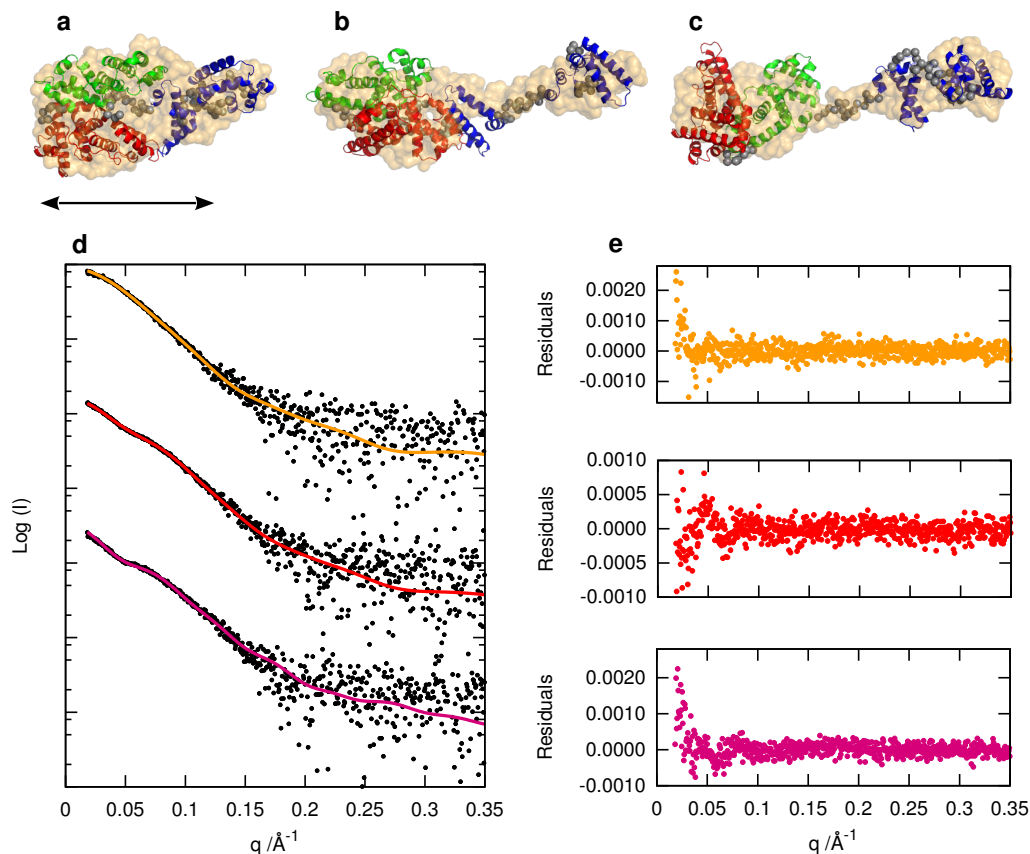


Figure 3.10. Results of the combined rigid-body/flexible-linkers modeling of oxidized HSA against the experimental SEC-SAXS data. The representative models fitting the SEC-SAXS data collected at the main elution peak of oxidized HSA samples with oxidant/protein ratio = 90 (a), 120 (b) and 170 (c) are shown. The three HSA domains I, II and III are colored in blue, green and red respectively and represented as cartoons. The most representative *ab-initio* model (Fig. 3.9) (bright orange surface) obtained for the same experimental data is superimposed for comparison. A dimensional bar corresponding to 72 Å is shown as a reference in (a). In (d) the theoretical scattering intensity for the models corresponding to oxidant/protein ratio = 90 (orange line), 120 (red line), 170 (purple line) is shown together with the experimental data (black dots). In (e) the residuals (difference between model and experimental intensities) are displayed.

The low-resolution models of the protein monomer recovered from the SAXS data for intermediate oxidant/protein ratio values provided intermediate structures showing a progressive extension of a lobe from a globular moiety upon increasing oxidation. A tentative analysis of the structural components possibly involved in the conformational transition is reported in a subsequent section (3.7).

3.5.1 Oxidation of HSA oligomers

Similar UV-vis spectral features of the oxidized protein were present both in the main monomer peak and in the minor peak of oligomers (Fig. 3.11), indicating that the oligomeric forms present in the samples underwent chemical damage to a similar extent. Even if notably noisier, the SAXS data collected in correspondence with the secondary maximum observed at lower elution volumes than the monomer clearly showed that the average shape of the oligomeric species (mainly dimers) was also distinctly different in the oxidized protein as compared to the native form (Fig. 3.12). These findings suggested that the chemical modification did not induce more protein oligomerization but rather triggered the structural transition in the already existing oligomers. The characteristic bimodal $P(R)$ could be interpreted as indicative of dimers formed by protein molecules in the elongated form seen for the highly oxidized monomers.

3.6 Zeta potential and circular dichroism results

Complementary experiments for the characterization of the secondary structure composition and average surface charge of the oxidized protein conformers were performed and the results are presented in Fig. 3.13.

The far-UV CD signal mostly due to the α -helical secondary structure of HSA progressively decreased with the increase of the oxidative damage without showing a sigmoidal variation resembling the low-resolution structural parameters (Fig. 3.13a). The CD-based estimation of the secondary structure composition implied that the full oxidation-induced denaturation caused a loss of about 35% of the native α -helical folding with advantage of the content of disordered structure. The zeta potential measurements pointed out that the hypochlorite-induced chemical modification caused an increase of the protein negative charge as soon as the doses reached a critical level (Fig. 3.13b). This threshold could be estimated around 80 in the oxidant/protein ratio scale, similarly to the SAXS dimensional parameter variation. The loss of roughly one third of the α -helical secondary structure could be probably attributable to the unfolding of the helices in the inter-domain regions and in one of the three main lobes of the protein, which drifted apart. The increase of negative charge accompanying the hypochlorite-induced damage would be the cause of relevant intramolecular repulsive interactions within the protein. The partial disruption of helical structure and the structural rearrangement towards an elongated form could be determined by this electrostatic effect. The growth of protein negative charge with oxidation would also be in agreement with the observations made about the SEC elution of the oxidized protein (section 3.5).

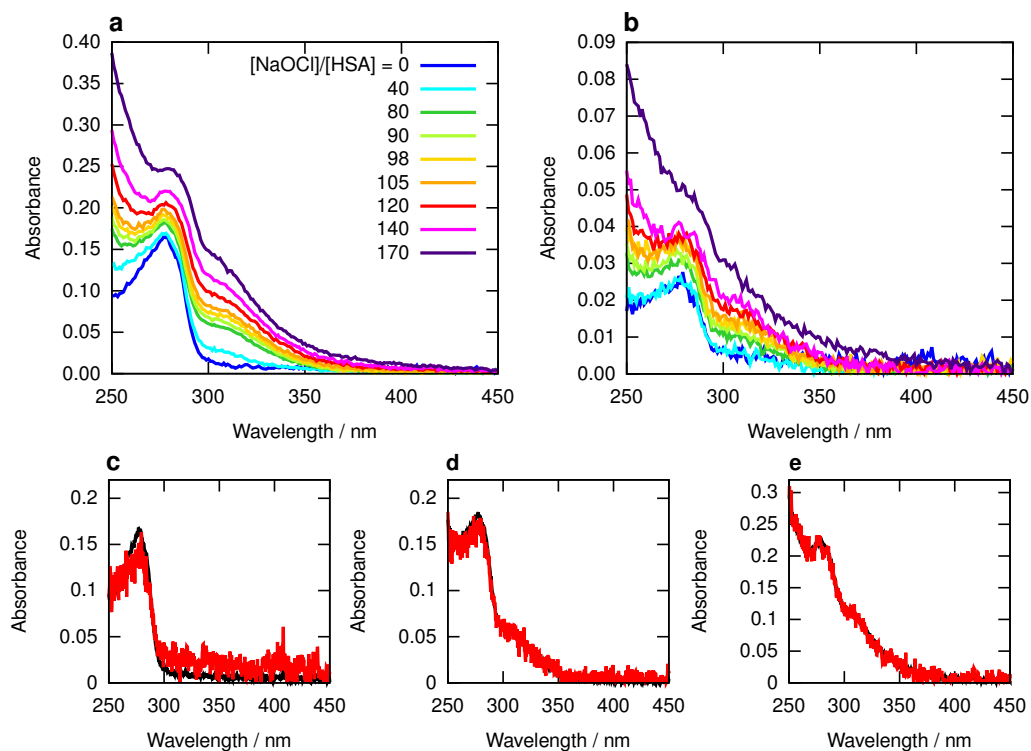


Figure 3.11. Comparison between the UV absorbance spectra of monomeric and oligomeric oxidized HSA collected in the SEC elution. The absorbance spectra at the maximum of the peaks of the monomer (a) and oligomer (b) for all oxidant/protein ratios are displayed. The UV-vis absorbance spectra corresponding to the maximum of the monomer (black line) and the oligomer (red line) peaks of samples with oxidant/protein ratio = 0 (c), 80 (d) and 140 (e) are superimposed in the lower panels. In these superimpositions the spectra collected at the oligomer peak were multiplied by a factor 5.5 to scale them to the monomer peak higher concentration.

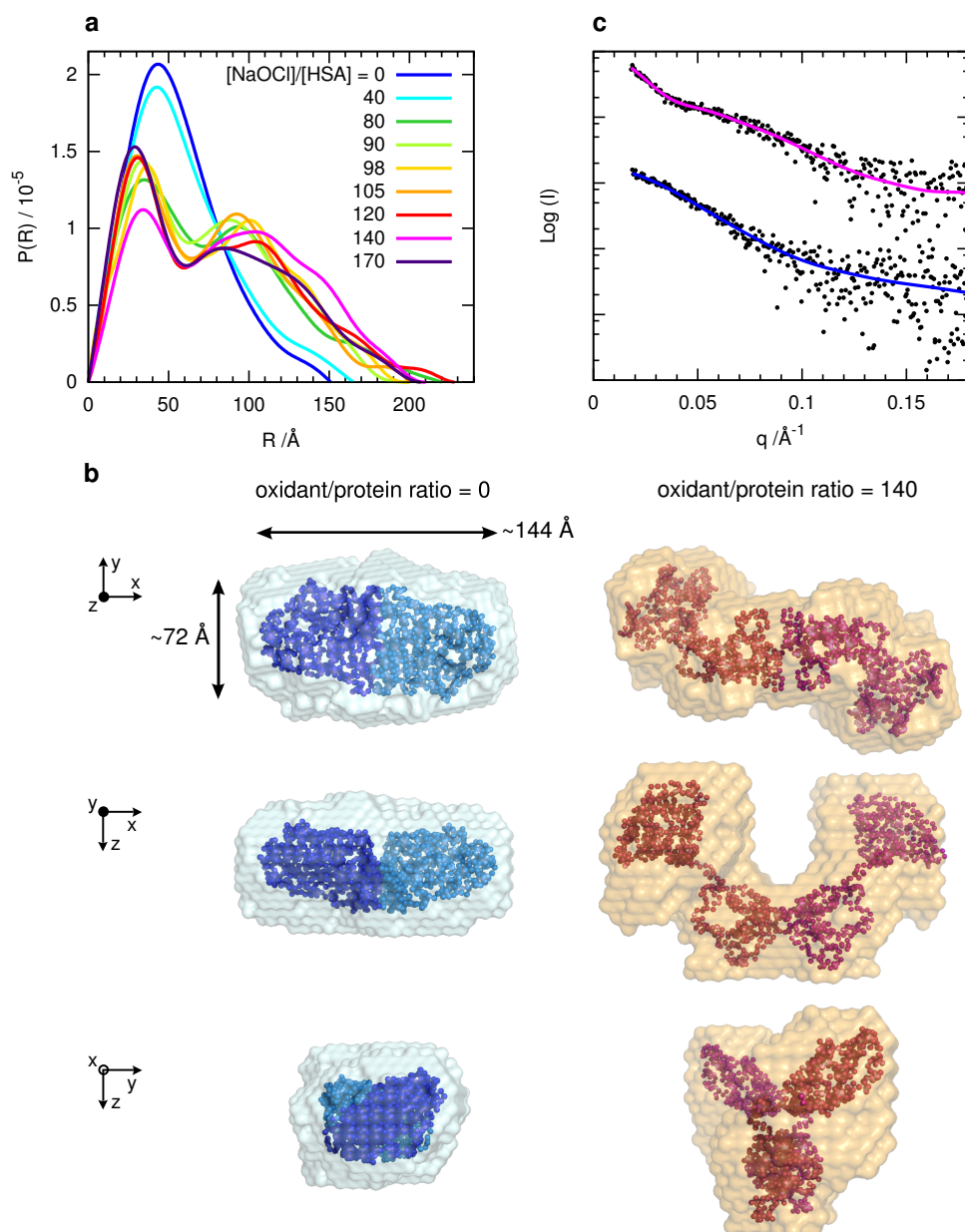


Figure 3.12. Structural information for the HSA dimers in both native and oxidized form derived from the SEC-SAXS data collected at the maximum of the oligomer elution peak. (a) $P(R)$ functions for all the oxidant/protein ratio values analyzed. (b) Ab-initio dummy-residue models for the samples with oxidant/protein ratio = 0 (reference, left) and 140 (oxidized, right). The most representative models (spheres) are shown superimposed to the overall probability maps from repeated calculations (transparent surfaces). Three orthogonal views are displayed. (c) The theoretical scattering intensity for the models corresponding to oxidant/protein ratio = 0 (blue line), 140 (purple line) is shown together with the experimental data (black dots).

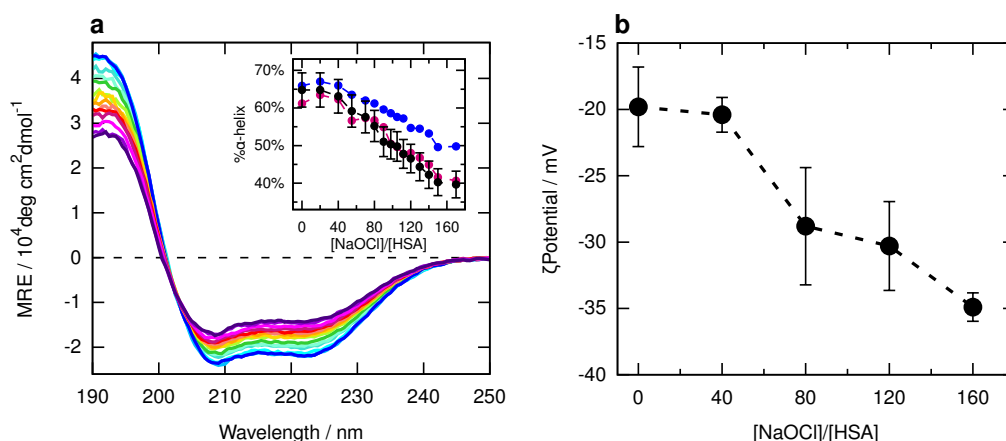


Figure 3.13. Results of experiments aimed at assessing the protein secondary structure and surface charge variation induced by the oxidation process. (a) Far-UV CD spectra of oxidized HSA samples with oxidant/protein ratios ($[\text{NaOCl}]/[\text{HSA}]$): 0, 20, 40, 55, 70, 80, 90, 98, 105, 112, 120, 130, 140, 150, 170 (line colors from blue to purple). In the inset, decrease of the α -helical content estimated from the CD spectra as a function of the oxidant/protein ratio. The percentage of α -helical structure estimated by means of the Chen equation [167] (black dots), the BeStSel web server [168] (purple dots) and the CONTIN/LL algorithm of the CDpro package [169] (blue dots) are displayed. (b) Decrease of the zeta potential measured on oxidized HSA samples.

3.7 Application of component analysis to the observed oxidation-dependent structural transition

From the experimental SAXS data obtained in correspondence of the main elution peak of the increasingly oxidized protein samples, two clearly distinguishable conformations could be pointed out, one typical of the native form (besides, stable even when the protein is oxidized with 40 molar equivalents of HOCl) and the other one, markedly elongated and with bimodal $P(R)$, typical of samples treated with an oxidant/protein ratio higher than a critical level (≥ 120). The correct way to describe the systems obtained by treating the protein with oxidant amounts intermediate between these two extremes could be debated.

The “cooperative” increase of the protein dimensional parameters observed upon increasing oxidative modification (sigmoidal shape of the experimental data) could resemble what is commonly described as an “all-or-none” transition. The presence of an isosbestic point in the evolution of the SAXS curves (Fig. 3.8d) could further support this hypothesis and in the light of this we tried to describe the SAXS profiles as a linear combination of scattering components. Indeed, according to Dill and colleagues [173], the possibility to model the changes through the transition midpoint as a simple linear combination of the signals from the two limiting states would be an indication of two-state behavior, whereas if the signals near the transition midpoint could not be fitted as a simple linear combination of the two limiting state signals, that would be evidence for one-state behavior. In this second case a significant population of “intermediate” molecules might contribute to the total signal. We considered reasonable to ask if the SAXS data collected for oxidant/protein

ratios half-way between the lower and upper extremes in terms of the observed particle dimensions could be described as a linear combination of the native and the fully elongated form, pointing out the possible underlying existence of two major conformations despite the step-by-step and irreversible nature of the dose-dependent oxidative modifications induced by increasing equivalents of the reactant. In order to assess this point, first we applied PCA in order to check how many independent components were needed to account for the experimentally observed variation of the scattering curves collected in correspondence of the main elution peak of the increasingly oxidized protein samples. The conclusion was that, even if the first two independent components were sufficient to explain at least 94% of the observed variance and to provide an acceptable fit (by visual inspection) to the experimental data (Fig. 3.14a,b), a significantly non-random third eigenvector was found to contribute (Fig. 3.14c). Interestingly, restricting the data analyzed to the oxidant/protein ratio range between 80 and 170, so excluding the native protein, a third component did not play a significant role anymore (Fig. 3.14c).

In the attempt to apply the MCR-ALS method to find three scattering components and the associated concentration profiles best fitting the data, the optimization calculated an intermediate component reaching 100% for the oxidant/protein ratio 90 (Fig. 3.15e,f), which together with the native and fully oxidized components could account for most of the variation observed, providing almost random residuals. A simpler approach involved the use of selected experimental scattering profiles as components, and the outcome was quite similar to the MCR-ALS optimization. We tried, first, to fit the data collected across the whole oxidant/protein molar ratio range as the linear combination of the two extreme samples ($[\text{NaOCl}]/[\text{HSA}] = 0$ and 170) only (Fig. 3.15a). Then, we tried to include as a third component one of the samples with intermediate oxidant/protein molar ratio, so dividing the experimental range in two parts, each describable with two components. The result was that the agreement between experimental and modeled data could be improved by including the sample with $[\text{NaOCl}]/[\text{HSA}] = 105$ (lowest average χ^2) or 90 (in agreement with the MCR-ALS result) as an intermediate form, providing a concentration profile of the three conformational states as a function of the oxidative damage undergone by the protein that could have a physical meaning (Fig. 3.15b,c,d).

Based on statistical considerations, no definitive choice between the suitability of a two-state model compared to a more gradual conformational variation involving intermediate forms could be made from this attempts to fit the scattering data by using models of fixed or optimized components. As it is often the case for data like scattering curves lacking clearly distinguished features, a two-state model could explain great part of the variance observed, but intermediate states seemed to be necessary for the more accurate reproduction of the most varying q range in the scattering data collected for the protein in all the experimental conditions. It was suggested [173] that the most definitive determination of two-state behavior would be the experimental observation of two distinct populations near the midpoint of the transition and that, for slowly exchanging systems, this could be done by transport methods such as SEC.

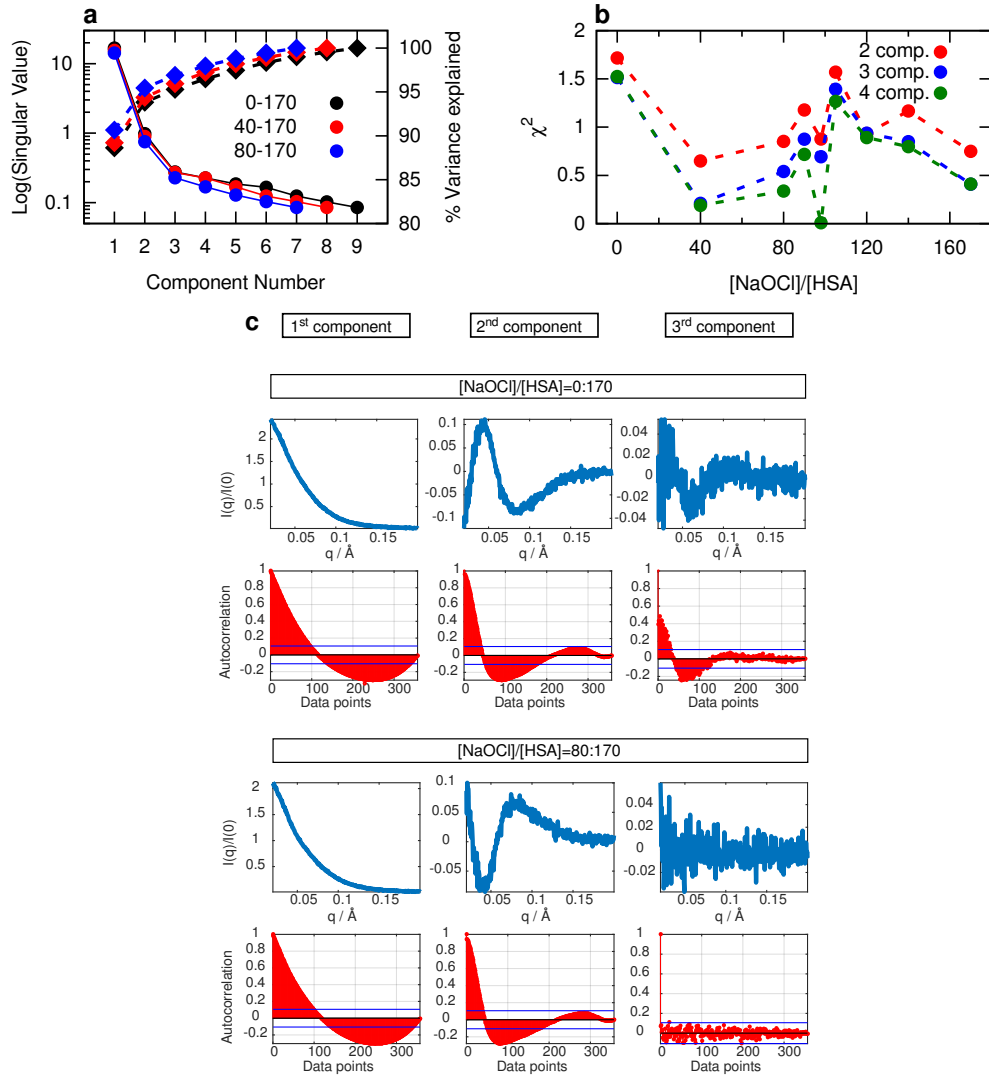


Figure 3.14. Results of the application of the PCA analysis to the SAXS data set from the SEC experiment on increasingly oxidized HSA samples. (a) Singular values on a logarithmic scale (left axis, dots) and percentage of the total variance explained by including each additional principal component (right axis, diamonds). The results obtained with the complete data set (oxidant/protein ratio $[\text{NaOCl}]/[\text{HSA}] = 0, 40, 80, 90, 98, 105, 120, 140, 170$, black) are compared to those obtained when excluding the $[\text{NaOCl}]/[\text{HSA}]=0$ sample (red) and the $[\text{NaOCl}]/[\text{HSA}]=0,40$ samples (blue). (b) Goodness of fit (χ^2 value) given by the PCA approximation including the first 2, 3, or 4 components for the SAXS data at the different oxidant/protein ratio. In (c) the first three eigenvectors (multiplied by the corresponding singular value) resulting from the PCA of the SEC-SAXS data of oxidized HSA in the $[\text{NaOCl}]/[\text{HSA}]$ ranges 0-170 (top panel) and 80-170 (bottom panel) are displayed as blue solid lines. In each case the autocorrelation function is displayed below the corresponding eigenvector and helps in assessing whether it represents only noise or is effectively important to reproduce varying features in the spectral series.

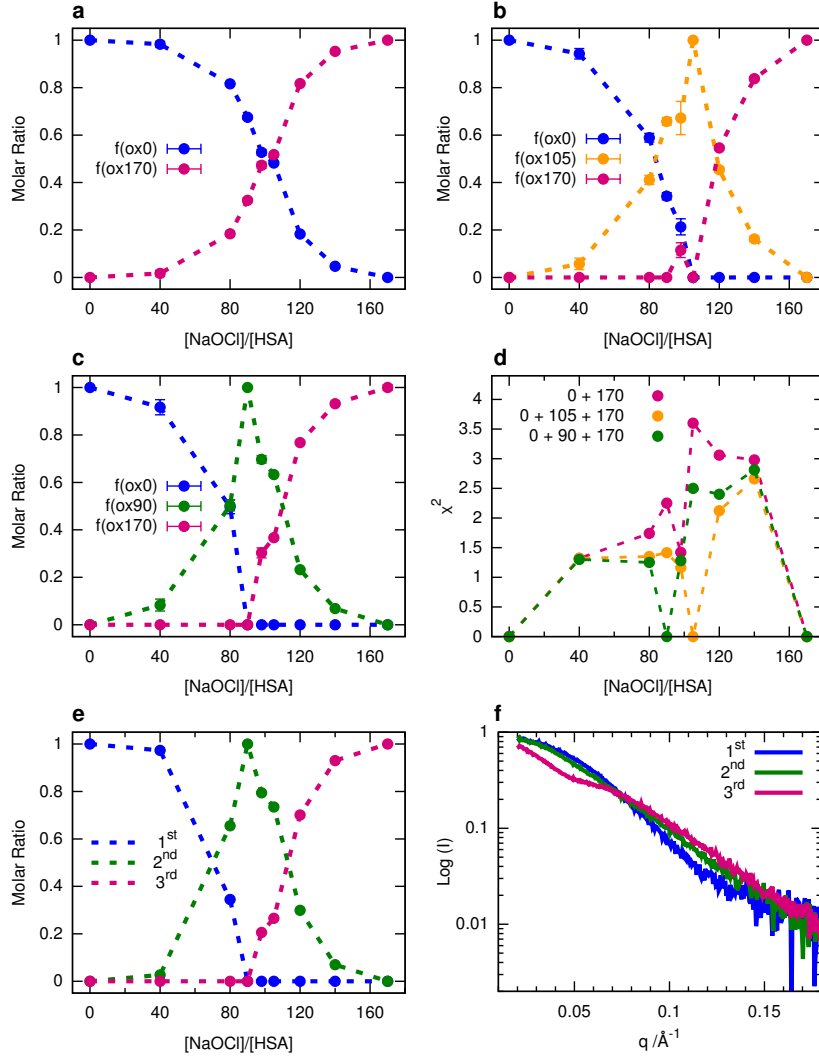


Figure 3.15. Results of the optimization of the SEC-SAXS dataset of increasingly oxidized HSA as a linear combination of components. In (a), (b), (c), (d) the results of the modeling by linear combination of selected experimental data are shown. In (a), (b), (c) the optimized molar fractions of each component are shown as a function of the oxidant/protein ratio. In (a) the two components were chosen as $[\text{NaOCl}]/[\text{HSA}] = 0$ and 170 (“ox0”, blue dots, and “ox170”, purple dots); in (b) three components were chosen as $[\text{NaOCl}]/[\text{HSA}] = 0$, 170 and 105 as intermediate state (“ox0”, blue dots, “ox170”, purple dots, and “ox105”, orange dots); in (c) three components were chosen as $[\text{NaOCl}]/[\text{HSA}] = 0$, 170 and 90 as intermediate state (“ox0”, blue dots, “ox170”, purple dots, and “ox90”, green dots). In (d) the χ^2 calculated between experimental and reconstructed data in each of the cases explained above is shown for the samples at different oxidant/protein ratio. It becomes 0 when the sample coincides with one of the components. In (e) and (f) the results of the MCR-ALS optimization with three components are summarized. In (e) the concentration profiles of the three components across the oxidant/protein ratio range are shown. The three optimized scattering components are displayed in (f).

Conformational heterogeneity in the intermediately oxidized samples A deeper exam of the protein conformation distribution could indeed be performed looking to the $P(R)$ functions obtained from data collected at different elution times within the SEC peak related to the monomer (Fig. 3.16).

Relatively well defined conformations were found in the native (oxidant/protein ratio=0) and fully oxidized (oxidant/protein ratio=140) protein samples. The reference sample provided the typical bell shape of native HSA at all the elution volumes (Fig. 3.16a) whereas the same characteristic bimodal shape of the $P(R)$ was found throughout the eluted peak of the highly oxidized sample at OPR=140 (Fig. 3.16d). In the case of two intermediately oxidized samples (Fig. 3.16b,c) a systematic variation towards more elongated forms was detected going from larger to smaller elution volumes (decreasing frame number). For the sample with oxidant/protein ratio=90 (Fig. 3.16b), the conformation with the smallest $D_{\max}$ could resemble the native form, while the form with larger $D_{\max}$ could be comparable with that present at the peak maximum of the sample with oxidant/protein ratio=105. On the lower volume side of this peak (Fig. 3.16c), a gradual appearance of the characteristic bimodal $P(R)$ could be associated to the dumbbell shaped conformer found in the more oxidized samples as a minor contribution. By an average chemical point of view, the sampling of the UV-vis spectra showed no clear dishomogeneity in the oxidation level throughout the peak for each oxidant/protein ratio value. However, it is obvious that in these experiments we did not probe the chemical modification at a sequence-specific detail.

In our experiment, we observed a notable peak broadening for the oxidized samples but clearly distinct populations could not be resolved. The variation of the $P(R)$ functions throughout the peak of the intermediately oxidized samples could be explained by both an unresolved mixture of two main distinct conformations and more complex mixtures of a larger number of intermediates. The coexistence of multiple conformational states could be intrinsic to the modified protein or, in a less speculative view, could depend on the specific nature and location of the modifications the protein underwent which could be not homogeneous. A possible interpretation is suggested in section 3.11.

3.8 Investigation of the effect of ligands

In addition to the ligand-free HSA, we investigated the effects of hypochlorite-induced damage in HSA bound to palmitic acid (HSAPalm) and HSA in the presence of the polyphenolic compounds quercetin (HSAQuerc) and curcumin (HSACurc).

Quercetin and curcumin are known as dietary antioxidant molecules belonging to the generic polyphenols family [174,175] and it is likely that they interact with HSA also *in-vivo* due to their scarce solubility at the physiological pH of plasma. The mechanism at the basis of the antioxidant properties of polyphenols could be simply deriving from the possibility to preferentially scavenge reactive species by sacrificial reaction. The two molecules considered have different properties in this context [176,177], reflecting their different molecular nature (Figure 3.17). In addition, the fact that they can interact specifically with albumin as a transporter could offer the opportunity to estimate possible effects originating from the stabilization of the

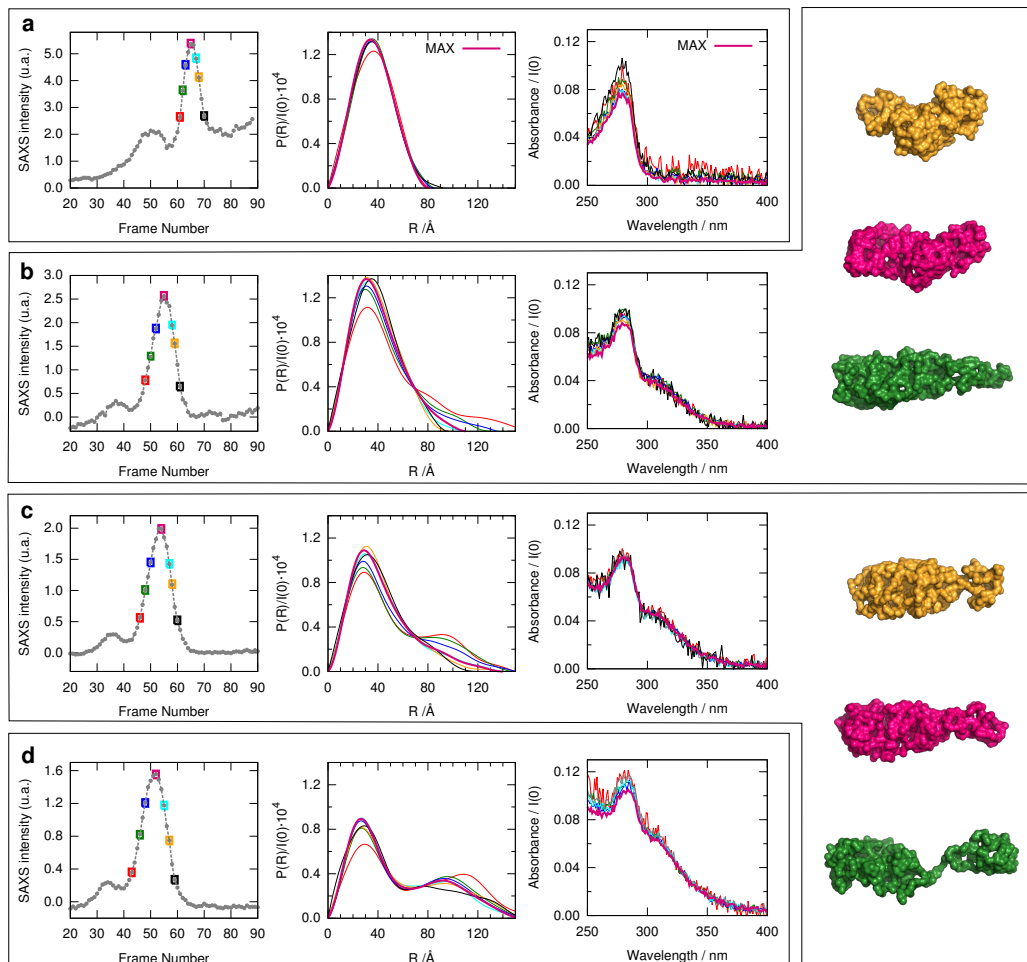


Figure 3.16. The comparison between the HSA monomer shapes derived from SAXS data collected at different points of the main SEC peak reveals different conformers can coexist in the intermediately oxidized samples. The results of this analysis for the samples with oxidant/protein ratio = 0 (reference), (a), 90 (b), 105 (c) and 140 (d) are shown. In each case the chromatogram trace plotted using the buffer-subtracted scattered intensity integrated between 0.04 and 0.1 Å⁻¹ is displayed on the left. The P(R) functions obtained for some of the experimental frames collected across the eluted peak are superimposed in the middle plot. In the third plot the UV spectra collected halfway during the SAXS exposure are also compared. The data are shown as solid lines of the same color as the box highlighting the corresponding frame. Both the P(R) and the UV spectra plots were rescaled according to the I(0) obtained from the P(R). On the right some dummy-residue reconstructions obtained by fitting the SEC-SAXS data of the intermediately oxidized samples in (b) and (c) are shown, with a color code corresponding to the highlighted points in the chromatogram trace.

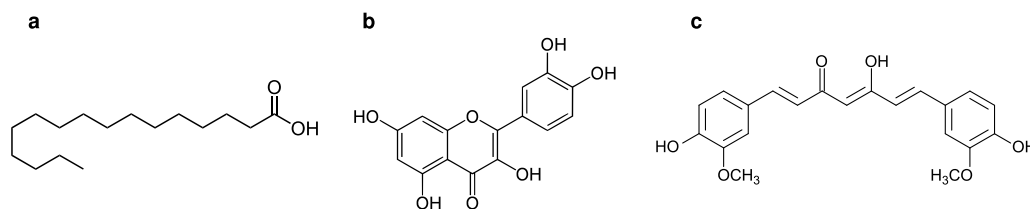


Figure 3.17. Chemical structures of the three molecules whose possible effect on the hypochlorite-induced oxidation of HSA was investigated: a) palmitic acid; b) quercetin; c) curcumin (enolic form).

native protein conformation. In the case of palmitic acid, on the other hand, the idea beyond the comparison was to check the possible protective effect of a physiological ligand known to oppose HSA unfolding [79] and delay thermal denaturation and aggregation [5]. In this case no “sacrificial” reaction with hypochlorite can be foreseen due to the lack of suitable functional groups.

A note about the interaction between HSA and these ligand molecules should be made. Palmitic acid as a fatty acid is a physiological ligand of albumin and the crystal structure of the protein showing how fatty acids bind specifically to seven different sites is known ([178], section 2.3). On the other hand the binding of curcumin and quercetin to HSA was only demonstrated by means of spectroscopic investigations [161,162,179–184] and possibly occurs in one of the two HSA drug binding sites; site I in subdomain IIA (see section 2.3) has been suggested as the one with higher affinity.

The two molecules had already been shown to undergo spectral variations when interacting with HSA, and we also performed a spectral characterization of the systems under study (Fig. 3.18). In the case of quercetin, a red shift of its absorption bands towards the visible region is observed (Fig. 3.18a). In the case of curcumin, a broadening of the band in the visible region occurs (Fig. 3.18b). In both cases a visible change of the color of the solution was observed when the ligand was added to HSA.

Unfortunately, the long-time evolution of the UV-visible spectra of HSAQuerc and HSACurc suggested that the molecules were not fully stable in the solution conditions over the time explored, and the conversion to some degradation products might have occurred [185–187]. The interaction with HSA should have a “preserving” effect according to previous observations [162], but maybe this can be true over shorter time frames. Storage at lower than room temperatures was observed to slow down the absorbance decay. In any case, in the experiments the solutions were freshly prepared in order to be analyzed as quickly as possible. The aspect of the stability over time of the spectral properties investigated is rarely clarified in the literature.

The fluorescence properties of HSAQuerc and HSACurc (Fig. 3.18c,d) involve a quenching effect of the ligands on the protein emission. In addition, the interaction between protein and ligands is further proved by an emission band in the visible range observable through excitation at the protein characteristic wavelength (280 nm) which is not observed for the molecules alone (Fig. 3.18c,d insets). These bands are also observed when directly exciting at a wavelength characteristic of the ligand

absorption and have a much higher intensity in the complex with HSA compared with the molecules alone, especially in the case of quercetin which is essentially non-fluorescent on its own.

In our experiments we noticed that the complex HSAQuerc freshly prepared (without any incubation) did not survive to chromatographic separation in the SEC, having probably quercetin a preferential interaction with the column stationary phase. If the separation was performed after 24-hour incubation at room temperature, on the other hand, the sign of the polyphenol binding in the eluted protein UV-vis spectrum was evident, but in this case the intensity of the characteristic quercetin band in the spectra had strongly decreased as shown above, suggesting a degradation of the quercetin molecule.

3.8.1 Spectroscopic characterization at low oxidant doses

In section 3.3 it was observed how the fluorescence emission of HSA quickly decayed as a function of increasing hypochlorite doses. In Fig. 3.19a the fluorescence spectra of HSA treated with oxidant/protein ratios between 0 and 20 allow for better appreciating the change in shape of the emission spectrum in this lower range of the oxidant doses. This shape variation could be interpreted as a consequence of the change of the relative contribution of tryptophan (only one, more readily reactive towards hypochlorite) and tyrosines (eighteen in total, less affected in this oxidation range).

Comparing the depletion of the fluorescence signal at increasing hypochlorite doses between HSA and HSA in the presence of its ligands, some differences could be appreciated, even if not very large (Fig. 3.19b). The presence of quercetin seemed to delay the drop of the protein emission to higher oxidant doses. The same effect but to a smaller extent could be noticed for HSAPalm and HSACurc. In addition, the fluorescence signal of quercetin and curcumin both directly excited at their absorption maxima and indirectly through energy transfer from HSA excitation was monitored (squares in Fig. 3.19b). Comparing the trend for the two antioxidant molecules we notice that the characteristic emission of curcumin was depleted already at the lowest oxidant ratio of the experiment. Conversely it seemed that the quercetin molecule was able to scavenge a larger number of hypochlorite equivalents. In hindsight this could be explainable by the different oxidant-scavenging properties and would also be in agreement with the fact that the presence of quercetin was able to preserve the HSA's tryptophan emission up to higher oxidant doses compared to both palmitic acid and curcumin.

3.8.2 Absence of effects on the structural transition

The possible effect of ligands was also investigated in the whole oxidant/protein ratio range explored in the SUrF experiments. The comparison among the behavior of ligand-free HSA, HSAPalm and HSAQuerc showed that all systems underwent roughly the same structural transition after a critical oxidant/protein ratio (Fig. 3.20a,b,c). A slight difference could be found for HSAPalm since the overall variation in the scattering pattern at the intermediate q values (about $0.11 \text{ \AA}^{-1}$) across the oxidant ratio range was less pronounced (Fig. 3.20c). This could be due to

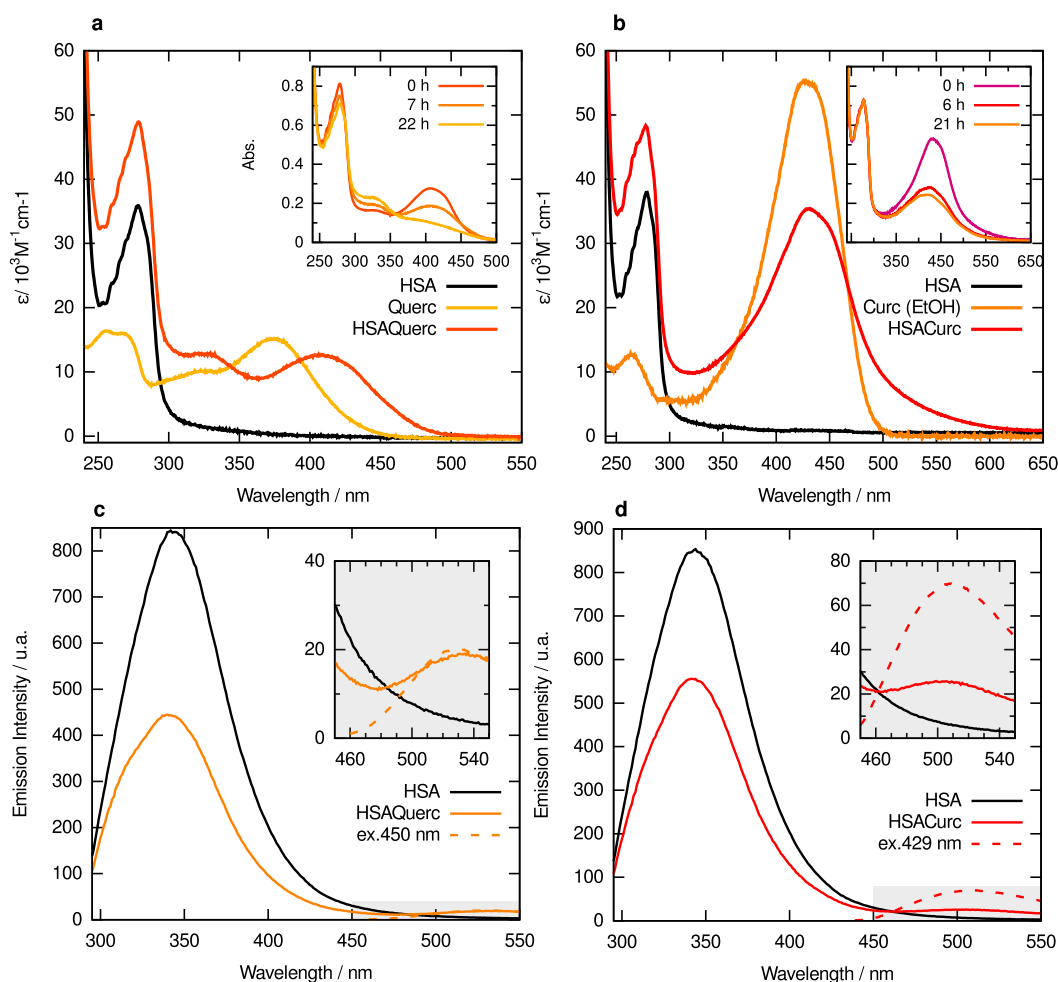


Figure 3.18. *Spectral properties of HSAQuerc and HSACurc.* In (a) and (b) the UV-vis absorption spectra of the polyphenol molecule alone, of the protein alone and of their complex are shown, for HSAQuerc and HSACurc respectively. In the inset, the variation of the absorbance spectra over time is shown. In (c) and (d) the fluorescence emission properties of HSAQuerc and HSACurc are presented, respectively. The spectra obtained with a 280 nm excitation are shown as solid lines, and the spectrum of HSA alone is shown in black for comparison. In the inset, an enlargement to show the emission band in the visible region characteristic of the ligand, both with indirect excitation through the protein (solid line, same spectrum as the main plot) and with excitation at the ligand absorption band (dashed lines).

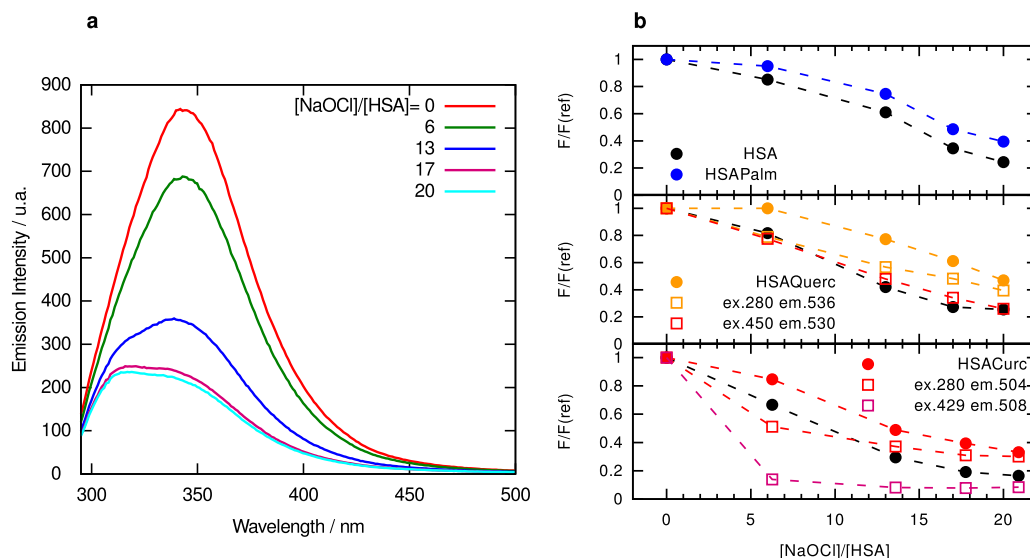


Figure 3.19. Effect of oxidation at lower hypochlorite doses on the fluorence of HSA and of HSA in the presence of ligands. (a) Emission spectra of HSA excited at 280 nm at different $[\text{NaOCl}]/[\text{HSA}]$ values between 0 and 20. (b) Relative decrease of the maximum fluorescence emission when excited at 280 nm as a function of the oxidant/protein ratio, in the case of HSA (black dots), HSAPalm (blue dots, top panel), HSAQuerc (orange dots, middle panel) and HSACurc (red dots, lower panel). For HSAQuerc and HSACurc the relative decrease of the quercetin and curcumin emission signals is also shown as a function of $[\text{NaOCl}]/[\text{HSA}]$ (empty squares).

some differences not in the overall but in the local shape of the oxidized conformer when some palmitic acid molecules are still bound to the protein. The fact that the presence of palmitic acid seemed to slightly delay the fluorescence emission decay to higher hypochlorite contents was also confirmed by the SURF experiment (not shown). A similar evaluation could not be easily assessed for HSAQuerc, since the thryptophan fluorescence, already quenched in the presence of quercetin (Fig. 3.18c), started from an already 60% lower value in the reference sample and sooner decayed at noise levels in these experiments. It could be noticed that for HSAQuerc the UV-vis absorbance at the wavelengths characteristic of oxidation increased to a lower extent compared to the ligand-free HSA (Fig. 3.20d,e,f). In fact, it is very likely that some of the hypochlorite equivalents were consumed by reaction with the quercetin, reducing only slightly the extent of chemical modification the protein underwent at a specific oxidant/protein ratio compared to the ligand-free HSA. There also could be in the spectra a certain compensation between absorbance decrease due to quercetin oxidation and absorbance increase due to protein oxidation, overall giving a less marked variation. From the SAXS/UV-vis correlation analysis was evident that the decay of quercetin absorption band in the visible range as a function of increasing hypochlorite doses, in contrast to the bands around 250 and 304 nm, was not strongly correlated with the variation of the scattering pattern related to the structural transition.

Overall, the experiments showed that the interaction of HSA with ligands (quercetin and palmitic acid) had no significant effect on the structural transition

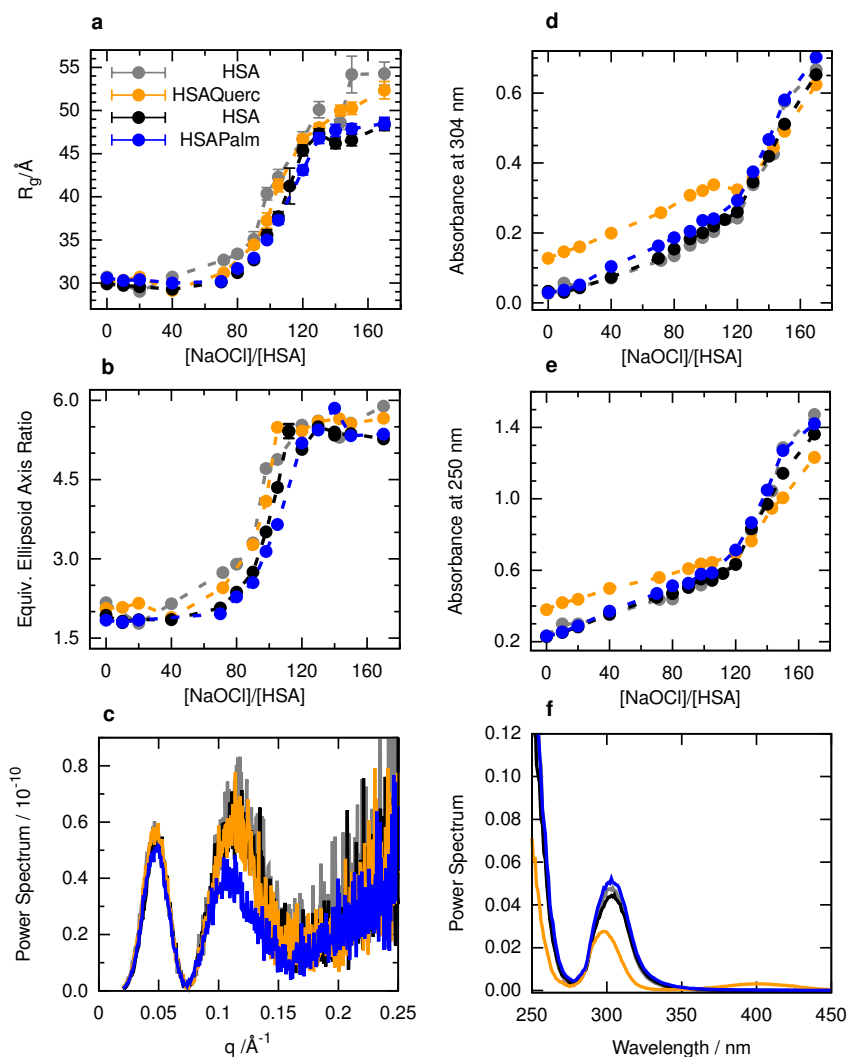


Figure 3.20. Results of the combined SAXS and spectroscopic characterization of HSAPalm and HSAQuerc at increasing oxidant/protein ratios and comparison with ligand-free HSA. The information derived by SAXS (a,b,c) and UV-vis absorbance analysis (d,e,f) are compared between HSAPalm (blue dots and lines), HSAQuerc (orange) and the ligand-free reference HSA. Two reference data sets are shown for HSAPalm (black) and HSAQuerc (gray) as deriving from two different experiments and protein batches. (a) Absorbance value at 304 nm as a function of oxidant/protein ratio ($[NaOCl]/[HSA]$). (b) Absorbance value at 250 nm. (c) Power spectrum of the series of UV-vis absorbance spectra.

induced by high hypochlorite doses. At low doses their presence seemed just to slightly delay the chemical modification (of tryptophan in particular) at higher molar equivalents of the oxidant. The possible mechanism would be based on the scavenging activity in the case of polyphenols. In the case of palmitic acid we could only suggest that the possible sites of modification could be slightly less accessible because of the steric hindrance given by the bound hydrophobic molecules or because of a decreased structural flexibility.

3.9 About the reversibility of the structural transition

In the literature about the study of oxidant-induced protein modification methionine was often added after oxidation in order to scavenge residual oxidants and “freeze” possible secondary reactions in order to have comparable samples irrespective of the effective time elapsed between preparation and characterization (e. g. in references [125,154,156,159,188–190]).

While showing the fluorescence spectra of HSA oxidized at low doses (section 3.8.1) it was not mentioned that the emission spectra, after being modified due to depletion of the tryptophan emission, could recover to a slightly higher intensity after some hours (Fig. 3.21). A control experiment on a solution of tryptophan alone showed in fact that, after being depleted due to oxidation, its fluorescence spectra increased again after some hours even reaching higher intensity than the original solution before oxidation. This behavior could be explained admitting that products of the reaction of tryptophan and hypochlorite at low doses have similar fluorescence properties as the native amino acid. In the literature, for example is reported that 5-hydroxytryptophan is fluorescent in a similar way but could be excited at longer wavelengths since its UV-absorption band extends up to 320 nm [191,192].

When methionine was added in excess to the HSA samples oxidized with a $[\text{NaOCl}]/[\text{HSA}]=20$, we observed clear spectral variations to an extent depending on the time elapsed from the initial oxidation (Fig 3.22a,b). The UV-vis spectra showed that methionine induced a decrease of the absorbance in the two wavelength ranges characteristic of the HSA oxidation (around 250 and 304 nm). This “recovery” after 6 hours from oxidation was as more pronounced as sooner methionine was added to the oxidized samples. The addition of methionine also lead to a recovery of the fluorescence emission, especially after some hours. The disappearance of the subsequent long-time evolution observed in the oxidized samples was also noticed (Figure 3.22b).

We talked about a partial “recovery” from hypochlorite-induced chemical damage due to reaction with methionine, but in practice it is not immediate to interpret it at a molecular level. It should involve the reaction of some modified residues with methionine to give the original residues back or maybe reaction products lacking the spectral features which distinguished the oxidized samples at first.

For example, Davies and colleagues [140] showed that the N-chlorinations induced by hypochlorite generate unstable species which can undergo chlorine-transfer reactions in the presence of a suitable substrate. Possible mechanisms were suggested and one of them could lead to recover the original amino or amide group [193,194].

It was interesting to assess if this partial “recovery” of spectral features more

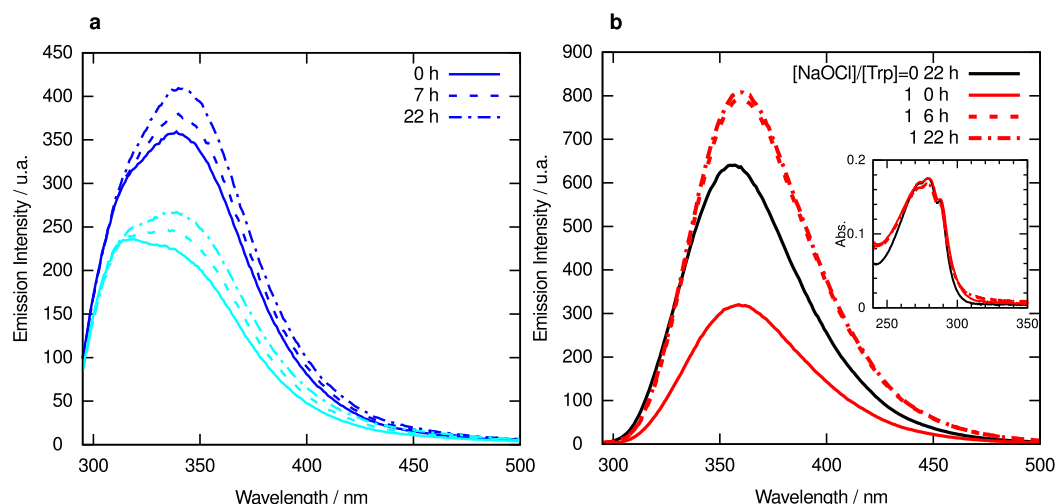


Figure 3.21. An increase of the fluorescence intensity was observed over time for HSA oxidized at low doses (a) Emission spectra with excitation wavelength 280 nm for two HSA samples oxidized with $[\text{NaOCl}]/[\text{HSA}] = 13$ (blue lines) and 20 (cyan lines) respectively, at three different times from the initial addition of oxidant. (b) Emission spectra with excitation wavelength 280 nm for a solution of tryptophan that was oxidized with a 1:1 molar equivalent of hypochlorite (red lines). The spectra collected at three different times from the initial reaction are compared. The spectrum of the non-oxidized tryptophan is shown in black. In the inset, the UV absorbance spectra are also presented.

similar to the non-oxidized protein could also affect the structural rearrangement induced by critical higher doses of hypochlorite (section 3.3). We indeed observed that the overall dimensions of the oxidized protein decreased after reaction with methionine (Fig. 3.22c). The characteristic bimodal shape of the P(R) was also lost (Fig. 3.22d), and a decrease in absorbance was observed in the UV-vis spectra, in a similar manner to the experiments at lower oxidant doses (Fig. 3.22a,inset). This suggested that the scavenger had been able to react with modified protein residues having a crucial role in determining the stabilization of the open conformer with a separated domain.

Since the zeta potential measurements (section 3.6) had suggested that the transition to an elongated conformation could have an electrostatic driving force, we investigated the effect of an increased ionic strength on the structural rearrangement induced by oxidation. The presence of NaCl 0.15 M induced more compact conformations of the protein monomer in the highly oxidized samples, being equal the extent of modification probed by the UV-vis absorbance (Fig. 3.23). The P(R) distributions were not clearly bimodal up to a $[\text{NaOCl}]/[\text{HSA}] = 120$, suggesting that in these conditions one of the protein end-domains was not as detached from the rest of the protein as when no screening of intra-molecular repulsive interactions was present.

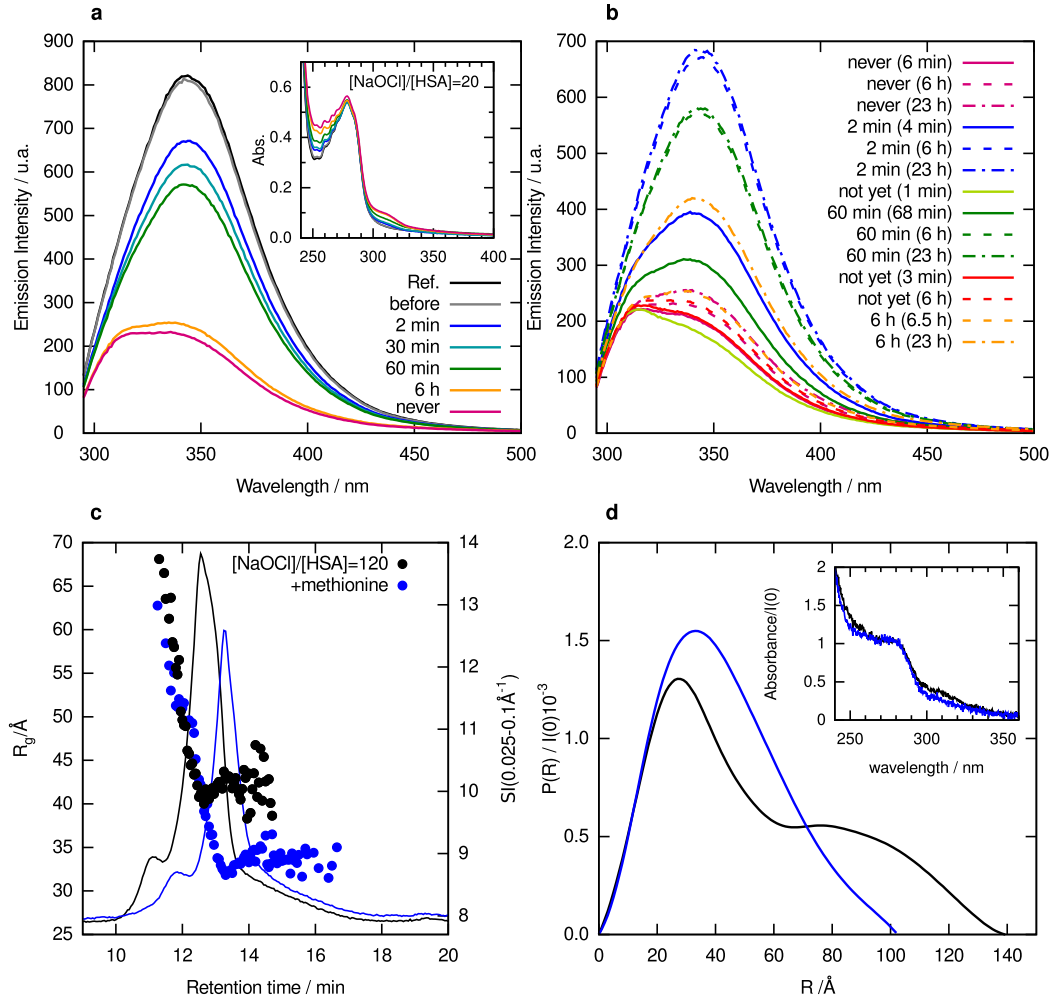


Figure 3.22. The reaction with methionine can cause a partial recovery from the spectral and structural modifications induced by hypochlorite in HSA. (a) Fluorescence spectra with excitation at 280 nm of samples of HSA oxidized with $[\text{NaOCl}]/[\text{HSA}]=20$. A 5:1 molar excess of methionine over nominal NaOCl concentration was added at different times from the initial preparation and the spectra shown were collected 6 hours after the initial addition of oxidant. A reference sample of non-oxidized HSA is also shown (black line). (b) Time evolution of the fluorescence spectra of oxidized HSA comparing the effect of methionine addition at different times from the initial oxidation. Each sample with a different methionine-addition time is represented with a different color. The different times of the measurements on a same sample are represented by different line styles and are indicated in parentheses in the key. For the samples in which methionine was added 1 hour or 6 hours after the initial reaction, the spectra measured before methionine addition are shown in a different color. (c) SEC elution profiles based on SAXS intensity for samples of HSA oxidized with $[\text{NaOCl}]/[\text{HSA}]=120$ without (black line) or with (blue line) the subsequent addition of methionine. The R_g obtained from the SAXS frames collected during the elution are also plotted as dots. (d) Comparison between the $P(R)$ functions obtained from the SAXS data at the elution maximum in the experiments shown in (c). In the inset, the UV-vis absorption spectra measured at the same time as the SAXS data are shown.

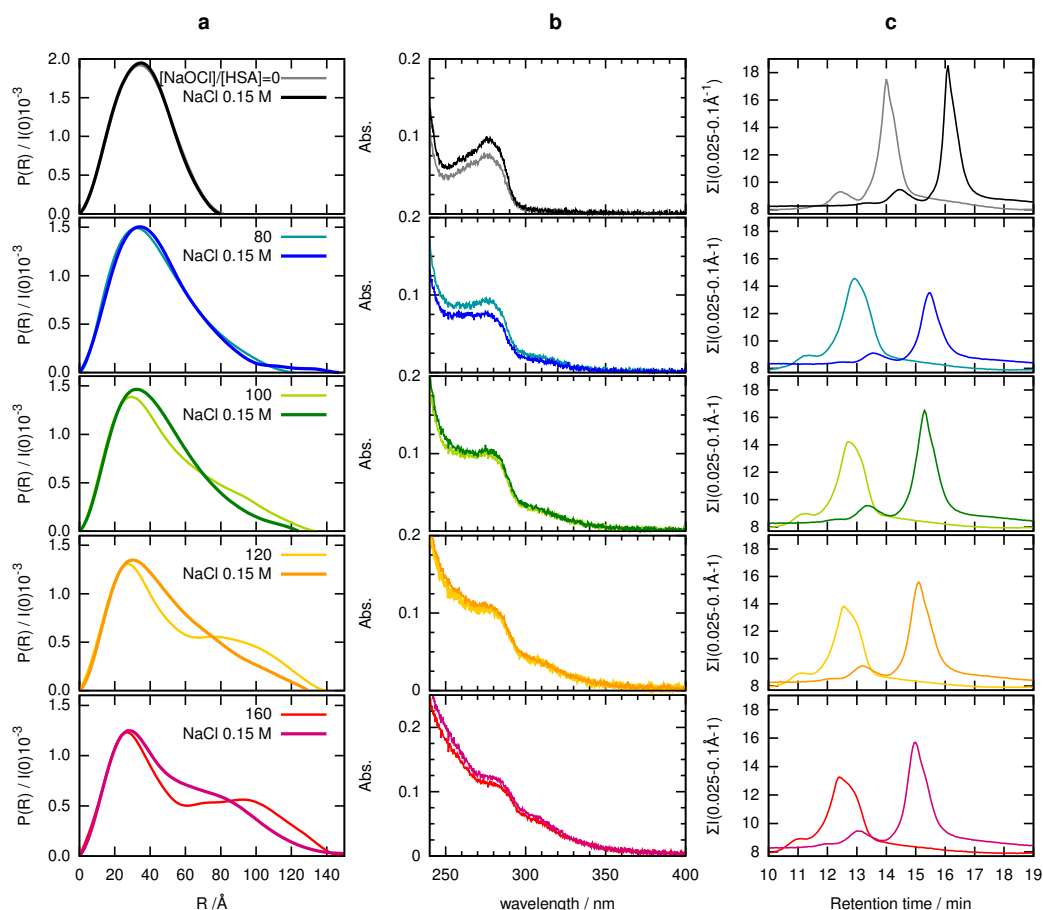


Figure 3.23. *Effect of higher ionic strength on the conformation assumed by oxidized HSA.* The results of SEC-SAXS experiments performed on HSA samples oxidized with $[\text{NaOCl}]/[\text{HSA}]=0$ (first row), 80 (second row), 100 (third row), 120 (fourth row) and 160 (fifth row) are shown in order to assess the effect of increased ionic strength given by the presence of NaCl 0.15 M. (a) The P(R) functions obtained from the SAXS data collected in correspondence of the elution maximum are compared. (b) The UV-vis absorbance spectra collected at the same time as the SAXS data in (a) are shown. (c) Elution profiles plotted using the integrated scattered intensity.

3.10 A note about mixing conditions

Mixing conditions constitute an important aspect of rapid chemical reactions. The method of sample preparation (section 3.2.1) was optimized to minimize inhomogeneity during mixing and avoid the uncontrolled and exclusively local exposure of the protein to concentrations of oxidant higher than the nominal value. We used SEC as a guide and assumed that if the chromatograms of the oxidized HSA samples showed the expected two peaks with the monomer/oligomer relative populations of a native sample (about 85% and 15%), then the mixing conditions ensured homogeneity (Fig. 3.8).

In order to assess how important this point could be for the characterization of a protein in an *in-vitro* model experiment, some tests were performed in which a different mixing proportion which did not ensure a homogeneous mixing was also employed.

In the case of a low oxidant ratio ($[\text{NaOCl}]/[\text{HSA}]=40$), which in the experiment with homogeneous mixing induced no effect on the protein structure as probed by SAXS, a dishomogeneous mixing determined an increased scattered intensity and slope at low angles, indicating the presence of higher molecular weight species (Fig. 3.24b). It seemed that this sample preparation method would induce oligomerization of the protein to a certain degree. In addition, the extent of depletion of the tryptophan fluorescence was different in the two cases (Fig. 3.24c) and being only one tryptophan present in each HSA molecule, this suggested that the mixing of not equal volumes determined the formation of distinguishable populations of protein molecules with different extent of chemical modification.

A further investigation by means of SEC-SAXS experiments was not able to clearly identify the hypothesized oligomers (Fig. 3.25a,inset). Possible explanations could be that the contribution of higher molecular weight species giving the increase in scattered intensity in the “single shot” experiment was very low in terms of mass fraction and below the limit to be clearly detected. Alternatively, we could also suppose that being these oligomers of non-covalent nature, a dissociation could be induced by the SEC process [195].

On the other hand, the SEC-SAXS characterization clearly showed that for hypochlorite/HSA molar ratios below 80 an inhomogeneous mixing determined the formation of populations distinguishable even by the SEC separation as multiple peaks. The $P(R)$ associated with the peaks showed the presence of more elongated conformers (pink lines in Fig. 3.25b,d) along with a main population with a similar $P(R)$ as in the case of homogeneous mixing. The more elongated conformers also had increased absorbance at the oxidation-characteristic wavelengths, suggesting that they underwent chemical modification to a higher extent (Fig. 3.25b,d,insets). With an oxidant/protein ratio as high as 80 no appreciable differences could be seen between a homogeneous and a inhomogeneous sample oxidation (Fig. 3.25be,f).

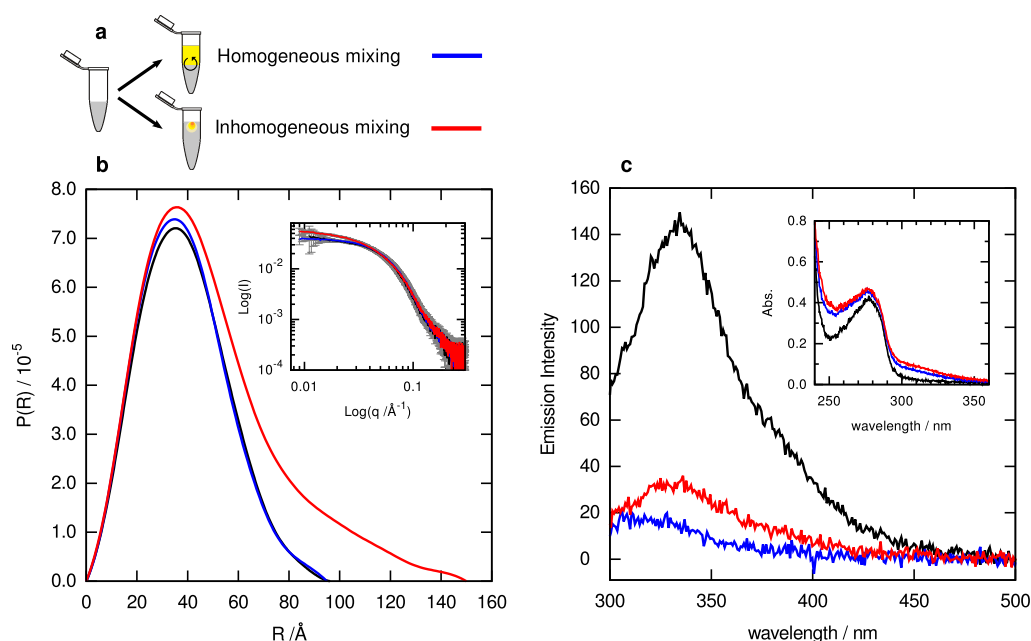
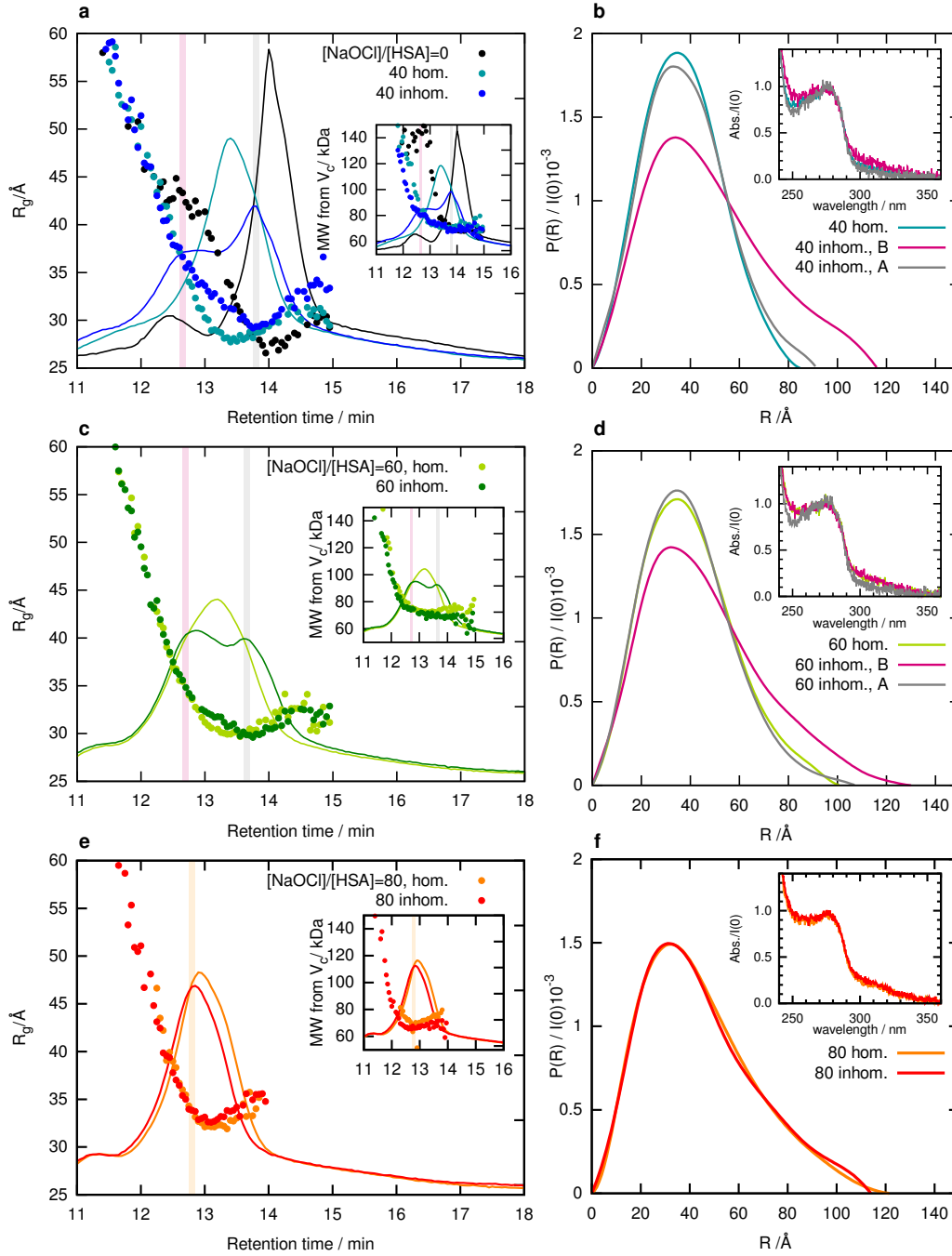


Figure 3.24. Effect of a dishomogeneous mixing of protein and oxidant solutions on the results of a “single shot” experiment (a) Scheme of the different mixing procedures. In the case labeled as “inhomogeneous” 50 μl of NaOCl solution were added to 350 μl of protein; in the “homogeneous” case 200 μl of NaOCl solution were added to 200 μl of protein. The final protein and nominal NaOCl concentration were the same. (b) The $P(R)$ functions obtained from the SAXS data collected on samples with oxidant/protein ratio = 40 prepared by “homogeneous” mixing (blue line) and “inhomogeneous” mixing (red line) are compared. The $P(R)$ of the reference non oxidized sample is also shown (black line). In the inset, the corresponding scattering profiles are displayed in double logarithmic scale. (c) The fluorescence emission spectra collected at the same time as the SAXS data in (b) are shown. In the inset the UV-vis absorbance spectra are displayed.



3.11 Summary and discussion

Remarkable findings about the effect of hypochlorite-induced oxidation on HSA came to light thanks to the structural characterization performed with SAXS, with the complementarity of the spectroscopic techniques allowing for the assessment of the extent of chemical damage.

The spectroscopic investigations demonstrated that the oxidant sodium hypochlorite determined a chemical modification of the protein from the very first addition. The damage proceeded to a significant extent when oxidant/protein molar ratios higher than 40 were reached as indicated by the complete loss of tryptophan fluorescence and the considerable absorbance increase. Despite this damage, the tertiary structure of the protein was only slightly affected and displayed an overall stability up to oxidation conditions with an oxidant/protein molar ratio < 80 . By further increasing the oxidant dose, a structural transition of HSA towards an elongated form occurred in the critical range of oxidant/protein molar ratio, i.e. from 80 to 120, and this was interpreted as a progressive detachment of one of the protein end-domains. In addition, the CD measurements showed that the structural modification implied the loss of roughly one third of the α -helical secondary structure, probably attributable to the unfolding of the helices in the inter-domain regions and in one of the three main lobes of the protein, which drifted apart. Zeta potential measurements suggested that an increase of negative charge on the protein was associated to oxidation. The intramolecular electrostatic repulsion generated with chemical modification was believed to be responsible for the protein structural transition, determining a driving force to partial local disruption of the secondary structure and to the domain separation. This would also explain why an increased ionic strength was able to induce a structural compaction of the oxidized conformers.

It is possible to draw a very simplified model to interpret these results. In order to explain the protein modifications occurring with an increasing amount of oxidant one could assume that the hypochlorite equivalents added are consumed to modify the oxidizable residues of the protein in decreasing order of reactivity. Kinetics constants of the reactivity of amino acid model systems towards hypochlorite were obtained by Pattison and coworkers [139] and they were also used to explain the amount of amino acid consumption at different reactant levels for proteins with known sequence [153]. In the case of BSA it was shown that the predictions based on the kinetic model, involving also secondary chloride transfer reactions from chloramines, were essentially in agreement with the experimental composition of the oxidized protein after 24 hours incubation, at least in terms of the average order of reactivity of the amino acids. Some discrepancies highlighted a role of the protein tertiary structure in modulating damage to specific residues.

Based on their results, one can build a rank of reactivity of the amino acids of HSA towards hypochlorite (Table 3.1) which can be used to predict where their oxidation is expected on an oxidant/protein molar ratio scale. The distribution of these potentially oxidizable amino acids in the HSA tertiary structure is shown in Fig. 3.26b.

According to this order, the initial variation of spectral features was due to the most reactive amino acids. For example tryptophan was expected to be oxidized at the lower oxidant/protein molar ratio regime of our experiments ($[\text{NaOCl}]/[\text{HSA}]$

Table 3.1. The residues of HSA oxidizable by hypochlorite are ranked in decreasing order of reactivity.

HOCl-oxidizable protein "component"	Kinetic constant* (HOCl) [$\text{M}^{-1}\text{s}^{-1}$]	Number in HSA	HOCl equivalents	Σ^{**}
Cysteine	$3.0 \cdot 10^7$	1	3	3
Methionine	$3.8 \cdot 10^6$	6	6	9
Disulfide Bonds	$1.6 \cdot 10^5$	17	17	26
Histidine	$1.0 \cdot 10^5$	16	16	42
N-terminal amine	$1.0 \cdot 10^5$	1	1	43
Tryptophan	$1.1 \cdot 10^4$	1	1	44
Lysine	$5.0 \cdot 10^3$	59	59(118) [#]	103 ^{##}
Tyrosine	44	18	18	121
Arginine	26	24	24	145
Backbone amide	10	584	584	729

*The kinetic constants values are taken from Ref. [153].

** Summation of the number of hypochlorite equivalents that can be consumed by reaction with more or equally reactive protein components.

[#] Dichloramines can also form. No separate kinetic constant reported in the reference.

^{##} The critical range of hypochlorite equivalents at which the protein denaturation was experimentally observed is highlighted in yellow.

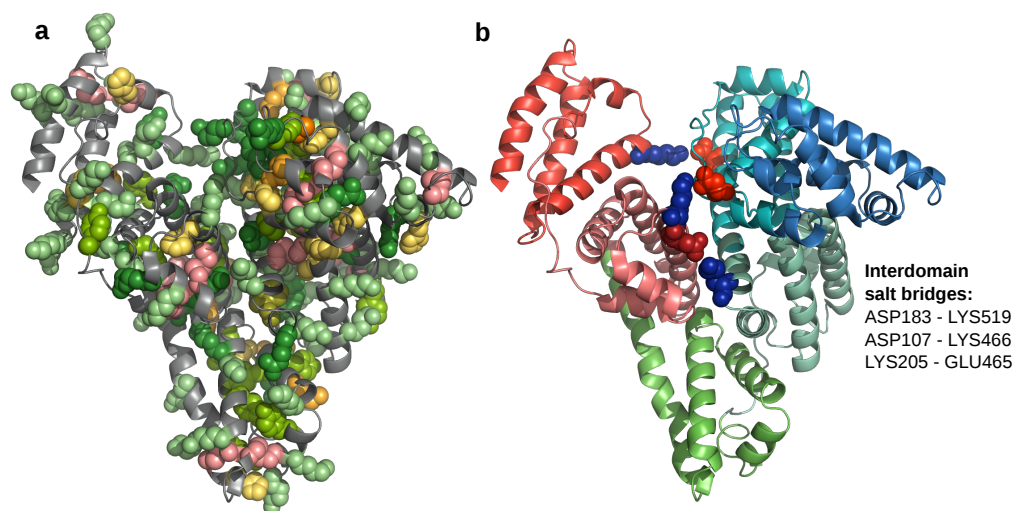


Figure 3.26. *Abundance and distribution of amino acids oxidizable by hypochlorite in the HSA tertiary structure. (a)* The oxidizable residues (spheres representation on the gray cartoon protein backbone) are highlighted in the HSA crystal structure (PDB entry: 1AO6) using a color code for each amino acid displayed in the Table 3.1. *(b)* In the HSA crystal structure in cartoon representation, three lysines (blue spheres) and three acidic residues (two aspartate, dark red spheres, and one glutamate, red spheres) are highlighted since they could interact through “salt bridge” at the interface between domains, potentially important for the stability of the protein tertiary structure. These figures were generated with PyMOL.

< 40), as manifested by the depletion of tryptophan fluorescence in this range of oxidant dose.

The critical region of oxidant doses where the progressive opening towards an elongated conformation was detected is highlighted in yellow in the Table 3.1. The lysine residues (very abundant in HSA) should be the main target of the chemical modification in this regime of oxidant/protein ratios. We could argue that these residues started a conversion to chloramines for molar ratios higher than 40 [171,196]. After a critical molar excess (around 80), the modification of some pivotal lysine residues triggered the disruption of crucial inter-domain salt-bridge interactions stabilizing the heart-shaped native form (Fig, 3.26b). Modification of additional lysine residues by higher hypochlorite doses (between 80 and 120) would to a greater and greater extent stabilize an “open” conformation in which one of the end domains drifted apart from the rest of the oxidized protein. At oxidant/protein molar ratios >120, where no further opening was observed, a progressive oxidation of chloramines to dichloramines was believed to occur as demonstrated by the formation of a more intense absorption band at 304 nm in the UV spectra. This band quickly decayed in time, as expected, due to the instability of the dichloramines [197]. It is interesting to note that a further negative charge increase and helical structure disruption occurred in this regime. Specific kinetic constants for the further conversion of the lysines from monochloramines to dichloramines were not measured, but the possibility of this conversion with an excess of hypochlorite was pointed out for the nitrogen-containing side-chains of glycosamines [198] and deduced for HSA in one of

the early studies about hypochlorite modification [199]. The longer-time evolution of the spectral features observed for the oxidized protein was rather complex and it is hard to suggest a detailed explanation at this stage. We should nonetheless take note that secondary reactions involving the modified protein residues themselves should be implicated.

All together, we could deduce that the modification of electrostatic interactions through transformation of positively charged amine groups in unstable chloramines represented the major factor in determining protein native shape destabilization. Even if not evidenced as frequently as the more common methionine oxidation products, probably due to the difficulty in experimentally capturing the formation of still reactive chloramines by means of the most used chemical essays and mass spectrometric methods [200], lysine oxidation could be a relevant and general mechanism by which hypochlorite-damaged proteins become unstable or undergo conformational changes. For example the correlation between the formation of protein chloramines and physiologically relevant functional variations induced by hypochlorite was demonstrated in the case of a protein acting as chaperone in Ref. [201].

3.12 Conclusions

Among all the reported results, the main findings highlighted that HSA can tolerate the oxidation effect of large doses of hypochlorite in its folded state, as well as unfold only partially at extremely high oxidative damage, thus showing a tuned structural response. In addition, the oxidized conformers were stable against aggregation and phase separation. A hypothesis suggested by this behavior would be that HSA can act as a biological “*hypochlorite buffer*” to control the local formation of this highly reactive oxidant. Indeed, these properties give the protein the capability to store a short-term oxidizing potential in the form of the reactive mono- and di-chloramines and to release it in the environment upon diffusion. A direct consequence could be that HSA would have the ability to spread the bactericidal activity related to the hypochlorite-induced modifications and at the same time reduce the local oxidative burst and protect more susceptible proteins. The observed structural stability suggested that the biological functions of albumin could be partially preserved when oxidized.

Furthermore the well-defined conformational variation of HSA triggered by the hypochlorite oxidation represents a new and fundamental addition to the structural rearrangements possible for this multi-domain protein in response to functional (section 2.5.3, [121]) and denaturing conditions (sections 2.5.1 and 2.5.2, [79, 86, 117]). For the several investigators working with albumin the reported biophysical characterization can be an important information to rationalize experiments in the presence of oxidants.

We could also speculate that the conformational change of HSA caused by the hypochlorite-induced chemical modification could help in rationalize on a structural basis the observations about an apparent implication of the “chlorinated” form of HSA in specific receptor recognitions with potential roles in physiological processes like clearance of the damaged protein or inflammation stimulation [144–148].

Chapter 4

Acid unfolding of Human Serum Albumin

4.1 Introduction

As mentioned in section 2.5.1, the protein serum albumin is known to undergo structural transitions in non-native pH conditions. According to a “textbook” classification [5] of the pH-induced conformational changes of albumin, four additional isomers of the normal, or N form, have been recognized: F, or fast, at pH 4; E, or extended, below pH 3; B, or basic, near pH 8; and A, or aged, near pH 10. These transitions are all considered to be reversible.

In particular, a huge amount of biophysical data have been collected throughout years about the serum albumin (mainly BSA) in acid conditions, where the N-F and F-E transitions should take place. The accompanying structural changes have been predicted through physical chemical evidence and have been only tentatively identified in relation to the known tertiary configuration of albumin. According to Peters [5], the isomerizations are probably of interest more for what they can tell of the dynamics of the albumin structure than for physiological significance. Nevertheless the conservation of the N-F transition and of the pH of the transition among various diverse species suggested a physiological role for this transformation. Given the lower pH measured on the membrane surfaces of several tissues, one could speculate that the affinity for ligands is reduced in the F conformation, perhaps facilitating ligand off-loading at various tissue interfaces [106].

The N-F and F-E transitions in the literature The first observation leading to hypothesize a conformational rearrangement of albumin in acid environment dates back to 1939 when Luetscher noticed distinguished bands in the electroforetic migration at pH 4.0 [105]. The meaning of these observations in terms of an isomerization was chiefly elicited by Foster and colleagues [102]. The abrupt discontinuity in the titration of BSA at pH 4-4.5 observed by Tanford was interpreted as coinciding with the appearance of a faster migrating, or F, form as seen on gel electrophoresis at pH 3-4. Their early experimental work based on a variety of hydrodynamic studies confirmed that there was a large increase in the hydrodynamic volume of the molecule at low pH which probably resulted primarily from coulombic repulsion

due to the large positive charge developed on the protein. The further increase in viscosity at pH 2.8 was interpreted as an additional structural expansion leading to the E form [202].

The acid unfolding studied through CD spectroscopy pointed clearly out that a first decrease of helical structure between pH 5/4.4 and 3.8/3.5 occurred, while additional secondary structure unfolding was observed between pH 3.8/3.5 and 2.7/2.5 [108, 203–205]. This further decrease of helicity was inhibited by high salt concentrations above 1 M [5, 204], supporting its interpretation as an acid expansion in which salt forces predominated. A clear blue shift of the protein fluorescence spectrum was also observed to occur in acid environment [204–207] and was interpreted as the tryptophan becoming more buried into an hydrophobic pocket when the protein assumed the F form.

It was also suggested that the F isomerization could proceed in two steps. In conditions with high salt content, a multi-step change of the CD signal was observed in the pH region 5–3.5 and the major variation in the fluorescence was already completed at pH 4, in the middle of the N–F transition as probed by other methods. These findings lead to the proposal that the conformational change could occur first to an intermediate F1 form, which then expands to the F form spontaneously [204, 208, 209].

Further investigations tried to gain more structural insight on the conformational transitions.

Nuclear magnetic resonance (NMR) studies pointed out that the N–F transition of BSA was characterized by a loss of spectral features characteristic of a folded protein and appearance of features typical of residues in unstructured regions, but since a similar proportion of resonances was visible in the NMR spectra of the N and F form, it was apparent that the unfolding interested only small regions of the protein. A general rearrangement of the three-domain structure with a certain loss of structure in particular at the N-terminus was compatible with the NMR findings [210, 211], rather than a specific unfolding of domain III as suggested on the basis of experiments with albumin fragments [205, 212].

When the high-resolution structure of HSA was resolved, Carter and Ho [106] proposed speculative models for the F and E isomers (Fig. 4.1) that would be in agreement with the deductions from the change of hydrodynamic properties, the decrease in helical structure and the exposure of proteolytic sites. The F form was thought as deriving from the separation of the two halves of the molecule, domains I + IIA and domains IIB + III, from each other, while in the E form the molecule was thought to be as fully expanded as its disulfide bonding structure allowed. In the case of the E form, electron microscopy pictures had already shown the molecule at pH 2 as a chain of globules, about 21×250 Å in size [213], and a similar model of 9 loops kept together by the disulfide bridges and interconnected by flexible regions was suggested later from the SAXS analysis [92]. An earlier study performed with SAXS at different pH values also proposed models based on subdomains with different spatial configurations depending on pH [214].

The partially unfolded conformers found in strongly acid conditions were also suggested as molten globule states of the protein, in which the individual domains preserved great part of the secondary structure and had a higher degree of compactness compared to an unfolded chain [207]. More specifically, due to the absence of hydration of the hydrophobic core of the protein domains in this partially unfolded

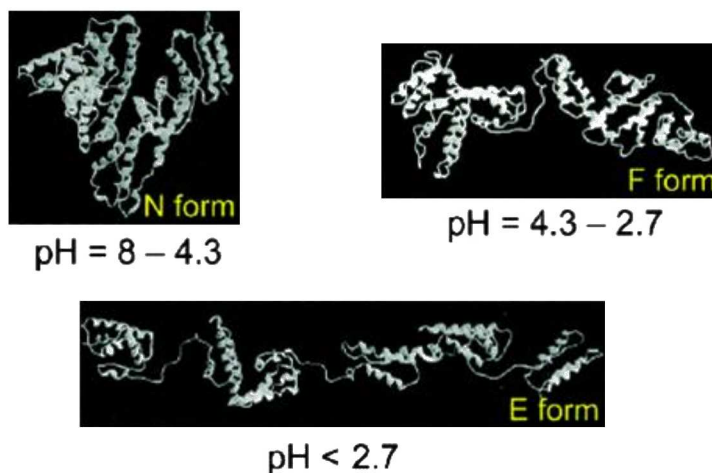


Figure 4.1. Proposed configurations of F and E isomeric forms of HSA. Reproduced from references [5,106].

forms, it was proposed that two out of the three HSA domains could be in the state of a “dry molten globule” in the acid induced form [215,216].

The possibility that the partial unfolded conformers of albumin could interact more easily in order to form aggregated structures had been suggested. Baler and colleagues derived a structural model for the F form from molecular dynamics simulation at pH 3.5 and on its basis suggested that hydrophobic interactions due to exposed regions and counterion binding could overcome the repulsion between the positively charged protein molecules and drive protein aggregation and gel formation in acid environment [217,218]. The influence of physiological concentration of NaCl was pointed out to increase the apparent molecular weight measured by LS or SAXS intensity, probably due to attractive interactions or formation of dimers [219]. Barbosa and colleagues [220] paid attention to the effect of the protein concentration on such transitions and suggested that at large amounts of protein extensive morphological changes as well as aggregation could be inhibited by an interaction potential.

It should be pointed out that a large variety of different experimental conditions were used in the huge number of works investigating the acid unfolding of albumin. These differences in the specific solution conditions adopted in the experiments determined a certain variability in the reported pH ranges to which the N-F or F-E transitions were assigned. For example the variation of the ionic strength and of the identity of the salts used could alter the equilibrium conformation assumed in each protonation state [217]. In addition the quantum yield and Stoke’s shift of the fluorophores like tryptophan, often used as probes of the acid transitions, could also be influenced by the specific buffer composition [221]. Moreover, in the very early literature, the uncontrolled presence of fatty acids bound to the protein as a result of the different purification methods was also a main factor determining heterogeneity in the characterization of the acid unfolding or protein precipitation and this was soon recognized [211]. In fact it is not immediate to correlate the

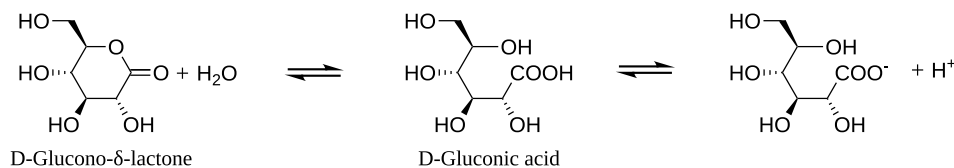


Figure 4.2. Chemical equilibria involving GdL in solution: hydrolysis to gluconic acid and subsequent acid dissociation.

information derived by all these investigations. The idea of this study was to apply the available possibility of a simultaneous SAXS and fluorescence characterization performed with the SURF platform (section 3.2.2) to follow the acidification of an HSA solution in time.

The two techniques are sensitive to local or global modifications of the protein structure to different extents. Also, the high sampling rate over the pH interval should help clarify the multi-step nature of the acid transition of serum albumin and identify the scattering profiles of characteristic conformers. The use of modeling procedures could allow the retrieval of low-resolution models highlighting how the multi-domain structure of the protein could explain the change in dimensions and flexibility of the protein in the process.

The hydrolysis of glucono- δ -lactone as acidification method In order to follow the conformational changes of HSA occurring with decreasing pH in a continuous manner, we decided to use a method allowing for a progressive acidification of the solution over time that had never been applied in the context of protein unfolding before. The acidification was caused by the slow hydrolysis of the glucono- δ -lactone (Fig. 4.2) in an initial neutral solution. The glucono- δ -lactone (GdL) is a food additive used mainly as acidifier, with high solubility [222]. The lactone in solution slowly hydrolyzes to gluconic acid, with which it is in equilibrium. The acid form in turn undergoes an acid dissociation equilibrium [223] increasing the H^+ concentration of the solution. Therefore with the advancing of the hydrolysis reaction a slow and uniform decrease of the solution pH can be attained. This method had been initially employed to gel alginate solutions containing calcium carbonate in order to release Ca^{2+} ions which could trigger the network formation [224,225]. It was then suggested for the preparation of uniform hydrogels through the self-assembly of small molecules triggered by the charge modification at low pH [226,227]. In the case of a fast gelation kinetics, in fact, the standard acid-addition technique provided a too slow kinetics of mixing compared to the rapid self-assembly process and this lead to visibly inhomogeneous gels with history-dependent properties. The GdL method instead, given the slow hydrolysis rate compared to the fast dissolution process, allowed for a uniform pH decrease providing reproducible homogeneous gels. The equilibria that occur during the hydrolysis of GdL are complex [223] and are further complicated in the presence of buffer salts [228], but the pH decrease in time is highly reproducible in given conditions.

4.2 Experimental procedures

4.2.1 Sample preparation

A stock solution of HSA was prepared as described in section 3.2.1, using degassed 10 mM sodium phosphate buffer with pH 7.4. The protein concentration was 3 g/l for the SUrF experiments and 1 g/l for the bench-top spectroscopic measurements. The acidification was started adding a known volume of the HSA stock solution to a weighted amount of GdL ($\geq 99\%$ type G4750, from Sigma) and mixing for 30 seconds in order to induce complete solubilization of the lactone. For the blanks, the same procedure was applied using the buffer rather than the protein solution. The GdL amount added in the samples ranged between 1-5% expressed as weight of GdL/volume of protein solution.

pH-time calibration

An optimization of the solution conditions was necessary in order to realize the most suitable pH profile as a function of time to monitor experimentally the HSA unfolding. The pH of the solutions prepared as described before was measured periodically over time with a Hamilton minitrode electrode connected to a Crison 2002 pH-meter and calibrated with Hamilton standard buffer solutions at pH 7.00 and 4.01. From this preliminary screening it was observed that the greatest drop in pH occurred in the first hour from the GdL addition, while a slow decrease towards a plateau value took place at longer times. The lowest value of pH reachable decreased with the increase of the % GdL (Fig. 4.3b, inset) and was slightly increased at higher protein concentration. A high reproducibility of the pH variation in time was also assessed, estimating error bars of ± 0.05 pH units as maximum absolute error.

A calibration of the pH variation over time in the conditions applied during the experiments was needed to estimate the pH of the protein solution at each point of time from the addition of GdL (Fig. 4.3). In this way the variation of the experimental parameters monitored could be expressed as a function of pH rather than as a function of time. In order to do this, samples identical to those which underwent the SAXS and spectroscopic measurements were prepared starting from the same protein stock solution. A pH-vs-time profile was measured as stated before. For the conversion of time to pH values, a spline interpolation of these profiles at the experimental time points of the measurements was performed using the Matlab *interp1* function.

4.2.2 Time-resolved SUrF experiments

The combined SAXS-spectroscopic experiments were performed using the SUrF platform (section 3.2.2) at the SWING beamline (section 3.2.6). Samples with GdL 1%, 2%, 3% and 5% were measured. For each of them, the HSA solution was injected into a 1-ml loop and pumped by means of the HPLC apparatus at a flow rate of 0.02 ml/min into the SUrF capillary. SAXS exposures of 500 ms spaced out by a waiting time of 29.5 s were collected while the sample was flowing and was undergoing progressive acidification, during a time frame of 40 minutes. At the same time 1-s exposures were continuously collected by means of the UV-vis and

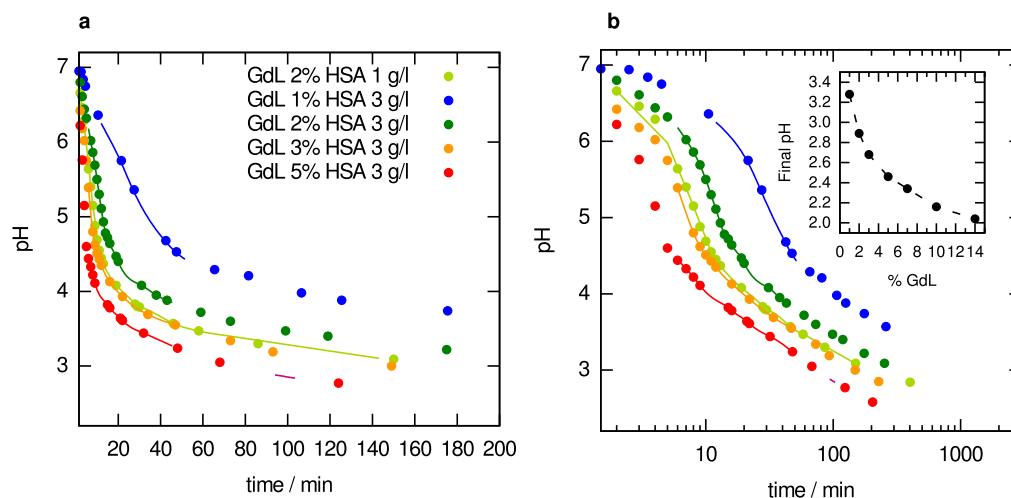


Figure 4.3. (a) Experimental pH-vs-time profiles measured on HSA samples acidified with different amounts of GdL. The dots represent the measured values, while the lines represent the interpolated values at the time points of the experiments. (b) The same data are presented with a logarithmic time axis to highlight the variation at small times. In the inset, the lowest pH value of a 3 g/l HSA solution reached at long times is plotted as a function of the % GdL.

fluorescence probes of the SURF platform. A total of 80 SAXS frames, 280 UV-vis spectra and 280 fluorescence spectra were acquired for each sample. The sample with GdL 5% was also measured again after 90 minutes from preparation, in order to acquire data of the protein at lower pH. The data for background subtraction and the conversion of the UV transmission to absorbance were collected on blank solutions directly loaded into the capillary. Ten SAXS exposures of 500 ms spaced out by 2.5-s waiting time, together with 5 fluorescence and UV-vis spectra were collected and averaged.

Since the albumin acid isomerization had been described as a process occurring in a time of the order of the milliseconds [229], the sampling frequencies adopted in these experiments were considered slow enough to ensure that the collected data were representative of the protein structure at the pH predicted on the basis of the known acidification rates.

4.2.3 Complementary measurements

UV absorption spectra and fluorescence emission spectra were collected on HSA samples undergoing acidification by means of bench-top instruments (section 3.2.5), in order to compare the SURF results to data with better signal-to-noise ratio. For both experiments a 1.4-ml sample of HSA 15 μ M (about 1 g/l) was acidified with GdL 2% and time-series of spectra were collected starting from 2 minutes after the solubilization of GdL and then every 3 minutes, for 2.5 hours. The UV spectra were collected between 240 and 360 nm with a step of 0.5 nm and a total acquisition time of 48 s. The sample was measured in a quartz cuvette with 1-cm path length and a maximum volume of 1.5 ml. In the fluorescence experiment a quartz cuvette

with 1-cm excitation path length and 0.5-cm emission path length was used. The intrinsic fluorescence was measured by exciting the protein solution at $\lambda_{ex} = 295$ nm with a band-pass of both excitation and emission monochromators of 5 nm. The emission spectra were collected between 305 and 500 nm with a step of 1 nm and an acquisition time of 0.2 s. Filters were used both in excitation (250-395 nm) and in emission (295-1100 nm).

4.2.4 Data analysis

SAXS data analysis

The SAXS data reduction was performed using FoxTrot software, developed in-house at Soleil. An automatic analysis of all the subtracted protein profiles was performed with a Matlab script as stated in section 3.2.7. In addition, the script included an alternative analysis of the SAXS data by means of the IFT-BayesApp method [17] in the fixed q range 0.02-0.3 $\text{\AA}^{-1}$. After the automatic pipeline analysis, a set of data spanning the available pH range was built including selected frames from the experiments at different % GdL. These were analyzed once more, manually choosing the most reliable q range for the Guinier approximation, and the optimal q range and maximum R for calculation of the pair distance distribution function $P(R)$. This selected subset of scattering profiles was then used for the component analysis as described in section 4.2.4

Spectroscopic data treatment

Due to the higher sampling frequency applied with the spectroscopic techniques compared with the SAXS measurements a “rebinning” was performed to match a representative spectrum with each scattering curve measured, in the following simplified way. The UV-vis absorbance spectra and fluorescence emission spectra collected between the time of one SAXS frame and the subsequent, were averaged and nominally assigned to the same time point of the first SAXS profile. In the case of the fluorescence emission, a heavy smoothing (Savitzky-Golay with a third-degree polynomial and a 41-point window) was applied in order to better estimate the wavelength of the emission maximum and prepare the data for correlation analysis.

2D-correlation spectroscopy

The variation of the scattering profiles and fluorescence emission spectra of HSA induced by the progressive acidification were correlated according to the method mentioned in the sections 3.2.7 and 1.3. This analysis was applied to the SAXS data in the form of Kratky plots ($I(q) \cdot q^2$ vs. q) in the q range 0.02-0.2 $\text{\AA}^{-1}$ and the smoothed fluorescence spectra between 298 and 379 nm, using pH as the perturbation variable.

Component analysis

In order to estimate the minimum number of structural components involved in the unfolding process captured in these experiments, the Principal Component Analysis

(PCA) or Singular Value Decomposition (SVD) method (sections 1.2 and 3.2.7) was applied to the data matrix built by the series of the SAXS data spanning the available pH range. More in detail, the SAXS data were normalized for the scattered intensity at zero angle value ($I(0)$) calculated from the $P(R)$ function and the q range was restricted to 0.02-0.2 $\text{\AA}^{-1}$. An analysis based on the data in the form of Kratky plots rather than simple intensity profiles was preferred since this representation provided the biggest relative variation with unfolding, making easier the identification of relevant components. The autocorrelation function of the eigenvector ($C(r)$) was the main discriminant between a component significantly contributing to characteristic features of the Kratky plot varying with the protein unfolding, or simply to noise, since in this second case the eigenvector's autocorrelation does not deviate significantly from zero. It was computed as:

$$C(r) = \frac{1}{N-r} \sum_{i=1}^{N-r} K(i)K(i+r) \quad (4.1)$$

where N is the total number of points and K is the value of the component [84].

The multivariate curve resolution - alternating least squares (MCR-ALS) [170] method was also applied in order to find scattering components which would satisfy the physical constraints of non-negativity of the scattering profile and of closure of the fractions, in order to interpret the process in terms of coexisting species and their relative amount. The optimization was performed by using the MCR-ALS toolbox implemented as a Matlab graphical interface [83].

3D-modeling based on SAXS data

3D-models representative of the average folding of HSA in selected conditions were calculated from the SAXS profiles.

The software GASBOR [57] was used, which searches a dummy-residue model in which the number of beads providing a chain-like assembly is constrained by the known protein sequence, and the average scattering factor of an amino acid is used for computing the theoretical scattering. In each case ten optimization runs were performed and the results compared by the software DAMAVER which superimposes the obtained models by means of the SUPCOMB code [64] and provides a parameter for the goodness of superposition (the normalized spatial discrepancy, NSD). The model most similar to the others was chosen as representative of the particular conditions, while the averaged envelope including all the superimposed models was visually indicative of the variability of the 3D reconstruction.

4.3 Monitoring structural and spectral features of HSA as a function of pH

The time-resolved analysis performed with the SURF platform of HSA samples containing GdL allowed for a continuous monitoring of the structural and spectral features of the protein in solution in a progressively more acid environment (Fig. 4.4). By analyzing samples with different GdL content it was possible to tune the pH-window accessible in a given experimental time and overall capture all the major variations observed with the available techniques in the acid unfolding process.

The scattering profiles very sensitively probed the conformational variations HSA underwent, in particular when visualized in the form of Kratky plots (Fig. 4.4, first row). Starting from the typical bell-shape in nearly-neutral conditions, a clear lowering of the maximum could be then seen at pH values approximately around 4.5. Below pH 4 the shape of the Kratky plot started to markedly change: the characteristic maximum position was lost and an increase of the tail at larger q values occurred. This could be interpreted as a clear sign of increased protein flexibility. In the pH range around 3.5 a Kratky plot with a characteristic shape endowed with a maximum around $0.1 \text{ \AA}^{-1}$ was recognizable, and going lower in pH a further decrease of the the Kratky plot ordinate at $0.05 \text{ \AA}^{-1}$ and further increase at q values greater than $0.1 \text{ \AA}^{-1}$ were observed, as evidenced also by the data collected after longer time from GdL addition, so that a lower pH value could be reached (purple line in Fig. 4.4).

In spite of the high noise level of the fluorescence emission spectra collected with the SURF, an evident blue shift with decreasing pH was observed to occur (Fig. 4.4, second row), as reported in the literature. This change was not accompanied by noticeable variations of the emission intensity and took place in the pH range between 6 and 4.5.

Regarding the UV spectra collected simultaneously with the SAXS and fluorescence data, they did not show much variations informative about the albumin unfolding. However it was important to check the correctness of the concentration of the protein solutions being analyzed, by looking at the constant value of the absorbance at 280 nm in the different experiments. In addition, we assessed that the spectra became unreliable below 260 nm, and off-line spectroscopic measurements clarified that this was due to the high UV absorbance of GdL below this threshold wavelength (Fig. 4.5). More in detail, it was observed that the absorbance of a GdL solution was not constant with the progression of the hydrolysis and acid dissociation, but an overall decrease in time could be seen (Fig. 4.5b), affecting also the experiments in which the HSA spectra were monitored during acidification (Fig. 4.5a). Therefore the unreliability of the absorbance values at low wavelengths in the SURF UV spectra was given, in addition to the low counts due to high absorption, also by the fact that a transmission spectrum of the blank which was collected a few minutes after the GdL solubilization was employed for the absorbance calculation, rather than taking the time-variation into account.

After the initial assessment of the results of the experiments, a deeper examination of the information that could be derived from the SAXS and fluorescence analysis of

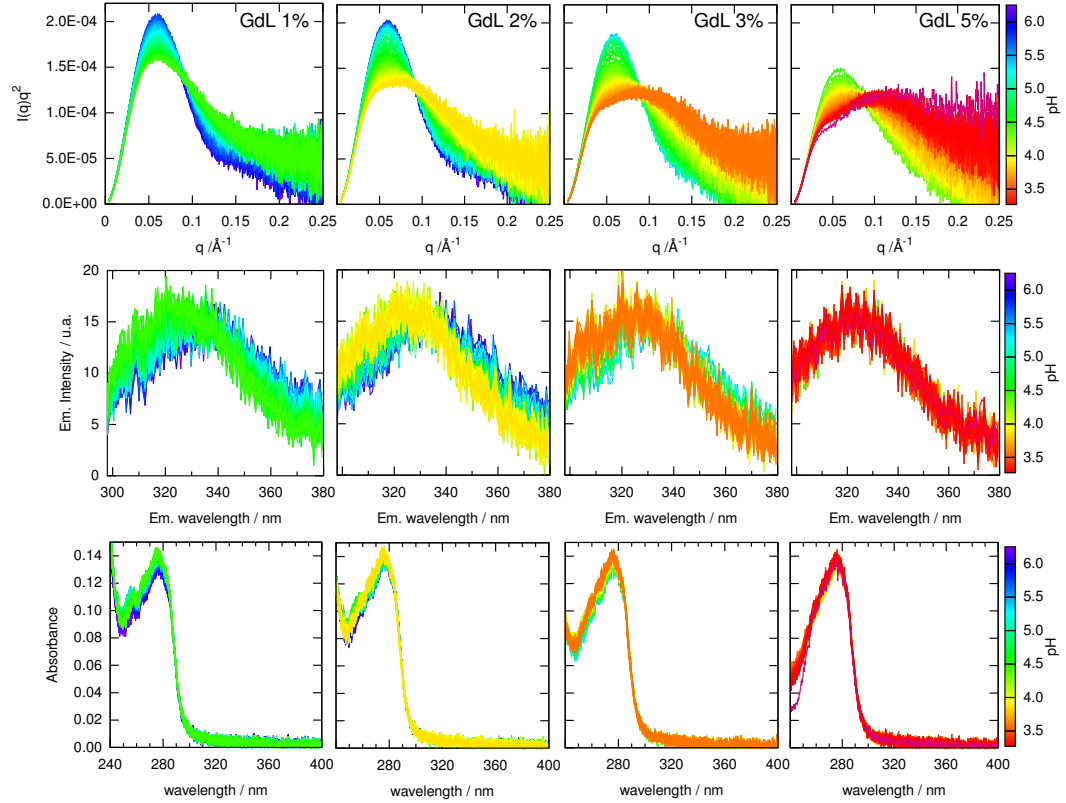


Figure 4.4. Acid unfolding of HSA as monitored by means of the the SURF platform. The experimental data collected during progressive acidification of the HSA samples with 1,2,3 and 5 % added GdL are presented. A color palette indicates the corresponding pH value. The SAXS data in the form of the Kratky plot (first row), the fluorescence emission spectra (second row) and the UV-vis absorbance spectra (third row) are shown. The pH value corresponding to the data collected after longer time for the 5% GdL sample is not included in the color bar and is 2.84. These additional data are shown as purple lines.

4.3 Monitoring structural and spectral features of HSA as a function of pH 99

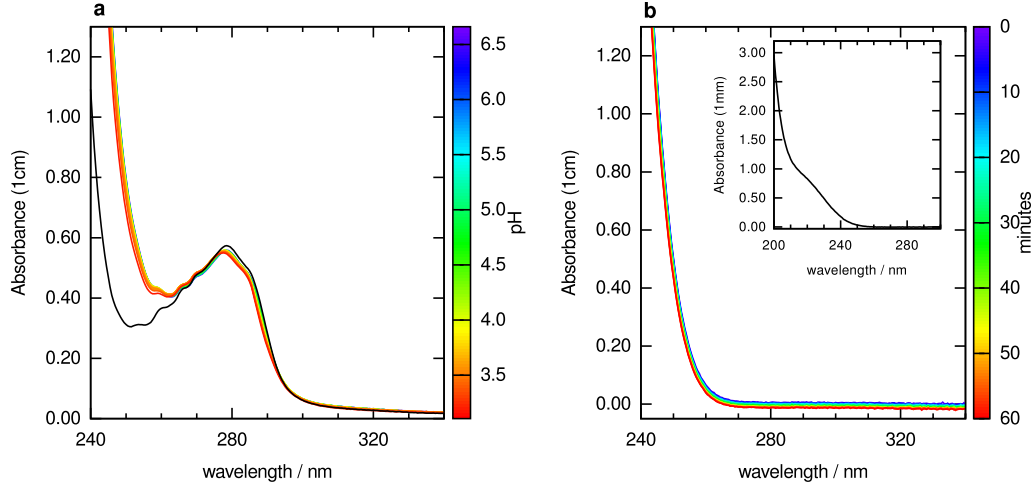


Figure 4.5. *Complementary characterization of the UV absorbance of GdL-containing samples. (a)* UV-vis spectra of a HSA sample acidified with 2 % GdL as a function of the decreasing pH, as indicated by the color bar. The spectrum of the same HSA sample before addition of GdL is shown for the sake of comparison (black line). *(b)* UV-vis absorbance spectra of a GdL solution as a function of the time elapsed from initial solubilization. The same cuvette and GdL concentration as for the spectra in *(a)* were used. In the inset, a UV spectrum of the same solution measured with smaller path length and up to a lower wavelength is shown.

the acid unfolding of albumin is possible.

From the large number of SAXS data the radius of gyration (R_g) and the scattered intensity at zero angle ($I(0)$) were estimated by means of the Guinier approximation at low q (Fig. 4.6, top panel). By inspection of the trend as a function of decreasing pH, it could be noticed that an increase of $I(0)$ occurred in a pH range around 5.5. Since the protein isoelectric point is located in this pH region [5], this effect could be ascribed either to the appearance of the attractive protein-protein interactions or to a real increase of the degree of dimerization, in both cases arising from the depletion of electrostatic repulsion when a condition of neutrality was reached [220].

In the same pH region, the R_g also appeared to increase from 32 Å to 37 Å and we noticed that its determination was less reliable in the pH range around 5.5. A plateau where the R_g value remained around 37 Å was observed approximately between pH 5.3 and 4.5. Then the protein's R_g significantly grew going from 37 Å at pH 4.5 to 46 Å at pH 3.5. At pH 2.8 it was only slightly incremented to 47.5 Å.

The $P(R)$ functions evidenced the increase of the protein's overall dimensions with decreasing pH, reaching a maximum distance around 180-190 Å at the lowest pH value of the experiments (Fig. 4.6, lower panels). We also noticed that the $P(R)$ function in nearly-neutral conditions already presented a tail at dimensions larger than 90 Å and this could be interpreted as the effect of the small amount of albumin dimers present in the solution, condition avoidable only when performing a SEC separation in-line with the SAXS measurement [39].

Summarizing the results of the four experiments at different % GdL by plotting them together, helped to assess the reproducibility of the process (Fig. ??). The fact

that the variation of the protein dimensions with decreasing pH had superimposable trends (especially in the pH region of steep increase between 4.5 and 3.5) in spite of the variable amount of GdL present in the sample, suggested that the presence of the lactone or its hydrolysis products at the concentrations used had not an impact on the protein conformation in acid conditions. This observation gave validity to the experimental conditions employed as a tool to actually probe the protein acid unfolding mechanism.

The main variation observed at different % GdL was in the absolute scattering intensity. An increased concentration of the additive determined in fact an increased scattering of X-rays by the solvent, as assessed by measuring the background scattering of the blanks. From the value of the density of an aqueous solution of GdL reported in the literature one could indeed determine that a 3.52% increase of electron density compared to pure water is expected for a GdL concentration of 11.16% [222]. Therefore a higher electron density of the solvent in the presence of GdL can be predicted, determining a decreasing contrast between the protein and the solvent at increasing GdL concentration in the buffer, as assessed by the decreased forward scattered intensity in our experiments.

Thanks to the simultaneity of the fluorescence and SAXS data acquisition, the variation of the structural features could be easily correlated with the observed blue shift of the tryptophan emission. By inspection of the wavelength of the emission maximum as a function of pH (Fig. 4.7a), one could assess that the blue shift was already completed when pH 4.5 was reached, in correspondence of the 37 Å R_g value and clearly before the major conformational variation probed by SAXS occurred. The aspect of the asynchronicity between the major unfolding process and the fluorescence blue shift could be pointed out also by the 2D-correlation analysis between SAXS and fluorescence data (Fig. 4.8).

Less noisy fluorescence emission spectra confirmed the occurrence of a blue shift with acidification and the pH range in which it took place (Fig. 4.7b). Furthermore the complementary experiment highlighted how no further relevant spectral changes were observed in the pH range below 4. A slight decrease of emitted intensity accompanied the blue shift, according to the spectra collected with a bench-top instrument. The fact that this effect was not evident in the SURF experiments could suggest that it was prevented by the exposure of a new volume of sample at each excitation of the fluorescence emission, or could be simply due to artifacts deriving from the non-optimal environment for performing spectroscopy (i.e. background environmental light and flow of sample).

4.3 Monitoring structural and spectral features of HSA as a function of pH

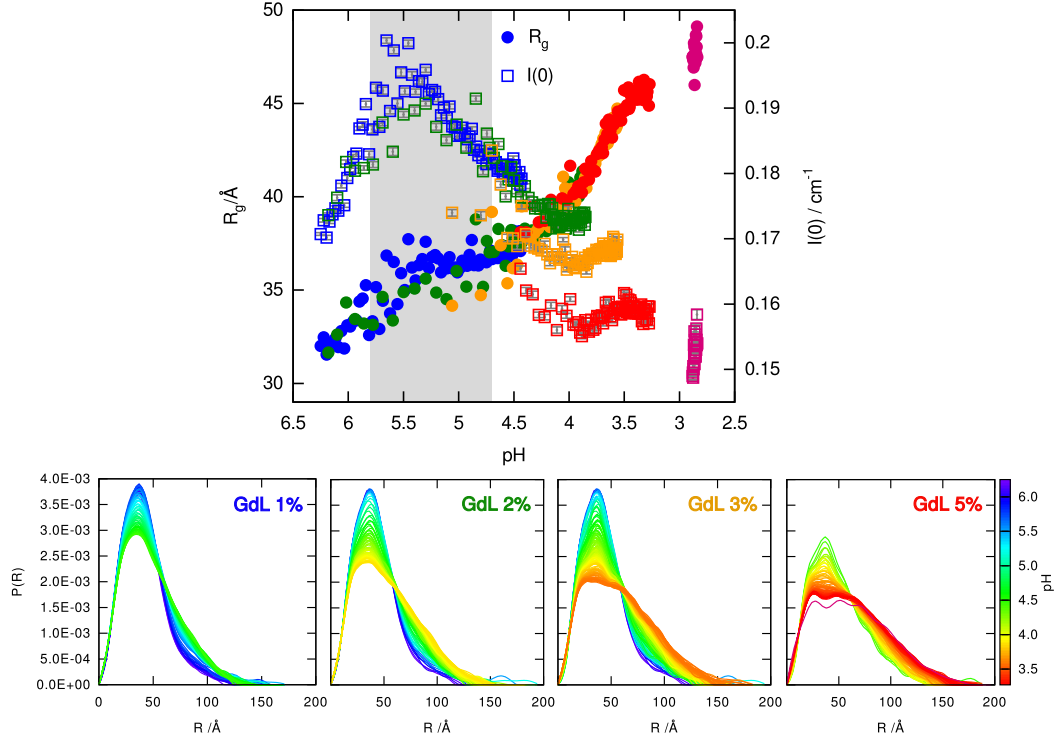


Figure 4.6. *Results of the SAXS analysis of HSA undergoing acid unfolding.* In the top panel, the R_g (dots) and $I(0)$ (empty squares) calculated by means of the Guinier approximation are plotted as a function of the pH value. The results from the SAXS data collected during the progressive acidification of the HSA samples with 1% (blue), 2% (green), 3% (orange) and 5% (red) added GdL are shown. For the 5% sample measured 90 minutes after the initial GdL solubilization, the purple color is used. The pH range approximately corresponding to the isoelectric point of HSA as reported in the literature is highlighted in gray. In the bottom panels, the $P(R)$ functions calculated by means of the IFT-BayesApp software are shown for the four samples, with a color palette indicating the pH.

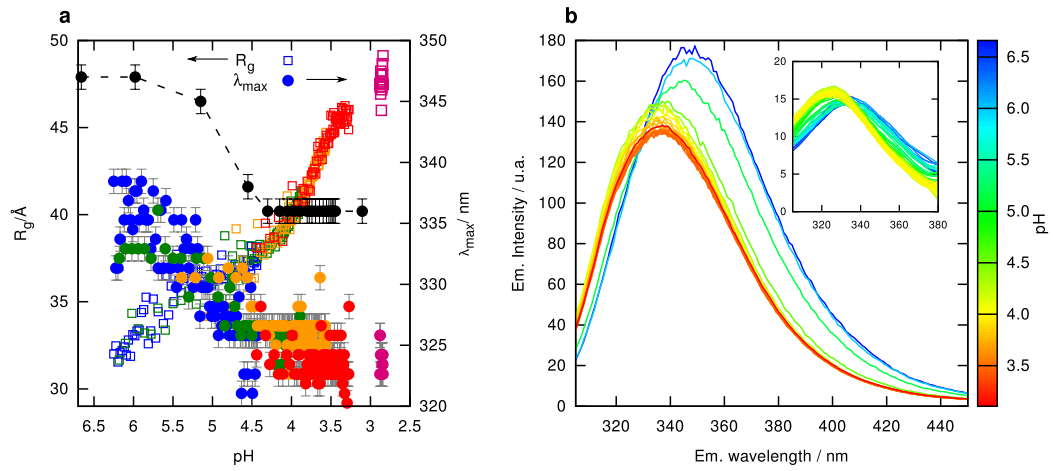


Figure 4.7. *Effect of the acid unfolding on the fluorescence of HSA.* (a) The variation of the maximum emission wavelength (dots) as a function of pH is shown, superimposed to the R_g values of Fig. 4.6 (empty squares) for ease of comparison. The color code is the same as Fig. 4.6; in addition the values obtained from an independent experiment with a bench-top instrument are shown as black dots. The absolute shift of the wavelength range was probably due to a not optimal calibration of the fluorescence optics of the SURF platform. The relative variation with pH is anyway the meaningful observation. (b) Fluorescence spectra of a HSA sample acidified with 2 % GdL as a function of the decreasing pH, as indicated by the color bar. The spectra were collected with a bench-top instrument. For comparison's purposes, the smoothed emission spectra collected by means of the SURF platform on an analogue sample are presented in the inset.

4.3 Monitoring structural and spectral features of HSA as a function of pH103

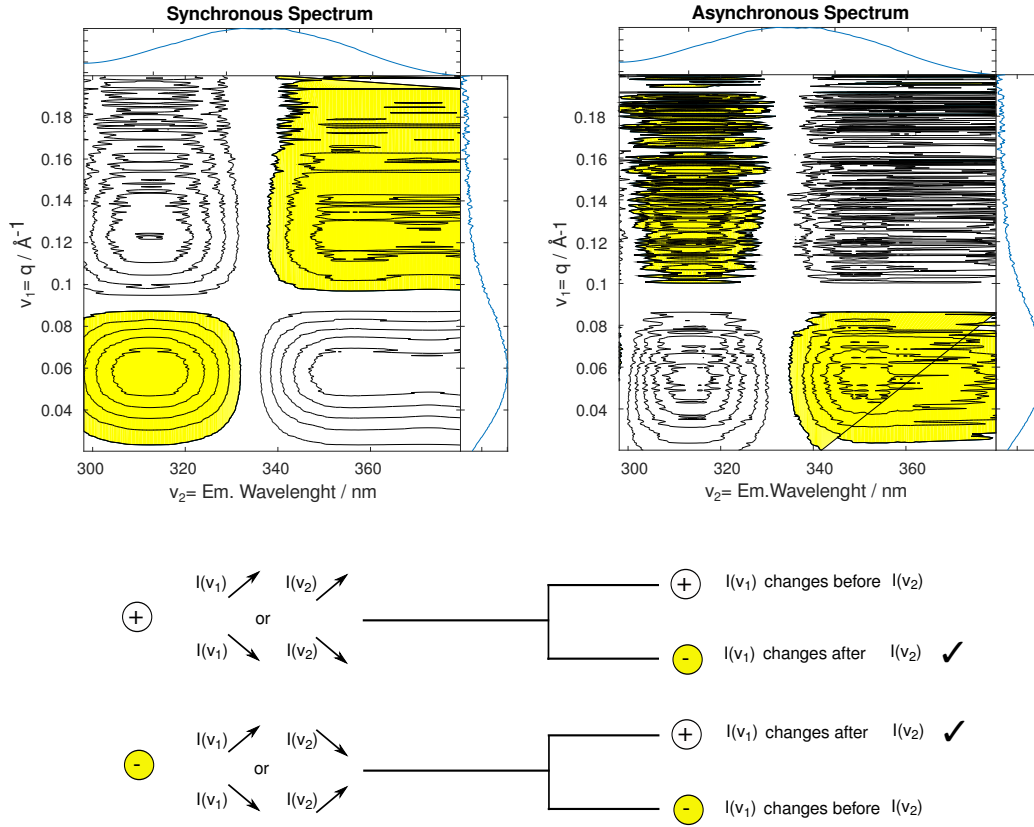


Figure 4.8. *2D-correlation between SAXS and fluorescence data pertaining to the HSA acid unfolding.* The hetero-correlation between the SAXS data (Kratky plots) and the smoothed fluorescence emission spectra collected during the decrease in pH of a HSA sample with 2% GdL is shown, presenting the synchronous correlation spectrum on the left and the asynchronous correlation spectrum on the right. In the bottom panel a guide to the interpretation of the relative signs of peaks found in synchronous and asynchronous correlation spectra is also presented. More explanations are given in section 1.3.

4.4 Looking for structural components in the acid unfolding

A selected subset of the SAXS data providing an evenly spaced sampling of the pH interval explored in the experiments was built, for the purpose of interpreting the conformational variations in term of a minimum number of structural components which could be characterized.

The more careful extraction of parameters from these selected SAXS data provided trends of the R_g , $I(0)$ and D_{max} as a function of pH which allowed for a cleaner interpretation compared to the results of the automatic analysis, sometimes affected by artifacts due to faulty background subtraction or excessive reduction of the considered q range (Fig. 4.9).

The presence of a pH interval located around pH 4.5 in which the R_g and D_{max} stabilized at values higher than those of the neutral form before the steep increase at lower pH, was more evident (Fig. 4.9a,c). This pH region corresponded to a shape of the dimensionless Kratky plot which testified a slightly increased flexibility and a higher anisotropy of the protein conformation [46] (Fig. 4.9d). The corresponding $P(R)$ function showed a peak only slightly shifted to lower distances compared with the native form (from 35 to 34 Å), but significantly lower, and a falling edge extending to larger distance values (60-140 Å), with a more asymmetric profile indicating a more elongated particle (Fig. 4.9e).

After the steep increase of the overall dimensions occurring between pH 4.5 and 3.5, the R_g and D_{max} trends seemed to start leveling off below pH 3.5. The dimensionless Kratky plot at this low pH values indicated a clearly unfolded conformer, though still presenting some features which would imply a residual structure. The comparison between the plots corresponding to pH 3.5 and 2.8 could suggest that the further lowering in pH determined a further disruption of local structure, with a minor impact on the overall dimensions. The $P(R)$ functions changed markedly in the pH range 4.5-3.5 indicating a protein expansion and reached a characteristic shape with an almost flat profile for distances between 20 and 60 Å at the lowest pH of the experiment.

The PCA decomposition (section 1.2) applied to the series of Kratky plots showed that all the first three independent components markedly contributed to the variation observed in the pH interval explored, being clearly not random. The fourth and fifth eigenvectors seemed to be less relevant and to mainly contribute to noise (Fig. 4.10). The same outcome was obtained when not including the additional data of the series corresponding to pH lower than 3. Even when restricting the pH range only to the interval where the “expansion” was observed (4.5-3.5), the number of significant components was still three. It is therefore probable that the acid unfolding could be better described as a coexistence of main characteristic conformations with contributions varying as a function of the protonation state, rather than assuming that a single conformer could exist at each pH value.

The constrained component analysis (MCR-ALS) helped find a possible model for the process, based on this assumption (Fig. 4.11). The best-fit model assuming three scattering components envisaged an intermediately unfolded conformation mainly populated between pH 4.5 and 3.5, while a more unfolded conformer appeared at

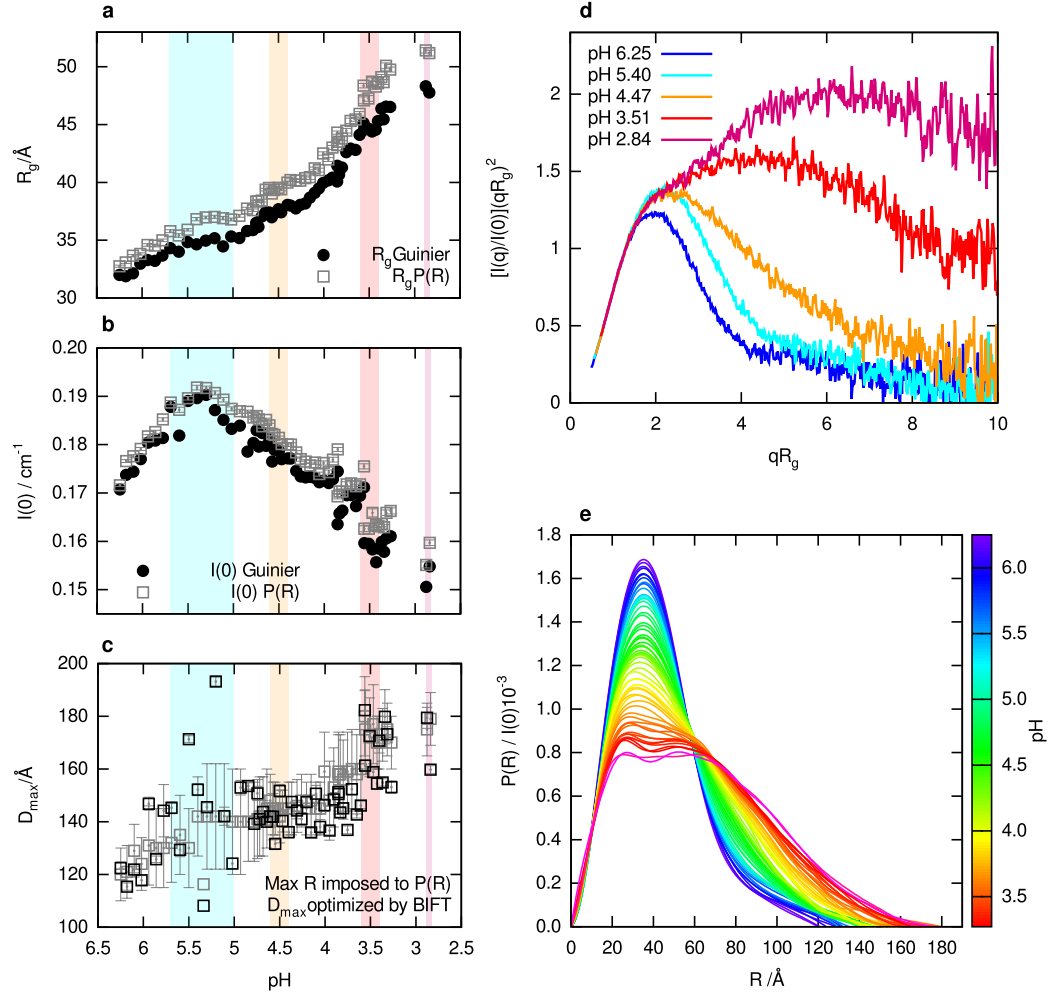


Figure 4.9. Results of the SAXS analysis of a selected subset of data from all the SURF experiments, spanning the available pH range. (a) R_g as a function of pH. (b) $I(0)$ as a function of pH. (c) D_{max} as a function of pH. (d) Dimensionless Kratky plots for five representative pH values: the reference at the highest pH of the experiment (blue), the region of the protein isoelectric point (cyan), the region where the blue shift of the fluorescence was completed (orange), the maximum expansion observed in the standard experiments (red) and the data collected at lower pH after additional waiting time (purple). Except for the starting point, the corresponding pH regions are highlighted with a colored background in (a), (b), (c). (e) $P(R)$ functions calculated with GNOM at the pH values indicated by the color bar. The $P(R)$ functions of the data collected after longer time in order for a lower pH to be reached are shown as purple and magenta lines. The corresponding pH values are not included in the color bar and are 2.88 and 2.84, respectively.

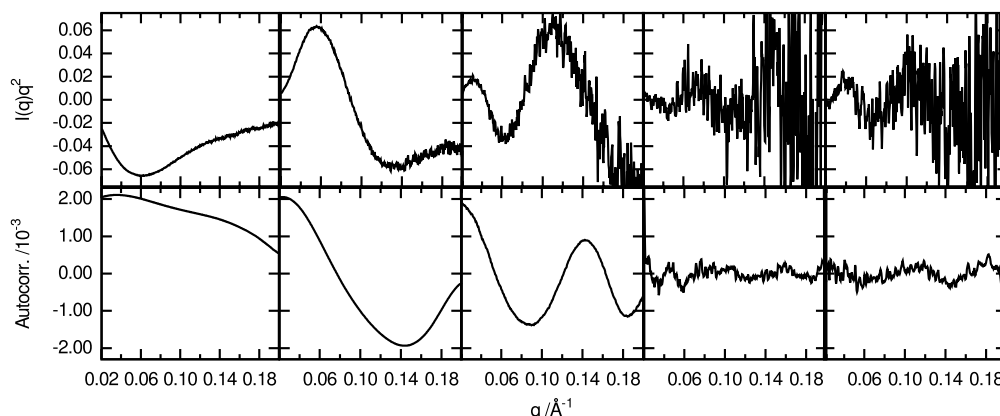


Figure 4.10. *PCA analysis of the HSA acid unfolding.* The first five eigenvectors (ordered from left to right) resulting from PCA of the pH-dependent SAXS data (in the form of Kratky plots) of HSA in the pH range 6.25-2.84 are shown. The bottom row corresponds to the vectors' autocorrelation functions.

the expense of the native folding. If a fourth component was included in the model, this contributed (together with the native form) only in the pH range around the isoelectric point, where attraction/dimerization could occur, with improvement of the fit of the data in this interval. The description of the unfolding occurring at lower pH values remained unchanged, with the third and fourth components indeed coinciding with the second and third of the three-component model.

The intermediate form retrieved by the self-modeling procedure could be approximately compared to the state of folding experimentally probed around pH 4.5, while the more unfolded form practically coincided with the conformation assumed at the lowest pH of the experiment (Fig. 4.11c,d).

The 3D-modeling based on these characteristic scattering profiles gave a better picture of what the main conformations identified in the acid unfolding could mean in term of the protein tertiary structure (Fig. 4.12 and 4.13).

The models retrieved from the intermediate form at pH 4.5 suggested a rearranged conformation more elongated compared to the hearth-shaped native form, in which the protein domains changed their relative positions but appeared to roughly maintain their identity and folding. In terms of the number of residues present in the distinguishable lobes of the *ab-initio* models (Fig. 4.13), it seemed that one half of the protein could remain more compact at the interface between two of the domains, while the other end-domain drifted apart from the compact region determining a more open conformation.

The unfolded conformer at lower pH was characterized by higher variability of the reconstructed model. It could be represented by means of a sort of flexible tube still preserving a certain degree of compactness at a local level. The protein could be described in these conditions as composed by partially folded loops kept together by the disulfide bridges and residual native interactions, interconnected by linker regions with flexible behavior, while a further disruption of local interactions could in principle occur if the pH would be further lowered below 3.

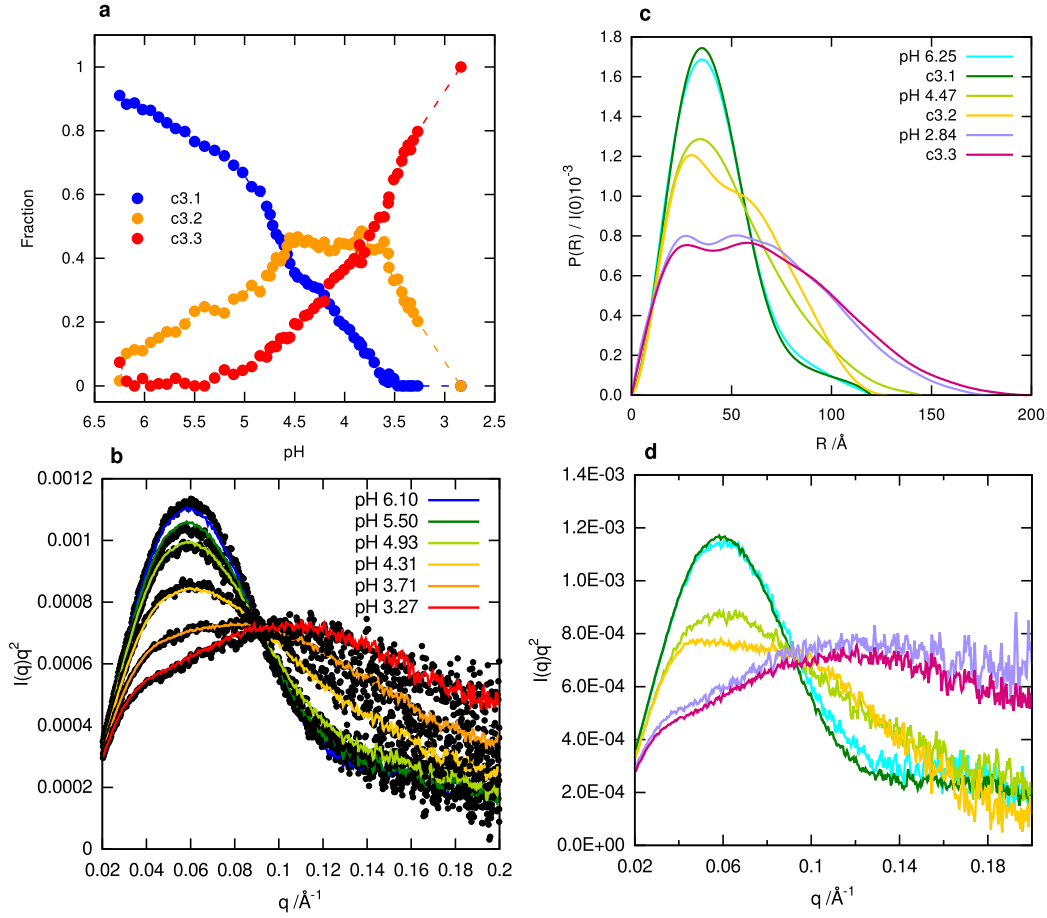


Figure 4.11. Results of the component analysis of the HSA acid unfolding by means of the MCR-ALS algorithm. (a) Relative contribution of the three structural components (c3.1, c3.2, c3.3) as a function of pH. (b) The approximation of the Kratky plot as a linear combination of the three components (solid lines) is compared to the experimental data (black dots) for six representative pH values. (c) The $P(R)$ functions of the three components (c3.1, c3.2, c3.3) and of selected experimental data at pH values where the components contribute at most are shown. (d) Kratky plots of the MCR-derived components and of the selected data shown in (c).

Table 4.1. Parameters of the ab-initio modeling of the partially unfolded conformers of HSA in acid conditions. The average of the normalized spatial discrepancy index (NSD) calculated from the superposition between the most representative model (Fig 4.12) and all the others obtained from repeated calculations is shown, together with the χ^2 value of the fit. The same parameters averaged over all the repeated calculations are presented, together with the standard deviation values.

ID of fitted data	$\langle \text{NSD} \rangle_{\text{repr.}}$	$\chi^2_{\text{repr.}}$	$\langle \text{NSD} \rangle_{\text{all}}$	$\langle \chi^2 \rangle_{\text{all}}$
pH 4.47	1.55	1.22	1.62 ± 0.04	1.26 ± 0.03
c3.2	1.57	1.31	1.67 ± 0.07	1.35 ± 0.06
pH 2.84	2.16	1.38	2.36 ± 0.09	1.43 ± 0.11
c3.3	2.39	0.75	2.56 ± 0.07	0.75 ± 0.06

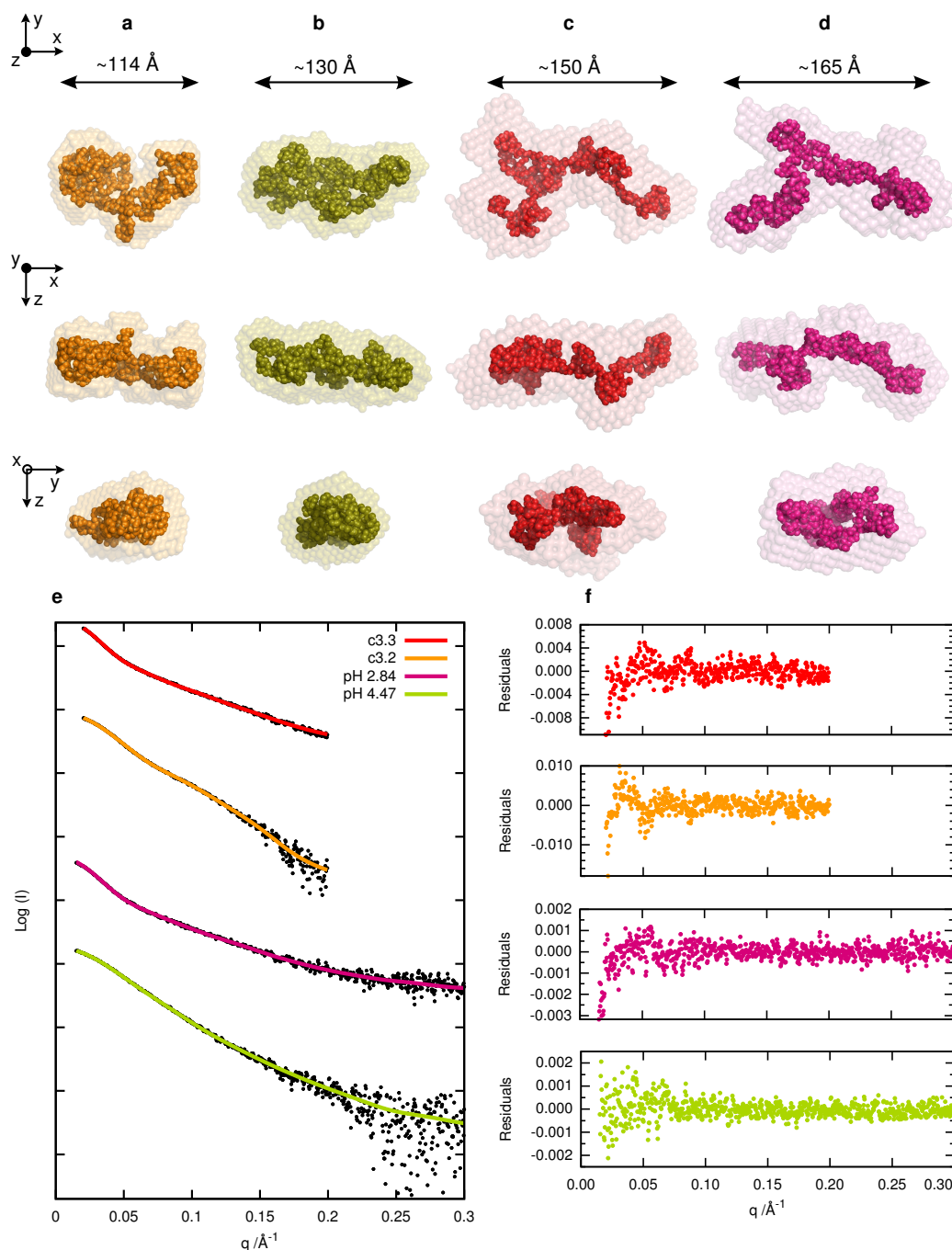


Figure 4.12. Results of the *ab-initio* modeling of the partially unfolded conformers of HSA in acid conditions. The *ab-initio* models were recovered by fitting both the scattering profiles obtained from the component analysis ((a) component 2, (c) component 3) and the experimental data collected at selected pH values ((b) pH 4.47, (c) pH 2.84). The most representative dummy-residue models (colored beads) are shown together with the envelopes produced by averaging the results of repeated calculations (transparent surfaces). Three orthogonal views as indicated by the reference axes are shown. In (e) the theoretical scattering intensity for the shown models corresponding to component 3 (c3.3, red line), component 2 (c3.2, orange line), pH 2.84 (purple line) and pH 4.47 (green line) is shown together with the data to be fitted (black dots). In (f) the residuals (difference between experimental and model intensities) are displayed.

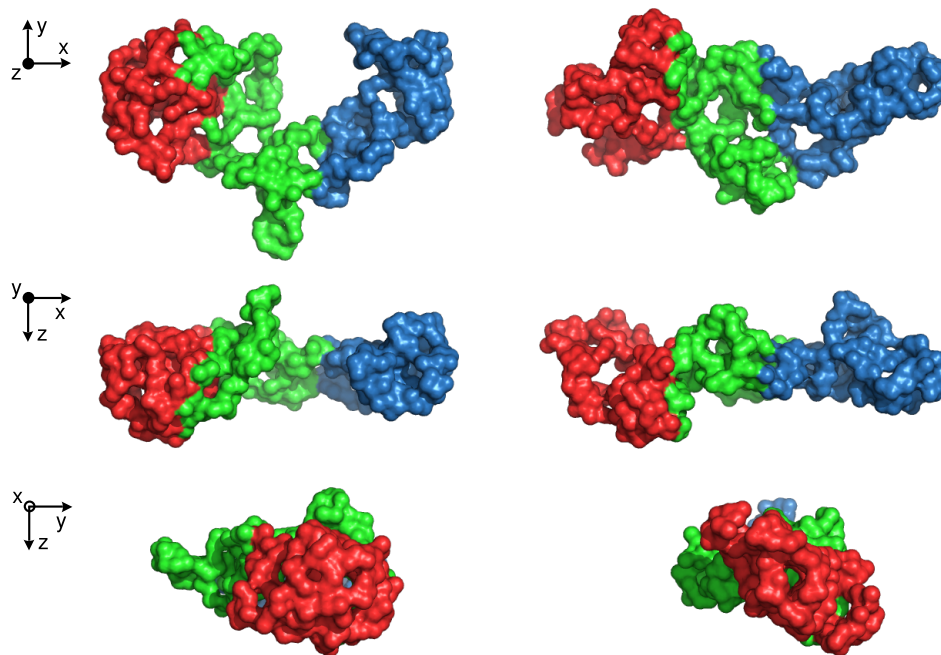


Figure 4.13. *Ab-initio* models of the intermediately unfolded conformer of HSA at pH 4.5

The *ab-initio* models recovered by fitting both the scattering profile of component 2 and the experimental data collected at pH 4.47 are reproduced here by coloring the dummy-residues in a way that could represent the three-domain structure of the protein.

4.5 Summary and discussion

In the experiments described in this Chapter, for the first time the process of acid unfolding of albumin was followed across a continuous pH decrease with the SAXS technique. The simultaneous collection of the fluorescence spectra allowed for a correlation between the structural modifications of the protein and the spectral changes known to occur during the albumin isomerization.

This allowed for a structural characterization of the two acid isomers of albumin, the F and E forms, the existence of which is widely established in the literature.

The picture emerging from our analysis was actually a rather continuous opening of the protein molecule with decreasing pH below 4.5. These elongation and increase of flexibility were interpreted as the gradual occurrence of conformers in which the subdomains of the protein could drift apart from each other as a consequence of electrostatic repulsion, and that reached a major contribution at pH lower than 3.5.

Before this happened, it seemed that a partial unfolded isomer was populated, starting from pH around 4.5. The three protein's domains rearranged in a more open and elongated flexible conformer maintaining their identity, or maybe with just a partial unfolding of one of the end-domains. This structure could be interpreted as the F form which would represent the first structural response of the protein molecule to the changed protonation conditions. This first rearrangement in acid environment determined an abrupt modification of the tryptophan surroundings, which became buried in an hydrophobic pocket, being rather exposed to the polar solvent in the

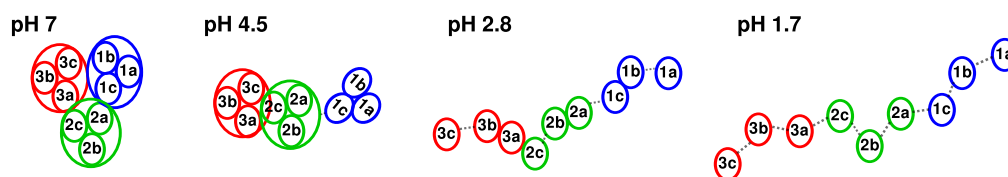


Figure 4.14. Schematic view of the HSA acid unfolding process as opening of the subdomain structure

native conformation at neutral pH. The tightly packed interface between the central domain and one of the end-domain deduced on the basis of the *ab-initio* model could provide an explanation for the occurrence of such an environment. In addition a general de-hydration of the protein pockets had been hypothesized to occur at low pH to explain measurements of compressibility [215]. The suggested model for the F form extracted from the SAXS data would be approximately in agreement with the speculative model which suggested the N-F isomerization as a separation of the two halves of the molecule [106]. It also resembled the conformation presented as a result of molecular dynamic simulations of HSA in acid conditions, in which the electrostatic repulsion triggered the increase of distance between domains I and III [217].

The E form had been already characterized by means of SAXS in the past [92], as existing at pH 1.7 reached with the addition of 20 mM HCl to a phosphate buffer solution. In that case larger dimensional parameters ($R_g = 66$ Å and $D_{\max} = 220$ Å) were found compared to the values measured in these experiments ($R_g = 52$ Å and $D_{\max} = 190$ Å) and the $P(R)$ function and Kratky plot seemed to show a slightly more pronounced unfolding of the protein. R_g values could be predicted for a completely unfolded albumin (585 amino acids) according to the Flory scaling laws of flexible polymer chains ($R_g = N^\nu$, where N is the number of residues, R_0 is a constant and ν is the scaling factor) [230]. According to the parameters found for denatured proteins [231] R_g values between 63 and 118 Å could be calculated, while a range between 66 and 76 Å could be estimated on the basis of the parameters for intrinsically unfolded proteins [232]. It was clear that the dimensions assumed by albumin in acid conditions are not compatible with such a model, as already pointed out [92]. In extremely acid conditions, the albumin E form should correspond to the maximum elongation allowed by the disulfide bonds pattern, with the protein behaving as a linear sequence of its nine loops probably still in a coiled conformation, interconnected by flexible linker regions. It could be reasonable to assume that the lowest pH value accessible in the time of the experiments performed did not determine the maximum expansion possible for albumin in acid conditions. A possible addition to the picture mentioned above, given by the experiments at varying pH, could be that when the electrostatic repulsion has not yet reached the highest possible level, more inter-subdomains interactions can still survive. A cross-section R_g could be measured for such elongated structures, and a thicker average R_g of the cross-section could be calculated at pH 2.8 (19.5 Å) compared to that found at pH 1.7 (11 Å) [92], indicating a flexible chain of loops presenting still more compact regions (Fig. 4.14).

The results presented in this Chapter were not considered to be a completely

definitive structural interpretation of the acid unfolding process of albumin, still provided an additional experimental insight. The simultaneous SAXS-spectroscopic *in-situ* characterization ensured by the time-dependent acidification constituted the main element of novelty in the study of the protein's conformational variations. Another important remark compared to some approaches presented in the past in the literature is that the method employed avoided the necessity to change the chemical composition of the buffer across the most critical pH interval [205]. These results could be enriched by means of scattering data at even lower pH values and also by the use of modeling techniques based on the known high resolution structure of the protein [233,234] also in the context of an ensemble-based approach [70] to represent the partially unfolded conformers.

Chapter 5

SAXS analysis of Calvin cycle enzymes and their regulatory complexes

5.1 Introduction

Oxygenic photosynthetic organisms (land plants, algae and cyanobacteria) are responsible for the fixation of atmospheric CO₂ in the form of organic compounds (mainly sugars) by means of the Calvin-Benson cycle, a process that is at the basis of our food chain. This metabolism takes place in the chloroplasts and it is tightly linked to the light reactions of photosynthesis since the reducing and chemical-energy equivalents produced by the electron transport in the form of NADPH and ATP are then consumed in the reactions of the Calvin-Benson cycle to synthesize organic molecules in which the atmospheric CO₂ is reduced [235].

In order to maintain its efficiency under the fluctuating environmental conditions, the regulation of the Calvin-Benson cycle must be responsive to changes both in the supply of enzyme cofactors regenerated thanks to the electron transport and in the demand for carbon skeletons. As a result, regulatory mechanisms exist allowing for a modulation of enzyme activities in response to changes in light intensity during the day and the complete suppression of the Calvin-Benson cycle at night [236]. Investigation of the molecular mechanism of this light-dependent regulation led to the identification of the ferredoxin/thioredoxin system [237], by means of which the photosystem I, one of the two photosystems of the light-dependent electron transport chains of higher plants, keeps reduced a pool of thioredoxins able to interact with some enzymes and induce dithiol/disulfide exchange reactions. The enzymes of the Calvin-Benson cycle regulated by light were shown to contain regulatory cysteine residues which can be oxidized to disulfides in the dark and reduced in the light by thioredoxin. This reduction allows for a transition from a form with low or no activity, to a fully active enzyme [238]. The picture of the regulation mechanism is far more complicated since a number of other redox post-translational modifications were suggested to contribute [239]. In addition protein-protein interactions provide an additional level of regulation over enzyme activities. The formation of supramolecular complexes, possibly influenced by the presence of enzyme ligands, was indeed observed

to accompany the transition to the inhibited states of the enzymes in particular cases [6], which are the object of the structural studies in solution presented in this Chapter.

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and phosphoribulokinase (PRK) are two enzymes involved in the photosynthetic carbon fixation. GAPDH is responsible for the single reducing step of the cycle. Differently from the analogous GAPDH present in the glycolytic pathway, the photosynthetic GAPDH can use for the reaction both NADH and NADPH as coenzymes, but the NADPH-dependent activity is the one regulated according to the illumination conditions [6]. PRK catalyzes the regeneration of the acceptor of CO₂, ribulose-1,5-bisphosphate. Through their activities most of the NADPH and ATP produced by the electron transport chain are consumed.

Though GAPDH and PRK catalyze two non-consecutive reactions of the Calvin-Benson cycle, they have been found to form, together with the small regulatory protein CP12, a multi-enzyme complex constituted by two GAPDH tetramers of the gapA isoform (A₄-GAPDH) and two PRK dimers, in addition to four molecules of the CP12 protein [240, 241]. CP12 is an intrinsically disordered protein of about 80 residues universally distributed in oxygenic photosynthetic organisms [242] which has two pairs of cysteine residues that can be reversibly oxidized to disulfide bridges, allowing for its regulatory function as scaffold for the complex formation, in dependence of the redox conditions [241]. PRK dimers are also redox-sensitive, being their activity reversibly inhibited upon oxidation of cysteine residues found in the active site [243]. The assembly and disassembly of this complex has a regulatory role, being the A₄-GAPDH and PRK activities switched off in response to complex formation, which occurs in the dark [244].

In the case of the proteins from *Arabidopsis thaliana* the pathway of the complex assembly has been elucidated. First, two CP12 molecules in the oxidized form and an A₄-GAPDH tetramer, only if binding NAD(H) as coenzyme [245, 246], interact together forming a *binary complex*. The crystal structure of the binary complex was solved [247]. Thanks to this structural insight, and in combination with nuclear magnetic resonance studies of the oxidized CP12 in solution, it has been proposed that the mechanism of interaction between A₄-GAPDH and CP12 begins with a conformational selection exerted by A₄-GAPDH on the most abundant conformer assumed by the oxidized CP12 in solution; upon binding, CP12 is induced to fold through a zippering process that leads its C-terminal tail to fit into the active sites of A₄-GAPDH [247]. The N-terminal regions should remain unstructured and flexible since it was not possible to assign the corresponding coordinates in the crystallographic analysis due to lack of suitable electron density. Two A₄-GAPDH/CP12 binary complexes can thereafter interact with two PRK dimers in the oxidized form, resulting in the final assembly of the *ternary complex* [241, 245, 248].

Besides the homotetrameric isoform A₄-GAPDH, the activity of which is regulated through the CP12-dependent mechanism described, higher plants possess a second isoform made up of gapA and gapB subunits (AB-GAPDH) [249]. The main difference between the A and B subunits is the presence in the B subunits of a C-terminal extension (CTE) of 30 amino acids, highly homologous to the C-terminal portion of CP12 [250]. Similarly, also the CTE is devoid of well-defined structure and contains a pair of redox-sensitive cysteines. When in the oxidized form, the CTE inhibits

the NADPH-dependent activity of the enzyme by hindering the coenzyme binding site [251], as suggested by the crystal structure of the A₂B₂-GAPDH tetramer that was solved in the case of the protein from spinach [252]. In addition, upon oxidation of the CTE and binding of the less cumbersome NAD(H) in place of NADP(H), AB-GAPDH was found to be self-associated in higher molecular weight oligomers observed to accumulate in chloroplasts in the dark, having suggested stoichiometry A₈B₈ and lower NADPH-dependent activity [253, 254]. The presence of the CTE, is therefore responsible for the autonomous (CP12-independent) regulation of the A₂B₂-GAPDH tetramer [250, 255].

On the other hand, a single isoform of the photosynthetic PRK enzyme is known from green algae to land plants. It is a dimeric form made of 40 kDa subunits, which has not been structurally characterized yet. The unique crystal structure available for a PRK protein is from the anoxygenic procariote *Rhodobacter sphaeroides*, which differently from eukaryote PRK is an octamer of 32 kDa subunits [256] sharing only 19% sequence identity and is not redox-regulated. Eukaryote PRK is instead directly regulated by thioredoxins through two cysteine residues located in the enzyme active site [257, 258] and it may be further regulated through the formation of the ternary complex with A₄-GAPDH and CP12 [244, 245, 248, 249, 259, 260].

A scheme summarizing these self-association behaviors regulated by the redox (and so illumination) conditions is depicted in Fig. 5.1, which was adapted from Ref. [6].

The detailed molecular interactions which determine the specific assembly of the ternary A₄-GAPDH/CP12/PRK complex and of the AB-GAPDH oligomers have not yet been revealed. The determination of the complexes' structure would be crucial for the understanding of the photosynthetic metabolism in the light/dark regime, but crystallization trials to produce single crystals for X-ray diffraction experiments failed. A structural study in solution by small angle X-ray scattering (SAXS) was then approached and some results are reported in the following sections.

In the case of the A₄-GAPDH/CP12/PRK complex, a SAXS analysis of both the supramolecular assembly and of its components alone was performed on the recombinant proteins of the land plant *Arabidopsis thaliana*. In particular, a characterization of the PRK dimer alone was needed in order to have a piece of structural information, even at low resolution, due to the lack of a credible homology model.

The characterization of the AB-GAPDH isoform from *Spinacia oleracea* in either inactivating or activating conditions revealed a more complex picture of its oligomerization behavior from the structural point of view than commonly assumed to explain the observed changes of enzymatic activity.

In both cases the structural study in solution showed the dynamic nature of the supramolecular complexes, which probably had hampered their study by means of protein crystallography.

5.2 Experimental procedures

5.2.1 Sample preparation

The preparation of the purified protein samples for SAXS analysis was performed by the laboratory of molecular plant physiology of Prof. Francesca Sparla, Department

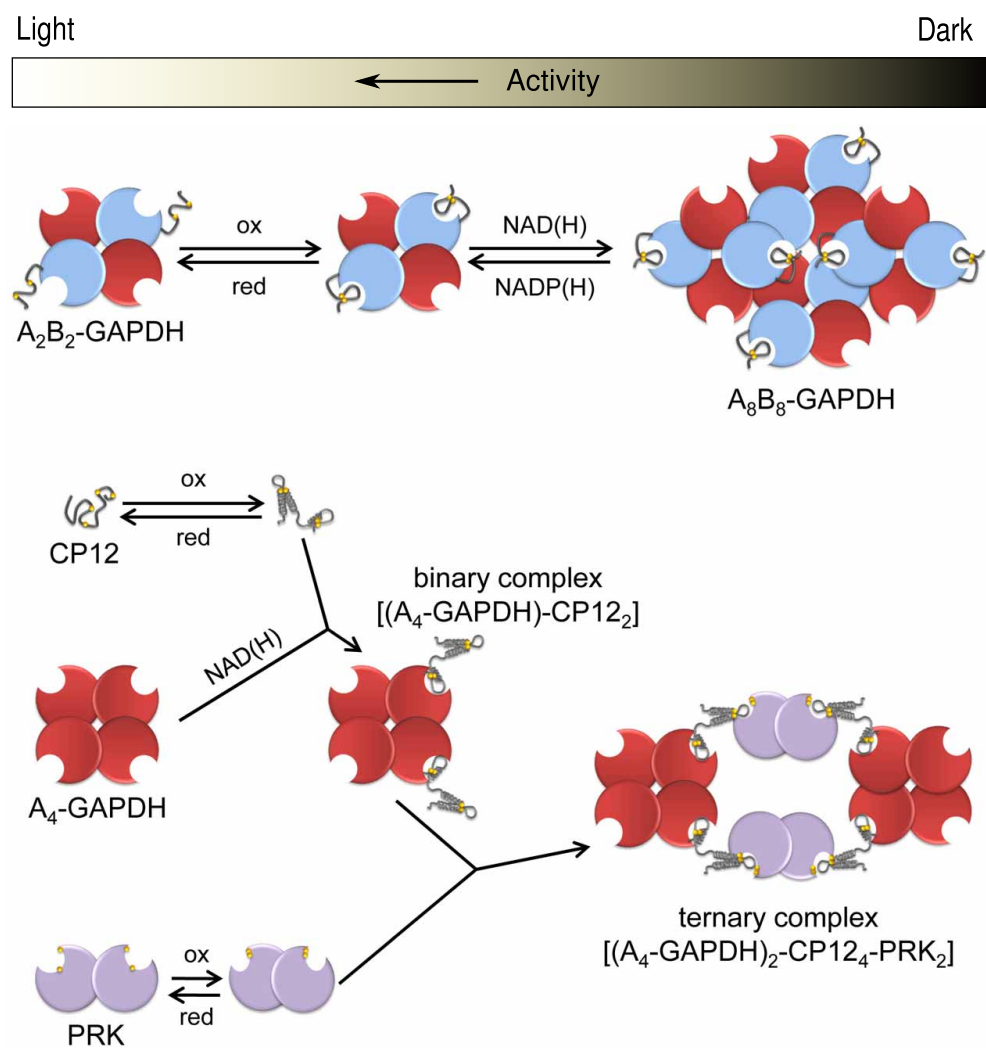


Figure 5.1. A schematic view of the regulation and interactions of chloroplast A₄-GAPDH, CP12, PRK and A₈B₈-GAPDH. The scheme was readapted from Ref. [6].

of Pharmacy and Biotechnology, University of Bologna.

A₄-GAPDH/CP12/PRK Recombinant GAPDH (EC 1.2.1.13), PRK (EC 2.7.1.19) and CP12 (isoform 2) of *Arabidopsis thaliana* without the signal peptide for localization in the chloroplasts which is absent in the protein mature forms, were expressed and purified as described in Ref. [241]. Different aliquots of pure PRK were treated with 5 mM reduced or oxidized dithiothreitol (DTT) to ensure the presence of a unique redox form. After overnight incubation at 4°C, samples were desalted and concentrated to 10 g/l using a Centricon (Millipore) device with a 10 kDa cutoff, in 25 mM potassium phosphate buffer at pH 7.5. The A₄-GAPDH/CP12 binary complex was obtained after overnight incubation at 4°C of pure A₄-GAPDH and CP12 in a 1:1 ratio (on a subunit basis, a twice-molar excess of CP12) in the presence of 5 mM NAD and 5 mM oxidized DTT. To remove excess CP12, the sample was concentrated to 200 μl using a Centricon (Millipore) device with a 3 kDa cutoff and then loaded onto a gel-filtration column (Superdex 200 10/300 GL; GE Healthcare) pre-equilibrated with 25 mM potassium phosphate buffer, pH 7.5. The eluted A₄-GAPDH/CP12 binary complex was collected, 1 mM NAD was immediately added and the sample was brought to a final concentration of 10 g/l. The A₄-GAPDH/CP12/PRK ternary complex was obtained after overnight incubation at 4°C of the three purified protein components in the presence of 5 mM NAD and 5 mM oxidized DTT. A₄-GAPDH, PRK and CP12 were present in a 2:1:2 ratio (on a subunit basis) and excess CP12 was removed by gel filtration as described above for the binary complex. The eluted ternary complex was added with 1 mM NAD and brought to a final concentration of 10 g/l. For the binary and ternary complexes, the protein concentration in the stock solutions was estimated by means of the BCA assay [261]. All stock solutions were stored at -80°C and then thawed on ice and diluted to selected concentrations in 25 mM potassium phosphate buffer at pH 7.5 (containing also 1mM NAD for the binary and ternary complexes) prior the measurements. The concentration of the PRK samples was measured by means of the UV absorbance at 280 nm with a nanodrop instrument, while it was estimated on the basis of dilution for the A₄-GAPDH/CP12/PRK and A₄-GAPDH/CP12 samples.

AB-GAPDH AB-GAPDH was purified from spinach chloroplasts as described in Ref. [262]. After initial purification, the sample was split in aliquots in which solution conditions that should either stabilize the inactive oligomeric forms or induce the dissociation to the fully active A₂B₂ form were maintained. In both cases a size exclusion chromatography (SEC) run (Superdex 200 10/300 GL; GE Healthcare) was performed, collecting selected fractions in the expected range of elution volume. Stock solutions were prepared by concentrating the selected fractions above 10 g/l, and the concentration was measured by means of the BCA assay [261].

The conditions inducing dissociation into A₂B₂ tetramers are referred in the following also as “A₂B₂” sample and were aimed to ensure: the binding of NADP in place of NAD; the reduction of the disulfide bridge in the CTE; the binding of the enzyme substrate 1,3-bisphosphoglycerate [249]. The “A₂B₂” sample conditions, unless otherwise specified, implied the presence in solution of: 20 mM NADP, 5

mM reduced DTT, and 1,3-bisphosphoglycerate. The last compound, substrate for GAPDH, is not a commercial product, therefore a reaction mixture had to be added in order to directly generate it in solution. This was given by 5 mM MgCl_2 , 15 mM 3-phospho-glycerate, 10 mM ATP and 20 units/ml (roughly 0.01 g/l) of commercial Phosphoglycerate kinase (product type P7634, from Sigma), the enzyme which catalyzes the formation of 1,3-bisphosphoglycerate.

In some cases only some of these conditions were applied. Indeed a certain variability was observed in the tendency of AB-GAPDH to oligomerize depending on the different batches of starting material (commercial fresh spinach): in the case of the samples directly measured in sample-changer mode the incubation conditions adopted to stabilize the dissociated form of the enzyme involved only the presence of the NADP cofactor at 1 mM concentration, whereas for the samples analyzed in the SEC-SAXS experiments the full set of disaggregating conditions described was needed in order to obtain an approximate hydrodynamic radius from the dynamic light scattering (DLS) quality control which indicated dissociation.

The inactive oligomeric form of AB-GAPDH was stabilized by addition of 1 mM NAD and these conditions are hereafter referred as “ A_nB_n ” sample. The samples were incubated overnight at 4°C in the selected conditions, checked by means of DLS and then stored at -80°C. A 25 mM potassium phosphate buffer at pH 7.5 was used.

Two “ A_2B_2 ” samples were analyzed by SEC-SAXS. The stock solution which had been incubated overnight in the dissociating conditions prior to freezing, was centrifuged at 15,000 rpm for 15 minutes. The obtained sample “ A_2B_2 -inc.” had an estimated concentration below 5 g/l. Another sample was kept in the dissociating conditions only for 2 hours before the SEC-SAXS experiment, at room temperature, and its final concentration was about 11 g/l (sample “ A_2B_2 -fresh”).

5.2.2 SAXS data acquisition and analysis

The SAXS data were collected at the BioSAXS beamline BM29 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France [263]. The sample-to-detector distance was 2430 mm and the X-ray wavelength 0.99 Å. The accessible q range was 0.005-0.45 Å⁻¹.

Sample Changer mode

The samples were analyzed by means of the automatized robot sample changer (SC) available at the beamline [42]. In the case of the study of the A_4 -GAPDH/CP12/PRK and its components, flushed volumes of 50 µl were used for each protein sample and a set of 10 consecutive 1-s exposures was made during sample flowing in the capillary with an extra flow-time of 10 s. In the case of the study of the AB-GAPDH isoform, volumes of 60 µl and 2-s exposures were used. The frames were automatically compared to assess radiation damage and then averaged. The scattering contribution of the capillary filled with buffer was subtracted and the intensity was divided by the protein mass concentration and converted to approximate kDa units by using water scattering as a standard and the protein specific volume value of 0.735 cm³g⁻¹ [31] (section 1.1.2). Two repetitions of the measurement procedure for each protein

concentration were run and the data averaged. From the subtracted scattering profiles the $I(0)$ and radius of gyration (R_g) were calculated using the Guinier approximation (eq. 1.23, page 9) as implemented in PRIMUS [30,66]. The model-independent analysis to obtain the $P(R)$ function, with an estimate of the maximum particle dimension ($D_{\max}$), in addition to an independent calculation of $I(0)$ and R_g , was performed using GNOM [14]. This fit of the experimental data was used as a good reference in the selected angular range for the model-dependent fitting attempts.

The molecular weight was estimated from (i) the $I(0)$ value already converted by the beamline pipeline in approximate kDa units as stated before, (ii) the Porod volume (V_P , section 1.1.2) according to the proportionality empirically found for roughly globular proteins ($MW \approx 0.625 \cdot V_P$) [30], (iii) the average excluded volume of dummy-atom models (V_d) computed using DAMMIF as $0.5 \cdot V_d$ [30], (iv) the invariant volume-of-correlation length (V_c) through a power-law relationship between V_c , R_g and MW that has been parametrized [25] and (v) a method based on an empirical relation to the Porod invariant estimated with a truncated integral [48].

Size Exclusion Chromatography mode

Combined SEC-SAXS experiments were performed for the study of the AB-GAPDH isoform in the conditions described in paragraph 5.2.1. For AB-GAPDH in “A_nB_n” conditions, 100 μ l of a 13.1 ± 2.6 g/l solution were injected into the column and eluted in buffer containing NAD 0.1 mM. For AB-GAPDH in “A₂B₂” conditions, for both the “A₂B₂-inc.” and the “A₂B₂-fresh” samples 200 μ l were injected into the column and eluted in buffer containing NADP 0.1 mM. The separation was run at a flow rate of 0.5 ml/min in a Superdex 200 10/300 GL column (GE Healthcare) attached to a HPLC system (Shimadzu), which had been pre-equilibrated with the appropriate buffer.

The chromatogram of the eluting samples was recorded by means of an UV-vis diode array detector before directing the elution path to the capillary for SAXS data collection. SAXS frames obtained by 1-s exposure of the capillary were acquired continuously. The automatic pipeline for SEC-SAXS data analysis implemented at BM29 helped in assessing the quality of the data collected during the experiment [264] and allowed for a preliminary identification of the chromatogram regions in which constant scattering profiles could be found. Afterwards a classification of the collected frames as buffer or protein frames was performed on the basis of the SAXS intensity trace; the statistical test implemented in CorrMap [27] aided by visual inspection was used to choose the superimposable buffer intensity profiles. The averaging of the buffer scattering data, the subtraction of the averaged buffer intensity from the protein data and an automatic analysis of the subtracted protein profiles was performed with a Matlab script as described in section 3.2.7. The frame numbers were converted into retention volumes considering the delay between the injection of the sample into the column and the starting time of the SAXS exposure series. Protein frames giving constant R_g values were scaled to the same intensity of the elution maximum, checked according to the statistical test [27] and then averaged in order to obtain a single representative scattering profile with better signal to noise ratio, presented in the results. For “A₂B₂-inc.” a scattering profile was also

obtained by merging the SEC-SAXS data (up to $q=0.05 \text{ \AA}^{-1}$) with the data acquired in sample changer mode on a particular fraction (between 14 and 14.5 ml) collected in the A_2B_2 SEC separation in the lab (starting from $q=0.079 \text{ \AA}^{-1}$), and by using the average of the two superimposed data set in the range $0.05\text{-}0.079 \text{ \AA}^{-1}$.

3D-modeling based on SAXS data

The programs of the ATSAS package [30,66] were used in order to build 3D-models in agreement with the SAXS data.

PRK For PRK, the *ab-initio* method implemented in the program GASBOR [57] was employed, since this software uses the average scattering factor of amino acids and enables the modeling of the particle as a protein chain-like assembly of dummy-residues, the number of which is constrained according to the known sequence of the protein. The sequence and the homodimeric state of PRK (349 amino acids $\times$ two monomers = 698 amino acids) were given as inputs and 20 models were generated imposing either P2 or no symmetry. The dummy-atom method implemented in DAMMIF [56] was also applied for fitting the data collected from both the oxidized and the reduced form of PRK, performing 20 calculations with both P1 and P2 symmetry.

A₄-GAPDH/CP12 Rigid-body modeling of the GAPDH/CP12 complex was performed by the program CORAL [30]. A twofold symmetry was applied to a pair of A-GAPDH subunits, one binding CP12 and one free. In order to build the complete complex, the missing CP12 amino acids were generated as dummy-residues. Crystallographic contacts between the A subunits highlighted by the Contact Map Analysis software [265] were imposed as constraints in the optimization. EOM [69,70]) was then used to assess the flexibility of the CP12 N-terminal region (residues 1-57) and to model the binary complex. A pool of conformers for the missing CP12 amino acids were randomly generated as a chain of dummy-residues. The structures with unacceptable superimpositions between dummy-residues and the high resolution A₄-GAPDH coordinates were discarded. From a starting random pool consisting of about 10,000 conformers, an ensemble was selected in order to best fit the experimental SAXS curve. A model recurrently selected in the final ensemble was chosen as representative. Also an *ab-initio* reconstruction was performed, using DAMMIN and generating 12 models.

A₄-GAPDH/CP12/PRK For the A₄-GAPDH/CP12/PRK complex, the software SASREF [63] and the implemented version SASREFMX [172] that takes into account partial complex dissociation were used to perform a rigid-body modeling using as subunits the A₄-GAPDH and PRK 3D-models obtained as described above. Some constraints were imposed in the calculations. The stoichiometry suggested a twofold symmetry of the system and mild contacts were imposed in order to keep every CP12 bound to GAPDH not further than $10 \text{ \AA}$ from one of the two halves of the PRK model. This condition is based on the knowledge that PRK has no significant affinity for A₄-GAPDH alone, but only for the A₄-GAPDH/CP12 complex [241,245]. An *ab-initio* model of the ternary complex was computed with

the software DAMMIN using a starting model obtained applying DAMAVER [65] to a number of similar rigid-body structures with a cutoff volume equal to 500 nm^3 . Also calculations with the default settings of DAMMIF [56] were performed. When required, the calculation of model intensities from atomic structures was performed by CRY SOL [60] using 24 spherical harmonics, 101 data points in the q range $0\text{--}0.35 \text{ \AA}^{-1}$ and the other parameters as default.

AB-GAPDH The scattering pattern ascribed to the A_8B_8 oligomer was modeled by means of a rigid-body approach using the structural information available for the A_2B_2 -GAPDH tetramer of *Spinacia oleracea* in oxidized form [252]. The software SASREF was employed applying different conditions, listed here in decreasing order of degrees of freedom: P2 symmetry applied to two A_2B_2 tetramers (tetramer formed by the chains O,P,Q, and R in the 2PKQ pdb entry [252]) provided as independent subunits in order to build $(A_2B_2)_4$; no symmetry imposed to two $(A_2B_2)_2$ rigid subunits; P2 symmetry applied to one $(A_2B_2)_2$ rigid subunit. The first minimization procedure started from a spatial arrangement of the subunits having already a degree of agreement with the experimental data, which had been obtained with the tool GLOBSYMM [63]. Then the optimization was repeated starting from different initial arrangements found from preliminary runs, performing ten minimizations in each condition. The same procedure was also applied using a rearranged tetramer as building block, obtained by fitting the scattering data of A_2B_2 as described below. In these preliminary attempts, the fact that the high resolution structures employed as subunits were missing up to 28 residues of the CTE, was neglected. Overall this led to have a maximum of 4% of the amino acid mass not represented in the models. The theoretical scattering curves corresponding to the optimized models was also calculated with an alternative version of the software CRY SOL (crysol_30), in which a layer of water molecules is generated surrounding the all-atom model in order to attempt a more realistic representation of the protein hydration and of the excluded volume in these multi-protein complexes, since the SASREF algorithm implicitly assumes each subunit to conserve its own solvation layer in the complex formation, in order to reduce the computing effort. The calculation was performed using 15 spherical harmonics, 61 data points in the q range $0\text{--}0.3 \text{ \AA}^{-1}$ and the other parameters as default.

The crystal structure of the A_2B_2 -GAPDH tetramer in oxidized form [252] was compared with the scattering pattern ascribed to A_2B_2 . In addition, an ensemble-optimization approach (EOM [70]) was also employed to model the CTEs in reduced form as dummy-residue chains potentially assuming multiple conformations. The coordinates of the 28-residue CTE, with the exception of a few residues involved in disulfide bridges or interacting with the clefts of the tetramer, were not clearly assignable even in the oxidized form of A_2B_2 from spinach whose crystal structure was solved, and the CTE was described as devoid of secondary structure [252]. Also, an attempt to optimize the positions of the subunits of the tetramer in order to improve the fit with the experimental data was performed. A rigid body modeling of the subunits O (gapB) and P (gapA) of the 2PKQ pdb structure, with imposition of P2 symmetry and contact conditions was performed with SASREF.

In the case of scattering data which were thought to be possibly due to the

co-existence of multiple oligomeric states of the protein, a tentative least square fitting was attempted to describe them as a linear combination of defined scattering components in which only the volume fractions are optimized (OLIGOMER [66]).

In addition, *ab-initio* modeling approaches both based on dummy-beads (DAMMIF) and on dummy-residues (GASBOR) were also applied.

In all cases repeated runs of the three-dimensional reconstruction algorithms were performed and the similarity of the obtained structures was checked by DAMAVER [65] and DAMCLUST [30], in which the superposition is performed by the SUPCOMB code [64] and a normalized spatial discrepancy (NSD) value is calculated and used as a parameter to determine the difference between two three-dimensional models. The superimposition procedure produces an average envelope, which incorporates all the models, and a filtered shape obtained by selecting the volume with maximum occupancy probability, which together visually inform about the variability of the solutions [65].

5.2.3 DLS measurements

In the case of the study of AB-GAPDH, samples prepared upon dilution of the protein stock solutions were analyzed by dynamic light scattering (DLS) using a MALVERN NanoZetaSizer (Malvern Instruments LTD, UK) instrument, in 1-cm path disposable cells. The intensity-based size distribution calculated by the instrument software was evaluated as a quality check of the sample and the hydrodynamic radius corresponding to the maximum was reported as a first characterization of the oligomerization state of the proteins.

5.2.4 Hydrodynamic calculations

The program HYDROPRO [266] was used to estimate the hydrodynamic radius (R_h) of the proposed models. The calculation was performed on the *ab-initio* model of oxidized PRK and on hybrid all-atom/dummy-residue models for the binary and ternary complexes. The radii of the beads used for building the hydrodynamic models were optimized as suggested by the authors [267], and values of 4.5, 3.3 and 3.7 Å for PRK and the binary and ternary complexes, respectively, were chosen. Room temperature and viscosity of water were considered in the calculations.

Table 5.1. Summary of the results of the SAXS analysis of the A₄-GAPDH/CP12/PRK ternary complex and of its free components.

	PRK		A ₄ /CP12	A ₄ /CP12/PRK
	oxidized	reduced		
Concentration range (g/l)	2.0 - 5.5	1.1 - 3.9	0.5 - 5.0	1.0 - 5.0
q interval for Fourier inversion ($\text{\AA}^{-1}$)	0.029 - 0.35	0.03 - 0.35	0.01 - 0.35	0.008 - 0.35
I(0) from P(R)	74.9 $\pm$ 0.5	81.4 $\pm$ 0.5	146.0 $\pm$ 0.1	281.6 $\pm$ 0.1
R _g ($\text{\AA}$) from P(R)	38.7 $\pm$ 0.1	40.1 $\pm$ 0.5	34.0 $\pm$ 0.1	57.2 $\pm$ 0.1
q interval for Guinier fit ($\text{\AA}^{-1}$)	0.020 - 0.033	0.025 - 0.032	0.013 - 0.037	0.010 - 0.023
I(0) from Guinier fit	77.2 $\pm$ 0.5	84.5 $\pm$ 0.5	146.0 $\pm$ 0.1	280.0 $\pm$ 0.5
R _g ($\text{\AA}$) from Guinier fit	39.4 $\pm$ 0.5	40.9 $\pm$ 0.5	34.0 $\pm$ 0.1	56.7 $\pm$ 0.1
D _{max} ($\text{\AA}$)	126 $\pm$ 5	130 $\pm$ 5	115 $\pm$ 2	175 $\pm$ 5
Porod volume estimate V _P (nm ³)	162 $\pm$ 5	165 $\pm$ 5	207 $\pm$ 10	732 $\pm$ 10
Dummy-atom excluded volume V _d (nm ³)	199 $\pm$ 2	201 $\pm$ 5	240 $\pm$ 2	675 $\pm$ 2
Dry volume from sequence and $\tilde{\nu}$ =0.735 cm ³ /g (nm ³)	96	96	197	584
MW from V _P (kDa)	81	103	129	458
MW from V _d (kDa)	99	100	120	338
MW from V _c (kDa)	96.4	95.6	120	355
MW from sequence (kDa)	78.4	78.4	161.6	480
R _h from model (nm)	4.2 $\pm$ 0.1	-	4.6 $\pm$ 0.1	7.5 $\pm$ 0.1
R _h from SEC-MALS-QELS (nm) [241]	3.3 $\pm$ 0.3	-	4.3 $\pm$ 0.5	7.0 $\pm$ 0.1

^sThe parameters refer to scattering profiles obtained by merging data collected at concentrations of 2.0 and 5.53 g/l for oxidized PRK and 0.50 and 5.0 g/l for A₄-GAPDH/CP12. For A₄-GAPDH/CP12/PRK data at concentrations of 2.0 and 2.86 g/l were averaged.

5.3 PRK, A₄-GAPDH/CP12 and their ternary complex

The parameters obtained from the SAXS analysis of PRK, either fully reduced or oxidized, of the binary complex A₄-GAPDH/CP12₂ and of the ternary complex also involving oxidized PRK were summarized in Table 5.1. The final scattering curves were reported as normalized Kratky plots (Fig. 5.2) that allow for a comparison between proteins with different dimensions and masses [46]. The symmetrical plots of the binary and ternary complexes suggested that they behave as globular proteins, while the positive skewness of the PRK plot (either reduced or oxidized) indicated a more elongated structure [46]. In the following, the results of the SAXS analysis of the three systems are reported in detail as separate sections.

5.3.1 Oxidized and reduced PRK

The experimental scattering patterns of oxidized PRK showed a slight increase of the slope of the Guinier plot at low q, when increasing the protein concentration from 2.0 to 5.5 g/l (Fig. 5.3a). This suggested that a partial protein aggregation could have occurred especially in the most concentrated samples, for which a non-negligible deviation from the linear Guinier behavior was recognized. When considering the

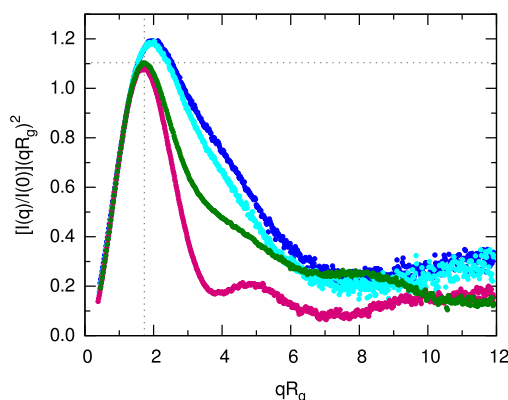


Figure 5.2. *Dimensionless Kratky plots of the ternary complex and its components.* Dimensionless Kratky plots calculated from the experimental data for oxidized PRK (cyan; $R_g = 39.4 \text{ \AA}$), reduced PRK (blue; $R_g = 40.9 \text{ \AA}$), the A_4 -GAPDH/CP12 complex (magenta; $R_g = 34.0 \text{ \AA}$) and the A_4 -GAPDH/CP12/PRK complex (green; $R_g = 56.7 \text{ \AA}$).

data at lower concentration a reliable Guinier linear fit was found neglecting the very first experimental points at the lowest q values which were deviating. In the q interval 0.02 - $0.03 \text{ \AA}^{-1}$ no dependence of the estimated R_g on the first angular point considered was found (Fig. 5.3b), and a R_g of $39.4 \text{ \AA}$ was obtained. From the forward scattering intensity $I(0)$ a molecular weight (MW) of 77.2 kDa was estimated, which is a value compatible with the presence of dimeric PRK in solution (Table 5.1). The scattering profile derived by merging low and high concentration data (2 g/l at low q and 5.53 g/l at high q , superimposing in the region 0.06 - $0.09 \text{ \AA}^{-1}$), was used for further analysis of oxidized PRK.

The SAXS data from reduced PRK samples were less reliable since no correlation was observed between the low- q behavior (expressed as R_g) of the scattering pattern and the protein concentration (Fig. 5.3a). In addition, the linear region of the Guinier plot was slightly shifted towards higher q values (Fig. 5.3d, inset).

In order to detect possible differences between reduced and oxidized PRK, the corresponding experimental curves (from $q_{\min} = 0.03 \text{ \AA}^{-1}$) were superimposed and found to be virtually identical (Fig. 5.3c). For this comparison, the experimental curves that least deviated from ideal behavior at low angles were chosen among the concentration series.

The pair-distance distribution functions ($P(R)$) computed from the data of oxidized and reduced PRK (Fig. 5.3d) showed an asymmetric bell-shaped profile, suggesting dimers with an elongated shape. By comparing the two $P(R)$ functions and the normalized Kratky plots (Figs. 5.2 and 5.3d), a slightly more compact conformation for oxidized PRK compared with the reduced form could be envisaged. However, a reliable assessment of the possible presence of conformational changes caused by disulfide-bridge formation/reduction could not be achieved because of the partial aggregation of the reduced PRK sample that probably affected the scattering curve, so the experiments rather indicated that no substantial conformational variation detectable by SAXS was induced by the different oxidation state of the enzyme.

A reliable high-resolution structure which could be used as a model of PRK

from *Arabidopsis thaliana* (AtPRK) is not available. Its sequence identity with *Rhodobacter spheroides* PRK [256] is less than 19% and in addition AtPRK is about 60 residues longer than the bacterial enzyme (349 versus 290 amino acids), with a C-terminal portion not homologous to any segments from other protein structures. In this context we chose to attempt an *ab-initio* modeling based on the scattering profile of oxidized PRK, with a dummy-residue approach (section 5.2.2). Repeated calculations converged to a satisfactory agreement between the intensity computed from the model and the experimental data. The goodness of fit provided by the most representative low-resolution structure obtained imposing P2 symmetry was comparable to that of the model independent analysis (Fig. 5.3e), reported as a reference for a good fit. According to this model, and other modeling attempts which lead substantially to the same results (Fig. 5.4), the PRK dimer, with dimensions of about 50 Å in width and 125 Å in length, has a bent and screwed shape, with two well-defined lobes possibly corresponding to the two monomers. The most representative dummy-residue model of oxidized PRK was employed for further modeling of the A₄-GAPDH/CP12/PRK complex.

5.3.2 The binary complex A₄-GAPDH/CP12

The experimental scattering profiles of the binary complex did not show significant concentration effects, except for a very slight downward deviation observed in the range of $q < 0.02 \text{ \AA}^{-1}$ for the most concentrated sample (5 g/l), possibly owing to weak repulsive interactions. Indeed, the net charge of the A₄-GAPDH/CP12 complex at pH 7.5, as calculated from the sequence, is highly negative and largely dependent on the N-terminal regions of the two CP12 proteins. A scattering pattern obtained by merging low-concentration (0.5 g/l) and high-concentration (5 g/l) data was used for further analyses. The corresponding P(R) function looked similar to that of a spherical particle with a radius of about 43 Å, with the addition of extended portions that caused the D_{max} to increase to 115 Å and made the curve slightly asymmetric (Fig. 5.5a).

In the crystal structure of the A₄-GAPDH/CP12 complex (PDB entry 3QV1, [247]) two CP12 fragments were bound to two of the four subunits of the A₄-GAPDH tetramer. The crystal structure included only 21 C-terminal amino acids of the CP12 peptide engaged in interactions with GAPDH, whereas the remaining 56 N-terminal residues were disordered and it was not possible to assign their coordinates. Two alternative occupancy patterns of the four possible binding sites were revealed (Fig. 5.5b, “same-side” and “opposite”) but they are indistinguishable from the point of view of SAXS as indicated by the computed P(R) functions (Fig. 5.5a, red and blue dashed lines).

The intensity calculated from the high-resolution models was roughly in agreement with the experimental data, even though a systematic deviation was observed around $q = 0.1 \text{ \AA}^{-1}$ (Fig. 5.5c, red line) similarly to that previously reported for cytosolic GAPDH alone [268]. This disagreement was probably owing to a slightly different arrangement of the GAPDH subunits in solution, leading to a less symmetric assembly. This hypothesis was suggested by the fact that a subunit rearrangement (Fig. 5.6) optimized by means of a rigid-body procedure (section 5.2.2) enabled the improvement of the agreement in this q -range (Fig. 5.5c, pink line). Repeated runs

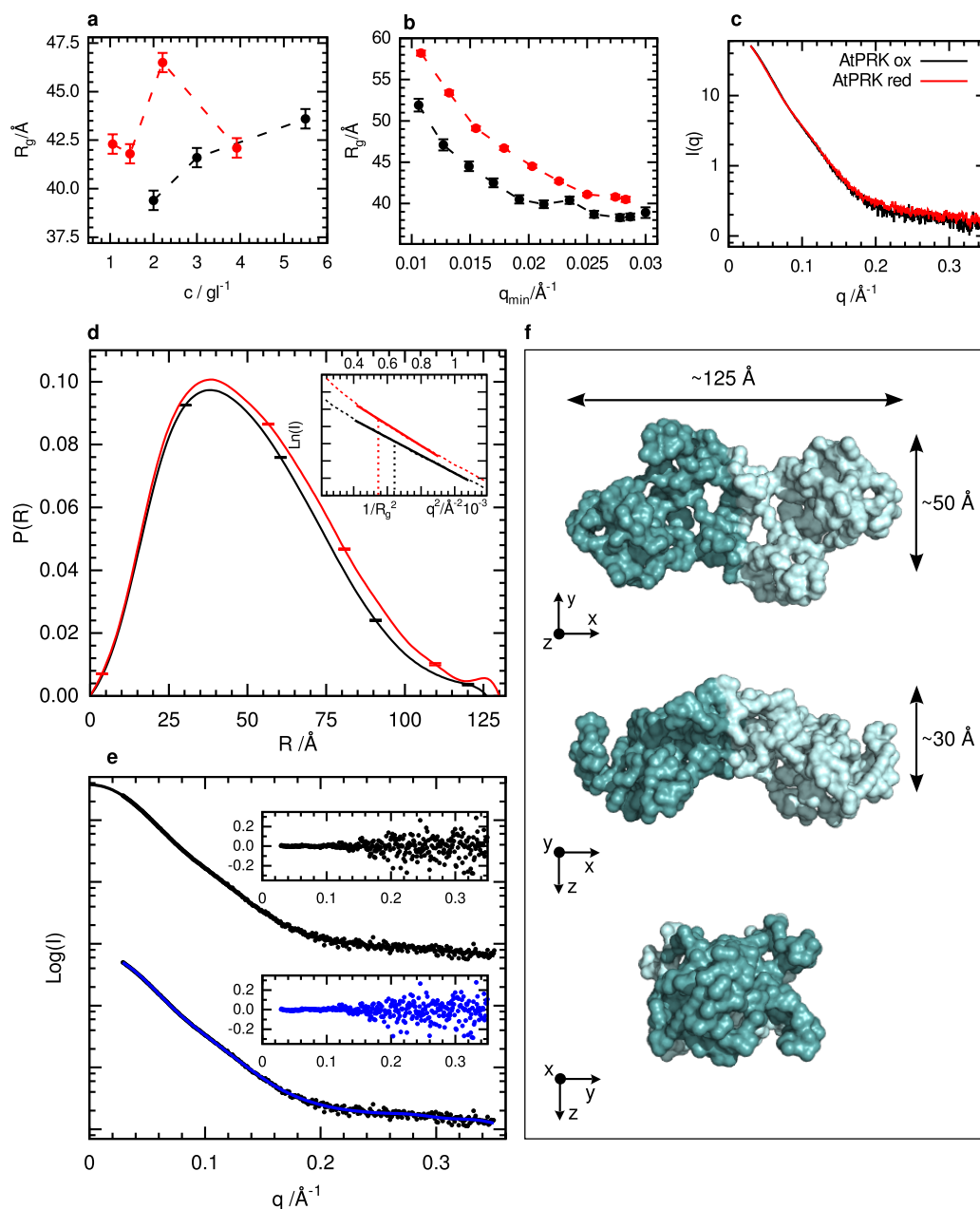


Figure 5.3. SAXS analysis and proposed *ab-initio* model of PRK. (a) R_g values obtained from the Guinier fit for oxidized (black) and reduced (red) PRK at increasing protein concentrations. (b) R_g values obtained from the Guinier fit of the experimental scattering profiles for oxidized (2.0 g/l) and reduced (3.92 g/l) PRK, considering different q -intervals, at increasing values of the minimum scattering vector included ($q_{\min}$). (c) Superimposition of the SAXS experimental curves of oxidized (black) and reduced (red) PRK. (d) $P(R)$ functions of oxidized PRK (black line) and reduced PRK (red line). Only a few error bars are shown. In the inset the Guinier plot (dots) and the linear fit (solid lines) are presented. (e) Scattering profile of oxidized PRK (black dots) compared to the calculated intensity by model-independent fit as a reference (black-line) and by the dummy-residue modeling. Normalized residuals are shown in insets. (f) The most representative dummy-residue model of oxidized PRK computed by applying P2 symmetry is represented by three orthogonal orientations. Dummy-waters bound to the surface of the molecule and introduced in the calculations are omitted. One half of the model is colored darker suggesting a possible dimerization interface. Lengths are shown for reference.

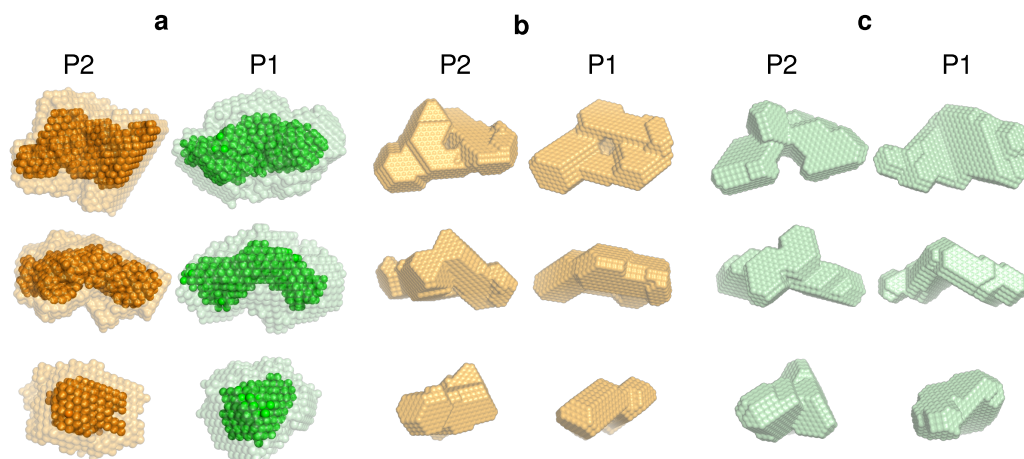


Figure 5.4. *Summary of the ab-initio modeling of the PRK dimer.* (a) Results of the dummy-residue approach for oxidized PRK, applied both with and without imposition of binary symmetry. The probability map of repeated reconstructions is displayed, showing the average envelope (transparent) and the filtered envelope having the same excluded volume as one solution (bright). (b) Results of the dummy-atom approach for oxidized PRK, applied both with and without imposition of binary symmetry. The most representative dummy-atom model is shown as a surface. (c) Results of the dummy-atom approach applied to the scattering profile of PRK in reduced form. In all cases three orthogonal views of the 3D-models are displayed.

showed almost identical subunit positions, as indicated by the computed normalized spatial discrepancy index ($\langle \text{NSD} \rangle = 0.611 \pm 0.054$). Moreover, the addition of the 112 residues of CP12 that the crystallographic model did not include (more than 10% of the overall mass) modeled by a dummy-residue approach resulted in a further improvement of the overall fit, especially in the Guinier region ($\chi = 1.36$; Fig. 5.5c, blue line). The rearranged models of the complex for the two possible occupancy alternatives showed identical $P(R)$ functions (Fig. 5.5a, green and pink dashed lines). On this basis, for further analyses only the “same-side” model was considered.

Since the N-terminal region of CP12 bound to GAPDH is predicted to be disordered, the best description of the protein complex in solution should involve the possibility for this flexible portion to explore different conformations. The pool selected after the genetic algorithm procedure showed an R_g distribution biased towards lower values compared with the randomly generated pool (Fig. 5.5d), indicating that more compact conformations of CP12 bound to GAPDH were preferentially populated in the complex. Since the complex was formed under oxidizing conditions, compact conformations are possibly favored by the N-terminal disulfide bridge of oxidized CP12. The selected ensemble, the calculated intensity of which fits the experimental data with $\chi = 1.10$ (Fig. 5.5c, green line), was composed of about 70% of conformers with $R_g < 35 \text{ \AA}$ and $D_{\text{max}} \approx 130 \text{ \AA}$ and about 30% of more extended structures with $R_g > 35 \text{ \AA}$ and $D_{\text{max}} > 150 \text{ \AA}$. A particularly recurring conformer (in repeated runs of the genetic algorithm) with $R_g = 33.25 \text{ \AA}$ and $D_{\text{max}} = 132.76 \text{ \AA}$ (Fig. 5.5b) was chosen as the best representation of the binary complex to be used in the modeling of the ternary complex. When superimposed with the

Table 5.2. Results of the SAXS analysis of the concentration series of the A₄-GAPDH/CP12/PRK ternary complex.

Conc. (g/l)	Guinier fit		P(R)		D _{max} (Å)	V _P (nm ³)	MW from V _c (kDa)
	R _g (Å)	I(0) (kDa)	R _g	I(0) (kDa)			
0.25	60.6 ± 4.0	258.2 ± 0.8	60.36 ± 0.25	256.3 ± 0.8	220.0	590	303
0.5	58.1 ± 0.9	265.2 ± 0.4	58.10 ± 0.14	264.5 ± 0.4	200.0	614	340
1.0	58.3 ± 0.2	300.2 ± 0.2	58.33 ± 0.10	300.2 ± 0.4	200.0	669	430
1.7	55.8 ± 0.8	277.2 ± 0.4	57.74 ± 0.10	282.8 ± 0.4	190.0	654	359
2.0	57.4 ± 0.3	283.7 ± 0.2	57.91 ± 0.05	285.0 ± 0.2	195.0	660	413
2.9	56.3 ± 0.1	276.5 ± 0.1	57.57 ± 0.03	279.9 ± 0.2	190.0	651	427
3.4	58.3 ± 1.0	328.7 ± 0.2	58.20 ± 0.03	328.6 ± 0.2	190.0	669	383
5.0	56.7 ± 2.6	252.8 ± 0.2	57.18 ± 0.03	254.5 ± 0.1	185.0	646	392
7.5	55.2 ± 1.4	237.1 ± 0.2	56.26 ± 0.02	240.7 ± 0.1	175.0	628	384

most representative *ab-initio* envelope a reasonable agreement was found (Fig. 5.6b). The envelope's lobes that are not occupied by the GAPDH tetramer structure may represent the conformational space sampled by the flexible N-terminal portions of the CP12 chains.

5.3.3 The ternary complex A₄-GAPDH/CP12/PRK

The SAXS analysis of a concentration series of the ternary complex did not show any clear systematic effect which could be ascribed to an association equilibrium (Table 5.2 and Fig. 5.7a). Since radiation damage and problems in background subtraction affected the data of the ternary complex at the lowest concentrations, whereas the highest concentrations seemed to induce a slightly lower R_g under the possible effect of repulsive interactions, a curve obtained by averaging the scattering profiles at concentrations of 2.0 and 2.86 g/l was used for further analysis. The mass estimated from the I(0) was lower than the value calculated from the amino acid sequences and the stoichiometry of the complex [241]. However, the values estimated using other methods not dependent on the assumed concentration of the sample were closer to the expected molecular mass.

A preliminary *ab-initio* modeling from the scattering profile of the ternary complex performed with default parameters showed a large variability of the optimized envelope which would fit the data (Fig. 5.8a). This anticipated that the retrieval of a 3D-model from the experimental scattering profile would have been affected by notable ambiguity [269], which we tried to ameliorate by means of additional information about the complex. A rigid-body modeling was so performed considering two A₄-GAPDH/CP12 complexes and two oxidized PRK dimers as structural components [241], using the models described in previous sections. The introduction of a volume fraction of free components (binary complex and PRK dimers) in the rigid-body modeling of the ternary complex improved the fit obtainable by using these subunit models.

Given the numerous degrees of freedom of the system, the minimization procedures were repeated a large number of times (overall around 200) also starting from different possible arrangements, obtaining an average agreement index of $\chi=1.28 \pm 0.02$ and a volume fraction of the associated complex of 0.66 ± 0.01 . The partial dissociation could be compatible with the reversible nature of this regulatory complex,

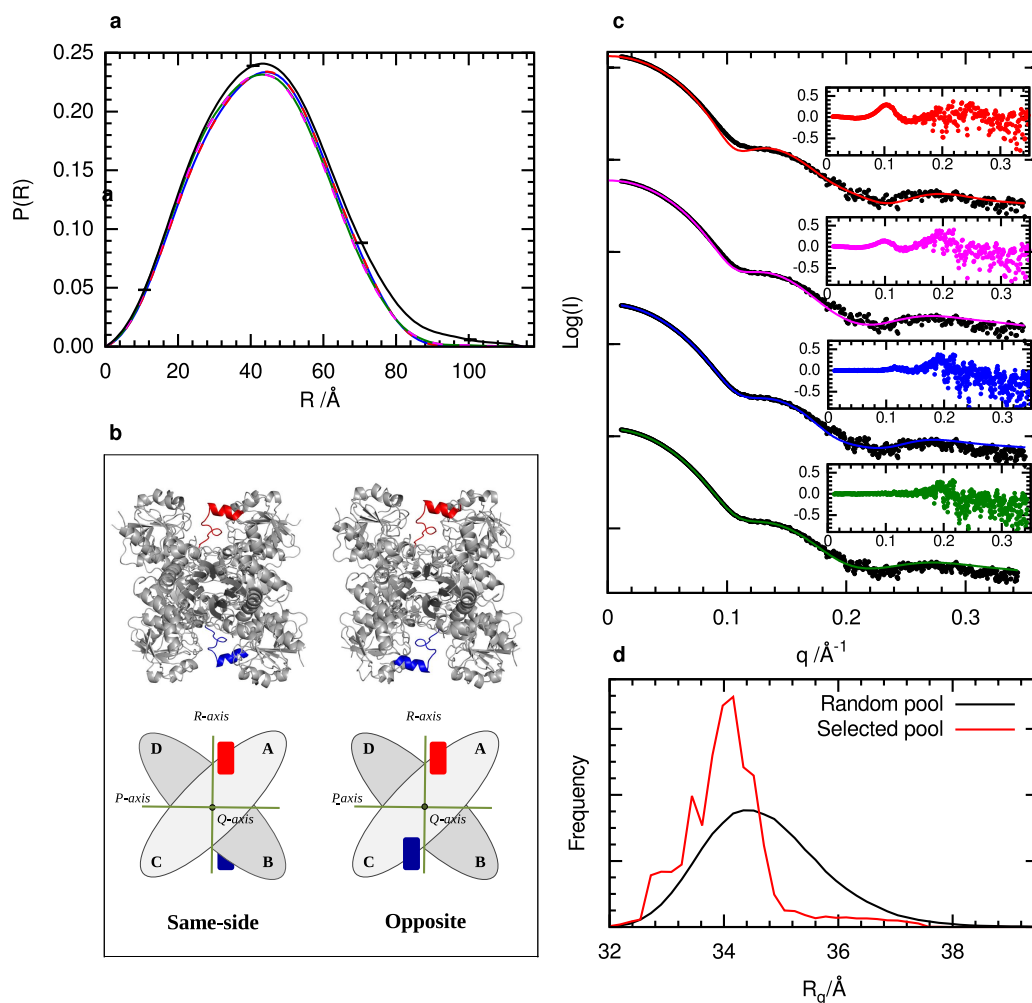


Figure 5.5. SAXS analysis of the binary complex. (a) $P(R)$ functions calculated from the experimental data (black solid line), from the crystal structures of the binary complex with "same-side" (blue dashed line) and "opposite" (red dashed line) occupancy of the CP12 binding sites, and from the rigid-body optimized models after removal of the dummy-residues representing missing CP12 amino acids, for both the "same-side" (green dashed line) and "opposite" (pink dashed line) conformations. The crystal structure of A₄-GAPDH/CP12 with the two alternative positions of the peptides in the possible binding pockets of the tetramer are taken from the pdb entry 3QV1 [247] and shown schematically in (b). (c) Fits between the experimental (black dots) and the theoretical intensities calculated from the high-resolution model of the binary complex ("same-side", red line), the optimized rigid-body/dummy-residue model without (purple line) and including (blue line) the dummy-residues representing missing CP12 amino acids, and the optimized ensemble of coexisting different conformations of CP12 residues (green line) are shown. In the insets normalized residues are reported. (d) The R_g distribution of the random starting pool of possible conformers (black line) and of the best-fitting ensemble (red line) obtained after the genetic algorithm procedure in the ensemble optimization [69] are compared.

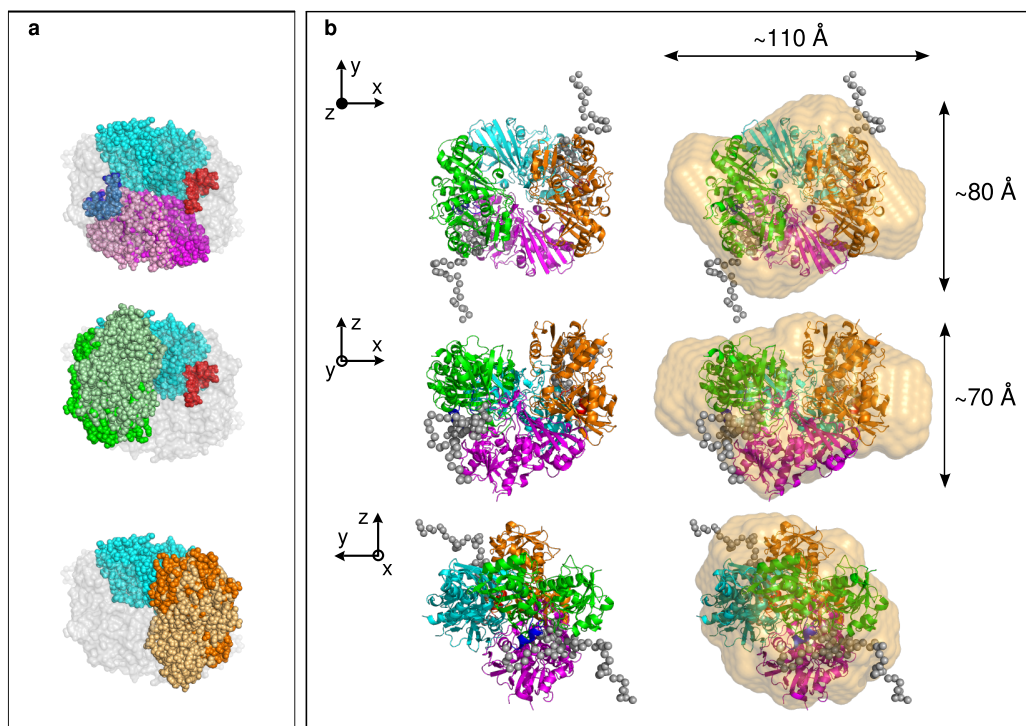


Figure 5.6. *Proposed model of the A_4 -GAPDH/CP12 complex in solution.* (a) Comparison between the positions of the four GAPDH subunits in the crystallographic structure and in the rigid-body optimized model. The optimized subunits (in brighter colors) are superimposed to the crystallographic positions (shown in lighter colors), assuming subunit A as reference. The GAPDH subunits are highlighted in couples for clarity, from the top: A and B, A and C, A and D and the chains are represented as spheres: chain A cyan, B magenta, C green, D orange, G(CP12) red, H(CP12) blue. (b) The binary-complex model obtained with a combination of rigid-body, dummy-residue modeling and the ensemble-optimization method, and its superimposition with the *ab-initio* model (orange envelope), are represented in the left and right columns, respectively. Three orthogonal orientations are shown. The GAPDH subunits and the C-terminal portion of CP12 are represented as cartoons, whereas the N-terminal dummy residues are shown as grey spheres. Lengths are shown for reference.

but was possibly also caused by the dilution required for the SAXS analysis. Consistently, it had previously been shown that the ternary complex from *Chlamydomonas reinhardtii* could partially dissociate upon dilution [270].

The fact that mass values lower than anticipated were estimated and that a partial dissociation was introduced in order to improve the fit of the data, could have suggested that the complex had a different composition with lower molecular weight than expected. On the other hand, the consistency of the SAXS fit results achieved by assuming a (A₄-GAPDH/CP12₂/PRK)₂ stoichiometry in the calculations was not achieved by using alternative heteromeric compositions with lower mass, thus strengthening the conclusion that the previously reported stoichiometry was the only one compatible with the experimental scattering data.

The optimized structures were compatible with a general assembly in which two PRK dimers tied together two binary complexes by interacting with the CP12s according to a previously proposed model based on biochemical grounds [241]. Notably, in the optimized structures the two GAPDH tetramers had to be close and possibly interacted with each other. The assembly of the two A₄-GAPDH constituted a region of electron density which could be possibly related to the filtered envelope with highest occupancy probability shown by the *ab-initio* models (Fig. 5.8a).

By the comparison of the final arrangements obtained by multiple runs of the optimization, three types of assembly were recognized, indicated in descending order of occurrence frequency as models 1, 2, 3, for which the most representative structures are reported (Fig. 5.8b,c,d, Table 5.3). The superimposition and the classification of the computed ternary-complex models within each type of assembly highlighted a reproducibility of the GAPDH tetramer assembly, while the orientation of the PRK dimers was more variable. However, a conserved feature in the general geometry of the 3D reconstructions was the fact that the two PRK dimers related by P2 symmetry always occupied the same half-space delimited by the plane perpendicular to the binary symmetry axis and bisecting the A₄ tetramers; in other words, the complex was not flat and the PRK dimers appeared as protuberances protruding on the same side.

The theoretical intensity was almost the same in all cases and showed a good agreement with the experimental data (Fig. 5.7c). We checked that the computed P(R) functions of the three representative models were also superimposed, and the P(R) function calculated from model 1, considering complex dissociation, was in agreement with the experimental P(R) (Fig. 5.7b). An *ab-initio* model was calculated with the aim of validating the ternary-complex assembly obtained by the rigid-body approach. The imposition of a starting model in the dummy-atom calculations (see paragraph 5.2.2) provided as good a fit to the data as the default calculations (Fig. 5.7c), and the superimposition between the optimized envelope and the rigid-body models was satisfying (Fig. 5.8b,1). The assembly 1, showing a higher frequency and a lower discrepancy among models, could be considered as the most representative.

The ambiguity of the 3D-reconstruction could possibly suggest that populations of the complex with conserved stoichiometry but different shapes could coexist in solution, implying an intrinsic flexibility in the assembly of ternary complex. However, it should be taken into account that the fact of having used a low resolution *ab-initio* model of the PRK dimer might impede a clear discrimination between the proposed

Table 5.3. *Summary of the parameters for the rigid-body modeling of the A_4 -GAPDH/CP12/PRK complex.* The average and standard deviation of the normalized spatial discrepancy index $\langle \text{NSD} \rangle$ calculated among the models classified as belonging to the same type, together with the $\langle \chi \rangle$ value of the fit and the associated volume fraction of the complex $\langle f_{\text{assoc}} \rangle$ are reported. The NSD calculated among the most representative solutions of the different assemblies (1-2-3) is also reported.

Model	$\langle \text{NSD} \rangle_{\text{repr.}}$	$\langle \chi \rangle$	$\langle f_{\text{assoc}} \rangle$
1	1.10 ± 0.09	1.28 ± 0.02	0.654 ± 0.005
2	1.45 ± 0.10	1.28 ± 0.01	0.658 ± 0.014
3	0.88 ± 0.24	1.30 ± 0.01	0.661 ± 0.003
1-2-3	1.80 ± 0.04		

ternary-complex models.

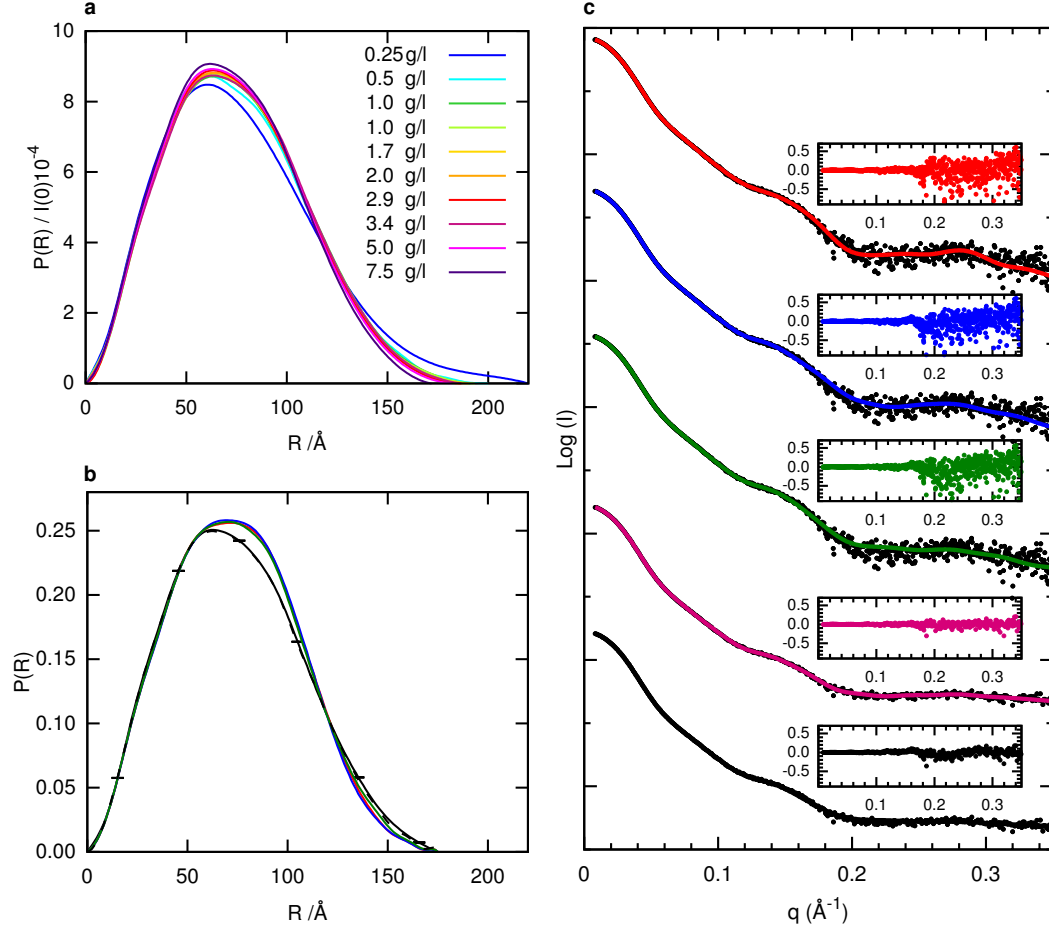


Figure 5.7. SAXS analysis of the ternary complex. (a) $P(R)$ functions calculated from the experimental data of A₄-GAPDH/CP12/PRK measured at different concentrations. (b) $P(R)$ functions calculated from: the experimental scattering profile (black line), the proposed models 1 (red), 2 (blue) and 3 (green) and a volume-fraction weighted mixture composed of model 1, the binary complex model and the PRK model (black dashed line). (c) Fit of the experimental scattering profile (black dots) obtained by means of the rigid-body optimization in the case of models 1 (red line), 2 (blue line) and 3 (green line), and by means of the *ab-initio* approach constrained by a starting arrangement of dummy-atoms (purple line) or with default parameters (black line). The normalized residuals $(I_{\text{exp}} - I_{\text{calc}})/I_{\text{exp}}$ are shown in the insets.

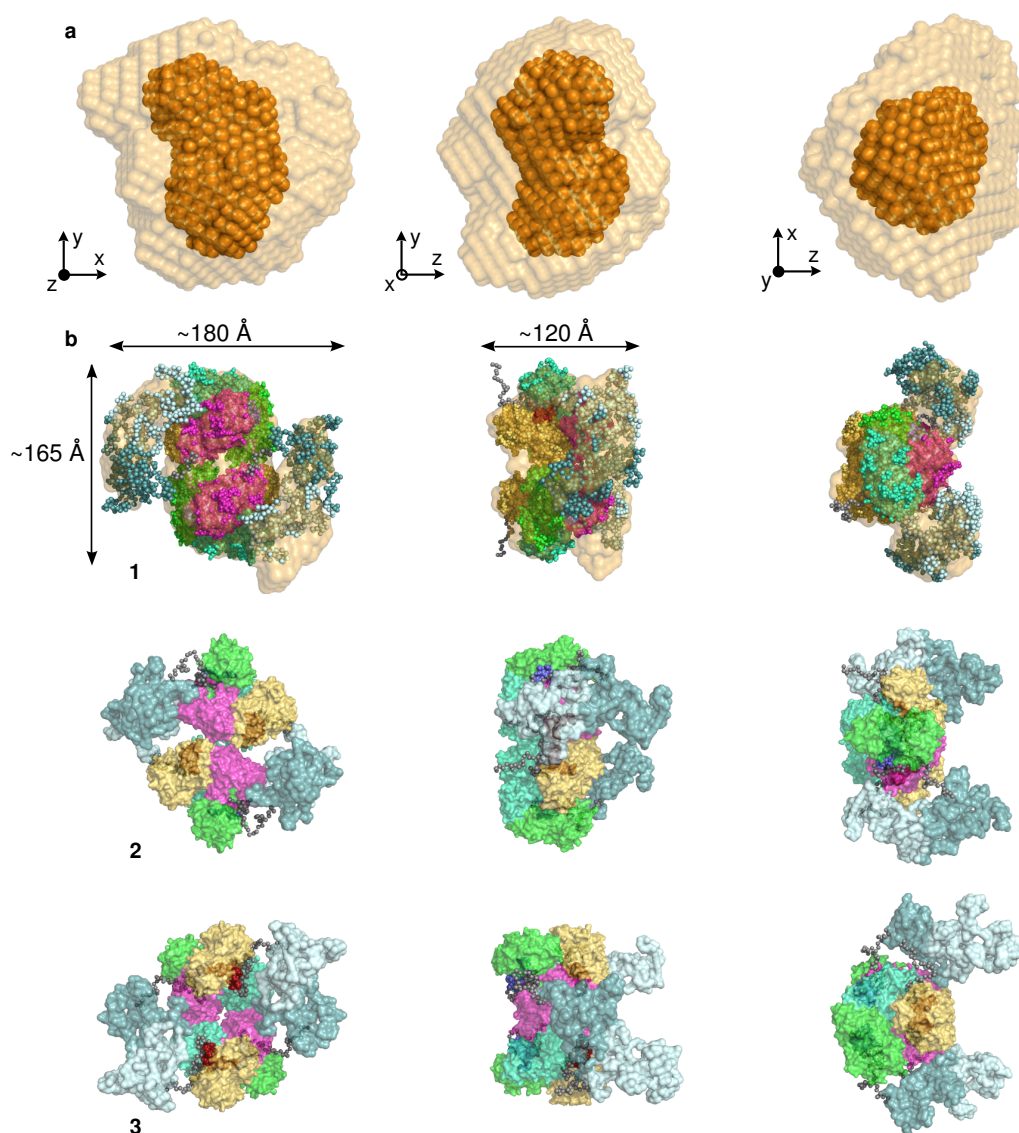


Figure 5.8. 3D-modeling of the A_4 -GAPDH/CP12/PRK complex from the SAXS data. (a) Results of multiple *ab-initio* reconstructions with default parameters; the overall envelope is shown as a transparent surface and the filtered highest occupancy volume is represented as orange spheres. (b) Three model assemblies indicated as 1, 2 and 3 were obtained by rigid-body modeling. The most representative structure for each assembly is shown. GAPDH tetramers (chain A in cyan, chain B in purple, chain C in green and chain D in orange) and PRK dimers (monomers in cyan and deep cyan) are represented as surfaces for 2 and 3, and as spheres in 1; CP12s are shown as spheres (chain G in red, chain H in blue and the N-terminal dummy residues in grey). Model 1 is shown superimposed to the *ab-initio* envelope optimized from a starting arrangement of dummy-atoms (transparent surface). Three orthogonal orientations are visualized. Lengths are shown for reference.

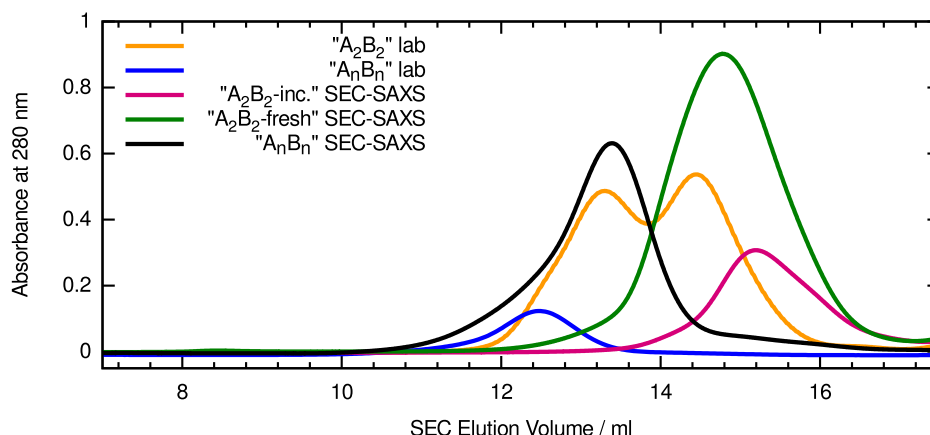


Figure 5.9. *SEC elution profiles.* The traces of UV absorbance at 280 nm as a function of the elution volume collected by the UV detector of the HPLC system during the SAXS experiments (purple, green and black lines) are compared to the SEC experiments performed in the lab during the preparation of the protein samples (orange and blue lines).

5.4 Oligomerization of AB-GAPDH

SAXS data were collected on samples of AB-GAPDH from spinach in which either the active “ A_2B_2 ” form was induced with long or short incubation time (“ A_2B_2 -inc.” and “ A_2B_2 -fresh” respectively, section 5.2.1) or oligomerized forms were stabilized (“ A_nB_n ” sample). For these three samples a SEC-SAXS experiment was performed collecting continuously SAXS frames of the eluate during the separation procedure (section 5.2.2). The “ A_2B_2 ” samples presented in both cases an elution peak centered at retention volumes approximately expected for the tetrameric species (around 15 ml). The “ A_nB_n ” sample presented a broader peak with a maximum around 12.5 ml (Fig. 5.9).

A systematic variation of the SAXS-derived dimensional parameters as a function of the elution volume was observed, testifying the underlying presence of other populations of oligomers in the samples, in addition to the expected species with stoichiometry A_2B_2 in the “ A_2B_2 ” samples and A_8B_8 in the “ A_nB_n ” sample (Fig. 5.10). Nevertheless it was possible to identify statistically superimposable frames in the regions where a constant R_g was observed and average them in order to obtain experimental data (Table 5.4) with satisfyingly high signal-to-noise ratio that could be considered for interpretation in terms of structural species of AB-GAPDH. In particular the distance distribution functions ($P(R)$) obtained guided towards the identification of the possible assemblies formed by the enzyme tetramer (Fig. 5.11).

Attempts to build structural models based on the crystal coordinates of the A_2B_2 tetramer that could explain the observed SAXS data were made. In the following, we first focus on the interpretation of the scattering profiles of the oligomeric species (“ A_nB_n ”) and then examine the outcome of the two SEC-SAXS experiments involving AB-GAPDH in disaggregating conditions (“ A_2B_2 -inc.” and “ A_2B_2 -fresh”).

As a first step we tried to compare the selected experimental scattering profiles and $P(R)$ functions to those obtained from different hypothetical oligomers (A_2B_2 ,

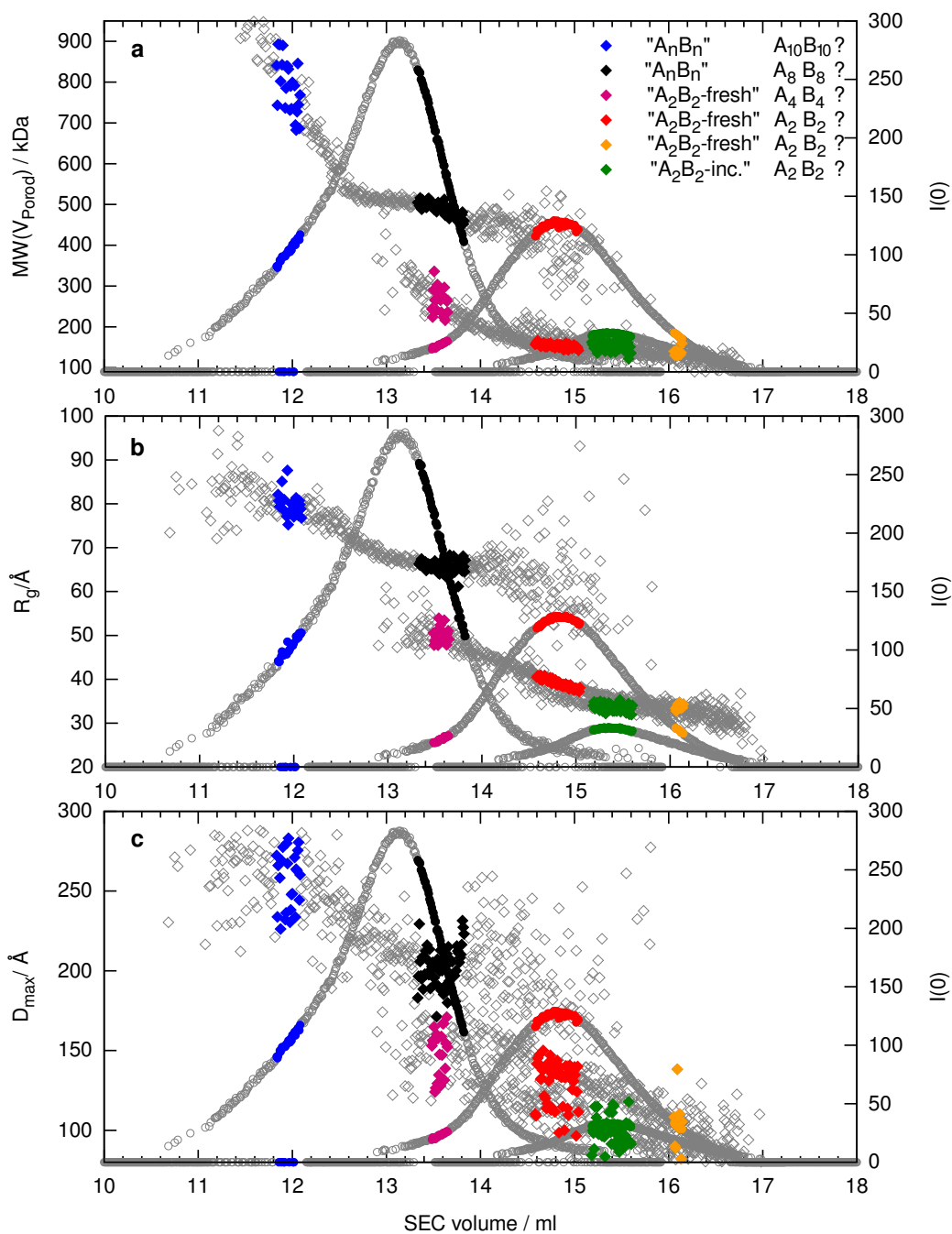


Figure 5.10. SEC-SAXS elution profiles. The parameters derived from the analysis of the SAXS frames for the three samples "A_nB_n" (maximum at 13 ml), "A₂B₂-inc." (maximum at 15.4 ml) and "A₂B₂-fresh" (maximum at 14.8 ml) are shown as a function of the SEC elution volume. The elution profile given by the scattering intensity at zero angle ($I(0)$, empty circles) is plotted together with dimensional parameters: (a) molecular weight estimated from the Porod volume ($MW(V_{\text{Porod}})$ empty diamonds); (b) radius of gyration from the Guinier approximation (R_g , empty diamonds); (c) maximum particle dimension from the $P(R)$ calculation (D_{max} , empty diamonds). In all plots the datapoints belonging to the frames averaged to obtain the selected scattering profiles are highlighted with a color code.

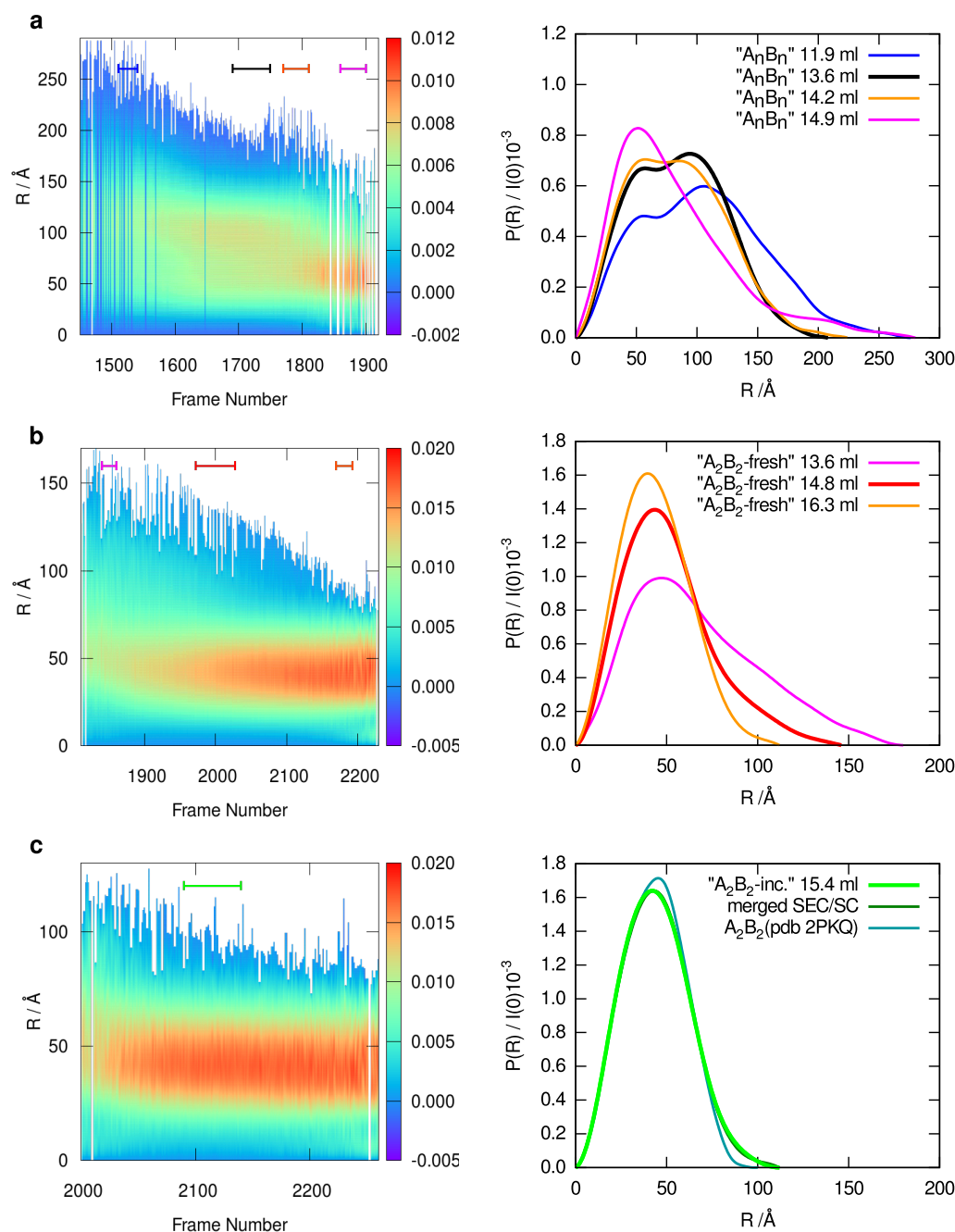


Figure 5.11. $P(R)$ functions of AB-GAPDH species. A 2D map showing the calculated pair distance distribution function normalized by the subtended area ($P(R)/I(0)$) as a function of the frame number during SEC elution is presented on the left. The $P(R)$ functions calculated from the selected scattering profiles are plotted on the right. The frames averaged in order to obtain the representative scattering profiles are highlighted in the $P(R)$ maps by means of bars whose color key corresponds to that of the $P(R)$ functions shown on the right. This information is shown for the three samples analyzed by means of SEC-SAXS: (a) "A_nB_n", (b) "A₂B₂-fresh", (c) "A₂B₂-inc.". In the case of the "A₂B₂-inc." sample the only characteristic $P(R)$ function derived from the SEC-SAXS experiment is compared to those calculated from the theoretical scattering curve of the A₂B₂ tetramer structure found in the 2PKQ pdb entry, and from the scattering profile obtained by merging this SEC-SAXS selected curve with data acquired in sample changer (SC) mode on a particular fraction (between 14 and 15 ml) collected in the A₂B₂ SEC separation in the lab.

Table 5.4. Results of the analysis of selected scattering profiles from the SEC-SAXS experiments on AB-GAPDH.

Sample	Frames	$\langle V \rangle_{\text{el.}}$ (ml)	Guinier fit	P(R)		V_{P} (nm ³)	MW [‡] (kDa)	MW [§] (kDa)
			R_{g} (Å)	R_{g}	D_{max} (Å)			
“A _n B _n ”								
13.1 g/l								
A ₁₀ B ₁₀ ?	1510 - 1540	12.0	80.6 ± 1.0	82.7 ± 0.4	276	1100	681	712
A ₈ B ₈ ?	1690 - 1749	13.6	65.6 ± 0.2	65.7 ± 0.1	208	808	546	536
A _n B _n ?	1770 - 1810	14.2	65.2 ± 0.4	65.5 ± 0.4	224	690	430	443
A ₄ B ₄ ?	1860 - 1900	14.9	59.6 ± 1.5	67.2 ± 3.3	280	508	231	222
“A ₂ B ₂ -fresh”								
10.9 g/l								
A ₄ B ₄ ?	1840 - 1860	13.6	50.9 ± 0.6	53.2 ± 0.6	180	346	225	208
A ₂ B ₂ ?	1972 - 2027	14.8	39.4 ± 0.1	40.3 ± 0.1	146	228	173	147
A ₂ B ₂ ?	2170 - 2193	16.3	34.0 ± 0.4	33.9 ± 0.3	112	183	122	112
“A ₂ B ₂ -inc.”								
< 5 g/l								
A ₂ B ₂ ?	2090 - 2140	15.4	33.9 ± 0.1	34.1 ± 0.1	112	201	148	132

[‡] According to the method of Fischer & Craievich [48].

[§] According to the method of Rambo & Tainer [25].

(A₂B₂)₂, (A₂B₂)₃, (A₂B₂)₄, (A₂B₂)₅, (A₂B₂)₆) built from the 2PKQ pdb coordinates with application of proper crystal symmetry operations (Fig. 5.12). The series of the P(R) functions showed how the highest probability peak characteristic of the distance histogram of a A₂B₂ tetramer located at 47 Å gradually decreased with the increase of the number of tetramers assembled together in the oligomers, while an additional maximum corresponding to the distance characteristics of the packing between two tetrameric units appeared around 90-100 Å.

Based on this analysis we observed that the predominant species in the elution of the “A_nB_n” sample ($R_g = 66$ Å) had indeed dimensions compatible with a (A₂B₂)₄ species. The rigid-body modeling of this construct using the crystal structure coordinates of the A₂B₂ tetramer (section 5.2.2) provided optimized arrangements which could reasonably fit the experimental scattering data. A “C-shaped” spatial arrangement of the subunits was reproducibly found in repeated optimizations and was overall confirmed by the most probable shape obtained by *ab-initio* modeling (Fig. 5.13a). In addition, a slight worse agreement was found for assemblies in which a cavity was formed by two (A₂B₂)₂ units in contact to each other (Fig. 5.13b). The agreement between the theoretical scattering profiles calculated for these models for the (A₂B₂)₄ construct and the experimental data (Fig. 5.14) was further checked by using an alternative method for describing the hydration layer (section 5.2.2), and this lead to a lower value of the discrepancy.

On the other hand the species identifiable from the SEC-SAXS data of the “A_nB_n” sample at lower elution volumes than the elution maximum had a distinguishable scattering profile corresponding to a bigger construct (R_g around 80 Å and D_{max} around 280 Å) having excluded volume and estimated MW roughly 1.3 times that of the A₈B₈ species. These volume and mass in principle would correspond to

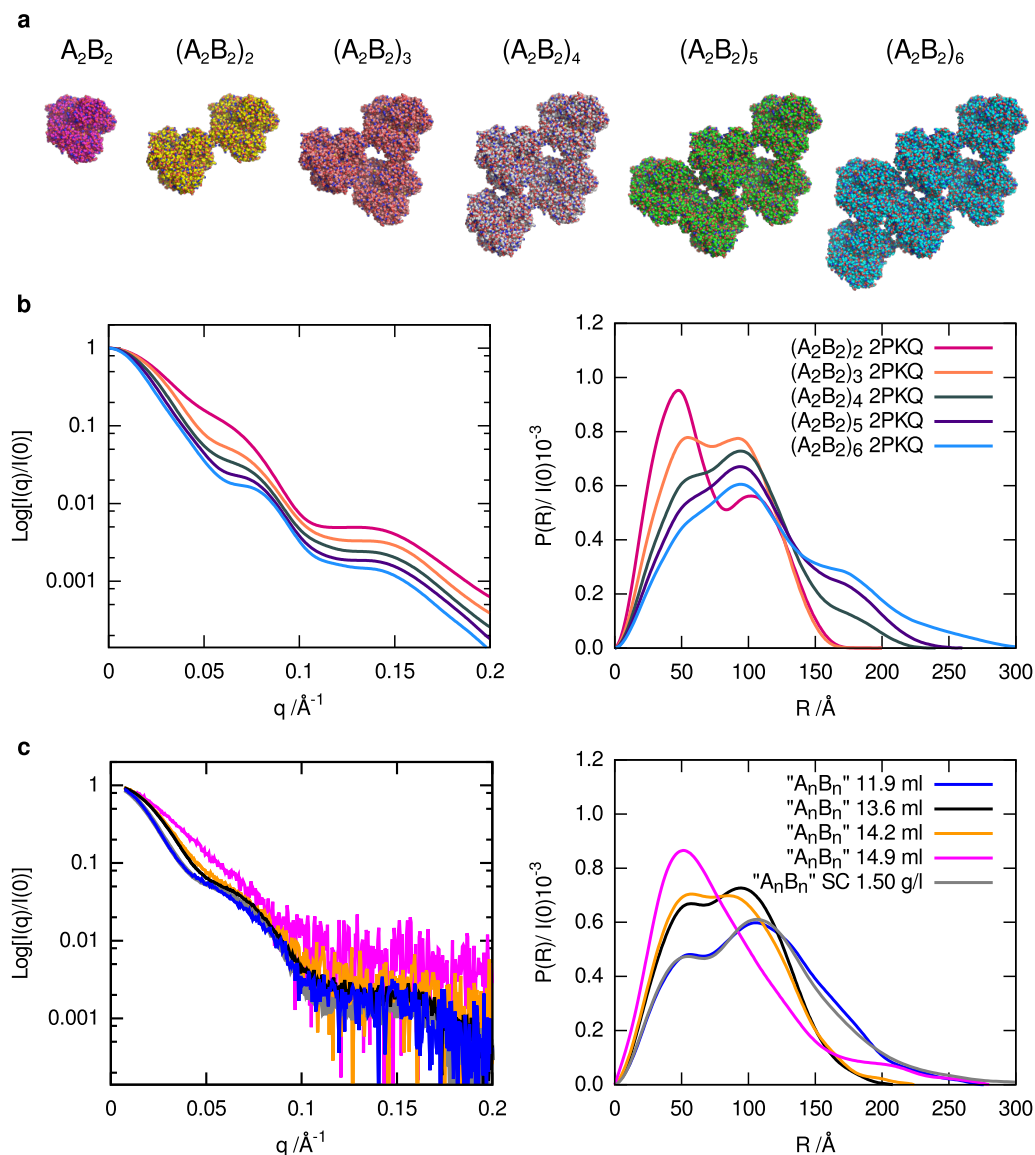


Figure 5.12. *AB-GAPDH oligomers suggested by crystal symmetry.* (a) Hypothetical oligomers of the A_2B_2 tetramer built from the coordinates of the 2PKQ pdb entry. A comparison is made between (b) the theoretical scattering intensities calculated from the hypothetical oligomer models shown in (a) and (c) the selected experimental scattering profiles from the SEC-SAXS experiments. The corresponding $P(R)$ functions are also presented in the panels on the right. In (c) the scattering profile collected on a "A_nB_n" sample in SC mode is also shown (gray line). It is almost superimposed to the scattering profile represented as blue line.

slightly more than 5 times a A_2B_2 tetramer, if a $(A_2B_2)_4$ stoichiometry for the main scattering profile in the elution of the “ A_nB_n ” sample is taken as a reference. This larger assembly could be interpreted as a $(A_2B_2)_5$ construct, or maybe the data could rather represent an underlying mixture of $(A_2B_2)_4$, $(A_2B_2)_5$ and $(A_2B_2)_6$ possible assemblies. So far only an *ab-initio* modeling was attempted. No symmetry constraints were imposed and the resulting most representative model out of a set of repeated optimizations with similar outcome shown in Fig. 5.15 appeared composed of globular moieties with an approximate diameter roughly equivalent to that of an A_2B_2 tetramer.

The data collected on “ A_nB_n ” samples in SC mode without separation (Table 5.7, gray lines in Fig. 5.12c) showed average dimensions closer to this larger assembly detected as a minor component of the elution rather than to the $(A_2B_2)_4$ model. We could argue that in conditions not perturbed by SEC bigger oligomers than $(A_2B_2)_4$ could form and contribute significantly to the scattering profile.

In the case of the “ A_2B_2 -inc.” sample analyzed after a long incubation in the “disaggregating” conditions (section 5.2.1), it was found that the scattering curve obtained by averaging the frames across the peak maximum had dimensions approximately in agreement with the crystal structure of the A_2B_2 tetramer. The calculated $P(R)$ function deviated slightly from that of the crystallographic structure of A_2B_2 (Fig. 5.11c, right panel) and the reduced χ^2 between this experimental curve and that computed from the tetramer model was 2.1. In addition, these SEC-SAXS data were compared to those acquired in SC mode on a particular elution fraction collected in the separation of the “ A_2B_2 ” sample in the lab. Since these two scattering profiles were found superimposable and the SEC-SAXS data were very noisy at $q > 0.1 \text{ \AA}^{-1}$ due to the low concentration of the sample, a scattering profile for further analysis was obtained by merging these two data-sets (details in section 5.2.2). With this curve the reduced χ^2 against the model intensity was 1.54 (Fig. 5.16a, red line).

The fact that the SAXS data of GAPDH tetrameric enzymes in solution slightly deviated from the theoretical scattering curves calculated from the crystal structures in particular in the q range around $0.1 \text{ \AA}^{-1}$ had already been reported for the commercial rabbit muscle enzyme [268] and then for the chloroplastic non-regulatory isoform A_4 -GAPDH bound to CP12 as described in the section 5.3.2 [271]. A small rearrangement of the subunits in the solution form leading to a less symmetric shape has been hypothesized in the latter case. In the attempt to find a better agreement with the experimental data of A_2B_2 we could confirm that the modeling of the two reduced CTEs as flexible chains of dummy-residues possibly assuming multiple conformations (EOM approach [70]) did not significantly contribute to a better fit in that specific q range, while improving the agreement in the Guinier region (Fig. 5.16a, blue line). On the other hand, a rigid-body optimization starting from the O and P subunits of the crystal structure and imposing P2 symmetry and inter-subunits contacts based on those present in the crystal, converged to a slightly rearranged tetramer whose scattering profile could better fit the experimental data in the $0.1 \text{ \AA}^{-1}$ q -range (Fig. 5.16b and a, green line).

Assuming this rearranged A_2B_2 tetramer as subunit in the rigid-body modeling of the $(A_2B_2)_4$ complex did not lead to an improvement of the disagreement with the experimental data seen for $q > 0.1 \text{ \AA}^{-1}$ (Fig. 5.14c,d), as assessed by visual inspection and by means of the statistical test [27]. In this case the optimized assemblies reasonably fitting the experimental scattering profile could also be classified according to the two main topologies identified before (Fig. 5.17b,d), with the hollow construct showing on average a better agreement with the data compared to the optimizations based on the crystal structure (Fig. 5.14).

Surprisingly none of the SAXS data collected along the elution of the “ A_2B_2 -fresh” sample analyzed after only 2-hour incubation in the “disaggregating” conditions were satisfyingly fitted by the crystal structure of the A_2B_2 tetramer in the oxidized form.

Two regions of “stability” of the SAXS profiles and of the dimensional parameters were recognized along the chromatogram by the automatic pipeline (Table 5.4), and an additional representative scattering profile was obtained by averaging data in the low elution volume side of the peak (13.6 ml). The species identified in correspondence of the elution maximum (14.8 ml) had an estimated MW roughly in agreement with the expected A_2B_2 tetramer (150 kDa) but dimensional parameters sensibly higher than the crystal model (R_g 39.8 $\text{\AA}$ vs. 33.4 $\text{\AA}$ and $D_{\max}$ 135 $\text{\AA}$ vs. 100 $\text{\AA}$). On the other hand, a minor species at larger retention volume (16.3 ml) had smaller overall dimensions (R_g 33-34 $\text{\AA}$, $D_{\max}$ 100 $\text{\AA}$) more in agreement with the crystallographic tetramer. We could try to explain this observation assuming that due to the tendency of the A_2B_2 tetramer to oligomerize in higher molecular weight forms, in these conditions we witnessed a fast-exchange dynamic equilibrium between A_2B_2 and higher order oligomers possibly involving $(A_2B_2)_2$ as intermediate species, giving rise to the impossibility to detect the two distinct populations even by SEC [195]. In this view the scattering data acquired in correspondence of the elution maximum (14.8 ml) could be already significantly affected by the higher order oligomers. Probably their contribution appeared more clearly in the increased overall dimensions determined mainly by the low- q behavior (R_g and $D_{\max}$) whereas the MW estimate revealed to be less sensitive, since it was evaluated by means of methods relying on the “Porod region” at higher q . Therefore the scattering profiles collected at larger elution volumes should be more reliably interpreted as related to the AB-GAPDH tetramer (Fig. 5.18, orange solid line).

In addition, we tried to compare the models of oligomers inspired by both the crystal coordinates and the optimized A_8B_8 assembly, with the data collected at lower elution volume than the maximum for the “ A_2B_2 -fresh” sample, in order to assess the possible identity of the higher molecular weight species (Fig. 5.18, purple solid line).

Besides, also in the elution of the “ A_nB_n ” sample one could infer from the map of the $P(R)$ functions in Fig. 5.11 that, together with the two previously described shapes, additional smaller species with characteristic $P(R)$ should be present at larger elution volumes across frame numbers 1800 (R_g around 65 $\text{\AA}$) and 1900 (R_g around 60 $\text{\AA}$), respectively (Fig. 5.18, light and dark green solid lines).

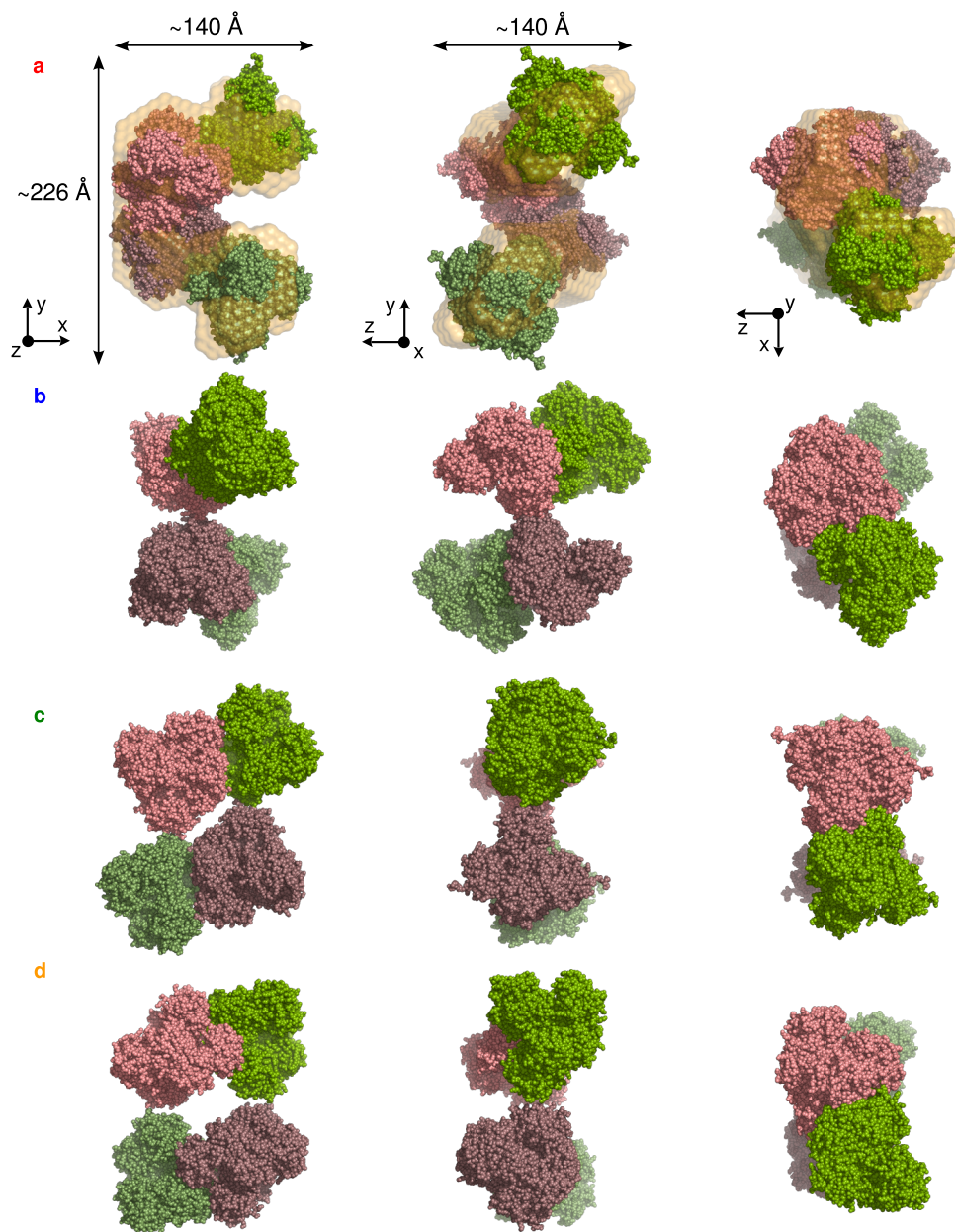


Figure 5.13. *Rigid-body modeling of the $(A_2B_2)_4$ oligomer based on the SEC-SAXS data.* (a) Best-fit model obtained from repeated runs of the optimization procedure using the crystallographic structure of the A_2B_2 tetramer. The most representative *ab-initio* envelope is also superimposed and shown as a light-orange surface. (b) Similar best-fit assembly found by rigid-body modeling in which a rearranged version of the A_2B_2 tetramer was employed, which was found by optimization of the A_2B_2 scattering profile. From the repeated optimization procedures an alternative arrangement of subunits fitting the data also emerged having a hollow structure. Representative models of this kind are shown in the case of modeling based both on the crystallographic (c) and rearranged (d) tetramer. Three orthogonal orientations are presented and lengths are shown as reference. The fit to the experimental data provided by the corresponding scattering intensities is shown in Fig. 5.14

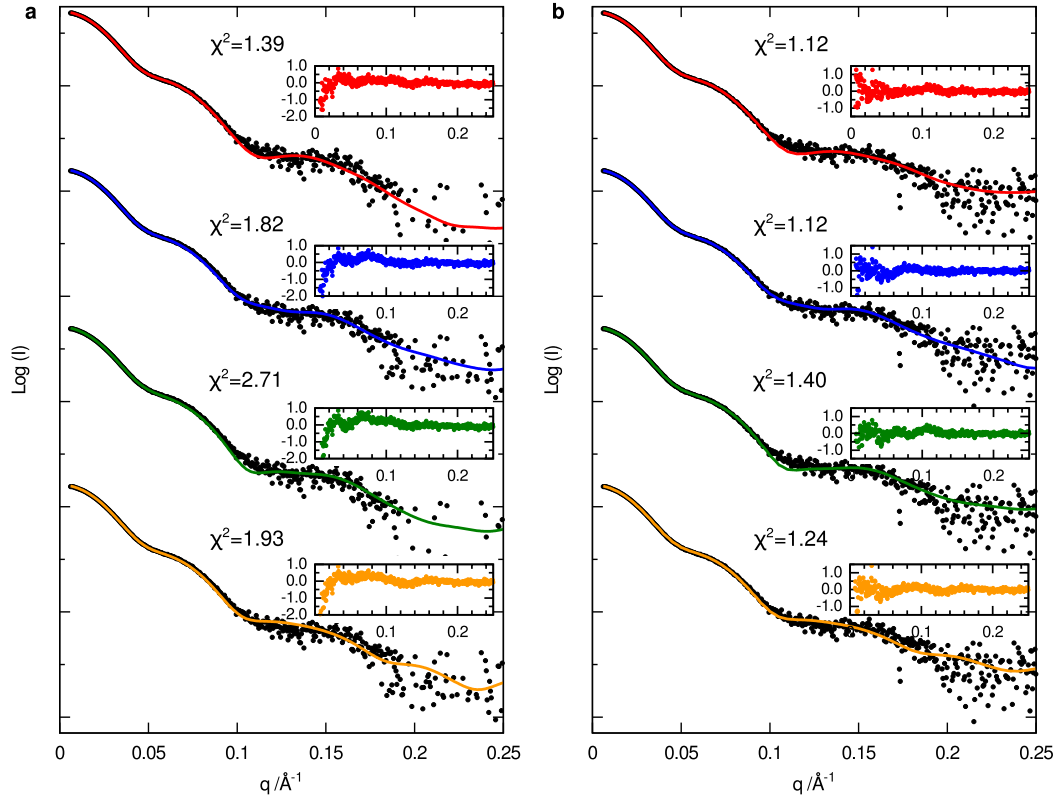


Figure 5.14. Theoretical scattering intensities calculated from the $(A_2B_2)_4$ models and agreement with the experimental data. (a) Fit provided by the SASREF optimizations to the experimental scattering intensity (black dots) in the case of the models *a* (red line), *b* (blue line), *c* (green line) and *d* (orange line) in Fig. 5.13. (b) Fit provided by the calculation of the theoretical scattering intensity of the optimized models with crysol₃₀ (section 5.2.2). The corresponding discrepancy parameter $\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[\frac{I_{\text{exp}}(q_j) - I_{\text{calc}}(q_j)}{\sigma(q_j)} \right]^2$ is also shown in the plots.

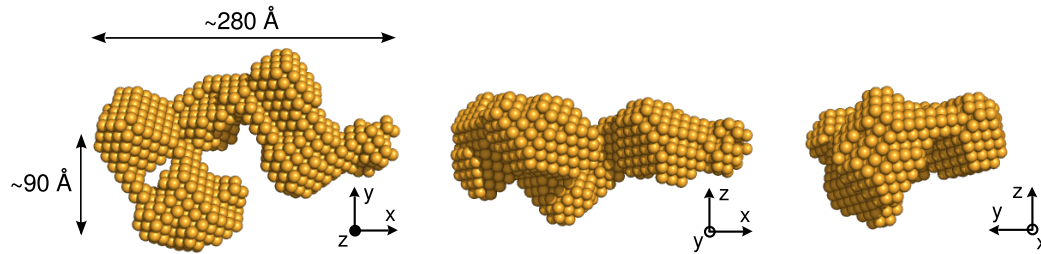


Figure 5.15. *Ab-initio* modeling of a possible $A_{10}B_{10}$ construct. The most representative model obtained by fitting the SAXS data collected on the low-retention volume side of the “ A_nB_n ” elution peak is shown as spheres in three orthogonal views. Lengths are shown as reference.

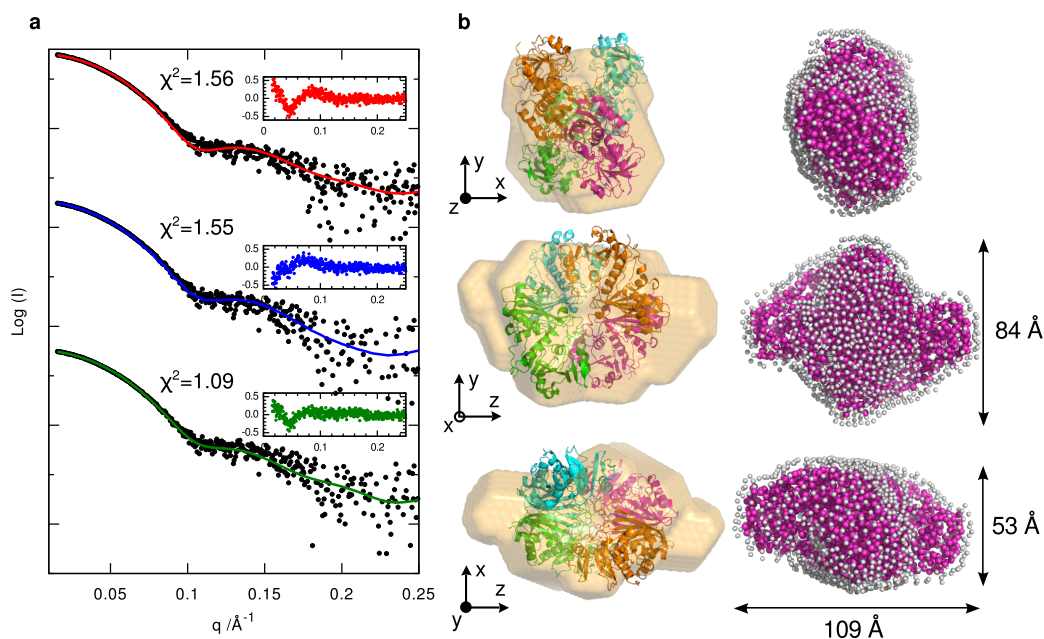


Figure 5.16. Modeling of the scattering profile ascribed to the A_2B_2 tetramer. (a) Fit to the experimental data provided by the theoretical scattering profiles of: the OPQR tetramer in the crystal structure of A_2B_2 (red line); the previous crystal structure in which the amino acids of the CTE in reduced form were represented as dummy-residues in different conformations to account for possible flexibility (blue line); a rigid-body model in which a certain degree of freedom was given to the AB-GAPDH subunits, being constraint to have P2 symmetry and to fulfill selected inter-subunit contacts as found in the crystal structure. The corresponding discrepancy parameter χ^2 is also shown in the plots. (b) The optimized model of the tetramer is shown in cartoon representation and compared to the *ab-initio* reconstructions based on the same SAXS data, both with the dummy-atom approach (transparent envelope) and with the dummy-residue approach (purple spheres). The white beads represent the hydration layer in the dummy-residue model.

Table 5.5. Dimensional parameters calculated from the theoretical scattering profile of models of the AB-GAPDH oligomers.

Model	Guinier fit	P(R)		V_P (nm ³)	MW [#] (kDa)	MW ^s (kDa)	MW _{seq} (kDa)
	R_g (Å)	R_g (Å)	D_{max} (Å)				
A_2B_2 2PKQ	33.1	32.9	100	162	149	137	147
$(A_2B_2)_2$ 2PKQ	56.0	56.7	200	356	313	253	294
$(A_2B_2)_3$ 2PKQ	59.5	59.3	200	619	473	433	441
$(A_2B_2)_4$ 2PKQ	69.4	70.1	240	1008	648	585	588
$(A_2B_2)_5$ 2PKQ	77.4	78.2	260	1400	823	737	735
$(A_2B_2)_6$ 2PKQ	87.5	89.3	340	1805	1013	880	882
Rigid-body optimized model							
$(A_2B_2)_2$ P1	50.5	50.8	157	461	312	271	294
$(A_2B_2)_2$ P2	44.8	44.9	154	439	297	260	294
$(A_2B_2)_3$	57.6	57.7	197	698	461	414	441
$(A_2B_2)_4$	66.2	66.1	223	956	622	589	588

[#] According to the method of Fischer & Craievich [48].

^s According to the method of Rambo & Tainer [25].

In these cases one could argue the presence of smaller oligomers $(A_2B_2)_3$ and especially $(A_2B_2)_2$. The existence of a $(A_2B_2)_2$ species with mass around 300 kDa had indeed been pointed out in different organisms, suggesting that this multimer whilst representing an intermediate step in the AB-GAPDH oligomerization, could be likely by itself represent an important complex in the regulation of the enzyme [260,272]. However it would be hardly possible to isolate the scattering profiles of these intermediate forms from the experimental data since each frame was probably the superimposition of different species maybe in fast dynamic equilibrium in the separation conditions, giving rise to a broad peak in which even an approach based on optimization of overlapping peaks as Gaussian functions [273,274] would probably fail due to the high degree of convolution.

An attempt to describe these SAXS data as a mixture of the hypothesized oligomers of A_2B_2 was made and a reasonable agreement was found especially for the scattering profiles acquired at large elution volumes for the “ A_nB_n ” sample when including $(A_2B_2)_2$ or $(A_2B_2)_3$ as intermediate species (Table 5.6).

SAXS measurements were also performed on “ A_2B_2 ” samples in SC mode. In this case the storage and dilution buffer contained only NADP 1mM in the role of dissociating factor. By analyzing the scattering profiles acquired on a dilution series a systematic increase of the average dimensions and forward scattered intensity was observed with the increase of the protein concentration (Table 5.7). This could be interpreted as deriving from an equilibrium between the A_2B_2 tetramer and higher molecular weight species, with the oligomerization being favored at higher total protein concentration. From the experimental P(R) functions pertaining to the “ A_2B_2 ” concentration series a systematic variation of the average shape could be clearly detected (Fig. 5.19). It could be noticed that a main maximum was present at 47 Å, in the same position as the P(R) calculated from the theoretical scattering

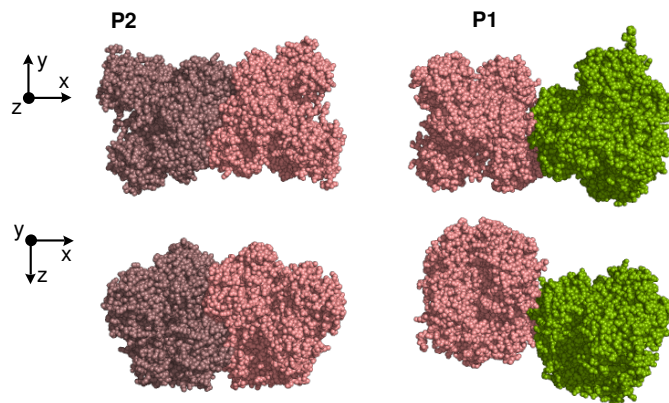


Figure 5.17. Models of possible $(A_2B_2)_2$ oligomers built from the $(A_2B_2)_4$ assembly best fitting the scattering data shown in Fig. 5.13a. From the overall $(A_2B_2)_4$ assembly having P2 symmetry two different pairs of tetramers in contact with each other could be extracted: one in which the two A_2B_2 units are related by binary symmetry (P2) and another without any symmetry relationship (P1). Both models are shown in two orthogonal views.

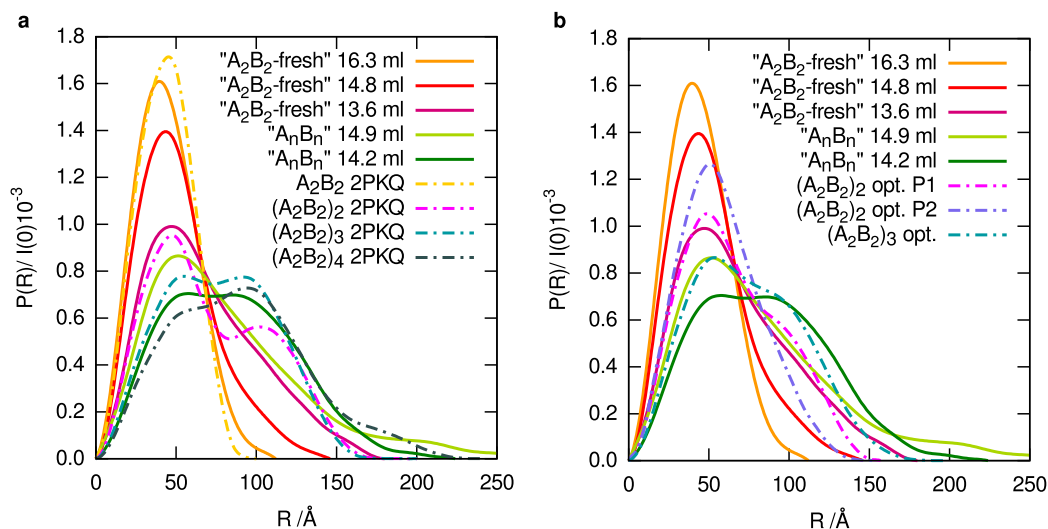


Figure 5.18. $P(R)$ functions of AB-GAPDH species. The $P(R)$ functions calculated from some scattering profiles of doubtful interpretation identified across the “ A_2B_2 -fresh” and “ A_1B_1 ” elutions are compared to those of models both derived directly from the crystal coordinates (a) and extracted from the optimized $(A_2B_2)_4$ construct (b).

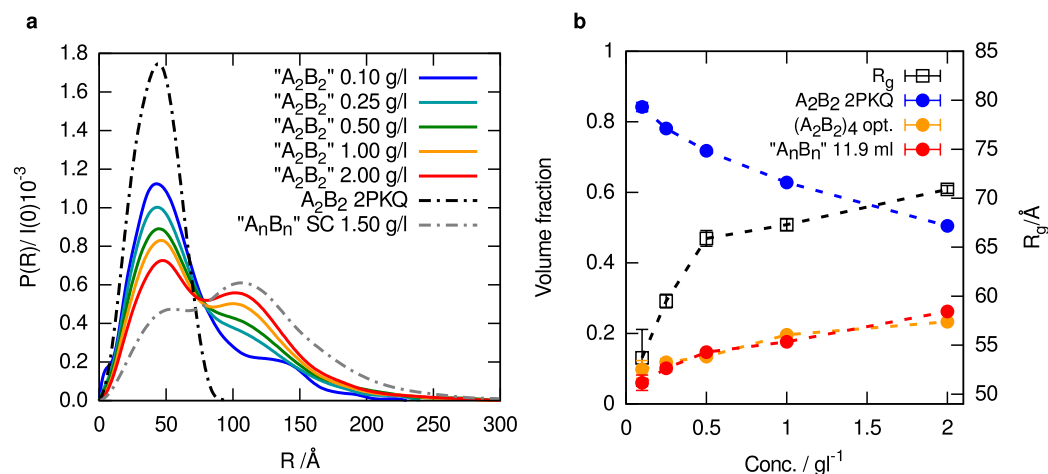


Figure 5.19. SAXS analysis of a concentration series of “ A_2B_2 ”. (a) $P(R)$ functions calculated from the experimental scattering profiles of “ A_2B_2 ” samples at different concentrations; the $P(R)$ of the A_2B_2 crystal model and of a “ A_nB_n ” sample measured in SC mode are also shown for comparison. (b) The volume fraction of possible structural components coexisting in the unseparated “ A_2B_2 ” samples estimated by least-square fit of the scattering profiles as a linear combination are shown as a function of the protein concentration. The variation of the R_g values found by means of the Guinier linear fit are also presented (black squares, scale on the right vertical axis).

curve of the crystallographic A_2B_2 structure, whereas another secondary maximum at $R \approx 105 \text{ \AA}$ became increasingly pronounced at higher protein concentration, which resembled the peak found in the $P(R)$ functions of the “ A_nB_n ” sample measured in SC mode or the scattering profile ascribable to a larger assembly (possibly $(A_2B_2)_5$) identified in the SEC-SAXS experiment at 11.9 elution volume (Fig. 5.12c).

We tried to model the data of the concentration series considering a mixture of the A_2B_2 tetramer and of its $(A_2B_2)_4$ assembly represented as the model obtained from the optimization of the SEC-SAXS data (Fig. 5.13a). In this case the introduction of intermediate components did not improve the fit since the best volume fraction found for the hypothetical oligomers $(A_2B_2)_2$ and $(A_2B_2)_3$ was identically zero for all the data of the series. On the other hand, the scattering profiles corresponding to higher concentrations seemed to be better described including as a third component even the larger assembly found in the “ A_nB_n ” elution, in addition to the A_2B_2 crystallographic model and the $(A_2B_2)_4$ construct (Table 5.6). Although the agreement between the suggested model mixtures of oligomers and the experimental data on these polydisperse samples was not yet fully satisfying, the overall SAXS analysis confirmed the capability of AB-GAPDH to convert between high and low molecular weight species and also the approximate correspondence of the structure found in the crystal and in solution for the free tetramer.

Table 5.6. Results of the analysis of the scattering profiles of AB-GAPDH in terms of mixtures of oligomers.

SEC-SAXS data							
Sample	$\langle R_g \rangle$ (Å)	χ^2	$f_{A_2B_2}$	$f_{(A_2B_2)_2\text{-opt.P1}}$	$f_{(A_2B_2)_3\text{-opt}}$	$f_{(A_2B_2)_4\text{-opt}}$	
“A ₂ B ₂ -fresh” 13.6 ml	52.8	0.88	0.368 ± 0.026	0.498 ± 0.039	0	0.135 ± 0.016	
“A ₂ B ₂ -fresh” 14.8 ml	41.1	25.97	0.761 ± 0.003	0.235 ± 0.004	0	0.004 ± 0.002	
“A ₂ B ₂ -fresh” 16.3 ml	33.31	1.83	0.999 ± 0.005	0	0	0.001 ± 0.003	
“A _n B _n ” 14.2 ml	63.5	0.9	0	0.287 ± 0.027	0.041 ± 0.044	0.672 ± 0.024	
“A _n B _n ” 14.9 ml	53.43	0.81	0.097 ± 0.048	0.796 ± 0.080	0	0.107 ± 0.037	
				$f_{(A_2B_2)_2\text{-opt.P2}}$			
“A ₂ B ₂ -fresh” 13.6 ml	53.28	0.91	0.561 ± 0.013	0	0.366 ± 0.030	0.073 ± 0.022	
“A ₂ B ₂ -fresh” 14.8 ml	42.52	28.85	0.865 ± 0.001	0	0.134 ± 0.003	0.001 ± 0.002	
“A ₂ B ₂ -fresh” 16.3 ml	33.31	1.83	0.999 ± 0.005	0	0	0.001 ± 0.003	
“A _n B _n ” 14.2 ml	63.59	0.91	0.100 ± 0.010	0	0.262 ± 0.027	0.638 ± 0.022	
“A _n B _n ” 14.9 ml	54.16	0.82	0.393 ± 0.023	0	0.602 ± 0.061	0.005 ± 0.049	
“A ₂ B ₂ ” SC mode							
conc. (g/l)	$\langle R_g \rangle$ (Å)	χ^2	$f_{A_2B_2}$	$f_{(A_2B_2)_2\text{-opt.P1}}$	$f_{(A_2B_2)_3\text{-opt}}$	$f_{(A_2B_2)_4\text{-opt}}$	
0.1	50.32	0.81	0.836 ± 0.011	0	0	0.164 ± 0.005	
0.25	53.9	1.04	0.764 ± 0.004	0	0	0.236 ± 0.002	
0.5	56.45	3.13	0.694 ± 0.002	0	0	0.306 ± 0.001	
1	59.01	7.61	0.600 ± 0.001	0	0	0.400 ± 0.001	
2	61.56	40.46	0.466 ± 0.001	0	0	0.534 ± 0.000	
			$f_{A_2B_2}$	$f_{(A_2B_2)_4\text{-opt}}$	$f_{\text{“A}_n\text{B}_n\text{”-11.9 ml}}$		
0.1	54.74	0.82	0.842 ± 0.013	0.098 ± 0.025	0.060 ± 0.022		
0.25	59.5	0.93	0.781 ± 0.005	0.118 ± 0.010	0.101 ± 0.009		
0.5	63.21	1.92	0.718 ± 0.002	0.134 ± 0.005	0.147 ± 0.005		
1	65.69	3.41	0.628 ± 0.001	0.196 ± 0.003	0.176 ± 0.003		
2	69.37	15.89	0.505 ± 0.001	0.233 ± 0.002	0.262 ± 0.002		

Table 5.7. Results of the analysis of the scattering profiles from the SAXS experiments on AB-GAPDH in SC mode.

Sample	$\langle MW \rangle_{DLS}$ (kDa)	Conc. (g/l)	Guinier fit		P(R)		V_P (nm ³)	$MW^\#$ (kDa)	$MW^\$$ (kDa)
			R_g (Å)	R_g (Å)	D_{max} (Å)				
“A _n B _n ”	537	1.5	83.7 ± 0.2	85.3 ± 0.3	330 ± 4	1507	725	719	
		2	83.6 ± 0.1	84.7 ± 0.3	320 ± 4	1440	689	680	
		3	83.9 ± 0.1	85.0 ± 0.2	320 ± 2	1479	706	700	
“A ₂ B ₂ ”	235	0.1	50.2 ± 1	53.7 ± 2.9	230 ± 5	283	66	130	
		0.25	54.2 ± 0.4	59.5 ± 0.7	240 ± 2	337	140	173	
		0.5	57.7 ± 0.3	65.9 ± 0.8	300 ± 2	394	177	200	
		1	62 ± 0.2	67.3 ± 0.6	320 ± 2	550	304	259	
		2	66.3 ± 0.1	70.9 ± 0.4	340 ± 3	650	371	324	

[#] According to the method of Fischer & Craievich [48].^{\$} According to the method of Rambo & Tainer [25].

5.5 Discussion and conclusions

The pathway of CO₂ fixation in oxygenic photosynthetic organisms, the Calvin-Benson cycle, is the major limiting factor to photosynthetic efficiency under saturating light conditions. A deep knowledge of the Calvin-Benson cycle and of its regulation mechanisms at the molecular level in the different species would be essential for understanding, and possibly improving, the photosynthetic conversion of light into chemical energy at the base of the production of food, biofuels and biomass by plants, algae and cyanobacteria. With the SAXS analysis presented in this Chapter we provided one of the first structural insights into the self-assembly behavior which contributes to the regulation of both A₄ and AB isoforms of the GAPDH enzymes of the Calvin cycle.

As a response to the changed redox conditions occurring at the onset of scarce illumination, the A₄ isoform is known to form together with the small intrinsically disordered protein CP12 and another enzyme of the Calvin cycle (PRK) a ternary protein complex, whereas the tetrameric AB isoform (A₂B₂) predominant in higher plants is able to self-associate into higher order oligomers with estimated stoichiometry A₈B₈. These complexes accumulate in chloroplasts during the night, when the metabolic cycle needs to slow down [249, 254, 260].

High resolution crystallographic information for the GAPDH tetramers was available since the crystal structures of the oxidized form of A₂B₂ from spinach and of the binary complex A₄-(CP12)₂ from *Arabidopsis* had been solved. The disordered moieties represented by the CTE of the AB isoform and by the N-termini of the CP12 peptides bound to A₄ were not found in the electron density maps. These crystallographic studies had given a structural explanation for the modulation of the NADP(H)-dependent activity of the enzyme upon formation/reduction of the disulfide bridges in the redox-sensitive CP12 or CTE [247, 252].

On the other hand, structural information about the supramolecular complexes which would suggest a molecular basis for the protein-protein interactions leading to the assembly process is scarce in the literature. The ternary complex purified from *Chlamydomonas reinhardtii* had been studied by cryoelectron microscopy [270] revealing particles with observed dimensions 200 Å × 100 Å. At that time the presence of CP12 in the *Chlamydomonas* complex was still under debate and also attempts to localize the PRK within the three-dimensional volume of the complex failed. Fitting two GAPDH tetramers within the complex was less problematic and resulted in two well separated GAPDH tetramers with no direct interactions between each other.

The SAXS analysis of the ternary complex formed by the recombinant proteins of *Arabidopsis thaliana* and of its components presented in section 5.3 showed that the PRK dimer had an elongated shape suitable to connect two A₄-GAPDH/CP12 binary complexes and form a compact ternary complex in which the A₄-GAPDH tetramers are much closer than previously suspected, suggesting a role of the GAPDH-GAPDH interactions in ternary-complex stabilization. So, besides depicting the shape and organization of the A₄-GAPDH/CP12/PRK complex with overall average dimensions 180 Å × 165 Å × 120 Å, the results suggested that the capability to build supramolecular complexes may be an intrinsic property of any type of photosynthetic GAPDH, either the A₄ or A₂B₂ isoforms. Since A₄-GAPDH/CP12 complexes do not further associate in the absence of PRK, we could speculate that the interaction

between binary complexes is sterically hindered and electrostatically unfavoured by the flexible and negatively charged N-terminal ends of CP12. In this view, the role of PRK in the formation of the ternary complex would consist in dislocating the CP12 N-terminal ends, clearing the interaction surfaces of GAPDH tetramers. In contrast, the assembly of AB-GAPDH would be independent of PRK because no dislocation of cumbersome polypeptides (like the N-terminal region of CP12) would be required.

In the case of the study of the oligomerization of the AB-GAPDH isoform, an approach based on SEC-SAXS could be adopted thanks to the larger availability of this protein for which an efficient method of purification from spinach existed.

Regarding the oligomeric species stabilized in deactivating conditions, the picture was far more complex than the simple formation of a A_8B_8 construct. The data rather suggested that the protein tetramer A_2B_2 is able to self-associate in steps, with formation of intermediate species which unfortunately it was not possible to fully characterize from a structural point of view in these limited SAXS experiments. In particular the likely existence of a $(A_2B_2)_2$ dimer of tetramers was pointed out, which could resemble the interaction between the two A_4 -GAPDH tetramers found in the ternary complex. A possible interpretation would be that the A_2B_2 hetero-tetramer fundamental unit could undergo a self-assembly to give several oligomeric species in a step-by-step process, being this oligomerization driven by the physiological redox regulation of the A_2B_2 isoform and mimicked in the “ A_8B_8 ” preparation.

The observation that a partial oligomerization spontaneously occurred in the “ A_2B_2 ” samples in the presence of only NADP suggested that the binding of this cofactor in place of NAD could not shift completely the equilibrium towards the tetrameric active form but also the reduction of the CTE and the binding of the substrate 1,3-bisphosphoglycerate played a crucial role.

The most remarkable and unambiguous finding was the assessment of the presence in deactivating conditions of oligomers of higher order than $(A_2B_2)_4$ which to our knowledge had not been previously described. We suggest that they could be formed by the addition of an additional tetrameric unit to the already formed $(A_2B_2)_4$, and even a possible further oligomerization in a hierarchical process could be hypothesized. We can suggest a scenario in which each tetramer possesses two possible interaction interfaces through which it can bind up to two other tetrameric A_2B_2 units. More symmetrical arrangements compared to the “C-shaped” low resolution models suggested, having P4 and P5 symmetry, would also be compatible with such an interaction model but apparently provided a worse agreement with the experimental SAXS profiles. Expanding this study by means of high-resolution electron microscopy experiments could help straighten out the intrinsic ambiguities of a SAXS analysis and gaining more information about the protein interfaces involved in the self-assembly. Preliminary attempts have been performed on “ A_8B_8 ” and seemed to suggest more symmetrical assemblies, but tomographic experiments are needed to give more reliability to the limited information available by means of only one 2D projection.

Concluding comments

Some general considerations could be drawn as a closure to this thesis work.

- The first general comment is about the behavior of a multidomain protein as HSA in “stress” conditions perturbing its native state. The partially unfolded F-form stable in acid conditions (Chapter 4), the conformational transition towards an open conformation induced by lysine oxidation (Chapter 3) and the denaturant-induced unfolding processes [79,120] all seem to reveal how intra-molecular electrostatic interactions are fundamental in the stabilization of the overall folding of this multi-domain protein:

- at pH 4 the protonation of carboxyl groups could occur disrupting potentially important salt bridges and also determining electrostatic repulsion between residues;

- the oxidation and chlorination of lysine residues could have a similar effect but determining an excess of negative charges and without affecting too much the protein flexibility. In addition, the screening provided by higher ionic strength of the aqueous solvent decreases the overall dimensions of the oxidized-protein open conformer.

- The fact that the charged denaturant guanidinium chloride (GdmCl) is able to determine HSA unfolding at lower molarity compared to urea suggests that inter-domains electrostatic interactions play a prominent role in the stabilization of the native heart-shape folding [275,276].

In addition, the fact that the binding of fatty acids has a strong stabilizing effect on the native protein conformation in the unfolding process caused by urea [79], but only a weak protective action against the GdmCl-induced unfolding [120], and has not effect at all in the oxidation-induced conformational variation, also reinforce this thought. In this view the interaction of fatty acids with the protein would confer stability due to enhanced hydrophobic effect but cannot prevent disruption of stabilizing inter-domain electrostatic interactions given by GdmCl or oxidation of crucial lysines.

- The second general comment is about methodological aspects. Despite the limited information given by the low-resolution and the intrinsic ambiguity of a structural model based on SAXS, this technique is still much valuable in the study of conformational alterations or variations of the oligomerization state of biological macromolecules in different solution conditions, potentially helping to understand on a structural basis physiological processes which occur in changing environment.

Especially the very short acquisition time needed in the case of SAXS experiments with synchrotron radiation allows for the possibility to follow the phenomena “*in-situ*” or with time-resolved approaches [277–281]. In addition, the relatively flexible sample environment of a SAXS experiment makes it possible to combine complementary techniques in real time by probing the same sample volume [3, 4, 41]. Therefore the task to detect the macromolecules’ state under varying conditions which can be achieved in a relatively quick and simple way with SAXS or simple spectroscopy experiments, is not equally easy to perform in the case of structural techniques with atomic resolution. The combination of information at atomic-scale with the larger length scales of overall structural transitions, supramolecular interactions or aggregation processes would be the most successful way to fully address the understanding of relevant phenomena involving biological macromolecules at a molecular level [34, 282].

List of abbreviations

SAXS	Small angle X-ray scattering
SEC	Size exclusion chromatography
UV-vis	Ultraviolet and visible light absorption spectroscopy
CD	Circular dichroism
DLS	Dynamic light scattering
NMR	Nuclear magnetic resonance
SVD	Singular value decomposition
PCA	Principal component analysis
MCR-ALS	Multivariate curve resolution alternating least square algorithm
2D(H)COS	Two-dimensional (hetero-)correlation spectroscopy
HSA	Human serum albumin
BSA	Bovine serum albumin
GdmCl	Guanidinium chloride, IUPAC: Carbamimidoylazanium chloride
OPR	Oxidant/protein molar ratio = $[\text{NaOCl}]/[\text{HSA}]$
ATP	Adenosine triphosphate
NAD(H)	Nicotinamide adenine dinucleotide, in either oxidized (NAD^+) or reduced (NADH) form
NADP(H)	Nicotinamide adenine dinucleotide phosphate, in either oxidized (NADP^+) or reduced (NADPH) form
DTT	Dithiothreitol, IUPAC: (2S,3S)-1,4-bis(sulfanyl)butane-2,3-diol
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gapA	Glyceraldehyde 3-phosphate dehydrogenase isoform A
gapB	Glyceraldehyde 3-phosphate dehydrogenase isoform B
CTE	Carboxy terminal extension
PRK	Phosphoribulokinase

Bibliography

- [1] J. Pérez and Y. Nishino. Advances in X-ray scattering: From solution SAXS to achievements with coherent beams. *Current Opinion in Structural Biology*, 22(5):670–678, **2012**.
- [2] B. Vestergaard. Analysis of biostructural changes, dynamics, and interactions - Small-angle X-ray scattering to the rescue. *Archives of biochemistry and biophysics*, 602:69–79, **2016**.
- [3] W. Bras and A. J. Ryan. Small-angle X-ray scattering and wide-angle X-ray scattering experiments combined with thermal and spectroscopic analysis techniques. *Journal of Molecular Structure*, 383(1-3):309–314, **1996**.
- [4] S. Haas, T. S. Plivelic, and C. Dicko. Combined SAXS/UV-vis/Raman as a diagnostic and structure resolving tool in materials and life sciences applications. *Journal of Physical Chemistry B*, 118(8):2264–2273, **2014**.
- [5] T. J. Peters. *All About Albumin*. Elsevier, **1995**.
- [6] L. Michelet, M. Zaffagnini, S. Morisse, F. Sparla, M. E. Pérez-Pérez, F. Francia, A. Danon, C. H. Marchand, S. Fermani, P. Trost, and S. D. Lemaire. Redox regulation of the Calvin-Benson cycle: something old, something new. *Frontiers in plant science*, 4:470, **2013**.
- [7] A. Guinier. La Diffraction des rayons x aux très petits angles: application à l'étude des phénomènes ultramicroscopiques. *Ann. Phys. (Paris)*, 12:161–237, **1939**.
- [8] O. Glatter and O. Kratky. *Small angle X-ray scattering*. Academic Press, London, **1982**.
- [9] D. I. Svergun and M. H. J. Koch. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, 66(10):1735–1782, **2003**.
- [10] J. P. Hansen and I. R. McDondal. *Theory of simple liquids*. Elsevier Academic Press, **2006**.
- [11] O. Kratky and G. Porod. Diffuse small-angle scattering of x-rays in colloid systems. *Journal of Colloid Science*, 4(1):35–70, **1949**.

- [12] O. Glatter. A new method for the evaluation of small-angle scattering data. *Journal of Applied Crystallography*, 10(5):415–421, **1977**.
- [13] P. C. Hansen. *Discrete Inverse Problems: Insight and Algorithms*. Society for Industrial and Applied Mathematics SIAM, Philadelphia, **2010**.
- [14] D. I. Svergun. Determination of the regularization parameter in indirect- transform methods using perceptual criteria. *Journal of Applied Crystallography*, 25:495–503, **1992**.
- [15] P. V. Konarev, M. V. Petoukhov, V. V. Volkov, and D. I. Svergun. ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 39(2):277–286, **2006**.
- [16] S. Hansen. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. *Journal of Applied Crystallography*, 33(6):1415–1421, **2000**.
- [17] S. Hansen. BayesApp : a web site for indirect transformation of small-angle scattering data. *Journal of Applied Crystallography*, 45(3):566–567, **2012**.
- [18] S. Hansen. Simultaneous estimation of the form factor and structure factor for globular particles in small-angle scattering. *Journal of Applied Crystallography*, 41(2):436–445, **2008**.
- [19] J. J. Müller, S. Hansen, and H.-V. Pürschel. The use of small-angle scattering and the maximum-entropy method for shape-model determination from distance-distribution functions. *Journal of Applied Crystallography*, 29(5):547–554, **1996**.
- [20] C. E. Shannon and W. Weaver. *The Mathematical Theory of Information*. University of Illinois Press, Urbana, Illinois, **1949**.
- [21] P. B. Moore. Small-angle scattering. Information content and error analysis. *J. Appl. Cryst.*, 13(2):168–175, **1980**.
- [22] D. Taupin and V. Luzzati. Information content and retrieval in solution scattering studies. I. Degrees of freedom and data reduction. *Journal of Applied Crystallography*, 15(3):289–300, **1982**.
- [23] M. C. Pedersen, S. L. Hansen, B. Markussen, L. Arleth, and K. Mortensen. Quantification of the information in small-angle scattering data. *Journal of Applied Crystallography*, 47(6):2000–2010, **2014**.
- [24] B. Vestergaard and S. Hansen. Application of Bayesian analysis to indirect Fourier transformation in small-angle scattering. *Journal of Applied Crystallography*, 39(6):797–804, **2006**.
- [25] R. P. Rambo and J. a. Tainer. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, 496(7446):477–81, **2013**.

- [26] P. V. Konarev and D. I. Svergun. A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems. *IUCrJ*, 2(Pt 3):352–360, **2015**.
- [27] D. Franke, C. M. Jeffries, and D. I. Svergun. Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nature Methods*, 12(5):419–22, **2015**.
- [28] J. Trehwella, W. A. Hendrickson, G. J. Kleywegt, A. Sali, M. Sato, T. Schwede, D. I. Svergun, J. A. Tainer, J. Westbrook, and H. M. Berman. Report of the wwPDB small-angle scattering task force: Data requirements for biomolecular modeling and the PDB. *Structure*, 21(6):875–881, **2013**.
- [29] G. Porod. Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen - I. Teil. *Kolloid-Zeitschrift*, 124(2):83–114, **1951**.
- [30] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, and D. I. Svergun. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 45(2):342–350, **2012**.
- [31] D. Orthaber, A. Bergmann, and O. Glatter. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *Journal of Applied Crystallography*, 33(2):218–225, **2000**.
- [32] E. Mylonas and D. I. Svergun. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of Applied Crystallography*, 40(s1):s245–s249, **2007**.
- [33] J. Lipfert and S. Doniach. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annual review of biophysics and biomolecular structure*, 36:307–327, **2007**.
- [34] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*, 40(3):191–285, **2007**.
- [35] D. I. Svergun. Small-angle scattering studies of macromolecular solutions. In *Journal of Applied Crystallography*, volume 40, pages s10–s17. International Union of Crystallography, **2007**.
- [36] G. L. Hura, A. L. Menon, M. Hammel, R. P. Rambo, F. L. Poole, S. E. Tsutakawa, F. E. Jenney, S. Classen, K. A. Frankel, R. C. Hopkins, S.-j. Yang, J. W. Scott, B. D. Dillard, M. W. W. Adams, and J. A. Tainer. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nature methods*, 6(8):606–12, **2009**.
- [37] D. A. Jacques and J. Trehwella. Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls. *Protein Science*, 19(4):642–657, **2010**.

- [38] H. D. T. Mertens and D. I. Svergun. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology*, 172(1):128–141, **2010**.
- [39] S. Skou, R. E. Gillilan, and N. Ando. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nature protocols*, 9(7):1727–39, **2014**.
- [40] C. M. Jeffries, M. A. Graewert, D. I. Svergun, and C. E. Blanchet. Limiting radiation damage for high-brilliance biological solution scattering: Practical experience at the EMBL P12 beamline PETRAIII. *Journal of Synchrotron Radiation*, 22(2):273–279, **2015**.
- [41] G. David and J. Pérez. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *Journal of Applied Crystallography*, 42(5):892–900, **2009**.
- [42] A. Round, F. Felisaz, L. Fodinger, A. Gobbo, J. Huet, C. Villard, C. E. Blanchet, P. Pernot, S. McSweeney, M. Roessle, D. I. Svergun, and F. Cipriani. BioSAXS Sample Changer: A robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. *Acta Crystallographica Section D: Biological Crystallography*, 71(1):67–75, **2015**.
- [43] S. Doniach. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chemical Reviews*, 101(6):1763–1778, **2001**.
- [44] V. N. Uversky. Natively unfolded proteins: a point where biology waits for physics. *Protein science : a publication of the Protein Society*, 11(4):739–756, **2002**.
- [45] P. Bernadó. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *European Biophysics Journal*, 39(5):769–780, **2010**.
- [46] D. Durand, C. Vives, D. Cannella, J. Perez, E. Pebay-Peyroula, P. Vachette, and F. Fieschi. NADPH oxidase activator P67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of Structural Biology*, 169(1):45–53, **2010**.
- [47] V. Receveur-Brechot and D. Durand. How random are intrinsically disordered proteins? A small angle scattering perspective. *Current protein & peptide science*, 13(1):55–75, **2012**.
- [48] H. Fischer, M. De Oliveira Neto, H. B. Napolitano, I. Polikarpov, and a. F. Craievich. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography*, 43(1):101–109, **2010**.
- [49] D. A. Jacques, J. M. Guss, D. I. Svergun, and J. Trehwella. Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta Crystallographica Section D: Biological Crystallography*, 68(6):620–626, **2012**.

- [50] D. I. Svergun and H. B. Stuhrmann. New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations. *Acta Crystallographica Section A*, 47(6):736–744, **1991**.
- [51] D. I. Svergun, V. V. Volkov, M. B. Kozin, and H. B. Stuhrmann. New developments in direct shape determination from small-angle scattering. 2. Uniqueness. *Acta Crystallographica Section A*, 52(3):419–426, **1996**.
- [52] P. Chacón, F. Morán, J. Díaz, E. Pantos, and J. Andreu. Low-resolution structures of proteins in solution retrieved from X-Ray scattering with a genetic algorithm. *Biophysical Journal*, 74(6):2760–2775, **1998**.
- [53] P. Chacón, J. F. Díaz, F. Morán, and J. M. Andreu. Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. *Journal of Molecular Biology*, 299(5):1289–1302, **2000**.
- [54] W. T. Heller, J. K. Krueger, and J. Trehwella. Further insights into calmodulin-myosin light chain kinase interaction from solution scattering and shape restoration. *Biochemistry*, 42(36):10579–10588, **2003**.
- [55] D. I. Svergun. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal*, 76(6):2879–2886, **1999**.
- [56] D. Franke and D. I. Svergun. DAMMIF , a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, 42(2):342–346, **2009**.
- [57] D. I. Svergun, M. V. Petoukhov, and M. H. Koch. Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal*, 80(6):2946–2953, **2001**.
- [58] D. I. Svergun, S. Richard, M. H. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proceedings of the National Academy of Sciences of the United States of America*, 95(5):2267–2272, **1998**.
- [59] D. Schneidman-Duhovny, S. Kim, and A. Sali. Integrative structural modeling with small angle X-ray scattering profiles. *BMC Structural Biology*, 12(1):17, **2012**.
- [60] D. Svergun, C. Barberato, and M. H. J. Koch. CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, 28(6):768–773, **1995**.
- [61] J. Pérez and A. Koutsioubas. Memprot: A program to model the detergent corona around a membrane protein based on SEC-SAXS data. *Acta Crystallographica Section D: Biological Crystallography*, 71(Pt 1):86–93, **2015**.
- [62] D. I. Svergun, V. V. Volkov, M. B. Kozin, H. B. Stuhrmann, C. Barberato, and M. H. J. Koch. Shape determination from solution scattering of biopolymers. *Journal of Applied Crystallography*, 30(5):798–802, **1997**.

- [63] M. V. Petoukhov and D. I. Svergun. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical Journal*, 89(2):1237–1250, **2005**.
- [64] M. B. Kozin and D. I. Svergun. Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography*, 34(1):33–41, **2001**.
- [65] V. V. Volkov and D. I. Svergun. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, 36:860–864, **2003**.
- [66] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun. PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36(5):1277–1282, **2003**.
- [67] F. Spinozzi, C. Ferrero, M. G. Ortore, A. De Maria Antolinos, and P. Mariani. GENFIT: Software for the analysis of small-angle X-ray and neutron scattering data of macro-molecules in solution. *Journal of Applied Crystallography*, 47(3):1132–1139, **2014**.
- [68] A. G. Kikhney and D. I. Svergun. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*, 589(19):2570–2577, **2015**.
- [69] P. Bernado, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*, 129(2):5656–5664, **2007**.
- [70] G. Tria, H. D. T. Mertens, M. Kachala, and D. I. Svergun. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, 2:207–17, **2015**.
- [71] G. J. Kleywegt. Validation of protein models from Calpha coordinates alone. *Journal of Molecular Biology*, 273(2):371–6, **1997**.
- [72] C. Eckart and G. Young. A principal axis transformation for non-hermitian matrices. *Bulletin of the American Mathematical Society*, 45(2):118–121, **1939**.
- [73] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *FORTRAN numerical recipes*. Cambridge University Press, **1996**.
- [74] E. R. Henry and J. Hofrichter. Singular Value Decomposition: Application to Analysis of Experimental Data. *Methods in Enzymology*, 210(1973):129–192, **1992**.
- [75] L. Chen, K. O. Hodgson, and S. Doniach. A lysozyme folding intermediate revealed by solution X-ray scattering. *Journal of Molecular Biology*, 261(5):658–71, **1996**.
- [76] D. J. Segel, A. L. Fink, K. O. Hodgson, and S. Doniach. Protein denaturation: A small-angle x-ray scattering study of the ensemble of unfolded states of cytochrome c. *Biochemistry*, 37(36):12443–12451, **1998**.

- [77] S. Akiyama, S. Takahashi, T. Kimura, K. Ishimori, I. Morishima, Y. Nishikawa, and T. Fujisawa. Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3):1329–1334, **2002**.
- [78] R. Russell, I. S. Millett, M. W. Tate, L. W. Kwok, B. Nakatani, S. M. Gruner, S. G. J. Mochrie, V. Pande, S. Doniach, D. Herschlag, and L. Pollack. Rapid compaction during RNA folding. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7):4266–4271, **2002**.
- [79] C. Leggio, L. Galantini, P. V. Konarev, and N. V. Pavel. Urea-induced denaturation process on defatted Human Serum Albumin and in the presence of palmitic acid. *Journal of Physical Chemistry B*, 113(37):12590–12602, **2009**.
- [80] R. Tauler. Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30(1):133–146, **1995**.
- [81] A. de Juan and R. Tauler. Multivariate Curve Resolution (MCR) from 2000: Progress in concepts and applications. *Critical Reviews in Analytical Chemistry*, 36(3-4):163–176, **2006**.
- [82] J. Jaumot, R. Gargallo, A. de Juan, and R. Tauler. A graphical user-friendly interface for MCR-ALS: A new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 76(1):101–110, **2005**.
- [83] J. Jaumot, A. de Juan, and R. Tauler. MCR-ALS GUI 2.0: New features and applications. *Chemometrics and Intelligent Laboratory Systems*, 140:1–12, **2015**.
- [84] J. Blobel, P. Bernadó, D. I. Svergun, R. Tauler, and M. Pons. Low-resolution structures of transient protein-protein complexes using small-angle X-ray scattering. *Journal of the American Chemical Society*, 131(12):4378–4386, **2009**.
- [85] I. Noda and Y. Ozaki. *Two-Dimensional Correlation Spectroscopy. Applications in Vibrational and Optical Spectroscopy*. John Wiley & Sons, Ltd., Chichester, U.K., **2005**.
- [86] G. Fanali, A. Di Masi, V. Trezza, M. Marino, M. Fasano, and P. Ascenzi. Human serum albumin: From bench to bedside. *Molecular Aspects of Medicine*, 33(3):209–290, **2012**.
- [87] X. M. He and D. C. Carter. Atomic structure and chemistry of human serum albumin. *Nature*, 358(6383):209–215, **1992**.
- [88] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, and K. Kobayashi. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein engineering*, 12(6):439–446, **1999**.
- [89] P. G. Squire, P. Moser, and C. T. O’Konski. The hydrodynamic properties of bovine serum albumin monomer and dimer. *Biochemistry*, 7(12):4261–4272, **1968**.

- [90] D. Bendedouch and S.-H. Chen. Structure and interparticle interactions of Bovine Serum Albumin in solution studied by small-angle neutron scattering. *Journal of Physical Chemistry*, 87(9):1473–1477, **1983**.
- [91] M. L. Ferrer, R. Duchowicz, B. Carrasco, J. G. de la Torre, and A. U. Acuña. The conformation of serum albumin in solution: a combined phosphorescence depolarization-hydrodynamic modeling study. *Biophysical Journal*, 80(5):2422–30, **2001**.
- [92] C. Leggio, L. Galantini, and N. V. Pavel. About the albumin structure in solution: cigar Expanded form versus heart Normal shape. *Physical chemistry chemical physics : PCCP*, 10(45):6741–6750, **2008**.
- [93] U. Kragh-Hansen. Molecular aspects of ligand binding to serum albumin. *Pharmacological Reviews*, 33(1):17–53, **1981**.
- [94] S. Curry. Lessons from the crystallographic analysis of small molecule binding to human serum albumin. *Drug metabolism and pharmacokinetics*, 24(4):342–357, **2009**.
- [95] M. Fasano, S. Curry, E. Terreno, M. Galliano, G. Fanali, P. Narciso, S. Notari, and P. Ascenzi. The extraordinary ligand binding properties of human serum albumin. *IUBMB life*, 57(12):787–96, **2005**.
- [96] J. Ghuman, P. A. Zunszain, I. Petitpas, A. A. Bhattacharya, M. Otagiri, and S. Curry. Structural basis of the drug-binding specificity of human serum albumin. *Journal of Molecular Biology*, 353(1):38–52, **2005**.
- [97] G. Sudlow, D. J. Birkett, and D. N. Wade. The characterization of two specific drug binding sites on Human Serum Albumin. *Molecular Pharmacology*, 11(6), **1975**.
- [98] G. Sudlow, D. J. Birkett, and D. N. Wade. Further characterization of specific drug binding sites on Human Serum Albumin. *Molecular Pharmacology*, 12(6), **1976**.
- [99] P. A. Zunszain, J. Ghuman, T. Komatsu, E. Tsuchida, and S. Curry. Crystal structural analysis of human serum albumin complexed with hemin and fatty acid. *BMC Structural Biology*, 3:6, **2003**.
- [100] C. Tanford. Preparation and Properties of Serum and Plasma Proteins. XXIII. Hydrogen Ion Equilibria in Native and Modified Human Serum Albumins. *Journal of the American Chemical Society*, 72(1):441–451, **1950**.
- [101] J. F. Foster. Plasma albumin. In F. W. Putnam, editor, *The Plasma Proteins*, chapter 1, pages 179–239. Academic Press, Inc., New York, **1960**.
- [102] J. F. Foster. Some Aspects of the Structure and Conformational Properties of Serum Albumin. In *Albumin: Structure, Function and Uses*, pages 53–84. Elsevier B.V., **1977**.

- [103] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, and T. Gray. How to measure and predict the molar absorption coefficient of a protein. *Protein science : a publication of the Protein Society*, 4(11):2411–2423, **1995**.
- [104] J. R. Lakowicz. *Principles of fluorescence spectroscopy*. Springer US, Boston, MA, **2006**.
- [105] B. Y. J. A. Luetscher. Serum Albumin. II. Identification of more than one albumin in horse and human serum by electrophoretic mobility in acid solution. *Journal of the American Chemical Society*, 61(10):2–4, **1939**.
- [106] D. C. Carter and J. X. Ho. Structure of serum albumins. *Advances in Protein Chemistry*, 45:153–203, **1994**.
- [107] M. Sogami, H. A. Petersen, and J. F. Foster. Microheterogeneity of plasma albumins. V. Permutations in disulfide pairings as a probable source of microheterogeneity in bovine albumin. *Biochemistry*, 8(1):49–58, **1969**.
- [108] K. Wallevik. Reversible denaturation of Human Serum Albumin by pH, temperature, and guanidine hydrochloride followed by optical rotation. *Journal of Biological Chemistry*, 248(8):2650–2655, **1973**.
- [109] M. Y. Khan, S. K. Agarwal, and S. Hangloo. Urea-induced structural transformations in bovine serum albumin. *J Biochem*, 102(2):313–317, **1987**.
- [110] J. Chmelík, P. Anzenbacher, J. Chmelíková, M. Matějčková, and V. Kalous. Mechanism of denaturation of human serum albumin by urea. *Collection of Czechoslovak Chemical Communications*, 53(D):411–422, **1988**.
- [111] K. Takeda, A. Wada, K. Yamamoto, Y. Moriyama, and K. Aoki. Conformational change of bovine serum albumin by heat treatment. *Journal of Protein Chemistry*, 8(5):653–659, **1989**.
- [112] J. K. A. Kamal, L. Zhao, and A. H. Zewail. Ultrafast hydration dynamics in protein unfolding: Human serum albumin. *Proceedings of the National Academy of Sciences of the United States of America*, 101(37):13411–13416, **2004**.
- [113] B. Ahmad, M. Z. Ahmed, S. K. Haq, and R. H. Khan. Guanidine hydrochloride denaturation of human serum albumin originates by local unfolding of some stable loops in domain III. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1750(1):93–102, **2005**.
- [114] Y. Moriyama and K. Takeda. Re-formation of the helical structure of human serum albumin by the addition of small amounts of sodium dodecyl sulfate after the disruption of the structure by urea. A comparison with bovine serum albumin. *Langmuir*, 15(6):2003–2008, **1999**.
- [115] G. Fanali, P. Ascenzi, and M. Fasano. Effect of prototypic drugs ibuprofen and warfarin on global chaotropic unfolding of human serum heme-albumin: A fast-field-cycling 1H-NMR relaxometric study. *Biophysical Chemistry*, 129(1):29–35, **2007**.

- [116] L. Galantini, C. Leggio, P. V. Konarev, and N. V. Pavel. Human serum albumin binding ibuprofen: A 3D description of the unfolding pathway in urea. *Biophysical Chemistry*, 147(3):111–122, **2010**.
- [117] A. Del Giudice, C. Leggio, N. Balasco, L. Galantini, and N. V. Pavel. Ibuprofen and propofol cobinding effect on Human Serum Albumin unfolding in urea. *The Journal of Physical Chemistry B*, 118(34):10043–10051, **2014**.
- [118] R. K. Mitra, S. S. Sinha, and S. K. Pal. Hydration in protein folding: Thermal unfolding/refolding of human serum albumin. *Langmuir*, 23(20):10224–10229, **2007**.
- [119] W. T. Heller. Comparison of the thermal denaturing of human serum albumin in the presence of guanidine hydrochloride and 1-butyl-3-methylimidazolium ionic liquids. *Journal of Physical Chemistry B*, 117(8):2378–2383, **2013**.
- [120] L. Galantini, C. Leggio, and N. V. Pavel. Human serum albumin unfolding: a small-angle X-ray scattering and light scattering study. *The Journal of Physical Chemistry B*, 112(48):15460–9, **2008**.
- [121] P. Ascenzi and M. Fasano. Allostery in a monomeric protein: The case of human serum albumin. *Biophysical Chemistry*, 148(1-3):16–22, **2010**.
- [122] P. Ascenzi and M. Fasano. Serum heme-albumin: An allosteric protein. *IUBMB Life*, 61(12):1118–1122, **2009**.
- [123] T. Komatsu, N. Ohmichi, A. Nakagawa, P. A. Zunszain, S. Curry, and E. Tsuchida. O₂ and CO binding properties of artificial hemoproteins formed by complexing iron protoporphyrin IX with human serum albumin mutants. *Journal of the American Chemical Society*, 127(45):15933–15942, **2005**.
- [124] B. Halliwell. Albumin-An important extracellular antioxidant? *Biochemical Pharmacology*, 37(4):569–571, **1988**.
- [125] M. Anraku, K. Yamasaki, T. Maruyama, U. Kragh-Hansen, and M. Otagiri. Effect of oxidative stress on the structure and function of human serum albumin. *Pharmaceutical research*, 18(5):632–639, **2001**.
- [126] M. Roche, P. Rondeau, N. R. Singh, E. Tarnus, and E. Bourdon. The antioxidant properties of serum albumin. *FEBS Letters*, 582(13):1783–1787, **2008**.
- [127] M. Bruschi, G. Candiano, L. Santucci, and G. M. Ghiggeri. Oxidized albumin. the long way of a protein of uncertain function. *Biochimica et Biophysica Acta - General Subjects*, 1830(12):5473–5479, **2013**.
- [128] G. Colombo, M. Clerici, D. Giustarini, R. Rossi, A. Milzani, and I. Dalle-Donne. Redox albuminomics: oxidized albumin in human diseases. *Antioxidants & redox signaling*, 17(11):1515–27, **2012**.
- [129] B. Halliwell and J. M. Gutteridge. The antioxidants of human extracellular fluids. *Archives of Biochemistry and Biophysics*, 280(1):1–8, **1990**.

- [130] S. J. Klebanoff. Oxygen Metabolites from Phagocytes. In R. Gallin, J. I. and Snyderman, editor, *Inflammation: Basic principles and clinical correlates*, pages 721–768. Lippincott Williams and Wilkins, Philadelphia, **1999**.
- [131] J. E. Harrison and J. Schultz. Studies on the chlorinating activity of myeloperoxidase. *The Journal of Biological Chemistry*, 251(5):1371–4, **1976**.
- [132] S. J. Klebanoff. Myeloperoxidase: friend and foe. *Journal of Leukocyte Biology*, 77(5):598–625, **2005**.
- [133] J. C. Morris. The acid ionization constant of HOCl from 5 to 35 °. *The Journal of Physical Chemistry*, 70(12):3798–3805, **1966**.
- [134] C. Bernofsky. Nucleotide chloramines and neutrophil-mediated cytotoxicity. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 5(3):295–300, **1991**.
- [135] S. J. Weiss. Tissue destruction by neutrophils. *The New England journal of medicine*, 320(6):365–76, **1989**.
- [136] R. Stocker and J. F. Keaney. Role of oxidative modifications in atherosclerosis. *Physiological reviews*, 84(4):1381–478, **2004**.
- [137] Y. W. Yap, M. Whiteman, and N. S. Cheung. Chlorinative stress: An under appreciated mediator of neurodegeneration? *Cellular Signalling*, 19(2):219–228, **2007**.
- [138] C. C. Winterbourn. Comparative reactivities of various biological compounds with myeloperoxidase-hydrogen peroxide-chloride, and similarity of oxidant to hypochlorite. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 840(2):204–210, **1985**.
- [139] D. I. Pattison and M. J. Davies. Absolute rate constants for the reaction of hypochlorous acid with protein side chains and peptide bonds. *Chemical Research in Toxicology*, 14(10):1453–1464, **2001**.
- [140] M. J. Davies. The oxidative environment and protein damage. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1703(2):93–109, **2005**.
- [141] C. L. Hawkins, D. I. Pattison, and M. J. Davies. Hypochlorite-induced oxidation of amino acids, peptides and proteins. *Amino Acids*, 25(3-4):259–274, **2003**.
- [142] K. Mera, M. Anraku, K. Kitamura, K. Nakajou, T. Maruyama, and M. Otagiri. The structure and function of oxidized albumin in hemodialysis patients: Its role in elevated oxidative stress via neutrophil burst. *Biochemical and Biophysical Research Communications*, 334(4):1322–1328, **2005**.
- [143] L. Musante, M. Bruschi, G. Candiano, A. Petretto, N. Dimasi, P. D. Boccio, A. Urbani, G. Rialdi, and G. M. Ghiggeri. Characterization of oxidation end product of plasma albumin ‘in vivo’. *Biochemical and Biophysical Research Communications*, 349(2):668–673, **2006**.

- [144] Y. Iwao, M. Anraku, M. Hiraike, K. Kawai, K. Nakajou, T. Kai, A. Suenaga, and M. Otagiri. The structural and pharmacokinetic properties of oxidized Human Serum Albumin, Advanced Oxidation Protein Products (AOPP). *Drug metabolism and pharmacokinetics*, 21(2):140–146, **2006**.
- [145] E. V. Mikhalechik, N. V. Smolina, T. S. Astamirova, I. V. Gorudko, D. V. Grigorieva, V. A. Ivanov, A. V. Sokolov, V. A. Kostevich, S. N. Cherenkevich, and O. M. Panasenko. Human serum albumin modified under oxidative/halogenative stress enhances luminol-dependent chemiluminescence of human neutrophils. *Biophysics*, 58(4):530–536, **2013**.
- [146] V. Binder, S. Ljubojevic, J. Haybaeck, M. Holzer, D. El-Gamal, R. Schicho, B. Pieske, A. Heinemann, and G. Marsche. The myeloperoxidase product hypochlorous acid generates irreversible high-density lipoprotein receptor inhibitors. *Arteriosclerosis, thrombosis, and vascular biology*, 33(5):1020–7, **2013**.
- [147] I. V. Gorudko, D. V. Grigorieva, E. V. Shamova, V. A. Kostevich, A. V. Sokolov, E. V. Mikhalechik, S. N. Cherenkevich, J. Arnhold, and O. M. Panasenko. Hypochlorous acid-modified human serum albumin induces neutrophil NADPH oxidase activation, degranulation, and shape change. *Free radical biology & medicine*, 68:326–34, **2014**.
- [148] L. Pasterk, S. Lemesch, B. Leber, M. Trieb, S. Curcic, V. Stadlbauer, R. Schuligoi, R. Schicho, A. Heinemann, and G. Marsche. Oxidized plasma albumin promotes platelet-endothelial crosstalk and endothelial tissue factor expression. *Scientific Reports*, 6:22104, **2016**.
- [149] R. Baker. Studies on the reaction between sodium hypochlorite and proteins: 1. Physico-chemical study of the course of the reaction. *Biochemical Journal*, 4:337–342, **1947**.
- [150] A. Temple, T.-Y. Yen, and S. Gronert. Identification of specific protein carbonylation sites in model oxidations of human serum albumin. *Journal of the American Society for Mass Spectrometry*, 17(8):1172–80, **2006**.
- [151] P. Salavej, H. Spalteholz, and J. Arnhold. Modification of amino acid residues in human serum albumin by myeloperoxidase. *Free Radical Biology and Medicine*, 40(3):516–525, **2006**.
- [152] S. Guedes, R. Vitorino, R. Domingues, F. Amado, and P. Domingues. Oxidation of bovine serum albumin: identification of oxidation products and structural modifications. *Rapid Communications in Mass Spectrometry*, 23(15):2307–15, **2009**.
- [153] D. I. Pattison, C. L. Hawkins, and M. J. Davies. What are the plasma targets of the oxidant hypochlorous acid? A kinetic modeling approach. *Chemical research in toxicology*, 22(5):807–17, **2009**.

- [154] E. Lissi, M. Alicia Biasutti, E. Abuin, and L. León. A fluorescence study of human serum albumin binding sites modification by hypochlorite. *Journal of photochemistry and photobiology. B, Biology*, 94(2):77–81, **2009**.
- [155] O. A. Azizova, A. V. Aseychev, E. M. Beckman, S. N. Moskvina, O. I. Skotnikova, N. V. Smolina, Y. A. Gryzunov, and G. E. Dobretsov. Studies of oxidant-induced changes in albumin transport function with a fluorescent probe K-35. Effect of hypochlorite. *Bulletin of Experimental Biology and Medicine*, 152(6):712–716, **2012**.
- [156] M. Petrônio, J. R. Fernandes, M. L. de Menezes, and V. F. Ximenes. Oxidation of bovine albumin by hypochlorous and hypobromous acids: Structural and functional alterations. *British Journal of Pharmaceutical Research*, 3:147–160, **2013**.
- [157] F. Monacelli and D. Storace. Structural alterations of Human Serum Albumin caused by glycative and oxidative stressors revealed by circular dichroism analysis. *International Journal of Molecular Sciences*, 14(6):10694–709, **2013**.
- [158] F. S. Graciani and V. F. Ximenes. Investigation of Human Albumin-Induced Circular Dichroism in Dansylglycine. *PLoS ONE*, 8(10):e76849, **2013**.
- [159] G. Sancataldo, V. Vetri, V. Foderà, G. D. Cara, V. Militello, and M. Leone. Oxidation enhances human serum albumin thermal stability and changes the routes of amyloid fibril formation. *PLoS ONE*, 9(1):e84552, **2014**.
- [160] A. R. Wyatt, J. R. Kumita, R. W. Mifsud, C. A. Gooden, M. R. Wilson, and C. M. Dobson. Hypochlorite-induced structural modifications enhance the chaperone activity of human α 2-macroglobulin. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20):E2081–90, **2014**.
- [161] F. Zsila, Z. Bikádi, and M. Simonyi. Probing the binding of the flavonoid, quercetin to human serum albumin by circular dichroism, electronic absorption spectroscopy and molecular modelling methods. *Biochemical pharmacology*, 65:447–456, **2003**.
- [162] F. Zsila, Z. Bikádi, and M. Simonyi. Unique, pH-dependent biphasic band shape of the visible circular dichroism of curcumin-serum albumin complex. *Biochemical and Biophysical Research Communications*, 301(3):776–782, **2003**.
- [163] A. Labrador, Y. Cerenius, C. Svensson, K. Theodor, and T. Plivelic. The yellow mini-hutch for SAXS experiments at MAX IV Laboratory. *Journal of Physics: Conference Series*, 425(7):072019, **2013**.
- [164] S. S. Nielsen, K. N. Toft, D. Snakenborg, M. G. Jeppesen, J. K. Jacobsen, B. Vestergaard, J. P. Kutter, and L. Arleth. BioXTAS RAW , a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *Journal of Applied Crystallography*, 42(5):959–964, **2009**.

- [165] T. N. Blanton, T. C. Huang, H. Toraya, C. R. Hubbard, S. B. Robie, D. Louër, H. E. Göbel, G. Will, R. Gilles, and T. Raftery. JCPD-International Centre for Diffraction Data round robin study of silver behenate. A possible low-angle X-ray diffraction calibration standard. *Powder Diffraction*, 10(2):91–95, **1995**.
- [166] G. Porod. General theory. In O. Glatter and O. Kratky, editors, *Small-Angle X-ray Scattering*, pages 17–51. Academic Press, London, UK., **1982**.
- [167] Y. H. Chen, J. T. Yang, and H. M. Martinez. Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry*, 11(22):4120–31, **1972**.
- [168] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, and J. Kardos. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 112(24):E3095–103, **2015**.
- [169] N. Sreerama and R. W. Woody. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical biochemistry*, 287(2):252–260, **2000**.
- [170] A. de Juan and R. Tauler. Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution. *Analytica Chimica Acta*, 500(1-2):195–210, **2003**.
- [171] C. Hawkins and M. Davies. Hypochlorite-induced oxidation of proteins in plasma: formation of chloramines and nitrogen-centred radicals and their role in protein fragmentation. *Biochemical Journal*, 340(Pt 2):539–548, **1999**.
- [172] M. V. Petoukhov, I. M. L. Billas, M. Takacs, M. A. Graewert, D. Moras, and D. I. Svergun. Reconstruction of quaternary structure from X-ray scattering by equilibrium mixtures of biological macromolecules. *Biochemistry*, 52(39):6844–55, **2013**.
- [173] H. S. Chan, S. Bromberg, and K. A. Dill. Models of cooperativity in protein folding. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 348(1323):61–70, **1995**.
- [174] E. V. Schmalhausen, E. B. Zhlobek, I. N. Shalova, O. Firuzi, L. Saso, and V. I. Muronetz. Antioxidant and prooxidant effects of quercetin on glyceraldehyde-3-phosphate dehydrogenase. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*, 45(10):1988–93, **2007**.
- [175] R. A. Sharma, A. J. Gescher, and W. P. Steward. Curcumin: The story so far. *European Journal of Cancer*, 41(13):1955–1968, **2005**.
- [176] S. V. Jovanovic, S. Steenken, C. W. Boone, and M. G. Simic. H-atom transfer is a preferred antioxidant mechanism of curcumin. *Journal of the American Chemical Society*, 121(41):9677–9681, **1999**.

- [177] O. Firuzi, P. Mladenka, R. Petrucci, G. Marrosu, and L. Saso. Hypochlorite scavenging in activity of flavonoids. *Journal of Pharmacy and Pharmacology*, 56(6):801–807, **2004**.
- [178] S. Curry, H. Mandelkow, P. Brick, and N. Franks. Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nature structural biology*, 5(9):827–835, **1998**.
- [179] B. Sengupta and P. K. Sengupta. The interaction of quercetin with human serum albumin: A fluorescence spectroscopic study. *Biochemical and Biophysical Research Communications*, 299(3):400–403, **2002**.
- [180] L. Di Bari, S. Ripoli, S. Pradhan, and P. Salvadori. Interactions between quercetin and warfarin for albumin binding: A new eye on food/drug interference. *Chirality*, 22(6):593–596, **2010**.
- [181] A. C. Pulla Reddy, E. Sudharshan, A. G. Appu Rao, and B. R. Lokesh. Interaction of curcumin with human serum albumin - A spectroscopic study. *Lipids*, 34(10):1025–1029, **1999**.
- [182] A. Barik and B. Mishra. Interaction of curcumin with human serum albumin: Thermodynamic properties, fluorescence energy transfer and denaturation effects. *Chemical physics letters*, 436(1-3):239–243, **2007**.
- [183] J. Mandeville, E. Froehlich, and H. Tajmir-Riahi. Study of curcumin and genistein interactions with human serum albumin. *Journal of Pharmaceutical and Biomedical Analysis*, 49(2):468–474, **2009**.
- [184] P. Bourassa, C. D. Kanakis, P. Tarantilis, M. G. Pollissiou, and H. A. Tajmir-Riahi. Resveratrol, genistein, and curcumin bind bovine serum albumin. *Journal of Physical Chemistry B*, 114(9):3348–3354, **2010**.
- [185] S. Ramešová, R. Sokolová, I. Degano, J. Bulíčková, J. Zabka, and M. Gál. On the stability of the bioactive flavonoids quercetin and luteolin under oxygen-free conditions. *Analytical and bioanalytical chemistry*, 402(2):975–82, **2012**.
- [186] S. Dall’Acqua, G. Miolo, G. Innocenti, and S. Caffieri. The photodegradation of quercetin: Relation to oxidation. *Molecules*, 17(8):8898–8907, **2012**.
- [187] H. H. Tønnesen and J. Karlsen. Studies on curcumin and curcuminoids - VI. Kinetics of curcumin degradation in aqueous solution. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung*, 180(5):402–404, **1985**.
- [188] E. Meucci, a. Mordente, and G. E. Martorana. Metal-catalyzed oxidation of human serum albumin: Conformational and functional changes: Implications in protein aging. *Journal of Biological Chemistry*, 266(8):4692–4699, **1991**.
- [189] E. Grippa, F. Pavone, M. T. Gatto, R. Petrucci, G. Marrosu, B. Silvestrini, and L. Saso. In vitro evaluation of antioxidant activity by electrophoresis and high performance liquid chromatography. *Biochimica et Biophysica Acta*, 1524(2-3):171–177, **2000**.

- [190] C. Capeillère-Blandin, V. Gausson, B. Descamps-Latscha, and V. Witko-Sarsat. Biochemical and spectrophotometric significance of advanced oxidized protein products. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1689(2):91–102, **2004**.
- [191] C. W. Hogue, I. Rasquinha, A. G. Szabo, and J. P. MacManus. A new intrinsic fluorescent probe for proteins. Biosynthetic incorporation of 5-hydroxytryptophan into oncomodulin. *FEBS letters*, 310(3):269–272, **1992**.
- [192] J. B. Ross, D. F. Senear, E. Waxman, B. B. Kombo, E. Rusinova, Y. T. Huang, W. R. Laws, and C. a. Hasselbacher. Spectral enhancement of proteins: biological incorporation and fluorescence characterization of 5-hydroxytryptophan in bacteriophage lambda cI repressor. *Proceedings of the National Academy of Sciences of the United States of America*, 89(24):12023–12027, **1992**.
- [193] W. A. Prütz. Interactions of hypochlorous acid with pyrimidine nucleotides, and secondary reactions of chlorinated pyrimidines with GSH, NADH, and other substrates. *Archives of biochemistry and biophysics*, 349(1):183–91, **1998**.
- [194] A. V. Peskin and C. C. Winterbourn. Kinetics of the reactions of hypochlorous acid and amino acid chloramines with thiols, methionine, and ascorbate. *Free Radical Biology and Medicine*, 30(5):572–579, **2001**.
- [195] R. Raffen and F. J. Stevens. Small zone, high-speed gel filtration chromatography to detect protein aggregation associated with light chain pathologies. *Methods in Enzymology. Amyloid, Prions, and Other Protein Aggregates*, 309:318–332, **1999**.
- [196] D. I. Pattison, C. L. Hawkins, and M. J. Davies. Hypochlorous acid-mediated protein oxidation: How important are chloramine transfer reactions and protein tertiary structure? *Biochemistry*, 46(34):9853–9864, **2007**.
- [197] M. S. A. Coker, W.-P. Hu, S. T. Senthilmohan, and A. J. Kettle. Pathways for the decay of organic dichloramines and liberation of antimicrobial chloramine gases. *Chemical research in toxicology*, 21(12):2334–43, **2008**.
- [198] D. I. Pattison and M. J. Davies. Reactions of myeloperoxidase-derived oxidants with biological substrates: gaining chemical insight into human inflammatory diseases. *Current medicinal chemistry*, 13(27):3271–3290, **2006**.
- [199] J. Arnhold, S. Hammersmidt, M. Wagner, S. Mueller, K. Arnold, and E. Grimm. On the Action of Hypochlorite on Human Serum-Albumin. *Biomedica Biochimica Acta*, 49(10):991–997, **1990**.
- [200] M. J. Raftery. Detection and characterization of N-alpha-chloramines by electrospray tandem mass spectrometry. *Analytical biochemistry*, 366(2):218–27, **2007**.
- [201] A. Müller, S. Langklotz, N. Lupilova, K. Kuhlmann, J. E. Bandow, and L. I. O. Leichert. Activation of RidA chaperone function by N-chlorination. *Nature communications*, 5:5804, **2014**.

- [202] J. Yang and J. Foster. Changes in the Intrinsic Viscosity and Optical Rotation of Bovine Plasma Albumin Associated with Acid Binding. *Journal of the American Chemical Society*, 5568(2):1588–1595, **1954**.
- [203] M. Sogami and J. F. Foster. Isomerization reactions of charcoal-defatted bovine plasma albumin. The N-F transition and acid expansion. *Biochemistry*, 7(6):2172–2182, **1968**.
- [204] M. Sogami, S. Nagaoka, K. B. Itoh, and S. Sakata. Fluorimetric studies on the structural transition of bovine plasma albumin in acidic solutions. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 310(1):118–123, **1973**.
- [205] M. Dockal, D. Carter, and F. Rüker. Conformational transitions of the three recombinant domains of human serum albumin depending on pH. *Journal of Biological Chemistry*, 275(5):3042–50, **2000**.
- [206] R. F. Chen. Fluorescence spectra of human serum albumin in the pH region of the N-F transition. *Biochimica et Biophysica Acta (BBA) - Biophysics including Photosynthesis*, 120(1):169–171, **1966**.
- [207] S. Muzammil, Y. Kumar, and S. Tayyab. Molten globule-like state of human serum albumin at low pH. *European journal of biochemistry / FEBS*, 266(1):26–32, **1999**.
- [208] M. Sogami, H. Inouye, S. Nagaoka, and S. Era. Conformational changes of bovine plasma albumin prior to the salting-out of the protein in concentrated salt solution. *International Journal of Peptide and Protein Research*, 20(3):254–258, **1982**.
- [209] S. Era, K. Kuwata, K. Kida, M. Sogami, and A. Yoshida. Circular dichroic and fluorometric studies on the acid-induced isomerization of bovine plasma albumin - 1-anilino-8-naphthalenesulfonate complex. *International Journal of Peptide and Protein Research*, 26(6):575–583, **1985**.
- [210] P. J. Sadler and A. Tucker. pH-induced structural transitions of bovine serum albumin. Histidine pKa values and unfolding of the N-terminus during the N to F transition. *European Journal of Biochemistry*, 212(3):811–7, **1993**.
- [211] S. Era and M. Sogami. ¹H-NMR and CD studies on the structural transition of serum albumin in the acidic region - The N-F transition. *The Journal of Peptide Research*, 52(6):431–442, **1998**.
- [212] M. Y. Khan. Direct evidence for the involvement of domain III in the N-F transition of bovine serum albumin. *The Biochemical Journal*, 236(1):307–310, **1986**.
- [213] E. M. Slayter. An electron microscope study of the conformational change in bovine serum albumin at low pH. *Journal of Molecular Biology*, 14(2):443–452, **1965**.

- [214] J. R. Olivieri and A. F. Craievich. The subdomain structure of human serum albumin in solution under different pH conditions studied by small angle X-ray scattering. *European Biophysics Journal*, 24(2):77–84, **1995**.
- [215] N. El Kadi, N. Taulier, J. Y. Le Hu  rou, M. Gindre, W. Urbach, I. Nwigwe, P. C. Kahn, and M. Waks. Unfolding and refolding of bovine serum albumin at acid pH: ultrasound and structural studies. *Biophysical Journal*, 91(9):3397–3404, **2006**.
- [216] N. Acharya, P. Mishra, and S. K. Jha. Evidence for dry molten globule-Like domains in the pH-Induced equilibrium folding intermediate of a multidomain protein. *Journal of Physical Chemistry Letters*, 7(1):173–179, **2016**.
- [217] K. Baler, O. A. Martin, M. A. Carignano, G. A. Ameer, J. A. Vila, and I. Szleifer. Electrostatic unfolding and interactions of albumin driven by pH changes: A molecular dynamics study. *Journal of Physical Chemistry B*, 118(4):921–930, **2014**.
- [218] K. Baler, R. Michael, I. Szleifer, and G. A. Ameer. Albumin hydrogels formed by electrostatically triggered self-assembly and their drug delivery capability. *Biomacromolecules*, 15(10):3625–3633, **2014**.
- [219] N. Varga, V. Hornok, D. Seb  sk, and I. D  k  ny. Comprehensive study on the structure of the BSA from extended-to aged form in wide (2-12) pH range. *International Journal of Biological Macromolecules*, 88:51–58, **2016**.
- [220] L. R. S. Barbosa, M. G. Ortore, F. Spinozzi, P. Mariani, S. Bernstorff, and R. Itri. The importance of protein-protein interactions on the pH-induced conformational changes of bovine serum albumin: A small-angle x-ray scattering study. *Biophysical Journal*, 98(1):147–157, **2010**.
- [221] W. Altek  r. Fluorescence of proteins in aqueous neutral salt solutions. I. Influence of anions. *Biopolymers*, 16(2):341–368, **1977**.
- [222] S. A. Parke, G. G. Birch, D. B. MacDougall, and D. A. Stevens. Tastes, structure and solution properties of D-glucono-1,5-lactone. *Chemical Senses*, 22(1):53–65, **1997**.
- [223] Y. Pocker and E. Green. Hydrolysis of D-glucono-  -lactone. I. General acid-base catalysis, solvent deuterium isotope effects, and transition state characterization. *Journal of the American Chemical Society*, 95(1):113–119, **1973**.
- [224] C. K. Kuo and P. X. Ma. Ionically crosslinked alginate hydrogels as scaffolds for tissue engineering: Part 1. Structure, gelation rate and mechanical properties. *Biomaterials*, 22(6):511–521, **2001**.
- [225] M. Dentini, G. Rinaldi, A. Barbetta, D. Risica, and G. Skj  k-Br  ek. Acid gel formation in (pseudo) alginates with and without G blocks produced by epimerising mannuronan with C5 epimerases. *Carbohydrate Polymers*, 63(4):519–526, **2006**.

- [226] D. J. Adams, M. F. Butler, W. J. Frith, M. Kirkland, L. Mullen, and P. Sanderson. A new method for maintaining homogeneity during liquid-hydrogel transitions using low molecular weight hydrogelators. *Soft Matter*, 5(9):1856–1862, **2009**.
- [227] A. Z. Cardoso, A. E. Alvarez Alvarez, B. N. Cattoz, P. C. Griffiths, S. M. King, W. J. Frith, and D. J. Adams. The influence of the kinetics of self-assembly on the properties of dipeptide hydrogels. *Faraday Discussions*, 166:101, **2013**.
- [228] R. Dowben. On the Hydrolysis and Oxidation of Glucuronolactone. *Biochimica et Biophysica Acta*, 31:454–458, **1959**.
- [229] R. P. Taylor, V. Chau, M. J. Zenkovich, and L. H. Leake. Stopped-flow studies of the N-F transition in serum albumin. *Biophysical Chemistry*, 7(4):293–299, **1978**.
- [230] P. J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, Ithaca, United States, first edition, **1953**.
- [231] J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, and K. W. Plaxco. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34):12491–6, **2004**.
- [232] P. Bernadó and M. Blackledge. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophysical Journal*, 97(10):2839–45, **2009**.
- [233] G. Evrard, F. Mareuil, F. Bontems, C. Sizun, and J. Perez. DADIMODO: A program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *Journal of Applied Crystallography*, 44(6):1264–1271, **2011**.
- [234] J. E. Curtis, S. Raghunandan, H. Nanda, and S. Krueger. SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Computer Physics Communications*, 183(2):382–389, **2012**.
- [235] M. Calvin. The Path of Carbon in Photosynthesis. *Angewandte Chemie International Edition in English*, 1(2):65–75, **1962**.
- [236] T. Pfannschmidt and C. Yang. The hidden function of photosynthesis: A sensing system for environmental conditions that regulates plant acclimation responses. *Protoplasma*, 249(SUPPL.2):125–136, **2012**.
- [237] B. B. Buchanan. Regulation of CO₂ assimilation in oxygenic photosynthesis: The ferredoxin/thioredoxin system. Perspective on its discovery, present status, and future development. *Archives of Biochemistry and Biophysics*, 288(1):1–9, **1991**.

- [238] P. Schürmann and B. B. Buchanan. The ferredoxin/thioredoxin system of oxygenic photosynthesis. *Antioxidants & redox signaling*, 10(7):1235–74, **2008**.
- [239] M. Zaffagnini, M. Bedhomme, H. Groni, C. H. Marchand, C. Puppo, B. Gontero, C. Cassier-Chauvat, P. Decottignies, and S. D. Lemaire. Glutathionylation in the photosynthetic model organism *Chlamydomonas reinhardtii*: a proteomic survey. *Molecular & cellular proteomics : MCP*, 11(2):M111.014142, **2012**.
- [240] N. Wedel, J. Soll, and B. K. Paap. CP12 provides a new mode of light regulation of Calvin cycle activity in higher plants. *Proceedings of the National Academy of Sciences of the United States of America*, 94(19):10479–10484, **1997**.
- [241] L. Marri, P. Trost, X. Trivelli, L. Gonnelli, P. Pupillo, and F. Sparla. Spontaneous assembly of photosynthetic supramolecular complexes as mediated by the intrinsically unstructured protein CP12. *The Journal of Biological Chemistry*, 283(4):1831–8, **2008**.
- [242] P. E. López-Calcano, T. P. Howard, and C. A. Raines. The CP12 protein family: a thioredoxin-mediated metabolic switch? *Frontiers in plant science*, 5:9, **2014**.
- [243] M. A. Porter and F. C. Hartman. Exploration of the function of a regulatory sulfhydryl of phosphoribulokinase from spinach. *Archives of biochemistry and biophysics*, 281(2):330–4, **1990**.
- [244] T. P. Howard, M. Metodiev, J. C. Lloyd, and C. A. Raines. Thioredoxin-mediated reversible dissociation of a stromal multiprotein complex in response to changes in light availability. *Proceedings of the National Academy of Sciences of the United States of America*, 105(10):4056–4061, **2008**.
- [245] L. Marri, P. Trost, P. Pupillo, and F. Sparla. Reconstitution and properties of the recombinant glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase supramolecular complex of *Arabidopsis*. *Plant Physiology*, 139(3):1433–1443, **2005**.
- [246] H. Matsumura, A. Kai, T. Maeda, M. Tamoi, A. Satoh, H. Tamura, M. Hirose, T. Ogawa, N. Kizu, A. Wadano, T. Inoue, and S. Shigeoka. Structure basis for the regulation of glyceraldehyde-3-phosphate dehydrogenase activity via the intrinsically disordered protein CP12. *Structure*, 19(12):1846–1854, **2011**.
- [247] S. Fermani, X. Trivelli, F. Sparla, A. Thumiger, M. Calvaresi, L. Marri, G. Falini, F. Zerbetto, and P. Trost. Conformational selection and folding-upon-binding of intrinsically disordered protein CP12 regulate photosynthetic enzymes assembly. *The Journal of Biological Chemistry*, 287(25):21372–83, **2012**.
- [248] E. Graciet, P. Gans, N. Wedel, S. Lebreton, J. M. Camadro, and B. Gontero. The small protein CP12: A protein linker for supramolecular complex assembly. *Biochemistry*, 42(27):8163–8170, **2003**.

- [249] R. Scheibe, N. Wedel, S. Vetter, V. Emmerlich, and S. M. Sauermann. Co-existence of two regulatory NADP-glyceraldehyde 3-P dehydrogenase complexes in higher plant chloroplasts. *European Journal of Biochemistry*, 269(22):5617–5624, **2002**.
- [250] E. Baalmann, R. Scheibe, R. Cerff, and W. Martin. Functional studies of chloroplast glyceraldehyde-3-phosphate dehydrogenase subunits A and B expressed in *Escherichia coli*: formation of highly active A4 and B4 homotetramers and evidence that aggregation of the B4 complex is mediated by the B subunit carb. *Plant molecular biology*, 32(3):505–13, **1996**.
- [251] F. Sparla, P. Pupillo, and P. Trost. The C-terminal extension of glyceraldehyde-3-phosphate dehydrogenase subunit B acts as an autoinhibitory domain regulated by thioredoxins and nicotinamide adenine dinucleotide. *Journal of Biological Chemistry*, 277(47):44946–44952, **2002**.
- [252] S. Fermani, F. Sparla, G. Falini, P. L. Martelli, R. Casadio, P. Pupillo, A. Ripamonti, and P. Trost. Molecular mechanism of thioredoxin regulation in photosynthetic A2B2-glyceraldehyde-3-phosphate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 104(26):11109–14, **2007**.
- [253] P. Pupillo and G. G. Piccari. The effect of NADP on the subunit structure and activity of spinach chloroplast glyceraldehyde-3-phosphate dehydrogenase. *Archives of Biochemistry and Biophysics*, 154(1):324–331, **1973**.
- [254] S. Scagliarini, P. Trost, P. Pupillo, and V. Valenti. Light activation and molecular-mass changes of NAD(P)-glyceraldehyde 3-phosphate dehydrogenase of spinach and maize leaves. *Planta*, 190(3):313–319, **1993**.
- [255] F. Sparla, M. Zaffagnini, N. Wedel, R. Scheibe, P. Pupillo, and P. Trost. Regulation of photosynthetic GAPDH dissected by mutants. *Plant Physiology*, 138(4):2210–2219, **2005**.
- [256] D. H. Harrison, J. a. Runquist, a. Holub, and H. M. Miziorko. The crystal structure of phosphoribulokinase from *Rhodobacter sphaeroides* reveals a fold similar to that of adenylate kinase. *Biochemistry*, 37(15):5074–85, **1998**.
- [257] M. A. Porter and F. C. Hartman. Catalytic nonessentiality of an active-site cysteinyl residue of phosphoribulokinase. *Journal of Biological Chemistry*, 263(29):14846–14849, **1988**.
- [258] H. K. Brandes, F. W. Larimer, and F. C. Hartman. The molecular pathway for the regulation of phosphoribulokinase by thioredoxin f. *Journal of Biological Chemistry*, 271(7):3333–3335, **1996**.
- [259] N. Wedel and J. Soll. Evolutionary conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/CP12/glyceraldehyde-3-phosphate dehydrogenase complex dissociation. *Proceedings of the National Academy of Sciences of the United States of America*, 95(16):9699–9704, **1998**.

- [260] T. P. Howard, J. C. Lloyd, and C. A. Raines. Inter-species variation in the oligomeric states of the higher plant Calvin cycle enzymes glyceraldehyde-3-phosphate dehydrogenase and phosphoribulokinase. *Journal of Experimental Botany*, 62(11):3799–3805, **2011**.
- [261] P. K. Smith, R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, and D. C. Klenk. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*, 150(1):76–85, **1985**.
- [262] S. Scagliarini, P. Trost, and P. Pupillo. The non-regulatory isoform of NAD(P)-glyceraldehyde-3-phosphate dehydrogenase from spinach chloroplasts. *Journal of Experimental Botany*, 49(325):1307–1315, **1998**.
- [263] P. Pernot, A. Round, R. Barrett, A. De Maria Antolinos, A. Gobbo, E. Gordon, J. Huet, J. Kieffer, M. Lentini, M. Mattenet, C. Morawe, C. Mueller-Dieckmann, S. Ohlsson, W. Schmid, J. Surr, P. Theveneau, L. Zerrad, and S. McSweeney. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *Journal of synchrotron radiation*, 20(Pt 4):660–4, **2013**.
- [264] M. E. Brennich, J. Kieffer, G. Bonamis, A. De Maria Antolinos, S. Hutin, P. Pernot, and A. Round. Online data analysis at the ESRF bioSAXS beamline, BM29. *Journal of Applied Crystallography*, 49(1):203–212, **2016**.
- [265] V. Sobolev, E. Eyal, S. Gerzon, V. Potapov, M. Babor, J. Prilusky, and M. Edelman. SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment. *Nucleic Acids Research*, 33(2):W39–W43, **2005**.
- [266] J. G. Garcia de la Torre, M. L. Huertas, and B. Carrasco. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophysical Journal*, 78(2):719–730, **2000**.
- [267] A. Ortega, D. Amoros, and J. de la Torre. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophysical Journal*, 101(4):892–898, **2011**.
- [268] C. M. Torres-Bugeau, C. L. Ávila, R. Raisman-Vozari, D. Papy-Garcia, R. Itri, L. R. S. Barbosa, L. M. Cortez, V. L. Sim, and R. N. Chehín. Characterization of heparin-induced glyceraldehyde-3-phosphate dehydrogenase early amyloid-like oligomers and their implication in alpha-synuclein aggregation. *Journal of Biological Chemistry*, 287(4):2398–2409, **2012**.
- [269] M. V. Petoukhov and D. I. Svergun. Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallographica Section D: Biological Crystallography*, 71(5):1051–1058, **2015**.
- [270] F. Mouche, B. Gontero, I. Callebaut, J.-P. Mornon, and N. Boisset. Striking Conformational Change Suspected within the Phosphoribulokinase Dimer Induced by Interaction with GAPDH. *Journal of Biological Chemistry*, 277(8):6743–6749, **2002**.

- [271] A. Del Giudice, N. V. Pavel, L. Galantini, G. Falini, P. Trost, S. Fermani, and F. Sparla. Unravelling the shape and structural assembly of the photosynthetic GAPDH-CP12-PRK complex from *Arabidopsis thaliana* by small-angle X-ray scattering analysis. *Acta Crystallographica Section D Biological Crystallography*, 71(12):2372–2385, **2015**.
- [272] E. Baalmann, J. E. Backhausen, C. Kitzmann, and R. Scheibe. Regulation of NADP-dependent glyceraldehyde 3-phosphate dehydrogenase activity in spinach chloroplasts. *Botanica Acta*, 107(5):313–320, **1994**.
- [273] E. Brookes, J. Pérez, B. Cardinali, A. Profumo, P. Vachette, and M. Rocco. Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan Solution Modeler (US-SOMO) enhanced SAS module. *Journal of Applied Crystallography*, 46(6):1823–1833, **2013**.
- [274] E. Brookes, P. Vachette, M. Rocco, and J. Pérez. US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *Journal of Applied Crystallography*, 49(5):36–39, **2016**.
- [275] O. D. Monera, C. M. Kay, and R. S. Hodges. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein science*, 3(11):1984–91, **1994**.
- [276] W. K. Lim, J. Rosgen, and S. W. Englander. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proceedings of the National Academy of Sciences of the United States of America*, 106(8):2595–2600, **2009**.
- [277] C. K. Chan, Y. Hu, S. Takahashi, D. L. Rousseau, W. A. Eaton, and J. Hofrichter. Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proceedings of the National Academy of Sciences of the United States of America*, 94(5):1779–1784, **1997**.
- [278] A. Martel, M. Burghammer, R. Davies, E. DiCola, P. Panine, J. B. Salmon, and C. Riekel. A microfluidic cell for studying the formation of regenerated silk by synchrotron radiation small- and wide-angle X-ray scattering. *Biomicrofluidics*, 2(2):024104, **2008**.
- [279] J. P. Lafleur, D. Snakenborg, S. S. Nielsen, M. Mller, K. N. Toft, A. Menzel, J. K. Jacobsen, B. Vestergaard, L. Arleth, and J. P. Kutter. Automated microfluidic sample-preparation platform for high-throughput structural investigation of proteins by small-angle X-ray scattering. *Journal of Applied Crystallography*, 44(5):1090–1099, **2011**.
- [280] M. Skou, S. Skou, T. G. Jensen, B. Vestergaard, and R. E. Gillilan. In situ microfluidic dialysis for biological small-angle X-ray scattering. *Journal of Applied Crystallography*, 47(4):1355–1366, **2014**.

-
- [281] F. Schwemmer, C. Blanchet, A. Spilotros, D. Kosse, S. Zehnle, H. Mertens, M. A. Gräwert, M. Roessle, N. Paust, D. Svergun, F. von Stetten, R. Zengerle, and D. Mark. LabDisk for SAXS: A centrifugal microfluidic sample preparation platform for small-angle X-ray scattering. *Lab Chip*, 16(7):1161–1170, **2016**.
- [282] B. Vestergaard and Z. Sayers. Investigating increasingly complex macromolecular systems with small-angle X-ray scattering. *IUCrJ*, 1(Pt 6):523–9, **2014**.