



This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

**Biochemical, biophysical, and structural studies of
a protein complex implicated in the erythrocyte
interaction with the Malaria parasite
*Plasmodium falciparum***

Manuel Blanc

Submitted in partial fulfilment of the requirements of the degree of
Doctor of Philosophy

June 2015

Research Institute for the Environment, Physical Sciences and Applied Mathematics

Keele University

SUBMISSION OF THESIS FOR A RESEARCH DEGREE

Degree submitted for PhD degree

Title of thesis Biochemical, biophysical, and structural studies of a protein complex implicated in the erythrocyte interaction with the Malaria parasite *Plasmodium falciparum*

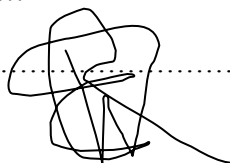
Date of submission September 2014
Original registration date 1st October 2010
Name of candidate Manuel Blanc
Research Institute I.L.L. / E.S.R.F. / EPSAM: Macromolecular Structure Group
Name of Lead Supervisor Professor V.T. Forsyth

I certify that:

- (a) The thesis being submitted for examination is my own account of my own research
- (b) My research has been conducted ethically. Where relevant a letter from the approving body confirming that ethical approval has been given has been bound in the thesis as an Annex
- (c) The data and results presented are the genuine data and results actually obtained by me during the conduct of the research
- (d) Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis
- (e) Where any collaboration has taken place with one or more other researchers, I have included within an 'Acknowledgments' section in the thesis a clear statement of their contributions, in line with the relevant statement in the Code of Practice (see Note overleaf).
- (f) The greater portion of the work described in the thesis has been undertaken subsequent to my registration for the higher degree for which I am submitting for examination
- (g) Where part of the work described in the thesis has previously been incorporated in another thesis submitted by me for a higher degree (if any), this has been identified and acknowledged in the thesis
- (h) The thesis submitted is within the required word limit as specified in the Regulations

Total words in submitted thesis (including text and footnotes, but excluding references and appendices)42649.....

Signature of candidate



Date01/09/2014...

ABSTRACT

This thesis describes biochemical and biophysical studies of two protein domains that are believed to be involved in the interaction between the merozoites of Plasmodium falciparum and human red blood cells. The parasite protein fragment derives from the erythrocyte binding antigen 181 (EBA-181) invasion protein, and the human protein fragment comes from the 4.1R erythrocyte skeletal protein. The initial goal of the PhD project was to derive structural information on the nature of this complex, with a perspective towards generating new therapeutic approaches.

Extensive biochemical and biophysical characterisation of the complex was carried out and is described in detail in Chapter 3 of the thesis: the results confirm the interaction, add insights to the stability of the complex and suggest the presence of significant disorder in both the individual proteins and the complex. Structural studies were carried out using small-angle neutron and X-ray scattering, used in conjunction with selective deuteration. These studies, which are described in Chapter 4, provide low resolution images of the individual proteins and of the complex; these have been compared to structure predictions using bioinformatics. In Chapter 5, solution state NMR studies were carried out, principally on the EBA-181 protein, but with preliminary results from titration work designed to further probe the nature of the interaction between the two proteins. Chapter 6 concludes the thesis with a summary of the work placed in context of the host-pathogen interaction, and proposes directions for future work.

TABLE OF CONTENTS

| | |
|---|------------|
| ABSTRACT | i |
| TABLE OF CONTENTS | ii |
| GLOSSARY | vii |
| ACKNOWLEDGMENTS | x |
| 1. INTRODUCTION | 1 |
| 1.1 Malaria: an emerging world disease | 1 |
| 1.2 Life cycle of the parasite | 3 |
| 1.3 Treatment and resistance | 7 |
| 1.4 The process of parasitic invasion and the role of erythrocyte binding-like (EBL) proteins..... | 9 |
| 1.5 Structural information: the state-of-the-art..... | 13 |
| 1.5.1 Protein complex between <i>P. falciparum</i> EBA-181 and erythrocyte 4.1R | 13 |
| 2. EXPERIMENTAL TECHNIQUES | 18 |
| 2.1 Molecular biology methods | 19 |
| 2.1.1 Plasmid DNA preparation | 19 |
| 2.1.2 <i>In-vitro</i> transposition | 19 |
| 2.1.3 Polyacrylamide gel electrophoresis..... | 21 |
| 2.1.4 Affinity purification | 22 |
| 2.1.5 Size-exclusion chromatography | 23 |
| 2.1.6 Protein concentration measurements..... | 23 |
| 2.1.7 Pull-down assays for purification and characterisation of the complex | 25 |
| 2.2 Biochemical and biophysical characterisation techniques | 26 |
| 2.2.1 Time-of-flight mass spectrometry (TOF-MS) | 26 |
| 2.2.2 N-terminal sequencing | 26 |
| 2.2.3 Dynamic light scattering (DLS)..... | 27 |
| 2.2.4 Multi-angle laser light scattering coupled to size exclusion chromatography and refractive index (SEC-MALLS-RI) | 28 |
| 2.2.5 Circular dichroism (CD) | 30 |

| | | |
|-----------|--|-----------|
| 2.2.6 | Fluorescence thermal shift assay (TSA) | 33 |
| 2.2.7 | 1D NMR spectroscopy | 35 |
| 2.2.8 | Heteronuclear single quantum coherence NMR spectroscopy | 36 |
| 2.3 | Small-angle X-ray and neutron solution scattering (SAXS and SANS) | 37 |
| 2.3.1 | Small-angle scattering: X-rays and neutrons as complementary probes | 37 |
| 2.3.2 | Deuteration for SANS | 39 |
| 2.3.3 | The ESRF synchrotron radiation source | 41 |
| 2.3.4 | The Institut Laue Langevin neutron source | 44 |
| 2.3.5 | Data treatment | 47 |
| 2.4 | Protein crystallisation | 48 |
| 2.4.1 | Fundamental principles: | 48 |
| 2.4.2 | Preparation of crystallography experiments | 50 |
| 3. | THE EBA181-4.1R PROTEIN COMPLEX: EXPRESSION AND CHARACTERISATION BY BIOCHEMICAL AND BIOPHYSICAL TECHNIQUES | 51 |
| 3.1 | Introduction | 52 |
| 3.2 | Production of individual EBA181 ₉₄₅₋₁₀₉₇ and 4.1R _{10kDa} proteins | 53 |
| 3.2.1 | Plasmid DNA: extraction and quality control | 53 |
| 3.2.2 | Change of resistance marker | 55 |
| 3.2.3 | Bacterial transformation and long-term storage of clones | 56 |
| 3.2.4 | Expression and solubility tests | 57 |
| 3.2.5 | Large-scale bacterial cultures | 58 |
| 3.2.6 | Protein purification | 60 |
| 3.3 | Preparation and purification of the complex | 73 |
| 3.4 | Biochemical characterisation of individual EBA181 ₉₄₅₋₁₀₉₇ and 4.1R _{10kDa} proteins and the complex | 76 |
| 3.4.1 | Quantification of nucleic acid contamination | 76 |
| 3.4.2 | Storage and degradation tests | 78 |
| 3.4.3 | Development of a 'micro pull-down' protocol | 80 |
| 3.4.4 | Inhibition assays | 81 |
| 3.4.5 | Application of the micro pull-down technique | 83 |
| 3.4.6 | Native polyacrylamide gel electrophoresis | 94 |

| | | |
|--|--|------------|
| 3.4.7 | Protein quantification using SDS-PAGE | 97 |
| 3.5 | Biophysical characterisation of individual EBA181 ₉₄₅₋₁₀₉₇ and 4.1R _{10kDa} proteins and the complex | 104 |
| 3.5.1 | Structural bioinformatics analysis..... | 104 |
| 3.5.2 | 1D NMR assessment of EBA181 ₉₄₅₋₁₀₉₇ , 4.1R _{10kDa} and the complex | 120 |
| 3.5.3 | Circular dichroism (CD) measurements..... | 123 |
| 3.5.4 | Thermal shift assay (TSA) measurements | 125 |
| 3.5.5 | Multi-angle laser light scattering coupled to size exclusion chromatography and refractive index (SEC-MALLS-RI) for determination of oligomeric state and molar mass calculation | 129 |
| 3.5.6 | Dynamic light scattering (DLS) measurements | 135 |
| 3.5.7 | Isothermal titration calorimetry (ITC) studies of the complex | 140 |
| 3.5.8 | Micro-scale thermophoresis (MST) studies of the complex | 142 |
| 3.5.9 | Crystallisation trials | 143 |
| 3.6 | Discussion and conclusions..... | 147 |
| 4. | SAXS AND SANS CHARACTERISATION OF EBA181₉₄₅₋₁₀₉₇, 4.1R_{10kDa} AND THEIR COMPLEX. | 152 |
| 4.1 | Introduction | 153 |
| 4.2 | Material and methods | 154 |
| 4.2.1 | Sample production for small-angle X-ray scattering | 155 |
| 4.2.2 | Sample production for small-angle neutron scattering | 156 |
| 4.2.3 | SAXS and SANS data collection | 163 |
| 4.2.4 | Data treatment | 166 |
| 4.3 | Structural characterisation using small-angle X-ray scattering..... | 167 |
| 4.3.1 | SAXS characterisation of tagged GST-4.1R _{10kDa} and..... | 168 |
| His-EBA181 ₉₄₅₋₁₀₉₇ proteins..... | 168 | |
| 4.3.2 | SAXS characterisation of untagged 4.1R _{10kDa} and EBA181 ₉₄₅₋₁₀₉₇ proteins | 178 |
| 4.3.3 | SAXS characterisation of the EBA181 ₉₄₅₋₁₀₉₇ / 4.1R _{10kDa} protein complex. | 185 |
| 4.4 | Structural characterisation of the EBA-181 / 4.1R complex using small-angle neutron scattering | 190 |
| 4.4.1 | SANS characterisation of the EBA181 ₉₄₅₋₁₀₉₇ / 4.1R _{10kDa} protein complex in 0 % D ₂ O | 191 |

| | | |
|-----------|---|------------|
| 4.4.2 | SANS characterisation of the EBA181 ₉₄₅₋₁₀₉₇ / 4.1R _{10kDa} protein complex in 40 % D ₂ O | 196 |
| 4.4.3 | SANS characterisation of the EBA181 ₉₄₅₋₁₀₉₇ / 4.1R _{10kDa} protein complex in 100 % D ₂ O | 201 |
| 4.4.4 | Validation of the data | 204 |
| 4.5 | Discussion and conclusions: SAXS and SANS consensus | 206 |
| 5. | SOLUTION STATE NMR STUDIES OF EBA181₉₄₅₋₁₀₉₇, 4.1R_{10kDa} AND THEIR COMPLEX | 213 |
| 5.1 | Introduction | 214 |
| 5.2 | Material and methods | 216 |
| 5.2.1 | Sample production and labelling for NMR solution studies | 216 |
| 5.2.2 | Solution-state NMR experiments | 218 |
| 5.3 | NMR solution spectroscopy measurements from EBA181 ₉₄₅₋₁₀₉₇ | 220 |
| 5.3.1 | Structural and dynamic characterisation of EBA181 ₉₄₅₋₁₀₉₇ | 220 |
| 5.3.2 | Structural ensemble description of EBA181 ₉₄₅₋₁₀₉₇ from experimental chemical shifts | 226 |
| 5.3.3 | Validation of structural ensembles by experimental residual dipolar couplings | 230 |
| 5.4 | NMR solution spectroscopy results of the complex | 232 |
| 5.5 | Discussion and conclusions | 236 |
| 6. | DISCUSSION, CONCLUSIONS and FUTURE WORK | 238 |
| | APPENDIX A: ADDITIONAL MATERIAL | 249 |
| A.1 | Bacterial media | 249 |
| | LB broth medium | 249 |
| | LB-agar | 249 |
| | Super optimal broth with catabolite repression | 250 |
| | pH adjustment | 250 |
| | Sterilisation | 250 |
| A.2 | Composition of buffers | 251 |
| | Buffers used for protein purification and storage | 251 |
| | Native polyacrylamide gel electrophoresis | 252 |
| | APPENDIX B: EBA-181 AND 4.1R PROTEINS | 255 |
| B.1 | The cloning strategy | 255 |

| | |
|--|------------|
| Protein EBA181 ₉₄₅₋₁₀₉₇ | 255 |
| Protein 4.1R _{10kDa} | 256 |
| B.2 EBA-181 / EBA-175 sequence alignment..... | 260 |
| B.3 Protein EBA181 ₉₄₅₋₁₀₉₇ sequence and parameters | 263 |
| B.4 Protein 4.1R _{10kDa} sequence and parameters..... | 264 |
| APPENDIX C: NMR supplemental information | 266 |
| C.1 Assigned chemical shifts of EBA181 ₉₄₅₋₁₀₉₇ | 266 |
| REFERENCES | 271 |

GLOSSARY

His-EBA181₉₄₅₋₁₀₉₇: tagged domain of EBA-181 encompassing the residues 945 to 1097

EBA181₉₄₅₋₁₀₉₇: untagged domain of EBA-181 encompassing the residues 945 to 1097

GST-4.1R_{10kDa}: GST-tagged 4.1R 10 kDa domain encompassing the residues 405 to 471

4.1R_{10kDa}: untagged 4.1R 10 kDa domain encompassing the residues 405 to 471

| | |
|---------------------------------|---|
| Amp: | Ampicillin |
| Cam: | Chloramphenicol |
| CD: | Circular dichroism |
| C-terminus / N-terminus: | Carboxy-terminal / amino-terminal part of a protein |
| Da / kDa: | Dalton / kilo Dalton |
| ddH₂O: | Double distilled H ₂ O |
| DLS: | Dynamic light scattering |
| D_{max}: | Maximum distance of a particle from the distance distribution function P(r) |
| DTT: | Dithiothreitol |
| <i>E.coli</i>: | <i>Escherichia coli</i> |
| EDTA: | Ethylenediaminetetraacetic acid |
| FPLC: | Fast protein liquid chromatography |
| GSH: | Reduced glutathione |
| GST: | Glutathione S-transferase |
| HSQC: | Heteronuclear single quantum coherence (NMR spectroscopy) |

| | |
|--------------------------|---|
| IDP(s): | Intrinsically disordered protein(s) |
| IMAC: | Immobilised metal affinity chromatography |
| IPTG: | Isopropyl- β -D-thiogalactopyranoside |
| ITC: | Isothermal titration calorimetry |
| Kan: | Kanamycin |
| LB: | Luria-Bertani broth medium |
| MM: | Molar mass |
| MW: | Molecular weight |
| MST: | Micro-scale thermophoresis |
| NaP: | Sodium phosphate buffer (NaH_2PO_4 / Na_2HPO_4) |
| nOe: | Nuclear Overhauser effect (transfer of nuclear spin polarisation from one nuclear spin population to another) |
| NMR: | Nuclear magnetic resonance |
| OD: | Optical density |
| PBS: | Phosphate buffered saline |
| pI: | Iso-electric point |
| PDB: | Protein Data Bank |
| p.p.m.: | Parts-per-million |
| P(r): | Distance distribution function |
| R_g: | Radius of gyration |
| R_h: | Hydrodynamic radius |
| rpm: | Rotations per minute |
| RT: | Room temperature (25°C) |

| | |
|----------------------|--|
| RT-PCR: | Reverse transcription polymerase chain reaction |
| SA: | Sialic acid |
| SANS: | Small-angle neutron scattering |
| SAXS: | Small-angle X-ray scattering |
| SDS-PAGE: | Sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| SEC-MALLS-RI: | Multi-angle laser light scattering coupled to size exclusion chromatography and refractive index |
| SOC: | Super optimal broth medium with catabolite repression |
| SP: | Sulphadoxinepyrimethamine |
| Tag: | Peptide sequence genetically fused to a recombinant protein (6x histidines or GST) |
| Tris: | 2-amino-2-hydroxymethyl-1,3-propanediol |
| TSA: | Thermal shift assay |

Remark: amino-acids and DNA/RNA bases are described using the international codes.

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my supervisors, Prof. Trevor Forsyth, Dr. Edward Mitchell and Dr. Michael Haertlein for their support, their expertise and their supervision. It was extremely valuable to work with the three of you. You have complemented each other with your various backgrounds and your suggestions; this was a strength for exploring different avenues and fresh ideas. I also want to warmly thank you for your patience and the helpful revisions of the current manuscript.

I am very grateful to the other people who helped me during this thesis work. Firstly, I would like to kindly acknowledge our South African collaborators, Dr. Theresa Coetzer and Dr. Kubendran Naidoo (University of Witwatersrand, South Africa) for initiating the project and for both their help and their enthusiasm during my trip to Johannesburg. Thank you for showing me your beautiful country! I also would like to thank Dr. Julien Perard (CNRS / UVHCI) for his advice and his strong expertise with the use of the PSB Biophysics Platform. So thank you very much Julien! Special thanks are given to Dr. Louiza Zerrad (ESRF) for her assistance with the ESRF beamlines and the SAXS data. I also want to thank Dr. Adam Round (EMBL) for his advice on data treatment during the 'Neutrons In Biology' course. Other particular thanks go to Dr. Joseph Zaccai (ILL). Thank you very much Jo for your help with the acquisition and the treatment of the SANS data. I am also grateful to Dr. Anne Martel, co-responsible of the ILL beamline D22. Finally, I would like to sincerely express gratitude to Dr. Malene Ringkjøbing Jensen and Dr. Martin Blackledge (IBS) for our successful collaboration with the NMR work and all the time you spent.

This work was supported by the provision of an ACORN award from Keele University, by the ANR MALZ TAUSTRUCT and the ANR ComplexDynamics, the Deuteration Laboratory (EMBL/ILL), originally established with funding from the UK EPSRC (grant ref EP/C015452/1), and by funding received from the ILL and the ESRF. The European Photon and Neutron (EPN) science campus is a very stimulating environment, I am thus particularly grateful to have received the possibility of doing my PhD work here.

Martine, je souhaite te remercier une nouvelle fois très chaleureusement. Merci pour ton aide précieuse, ton soutien, ta bonne humeur et pour nos échanges et nos discussions au long de ces trois années. Tu es compétente et appliquée, mais pas seulement. Reste comme tu es surtout, ne change rien !! Then, I do want to thank Dr. Samuel Lenton, particularly for his organisation and planning skills. You are more than my (bearded) hiking partner, you do know it! Other big thanks to my best Franglais'belgirishtalien friends Charlotte, Estelle, Juliette, Orla, Ashley (male) and Filippo. Without you O'callaghan and Bukana would probably have closed down already! And Ash, you must stop sleepwalking!! Un grand merci à toute l'équipe du D-lab passée et présente, que j'ai eu le plaisir de côtoyer, en particulier à Val, MT, FX, JB et Max.

La fin de mes remerciements ira évidemment à toute ma famille. A mes parents, qui intelligemment ne m'ont jamais forcé mais encouragé, ce qui m'a poussé à me dépasser. Je vous aime. Une pensée toute particulière pour mes beloved grands-parents Denise et Paul (mon héro), ainsi qu'à ma chère Mamine. Je te dois tant et te suis tellement reconnaissant. Vous avez tout suivi depuis le début et votre sollicitude me touche beaucoup. Je vous aime fort ! And last but not least, à la petite princesse qui partage maintenant ma vie <3.

*“Your mind is a powerful thing.
When you fill it with positive thoughts your life will start to change.”*

Anon

1. INTRODUCTION

1.1 Malaria: an emerging world disease

Malaria is a devastating parasitic disease that affects approximately one third of the human population. Every year up to five hundred million people contract the disease and several hundred thousand die, mainly African children under the age of five (*World Health Organization: The World Health Report, 2005*). This enormous mortality/morbidity burden has dramatic social and economic implications. Attempts to eradicate malaria have failed and current prevention and treatment programmes are compromised by the emergence of drug resistant parasites and the resistance of mosquitoes to insecticides. In addition, our lack of knowledge of the structural biology underlying vital aspects of the parasite-host interaction hinders effective drug and vaccine development.

Malaria is widespread in tropical and subtropical regions, including much of sub-Saharan Africa, the Americas and Asia (Figure 1.1). Malaria is also often a problem in co-infection with the human immunodeficiency virus (HIV). Compared with HIV-negative non-immune patients, more severe malaria is observed in HIV-infected non-immune and severely immune-compromised HIV-infected patients, where higher parasite densities are found (*Chirenda et al., 2000; Grimwade et al., 2003*). The World Health Organisation (WHO) has estimated that the number of cases of malaria rose from 233 million in 2000 to 244 million in 2005, but decreased to 225 million in 2009 and 198 million in 2013 (with,

according to WHO, high uncertainties). The number of deaths due to malaria is estimated to have decreased from 985 000 in 2000 to 781 000 in 2009, and more recently, in 2013, to 584 000 (uncertainty range 367 000 – 755 000). These numbers are however controversial and, according to Robert W. Snow seem to be underestimated (*Snow et al., 2005*). A combination of epidemiological, geographical and demographic data has allowed Snow and colleagues to estimate the number of clinical events around the world. As a result, it was estimated that there were 515 million episodes of clinical malaria in 2002 (range 300–660), which is up to 50 % higher than those reported by the WHO and 200 % higher for areas outside Africa. This reflects the reliance of the WHO upon weak national reporting in these countries.

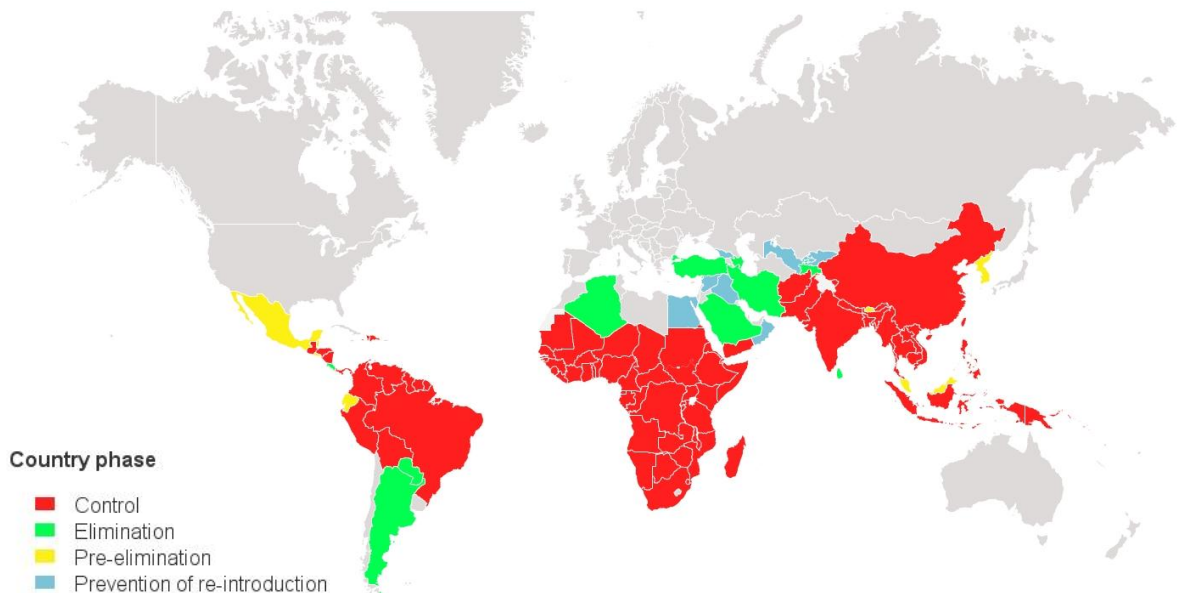


Figure 1.1: Malaria endemic worldwide distribution (map built from WHO data, *World Malaria report, 2014*)

1.2 Life cycle of the parasite

It has been over 100 years since the discovery that malaria is caused by infection with the protozoan parasite *Plasmodium*, a member of the Apicomplexan phylum that also includes parasites responsible for important human and animal diseases such as toxoplasmosis or coccidiosis. Evolution from ancient evolutionary host-vertebrate lineage specialisations has given rise to some 200 *Plasmodium* species transmitted by different mosquitoes of divergent genera (Levine *et al.*, 1988; Bruce-Chwatt *et al.*, 1993). Various animals such as reptiles, amphibians, birds, rodents and primates are susceptible to infection from these species, whereas other vertebrates are mysteriously free of *Plasmodium*, even if they can be hosts of other apicomplexan parasites. Malaria parasites infecting humans are *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* and *Plasmodium knowlesi*. Of these parasites, *P. falciparum* and *P. vivax* cause the greatest morbidity (disability-adjusted life years) and *P. falciparum* causes the most mortality, mainly amongst young African children (World Health Organization: *The World Health Report*, 2005).

The infection starts when a *Plasmodium*-infected mosquito injects sporozoite forms into the human host (Figure 1.2). They migrate to the liver, where they can pass through Kupffer cells, invade hepatocytes and thus develop into liver merozoites. After being released into the bloodstream where all clinical symptoms arise, they invade erythrocytes and develop through the ring, trophozoite and schizont stages (Figure 1.3). They replicate to produce from 16 to 32 daughter merozoites which are released during egress. To continue the asexual blood-stage life cycle, the free merozoites are then able to invade

other erythrocytes. Once in the blood, the sexual form of the parasite starts with the development of some intra-erythrocytic stages into male or female gametocytes. They are then taken up by the mosquito during feeding to develop into gametes in the insect gut. They then fuse to form zygotes and finally develop to form an invasive ookinete, which can traverse the midgut to transform into an oocyst. From these midgut cells, sporozoites are released and migrate to the salivary glands for injection into a human host during the next blood meal.

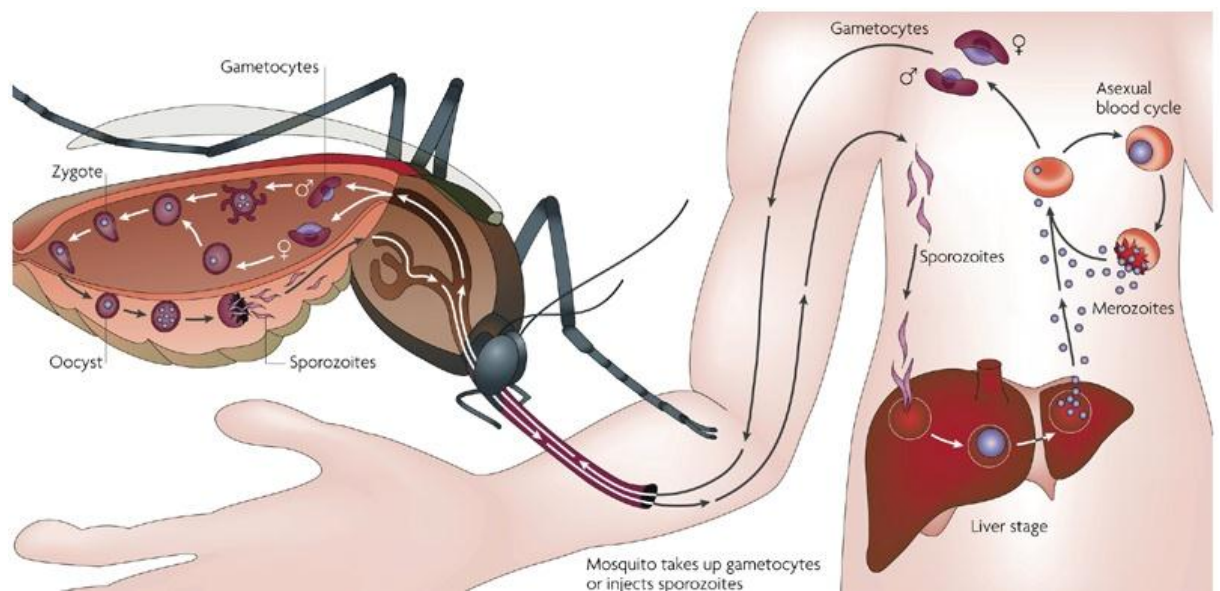


Figure 1.2: The life cycle of *P. falciparum* in the human host and the female *Anopheles* mosquito vector (from *Xinzhuan et al., 2007*).

The disease itself results from the multiplication of malaria parasites within red blood cells (Figure 1.3), causing flu-like symptoms that typically include fever, headache, nausea and vomiting. Moreover, cerebral malaria is the most important complication of *P. falciparum*

malaria, inducing changes in mental status and coma, with a mortality rate between 25-50 %. If a patient is not treated, cerebral malaria can be fatal in 24-72 hours because of the sequestration within cerebral capillaries and venules of parasitised red blood cells.

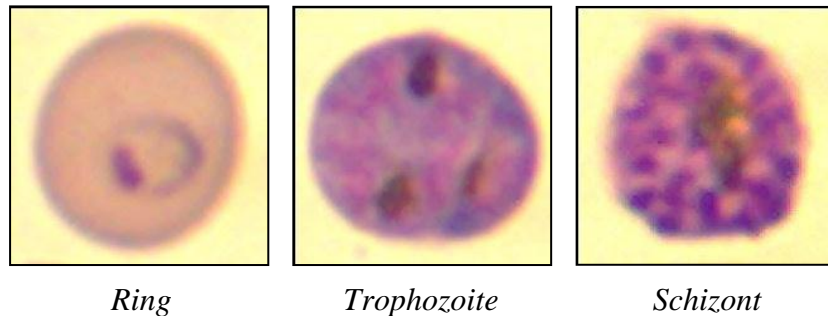
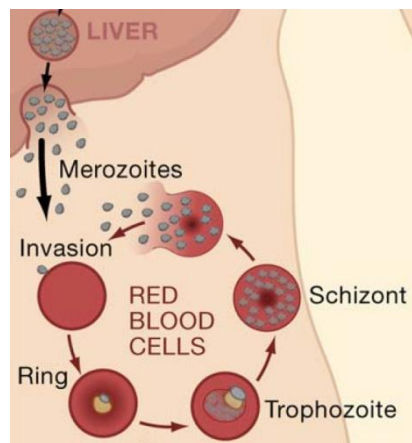


Figure 1.3: Merozoites differentiation in red blood cells. **Top:** Blood stage infection by *P. falciparum* (modified from Cowman *et al.*, 2006). **Bottom:** Optical microscopy pictures from red blood cell cultures that were infected with *P. falciparum* (M. Blanc, unpublished).

After rupture of an erythrocyte, each released merozoite can follow two developmental pathways (Figure 1.4). It can either follow another round of schizogony, or it can be

converted to a single male or female gametocyte (Diebner *et al.*, 2000). Recently, it has been demonstrated that expression levels of the DNA-binding protein PfAP2-G correlate strongly with levels of gametocyte formation, proving therefore that PfAP2-G would be essential for parasite sexual differentiation (Kafsack *et al.*, 2014). To increase its own chance of survival, *Plasmodium* is one of the few Apicomplexa to exhibit asexual schizogony, probably as a means of extending the period of its transmissibility (Dyer and Day, 2000), but this implies an elaborate balance between the asexual and sexual cycles. Prolonging the asexual cycle (and so reducing the exposure of the antigenic gametocytes) must be balanced with ensuring the availability of gametocytes when a mosquito bites (Dyer and Day, 2000). Presently the life cycle parasite switch mechanisms have not been fully elucidated, but an understanding of sexual development is the key in the development of transmission blocking new drugs and vaccines.

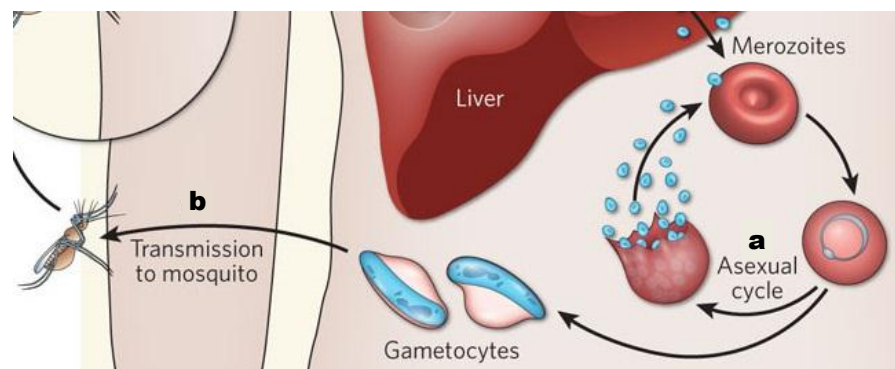


Figure 1.4: The processes from asexual merozoites to sexual gametocytes. **a:** asexual cycle within red blood cells followed by the production of male and female gametocytes, **b:** Transmission to a mosquito during a blood meal, where they fuse to form oocysts and divide to create sporozoites. These migrate to the salivary glands, where the cycle of infection starts again (modified from Michalakis & Renaud, 2009).

A coordinated expression of sexual stage-specific genes leads to the development of gametocytes, in addition to the up or down regulation of the expression of many genes (*Janse and Waters, 2004*). By using a high density microarray it is possible to identify genes showing elevated levels of expression in gametocytes, and genes expressed uniquely in sexual stages (*Le Roch et al., 2003*).

In addition, proteomic analysis of this stage has confirmed the existence of stage-specific gene expression, with many proteins found to be over-represented in the gametocyte (*Florens et al., 2002*). It includes proteins required to respond immediately to the stimuli (so involved in cell cycle/DNA processing) in order to initiate gametogenesis. It can be noticed that proteins involved in protection against oxidative stress are up-regulated in the sexual stages to compensate for the elevated levels of oxidative agents associated with the erythrocytic stages (*Lasonder et al., 2002*). The number of proteins expressed uniquely in sexual stages, and the diversity of their functions, reflects the fundamental changes and adaptations that the parasite must make to ensure its survival in a new environment.

1.3 Treatment and resistance

The emergence of drug resistance, mostly in *P. falciparum*, has been a major cause of the global reappearance of malaria in the last three decades (*Marsh et al., 1998*). Resistance to malaria treatment is the most probable explanation for a doubling of child mortality from malaria in eastern and southern Africa (*Korenromp et al., 2003*). The overall toll caused by

malaria has changed little since the 1950s (*WHO data*), when the first big programmes based on elimination of the mosquito vector and anti-malarial prophylaxis were performed. So whilst there have been large reductions in mortality and morbidity in some regions (for example South Asia), malaria continues to be a major paediatric killer in many parts of sub-Saharan Africa.

Control has traditionally relied on two approaches: 1) control of the mosquito through removal of reproduction sites, use of insecticides (such as DDT) and prevention of contact with humans (particularly by using screens and bed nets impregnated with insecticides); and 2) effective case management. A third approach, a fully effective malaria vaccine, is not expected for a least five to ten years.

Case management has been mainly through the use of inexpensive anti-malarial compounds, primarily chloroquine and more recently sulphadoxinepyrimethamine (SP) which are widely available. Together with antipyretics to decrease fever, anti-malarial compounds are among the most commonly used drugs in tropical areas. But overuse is widespread and as a consequence, detectable concentrations of chloroquine are found in the blood of a majority of the tropical population. The wide deployment of these anti-malarial drugs has in fact provided a tremendous selection pressure on human malaria parasites, and so selects for resistance mechanisms. Unfortunately *Plasmodium falciparum* is now highly resistant to chloroquine in most malarial zones. Resistance to SP is also widespread and has developed much more quickly. Resistance to mefloquine is confined to where it has been used extensively (Vietnam, Cambodia and Thailand) but arose within six years of systematic deployment (*Nosten et al., 2000*). For now and according to the WHO,

the most effective strategy is the use of artemisinins, in combination with other anti-malarial drugs (like quinine sulphate, hydroxychloroquine, mefloquine) to avoid the development of drug resistance against artemisinin based therapies.

In contrast to many other pathogens, *Plasmodium* subspecies can efficiently infect their host cells by using their own invasion machinery to penetrate and establish themselves. Host cell entry by *Plasmodium* represents a parasite-specific process, an Achilles' heel which can and will be exploited for development of new treatments. It is increasingly clear that a better understanding of the host-parasite interaction is essential for the development of new treatments and, hopefully and most importantly, a vaccine for long-term sustainable decrease in the global burden of malaria.

1.4 The process of parasitic invasion and the role of erythrocyte binding-like (EBL) proteins

The process of *Plasmodium falciparum* parasites invading erythrocytes involves a highly specific receptor–ligand recognition requiring multiple steps and complex interactions: binding, reorientation, apical junction, apical organelle content release, and erythrocyte entry. Different erythrocyte receptors facilitate parasite penetration during this process (Chitnis *et al.*, 2000; Chitnis *et al.*, 2001; Miller *et al.*, 2002; Cowman *et al.*, 2006), which takes only about one minute to occur (Gilson *et al.*, 2009). It is supposed that components of the initial attachment include antigens on the surface of the merozoite, such as

merozoite surface protein 1 (MSP1) (Goel *et al.*, 2003; Gilson *et al.*, 2006), which is probably the most abundant merozoite protein (O'Donnell *et al.*, 2001; Goel *et al.*, 2003). After attachment to an erythrocyte, the merozoite undergoes apical re-orientation involving the integral membrane protein called apical membrane antigen 1 (AMA1) (Mitchell *et al.*, 2004). It is also known that actin-based motility is used by the parasite to enter the erythrocyte (Baum *et al.*, 2008). Moreover, secondary interactions are required to complete the invasion process, with the participation of two families of invasion ligands: the erythrocyte binding-like (EBL) and reticulocyte binding-like (RBL) homologues which have been shown to play a central role (Sim *et al.*, 1990; Triglia *et al.*, 2001; Rayner *et al.*, 2001; Duraisingh *et al.*, 2003; Maier *et al.*, 2003; Stubbs *et al.*, 2005; Triglia *et al.*, 2005; Gaur *et al.*, 2006; Gao *et al.*, 2008; DeSimone *et al.*, 2009; Tham *et al.*, 2009; Triglia *et al.*, 2009).

Five genes that would encode EBL proteins have been identified. The three principal proteins encoded (see the sequence alignment and comparison in Appendix B) are the erythrocyte binding antigen 175 (EBA-175, Orlandi *et al.*, 1990; Sim *et al.*, 1990), EBA-181 (also called *JESEBL*, Gilberger *et al.*, 2003; Mayer *et al.*, 2004), and EBA-140 (also identified as *BAEBL*, Mayer *et al.*, 2001; Maier *et al.*, 2003). The remaining two genes code for EBL-1 and EBA-165. EBL-1 is not expressed in some parasite lines because of the presence of missense mutations within its coding region. It is nevertheless functional in some lines and thus could play an important role in *P. falciparum* invasion (Mayer *et al.*, 2009). The gene encoding EBA-165 (PFD1155w) carries also missense mutations, probably suggesting that it is a pseudogene. Its expression has not yet been detected in any parasite line (Triglia *et al.*, 2001).

Invasion phenotypes can be generally classified into two main groups of sialic acid (SA)-dependent invasion and sialic acid-independent invasion. SA-dependent invasion involves the EBA proteins and PfRh1. EBA175 and EBA140 bind to the erythrocyte surface molecules glycoporphin A (Orlandi et al., 1992; Sim et al, 1994) and glycoporphin C (Duraisingh et al, 2003; Maier et al., 2003) respectively, whilst EBA181 binds to SA and probably to an unknown second receptor on the erythrocyte surface as well as to the 4.1R protein (Figure 1.5) which is a critical protein in the erythrocyte cell membrane skeleton (Gilberger et al., 2003; Lanzilotti et al., 2005).

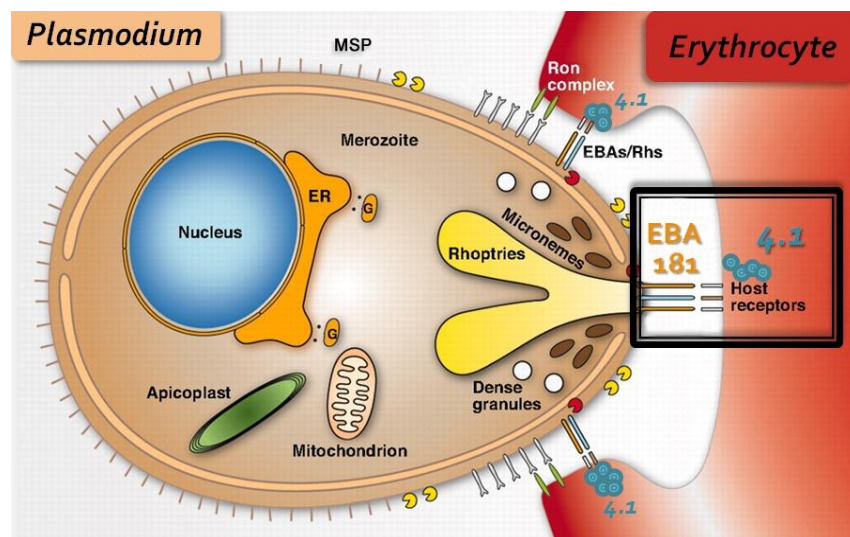


Figure 1.5: *Plasmodium falciparum* adhesion process to erythrocytes initiating the pathogenic phase essential for parasite survival and progression of clinical malaria (modified from Kappe et al., 2010).

On the other hand, *PfRh2b* and *PfRh4* which have been shown to bind to erythrocytes, are important in SA-independent invasion (*Duraisingh et al., 2003; Stubbs et al., 2005; Gaur et al., 2006 and 2007*), although their specific binding ligands are not known. EBL and *PfRh* proteins are found in the apical organelles of the merozoite and are released onto the surface prior to the erythrocyte invasion (*Duraisingh et al., 2003; Singh et al., 2005; Triglia et al. 2009*). Whilst the DBL and *PfRh* proteins are important in merozoite invasion, they are likely not to be essential. *PfRh5* would be the only member of either family demonstrated to be necessary for erythrocyte invasion, through its interaction with the erythrocyte surface protein basigin (*Baum et al., 2008; Crosnier et al., 2011*). Moreover, antibodies against this epitope also elicit high anti-invasion activity, suggesting this epitope could form the basis of an effective vaccine against malaria (*Crosnier et al., 2011; Douglas et al., 2011; Williams et al., 2011; Bustamante et al., 2013; Douglas et al., 2014; Reddy et al., 2014*). In contrary, the genes expressing the other invasion proteins can be independently disrupted in different *P. falciparum* lines without an obvious effect on blood-stage growth rates (*Duraisingh et al., 2003 and 2003; Maier et al., 2003; Stubbs et al., 2005; Triglia et al., 2005*). However, they each clearly mediate specific routes of invasion through independent receptors and termed invasion pathways, providing the merozoite with a broad array of ligands.

In summary, mediators of invasion have been identified on the merozoite but the precise functioning of these proteins and mechanisms governing the entry process remain poorly understood.

1.5 Structural information: the state-of-the-art

1.5.1 Protein complex between *P. falciparum* EBA-181 and erythrocyte 4.1R

The invasion of erythrocytes by *Plasmodium falciparum* initiates the pathogenic phase of the parasite's life cycle (Miller *et al.*, 2002). This complex process is essential for its survival and the progression of clinical malaria, and involves a series of molecular interactions between *P. falciparum* merozoites and erythrocyte membrane receptors. As mentioned above, one of them is the erythrocyte binding antigen-181 of *P. falciparum*, an important invasion protein mediating a unique host cell entry pathway by interacting with the human erythrocyte protein 4.1R (Lanzillotti *et al.*, 2006 and Figure 1.6).

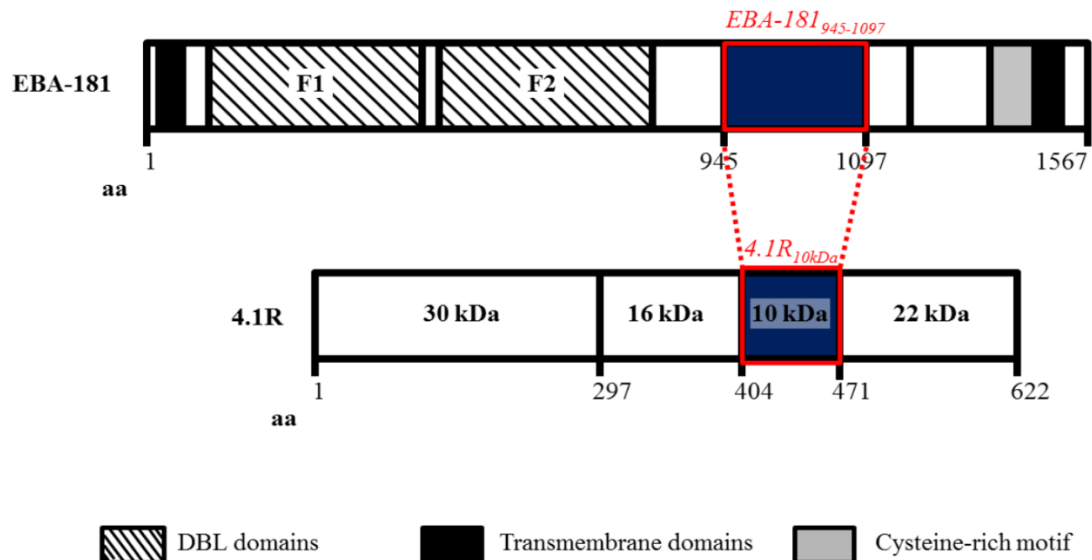


Figure 1.6: Interaction and schematic domains of EBA-181 and 4.1R (redrawn, original diagram from Lanzillotti *et al.*, 2006. See Appendix B.2 for more details). The EBA-181 protein interacts with the 10 kDa domain of 4.1R.

1.5.1.1 Protein EBA-181 (*Uniprot sequence accession number Q812B4*)

The structure of full length EBA-181 is as yet unknown, but the structure of two domains (the RII and RVI domains, Figure 1.7 A) has been determined by X-ray crystallography for the homologue EBA-175 (*Tolia et al., 2005*). These structures may be of value in building homology models for EBA-181 and EBA-140 (RII domains, Figure 1.7 B). The protein sequence comparison of EBA-175 and EBA-181 reveals 25% identity for the whole sequence (see Appendix B.2), 25% identity in the RII domain, and 20% identity in the 945-1097 region in EBA-181. The previous structural study of the erythrocyte-binding domain of EBA175 revealed an all-helical domain, forming dimers in a handshake arrangement. Each monomer consists of two Duffy-binding-like (DBL) domains (*Tolia et al., 2006*). This region is referred to as the 'F region' or 'region II' and is composed of two related domains (domains F1 and F2) involved in receptor binding (*Orlandi et al., 1990; Sim et al., 1990 and 1994*) and appears to be polymorphic (*Baum et al., 2003*). In contrast, EBA-140 and EBA-181 have only a few polymorphisms in the equivalent RII domain (*Mayer et al., 2001*). This low polymorphism in the binding regions of EBA-140 and EBA-181 has been shown to decrease the binding affinity to their receptor but without any changes in their specificity (*Maier et al., 2009*). Maier and colleagues have demonstrated a correlation between decreased EBA-140 binding and its functional contribution in merozoite invasion (*Maier et al., 2003 and 2009*), suggesting that the sequence variability in the EBL proteins is driven by immune selection and not by receptor selectivity. Changes in expression and/or use of EBA and PfRh proteins can result in utilisation of different erythrocyte invasion pathways (*Dolan et al., 1990; Reed et al., 2000; Baum et al., 2003;*

Duraisingh et al., 2003; Stubbs et al., 2005; Nery et al., 2006; Bei et al., 2007) allowing the parasites to escape antibody-mediated host immune responses.

Like the other EBL proteins, EBA-181 consists mainly of the N-terminal F region, (the region II with the two DBL domains F1 and F2), a C-terminal cysteine-rich domain (RVI) and a transmembrane domain. The central section between RII and RVI is composed of the RIII to RV regions.

1.5.1.2 **Protein 4.1R** (*Uniprot sequence accession number P11171*)

The human erythrocyte membrane is a complex protein network consisting of a membrane skeleton underlying a lipid bilayer, which contains transmembrane proteins. Erythrocyte protein 4.1 (4.1R) is a critical component of the membrane skeleton. It enhances the interaction of spectrin tetramers with F-actin thereby stabilising horizontal membrane protein connections (Figure 1.8 lower rectangle). It also links the underlying skeleton to the lipid bilayer through vertical interactions with glycophorin C (Figure 1.8 top rectangle).

Limited proteolysis of 4.1R generates four structural domains with distinct functions. Amongst them, a crystal structure is only available for the 30 kDa domain (Figure 1.9). Of particular relevance to this study is the 10 kDa domain that binds to spectrin and actin (*Schischmanoff et al., 1995*) and stabilises this ternary complex, which is crucial for membrane integrity. In addition, this domain interacts with myosin and regulates its activity.

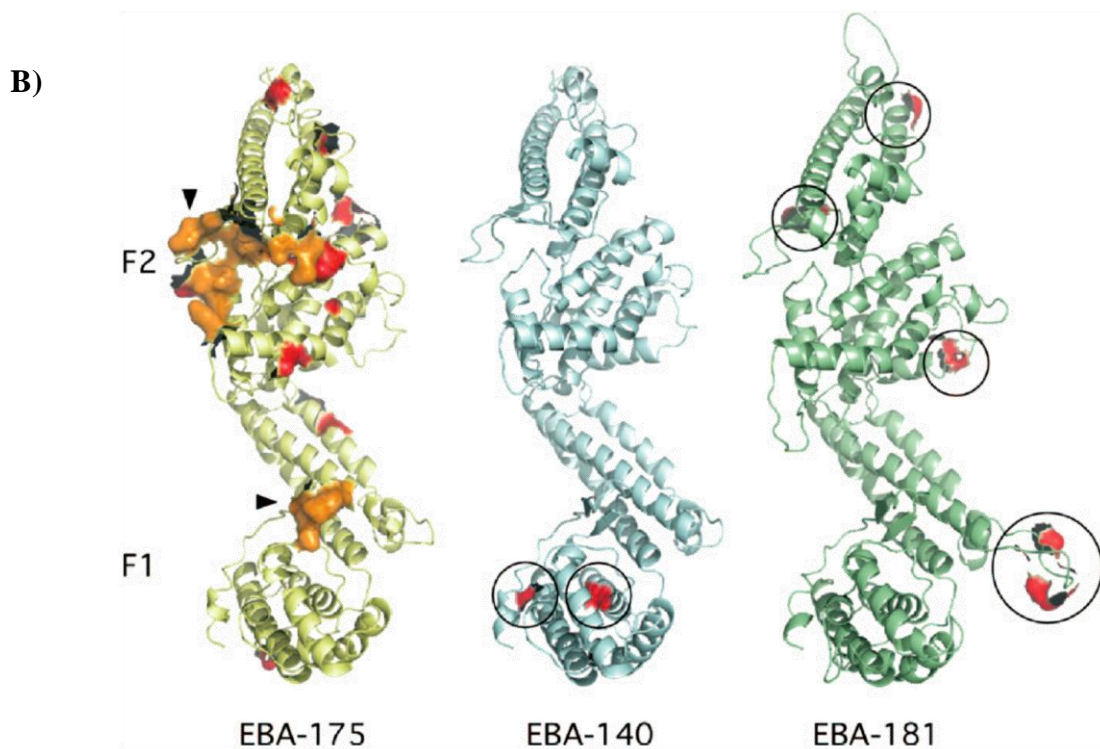
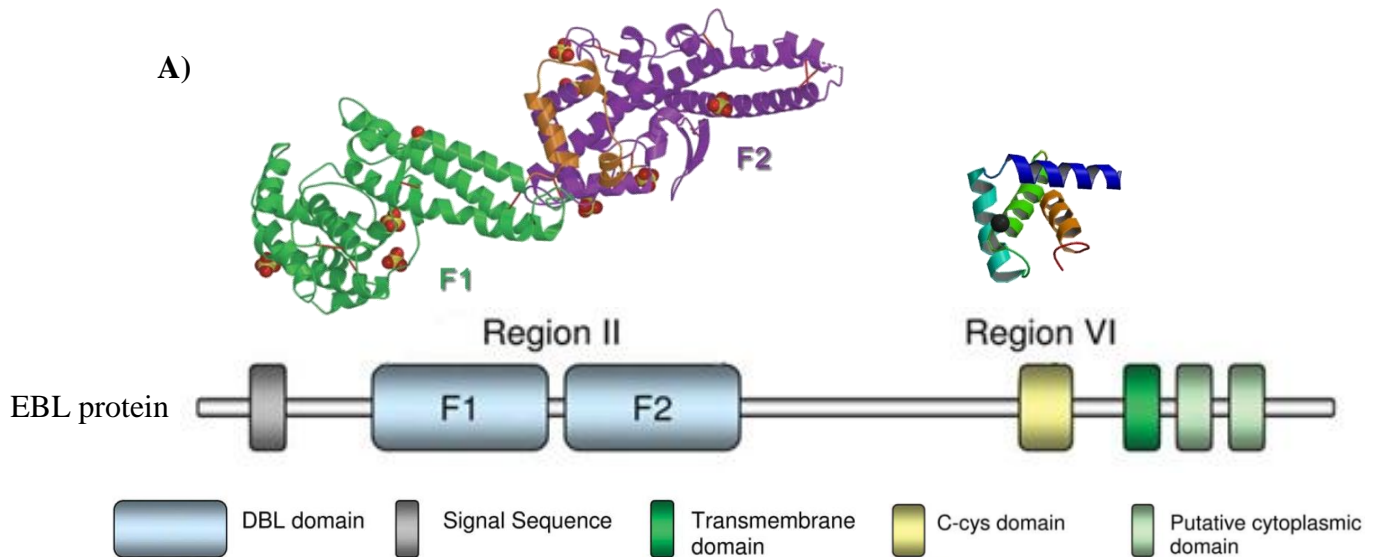


Figure 1.7: A) EBA-175 X-ray crystal structures of domains RII and RVI (pdb files 1ZRL at 2.3 Å resolution and 2RJI at 1.8 Å resolution respectively) and **B)** RII domain X-ray crystal structure of EBA-175 and homology models for EBA-140 and EBA-181. The polymorphic residues are shown in red, and the putative glycan contact residues with glycophorin A on EBA-175 are shown in orange. The polymorphism is highlighted on the structures of EBA-140 and EBA-181 with the circles (modified from *Maier et al., 2009*).

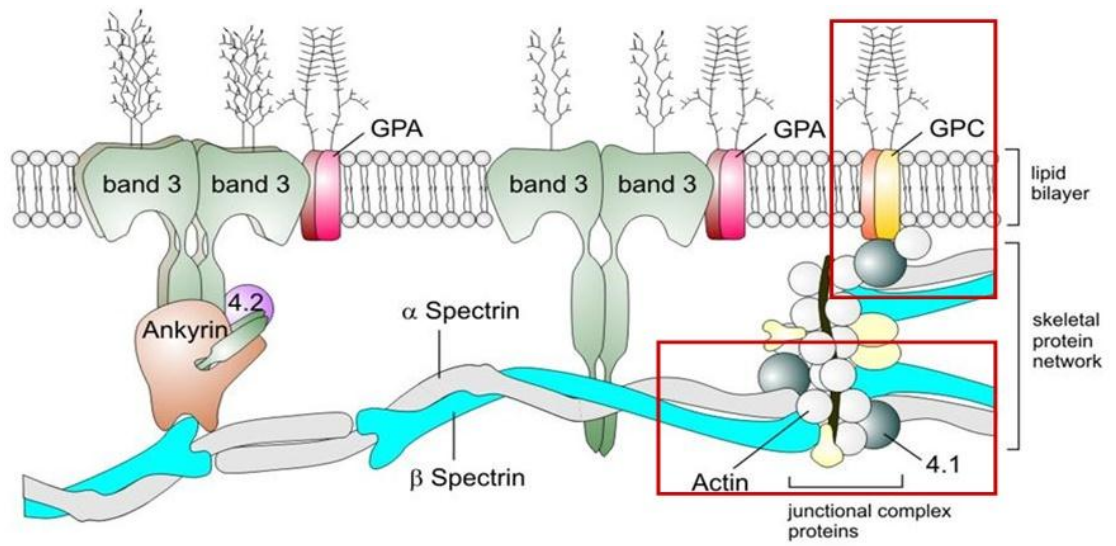


Figure 1.8: Model of the erythrocyte membrane, with the protein 4.1R (*T. Coetzer personal communication*).

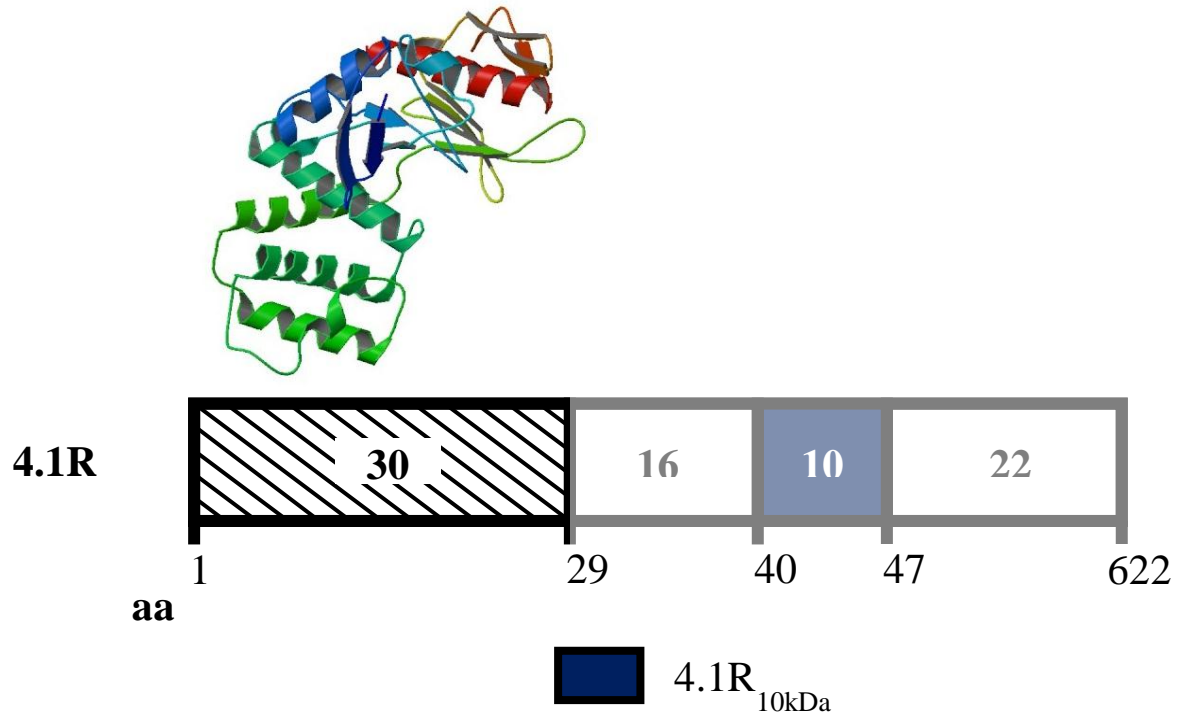


Figure 1.9: Crystal structure of the 30 kDa core domain of 4.1R (*pdb file 1GG3*) and the four different structural domains of 4.1R.

2. EXPERIMENTAL TECHNIQUES

The work described in this thesis employed a variety of methods including sample preparation techniques, biochemical and biophysical characterisation techniques as well as several multi-lengthscale light, neutron and X-ray scattering approaches for structural studies (Figure 2.1) available at the ‘European Photon and Neutron’ (EPN) science campus (Grenoble, France). This chapter describes these techniques in general terms.

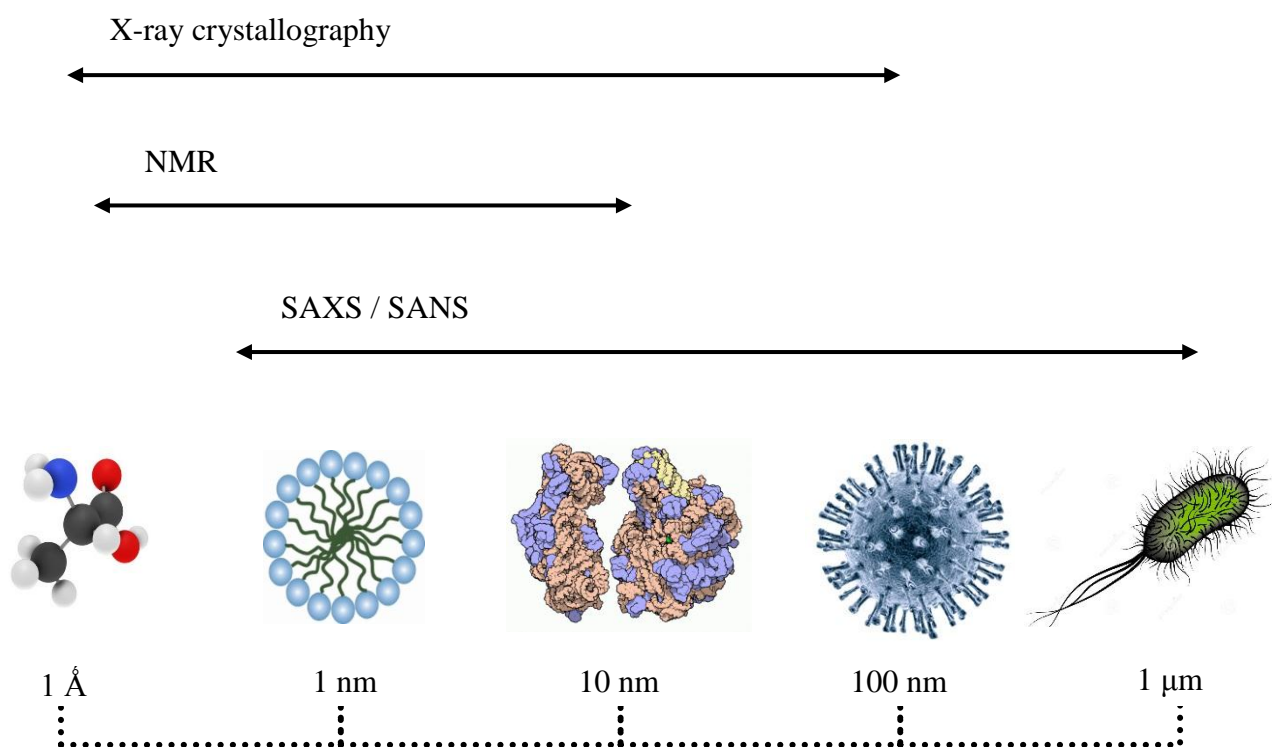


Figure 2.1: Sizes of different biological structures and examples of techniques used to probe them at different resolutions. From left to right: an alanine molecule, a micelle, the ribosome, the human immunodeficiency virus and the *E.coli* bacterium (images from Wikipedia).

2.1 Molecular biology methods

Molecular biology methods are day-to-day tools to isolate, characterise and manipulate the molecular components of cells and organisms, from DNA and RNA to proteins and their functional complexes. A range of these tools were used during the work of this thesis and are summarised below.

2.1.1 Plasmid DNA preparation

Plasmid preparation was used to extract and purify the DNA (plasmids are small and circular self-replicating DNA molecules) containing the genes (cDNA) for the various proteins of interest. It involves three steps, starting with the growth of the bacterial culture transformed beforehand with the desired plasmid, then the harvesting and lysis of the bacteria cells, finishing with the purification of the plasmid DNA. This was performed using a mini-prep kit from SIGMA-ALDRICH™ (GenElute™ Plasmid Miniprep Kit).

2.1.2 *In-vitro* transposition

Transposition is found in nature in prokaryotes and eukaryotes as a mutation where a DNA segment is transferred to a new position in the genome using transposable elements, also called transposons (*McClintock et al., 1950*). In molecular biology, transposition can be used *in vitro* as an alternative method to the traditional re-cloning technique for changing the antibiotic resistance carried by a plasmid (*Goryshin et al., 1998*). To perform the change of the antibiotic resistance from ampicillin to kanamycin, the kit *EZ-Tn5™ <KAN> Insertion Kit* from Epicentre Technologies Corporation® was used (Figure 2.2).

This system uses a transposase enzyme and a transposon element containing the kanamycin resistance gene to be randomly inserted into the target DNA (the plasmid). The crucial step is to select positive kanamycin resistant clones and screen for clones where the transposon element has been successfully inserted into the ampicillin resistance gene in order to inactivate it (see Section 3.2.2 for further technical details). Finally, the clone DNAs were sequenced to insure their quality.

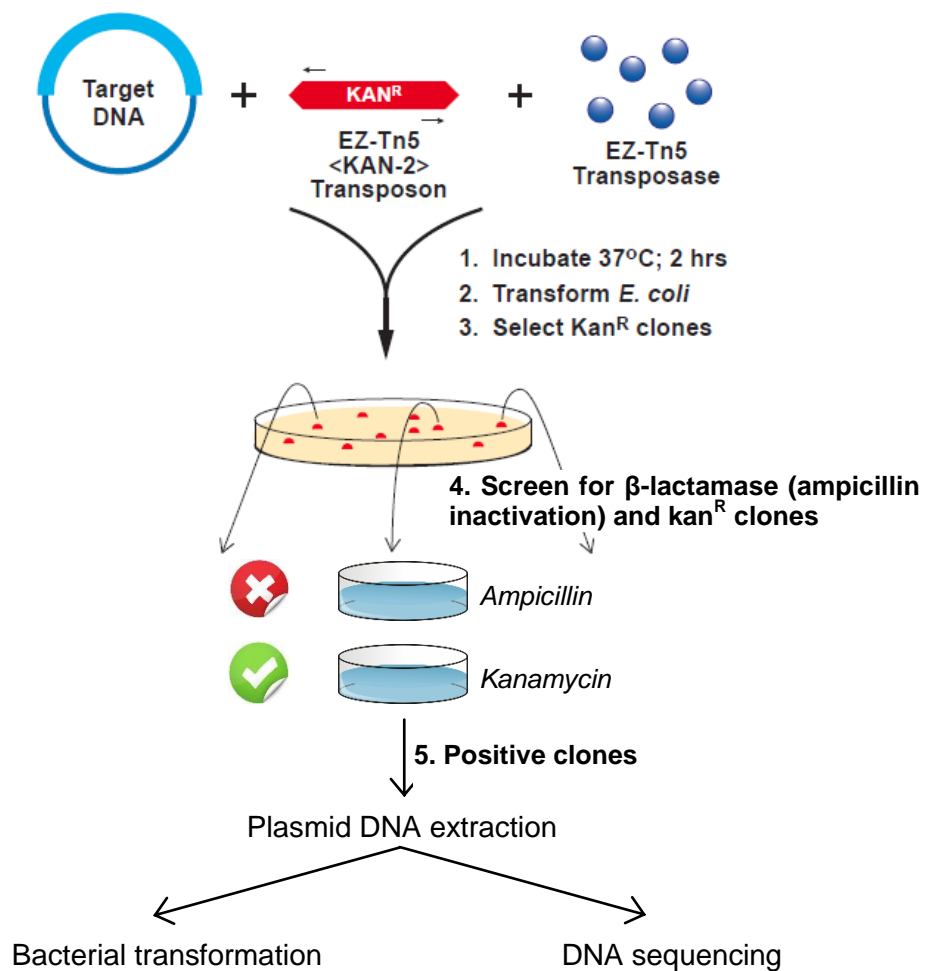


Figure 2.2: Transposon insertion protocol (modified from Epicentre®).

2.1.3 Polyacrylamide gel electrophoresis

2.1.3.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein expression and purity levels by size characterisation was performed using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), with Tris-glycine conditions (*Laemmli, 1970*) or Tris-tricine conditions, which increases the resolving power of the gel for lower (5-20 kDa) molecular weight proteins (*Schaegger et al., 1987*). For the work of this thesis, the gels were run at 165 V and stained with Coomassie blue (unspecific staining of the proteins, mainly arginine residues) and then destained with deionised water containing 5% v/v acetic acid and 7.5% v/v ethanol.

2.1.3.2 Native polyacrylamide gels

One-dimensional electrophoresis using non-denaturing (or 'native') conditions is an electrophoresis performed in the absence of denaturants such as detergents (for example SDS). Because mobility depends on the size, shape and intrinsic charge of a protein, non-denaturing electrophoresis provides a set of separation parameters distinctly different from mainly size-dependent denaturing sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Separation of proteins by non-denaturing electrophoresis requires the same type of equipment used for denaturing slab gels. In a continuous system, the same buffer is used to prepare the acrylamide solution and filling electrophoresis chambers, and separation is governed by pH (four different types of buffers in function of the chosen pH, from pH 3.7 to pH 10.6). This choice of pH and thus the buffer system

depends on the protein being studied (its isoelectric point). See Appendix A.2 for further details.

2.1.3.3 Peptide gels

In order to increase the resolution and separation of very small proteins (< 10 kDa) and peptides, commercially pre-prepared polyacrylamide gels were used (*Bio-rad 16.5% Criterion™ Tris-Tricine Gel*).

2.1.4 Affinity purification

Affinity chromatography is a technique used to separate biochemical components using highly specific interactions. Here, immobilised metal ion chromatography (IMAC) was used to purify the proteins of interest from the bacterial crude extract. IMAC is based on the specific binding affinity of histidine side-chains to metal ions (nickel or cobalt) covalently immobilised to the resin beads. This allows tagged proteins to be retained by the column and thus to be physically separated from the other proteins.

In this thesis two different IMAC were used: nickel NiNTA resin to retain the proteins expressed with a histidine tag, and glutathione-agarose resin to bind the GST tagged proteins.

2.1.5 Size-exclusion chromatography

For all proteins the final purification stage was a size-exclusion chromatography, also called gel filtration. This method allows the proteins to be separated as a function of their size by the use of a chromatography column packed with beads with pores of different sizes. The columns used were the *Superdex S200 10/300 GL* and the *Superdex S75 10/300 GL*, *GE Healthcare*.

2.1.6 Protein concentration measurements

To measure protein concentrations, four different techniques were used. Two direct UV measurements: near UV (280 nm) and far UV (190 nm) absorbance; and two colorimetric methods using dyes: Bradford (595 nm) and bicinchoninic acid (562 nm) assays.

2.1.6.1 Near UV absorbance (280 nm)

UV absorption spectroscopy is a direct measurement using natural UV absorption by proteins. The near UV absorption at 280 nm is mainly due to tyrosine and tryptophan side chains with a small contribution of phenylalanine and disulphide bonds. This method is simple, direct, and the sample can be recovered. It has also some disadvantages, including possible interference with the protein buffer components or with any chromophore and nucleic acid contaminants. Moreover the specific absorption value must be calculated theoretically for each protein.

2.1.6.2 Far UV absorbance (190 nm)

Another characteristic UV absorption is due to the peptide bond and occurs in the far UV (maximum at about 190 nm). Due to the difficulties caused by the low output of conventional spectrophotometers at 190 nm, measurements are usually made at 205 nm with various side chains making contributions, including tryptophan, phenylalanine, tyrosine, histidine, cysteine, methionine and arginine. Also for this method the extinction coefficients are needed. The advantages include simplicity and sensitivity, and the sample can be recovered. The principle disadvantage is the necessity for accurate calibration of the spectrophotometer in the far UV.

2.1.6.3 Bradford assay (595 nm)

The Bradford protein assay is a spectroscopic analytical procedure commonly used to measure the concentration of protein and developed by Marion M. Bradford (*Bradford, 1976*). The Bradford method is a colorimetric protein assay based on an absorbance shift of the dye Coomassie Brilliant Blue G-250. Under the acid conditions the dye is normally brownish/red colour but on binding to the protein the blue form of the dye is produced (binding mainly to arginine residues). The bound form of the dye has an absorption spectrum maximum historically held to be at 595 nm.

When used in this thesis, a standard reference was measured from bovine serum albumin.

2.1.6.4 **Bicinchoninic acid assay (562 nm)**

The principle of the bicinchoninic acid (BCA) assay (also called Smith assay) was developed by Paul K. Smith (*Smith et al., 1985*) and is similar to the well-known Lowry method. It relies on the formation of a Cu^{2+} / protein complex under alkaline conditions followed by reduction of the Cu^{2+} to Cu^+ proportionally to concentration of protein (tyrosine, tryptophan, cystine, cysteine and the peptide bond are able to reduce Cu^{2+} to Cu^+). The BCA forms a purple-blue complex with Cu^+ in alkaline environments, thus providing a basis to monitor the reduction of alkaline Cu^{2+} at the absorbance maximum of 562 nm, and thus monitor protein concentration. The BCA assay is more sensitive than the Lowry method and less variable than the Bradford assay. Moreover, since EBA181₉₄₅₋₁₀₉₇ does not contain any aromatic residues, it is therefore the only available method to quantify the concentration of EBA181₉₄₅₋₁₀₉₇ samples.

2.1.7 **Pull-down assays for purification and characterisation of the complex**

The complex formed by His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} was purified using pull-down assay. This is an *in vitro* method used to establish the existence of an interaction between two or more proteins. It can therefore be used to confirm the existence of a protein-protein interaction after being firstly predicted by other techniques (for example co-immunoprecipitation) or as an initial screening assay to identify unknown protein-protein interactions.

In a pull-down assay, a tagged protein is captured on an immobilised affinity ligand, specific to the protein tag. This allows generation of a 'secondary affinity support' which is

then incubated with a protein source (for example a bacterial crude extract) that contains the putative ‘prey’ proteins.

2.2 Biochemical and biophysical characterisation techniques

In order to improve the expression and purification conditions of the different proteins and to study their intrinsic characteristics and properties, a combination of biophysical and biochemical techniques were used. Their principles are described below.

2.2.1 Time-of-flight mass spectrometry (TOF-MS)

Mass spectrometry (MS) allows the calculation of the mass of a protein by ionising it and measuring its time of flight. MS measurements were carried out at the *Institut de Biologie Structurale* in collaboration with Lucas Signore.

2.2.2 N-terminal sequencing

This technique uses a chemical process based on Pehr Edman degradation (*Edman et al., 1950*) in order to sequence the first amino-acids in the N-terminus of a protein. N-terminal sequencing was carried out at the *Institut de Biologie Structurale* in collaboration with Jean-Pierre ANDRIEU.

2.2.3 Dynamic light scattering (DLS)

Dynamic light scattering (DLS), also called quasi-elastic light scattering (QELS) or photon correlation spectroscopy (PCS), allows the size of the protein of interest to be estimated and the homogeneity of a protein solution to be assessed. This technique is based on the interference generated by particles in solution when illuminated by a coherent light source (laser). A detector measures the intensity of the light scattered by the particles in the sample. The Brownian motion (*Brown, 1827*) of particles in solution makes the scattered intensity vary as a function of time and, using an autocorrelation function, allows the diffusion coefficient to be measured. The relation between this coefficient and the size of particles is based on the theory of Einstein concerning the Brownian motion of spherical particles (an example of a DLS result is presented in Figure 2.3). The Stokes-Einstein equation (see below) allows the hydrodynamic radius of the observed particles to be measured. With assumption of the protein as a sphere it is possible to estimate the molecular mass via a predefined calibration. However, proteins are not always comparable to spheres and can adopt various shapes so the calculated hydrodynamic radius can differ from the real size of the protein and the calculation of the molecular mass may remain approximate.

$$D = \frac{kT}{6\pi\eta R}$$

Stokes-Einstein equation: D is the diffusion coefficient for a particle in a free volume and depends on the Boltzmann constant (k) ($k = 1.3806503 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2}$, relates energy at the individual particle level with temperature), the absolute temperature (T), the viscosity of the solution (η), and the hydrodynamic radius (R) of the particle.

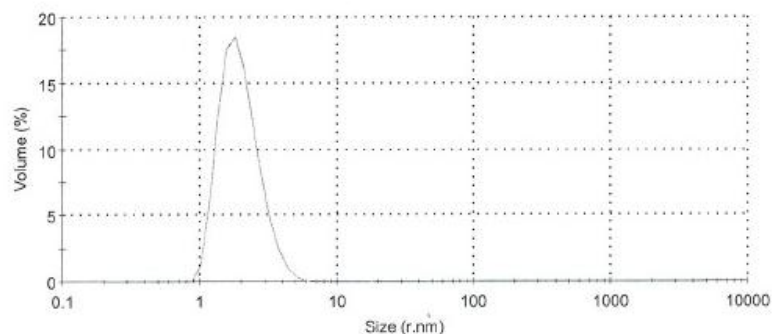


Figure 2.3: Example of a DLS result for a mono-disperse (meaning a single population) protein.

2.2.4 Multi-angle laser light scattering coupled to size exclusion chromatography and refractive index (SEC-MALLS-RI)

This method allows the molecular mass of a particle to be determined as well as its hydrodynamic radius using a combination of several techniques (Figure 2.4): size exclusion chromatography (SEC), multi-angle laser light scattering (MALLS) as well as the measurement of the refractive index (RI). The homogeneity and the polydispersity of a sample can also be estimated by this technique. Specifically for the work presented here, a size exclusion chromatography column *SuperdexTM 75* (Amersham) was used with a refractometer *RI2000b* (Schambeck SFD) and a *DAWN-EOS* (Wyatt Technology Corp.) MALLS instrument using a laser emitting at 690 nm.

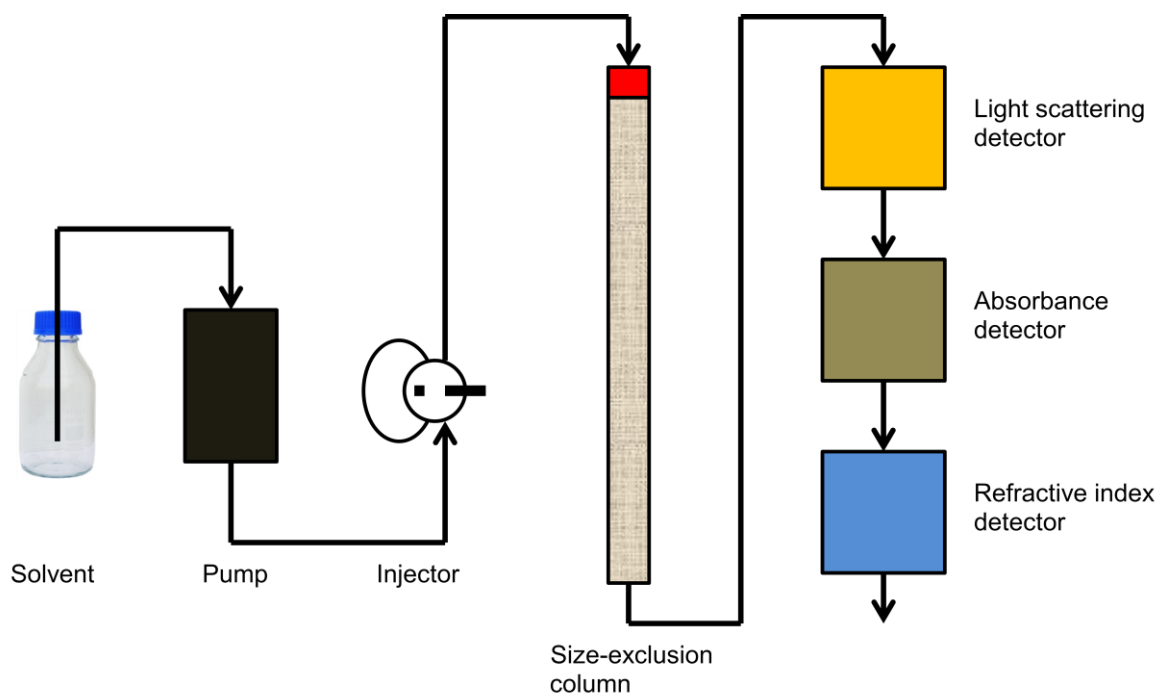


Figure 2.4: Representation of a SEC-MALLS-RI device.

The molecular mass is calculated from the intensity of light scattered in different angles and from the calculation of the concentration obtained from the refractive index. An average molecular mass is calculated on all the width of the elution peak. For every point on the chromatogram, the scattering light intensity is directly proportional to the concentration of proteins and to the molecular mass, in agreement with the Zimm formula for diluted solutions of polymers:

$$\frac{R_{\theta}}{K^*} = M * C$$

R_{θ} : measured Rayleigh ratio; C : protein concentration (g/ml); M : molar mass (g/mol); K^* : optical constant given by the following equation:

$$K^* = \frac{1}{N_a} \left(\frac{2\pi * n_o}{\lambda^2} \right) \left(\frac{dn}{dc} \right)^2$$

N_a : Avogadro constant; n_o is the refractive index of the solvent at the wavelength of the incident wave (1.33 for the diluted aqueous buffers); dn/dc : refractive index of the solute (ml/g); λ : wavelength of the incident wave.

In addition, the homogeneity and polydispersity of samples can be estimated by this technique. The key advantages lies in the use of the size exclusion chromatography which allows possible aggregates and other protein species to be separated, and the possibility to determine the molecular mass without making any hypothesis for the shape as is the case for DLS.

2.2.5 Circular dichroism (CD)

Circular dichroism (CD) measures the differences of absorption of left- and right-handed circularly polarised light. These differences result from structural asymmetries of proteins which are chiral molecules and which absorb differently both components of the polarised light. CD is particularly used to study the folding state of a protein and to characterise its

global secondary structure. The study is made by measuring the difference in left- and right-handed polarised light absorption in the far UV, from 190 to 260 nm. Each type of secondary structure (α -helix, β -sheet, turn or random coil) results in a characteristic spectrum (Figure 2.5). The average composition of the sample in terms of secondary structure elements can be deduced from the resulting curve.

For this thesis work, each experiment was repeated three times and at a temperature of 4°C or 20°C. For each protein sample, a blank sample which contains only the dialysis buffer of each protein (see Appendix A.2 for buffer compositions) was also measured to be subtracted from the final measurement.

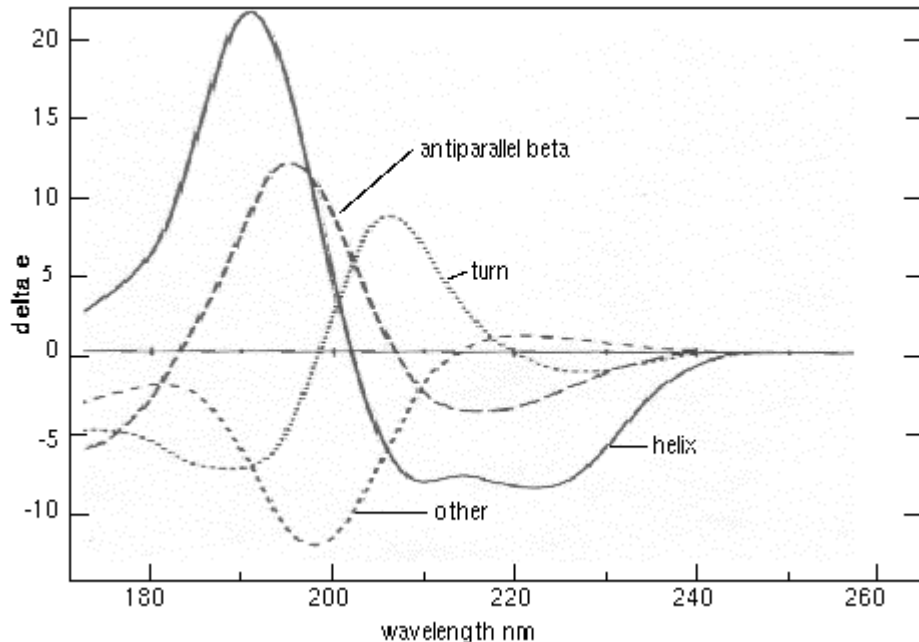


Figure 2.5: Circular dichroism spectra of 'pure' secondary structures (adapted from *Brahms and Brahms, 1980*).

When circularly polarised light passes through an optically active medium (chiral molecules like proteins), the speeds between right and left polarisations differ ($c_L \neq c_R$), as well as their wavelength ($\lambda_L \neq \lambda_R$) and the extent to which they are absorbed ($\epsilon_L \neq \epsilon_R$). By definition delta absorbance is the difference between absorbance of left and right circularly polarised light (LCP/RCP) as a function of the wavelength:

$$\Delta A = A_L - A_R$$

It can also be expressed as the molar circular dichroism, by applying Beer's law, as:

$$\Delta A = (\epsilon_L - \epsilon_R) \times C \times l$$

Where ϵ_L and ϵ_R are the molar extinction coefficients for LCP and RCP lights, C is the molar concentration and l is the path length in centimetres.

With the molar circular dichroism of the substance defined by:

$$\Delta \epsilon = \epsilon_L - \epsilon_R$$

Although ΔA is usually measured, for historical reasons most measurements are reported in degrees of ellipticity. Fortunately it is very easy to inter-convert between θ and ΔA :

$$\Delta A = \frac{\theta}{32.982}$$

The final CD curve in mean residue ellipticity is obtained using the measured rotation in mdeg applied to the following equation:

$$[\theta] = \frac{\theta_{measured}}{10 * n * c * l} \text{ (deg.cm}^2\text{.decimol}^{-1}\text{.residue}^{-1}\text{)}$$

Where $\theta_{measured}$ is the experimental rotation in mdeg, n the number of amino-acid residues in the protein, c the concentration in molarity and l the path length in cm.

2.2.6 Fluorescence thermal shift assay (TSA)

Fluorescence thermal shift assay technique (TSA, also called ThermoFluor™) allows the thermal stability of a protein to be measured using a fluorescent marker interacting with the hydrophobic core of proteins. The probe used (Sypro Orange®, Invitrogen) is sensitive to the surrounding medium and its fluorescence signal is limited when it is in aqueous solution due to the phenomenon of ‘quenching’. TSA is based on two principles. The first one is that a folded protein does not interact with the fluorescent agent. The second principle comes directly from the nature of the sample. Upon increasing temperature the protein unfolds and becomes denatured, so exposing its hydrophobic core to which the fluorescent probes can then bind and emit a fluorescent signal proportional to the protein denaturing. In effect, TSA gives an optical readout of protein melting making it possible to determine the melting temperature (T_m) of a protein (Figure 2.7). After emission of a maximum signal, the signal decreases because the intrinsic fluorescence of the probe quenches with increasing temperature.

As the melting point is linked to protein stability, applications of TSA are very broad:

- Optimisation of solution conditions (pH, salt concentration, cofactors, metals,) for protein stabilisation (protein preparation and biochemistry and optimisation of crystallisation conditions).
- Measurement of ligand binding affinity (drug screening, analysis of binding cooperativity between multiple ligands, inhibitors or cofactors).
- Use of libraries to find proteins of unknown function.
- Use of structural analogue libraries to determine detailed energetics of cofactor binding.

The assays are realised on a reverse transcription polymerase chain reaction (RT-PCR) device (specifically in the work here an Mx3005P Q-PCR[®], Stratagene) with protein sample concentration always included between 20 nM and 100 nM. The wavelengths of excitement and emission are respectively 470 nm and 570 nm.

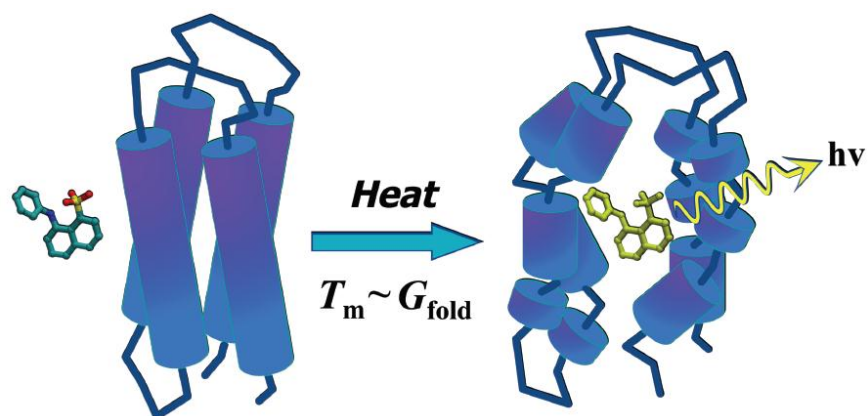


Figure 2.6: Principle of thermofluor (image from *Imiplex*, www.imiplex.com).

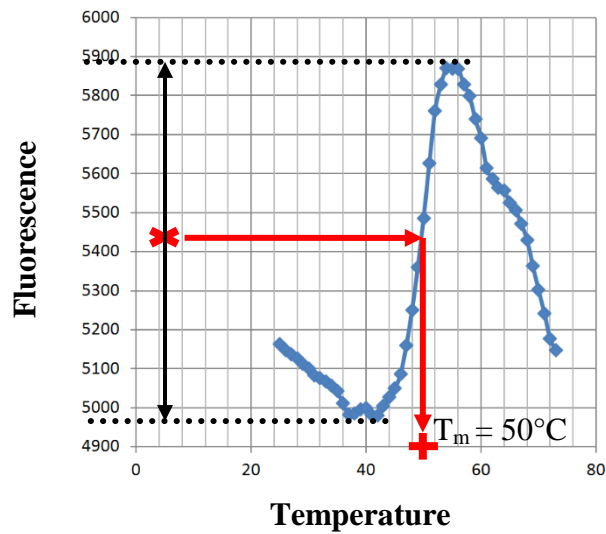


Figure 2.7: Graphical representation of a TSA measurement. The fluorescent signal increases as a function of the temperature which leads to the unfolding of the protein. The experimental T_m is calculated at the inflexion point of the curve.

2.2.7 1D NMR spectroscopy

One dimensional liquid state proton NMR spectroscopy (1D ^1H NMR) can be used to quickly characterise biological macromolecules in solution (*Silverstein, 1991*). In particular the degree of folding of a protein can be determined by observing the presence or not of proton NMR peaks in specific regions of chemical shifts (an example of a spectrum is given in Figure 2.8). 1D NMR was carried out at the *Institut de Biologie Structurale* with the ‘Characterisation of Proteins’ platform.

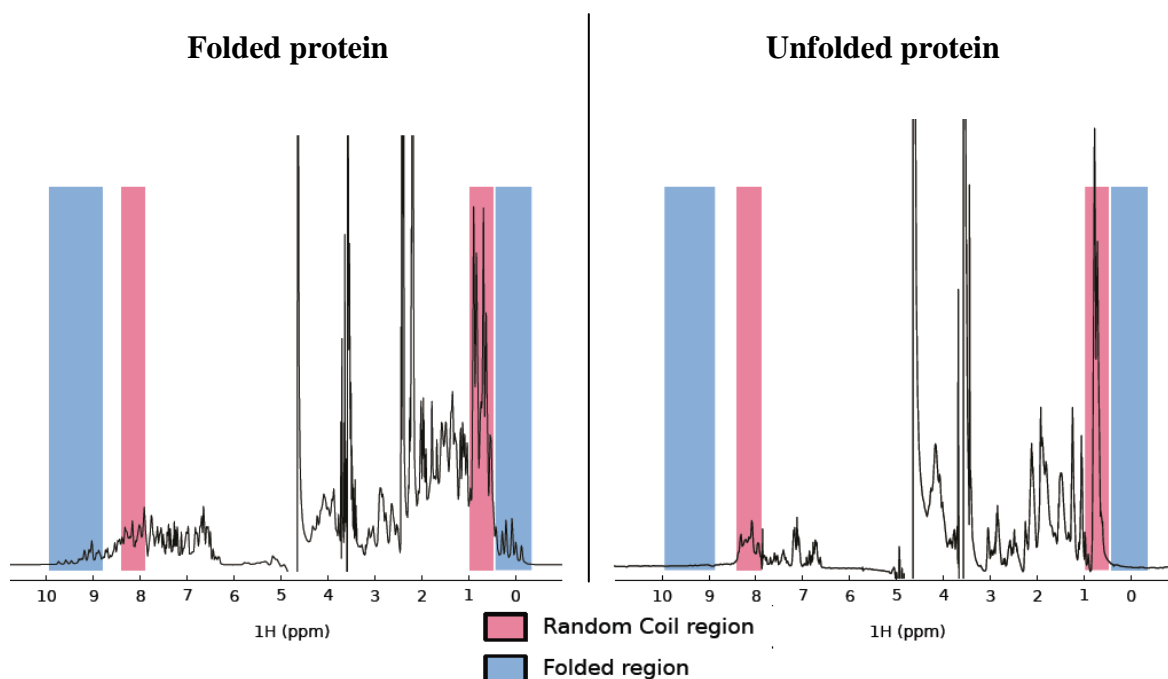


Figure 2.8: Example of a 1D-NMR measurement of a folded and an unfolded protein (taken as example from an IBS experimental NMR report).

2.2.8 Heteronuclear single quantum coherence NMR spectroscopy

The heteronuclear single quantum coherence (HSQC) or heteronuclear single quantum correlation experiment is used in NMR spectroscopy and particularly in protein NMR (*Bodenhausen et al., 1980*). The result is a two dimensional spectrum with one axis for ^1H and the other for a heteronucleus (an atomic nucleus other than a proton), most often ^{13}C or ^{15}N . The spectrum contains a peak for each unique proton attached to the heteronucleus being considered (see Chapter 5 for more details).

2.3 Small-angle X-ray and neutron solution scattering (SAXS and SANS)

The concepts and mathematical framework underlying the small-angle scattering of X-rays and neutrons, known as SAXS and SANS respectively, have been described in a number of classical texts and reviews (non-exhaustive list: *Baruchel et al., 1993* ; *Fanchon et al., 2002* ; *Jacques et al., 2010* ; *Kokhanovsky et al., 2014* ; *Serdyuk et al., 2007*) and full details will not be repeated here. However the studies carried out for this thesis work have made extensive use of X-ray and neutron scattering techniques, therefore it is appropriate to summarise the associated principles.

The added value for this work was to combine the strength of both X-ray and neutron scattering techniques in order to obtain a consensus on the structural properties of the proteins. The high X-ray flux of the ESRF allowed a wide set of buffer conditions to be screened to find the most suitable one for each of the protein samples to be studied and an overall molecular envelope to be determined, allowing even low protein concentration to be used. Small-angle neutron scattering was used to distinguish and discriminate between the two proteins involved in the complex, with contrast variation being used in conjunction with protein deuteration to visualise individual proteins whilst in complex.

2.3.1 Small-angle scattering: X-rays and neutrons as complementary probes

Light scattering is one of a few methods that allow the determination of the radius of gyration and molecular mass. The principles behind both neutron and X-ray techniques are

similar except for how the incident radiation interacts with the sample: X-rays interact with the electrons of each atom whilst neutrons interact with the nucleus. X-ray scattering depends upon the number of electrons present – and thus increases with increasing atomic weight. The interaction of neutrons with matter however, depends on a completely different interaction between the incident neutron and the nuclei of the atoms in the sample. The scattering powers for neutrons and X-rays for the atoms of a biological system are very different (Figure 2.9) and this leads to a strong complementarity of X-ray and neutron scattering, particularly in the case of using the strong difference in hydrogen and deuterium scattering for application in neutron scattering experiments. In the case of solution scattering, it is convenient to calculate a scattering length density (SLD) which quantifies the scattering power of particular parts of the structure.

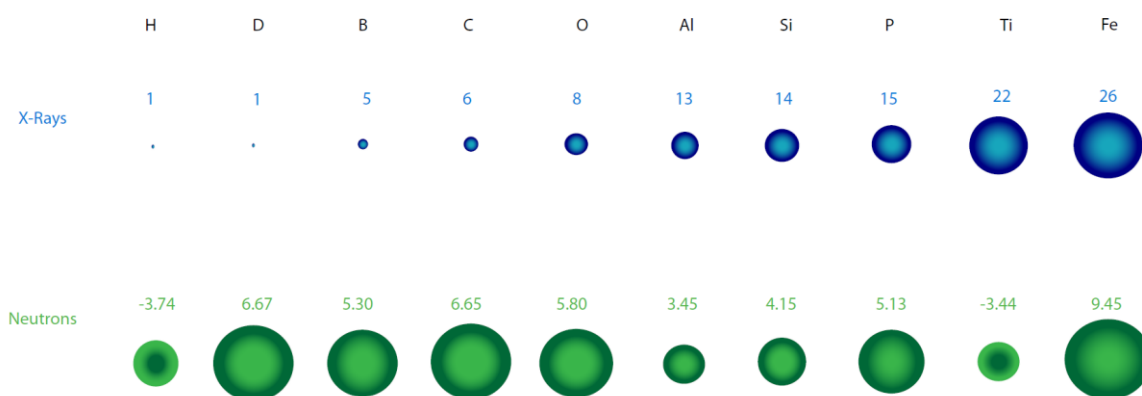


Figure 2.9: Representation of scattering lengths for X-rays (blue, proportional to Z) and neutrons (green).

2.3.2 Deuteration for SANS

The large difference in the scattering lengths of hydrogen and its isotope deuterium means that it is possible to label molecules in such a way that their scattering length densities (SLDs) are significantly different. Sample preparation using deuteration is well known for biological systems; and while perdeuteration (100 % deuteration) is commonly used to improve the signal-to-noise ratio in neutron crystallography, selective labelling (for example where only one particular subunit of a protein complex is labelled) can provide information about a specific part of a protein complex in neutron solution scattering studies (Figure 2.10). Moreover, combining the use of selective deuterium labelling of proteins with different ratios of D₂O/H₂O in the buffer means that it is possible to create different contrasts such that specific protein components can be matched in or out (*Stegmann et al., 1998*). Indeed, the interesting features of the curves shown in Figure 2.10 are the points where the lines intersect. At these points, the molecules involved are matched out and the point at which this occurs (in terms of % D₂O/H₂O) is known as the *match point*. Hence *contrast variation* can be used to highlight the individual components of a complex.

In the case of a two protein complex, by random fractional deuteration of one component to a global level of 75% (in D₂O) and keeping the other partner unlabelled, and by using three different concentrations of D₂O in the protein buffers, it is possible to alternatively match out one or the other protein, and also to observe the full complex. Briefly: in 0 % D₂O the whole complex will stay visible, and the unlabelled protein and the 75 % deuterated protein will be individually matched out in respectively 40 % and 100 % D₂O (Figure 2.11).

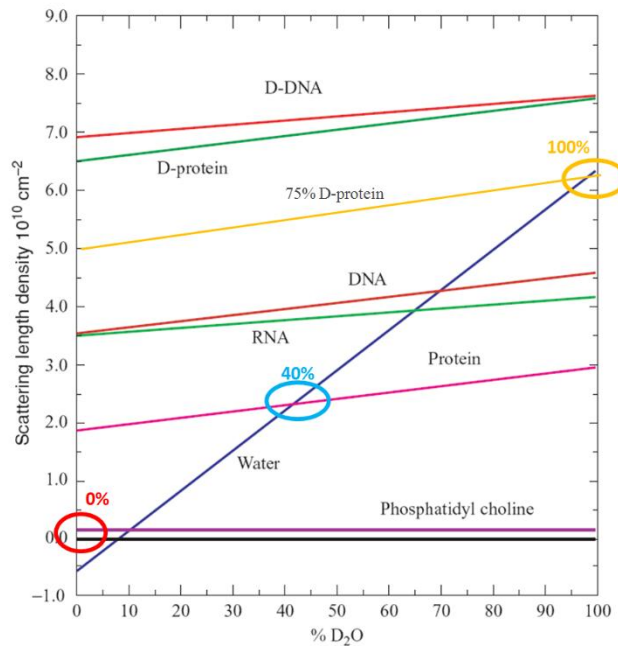


Figure 2.10: Plot of the variation of neutron scattering density with the percentage of D_2O in an aqueous solution of biological molecules (such as H or D proteins, phospholipids and nucleic acids). Match points for a 75 % deuterated protein and an unlabelled protein are shown at 40% and 100% respectively.

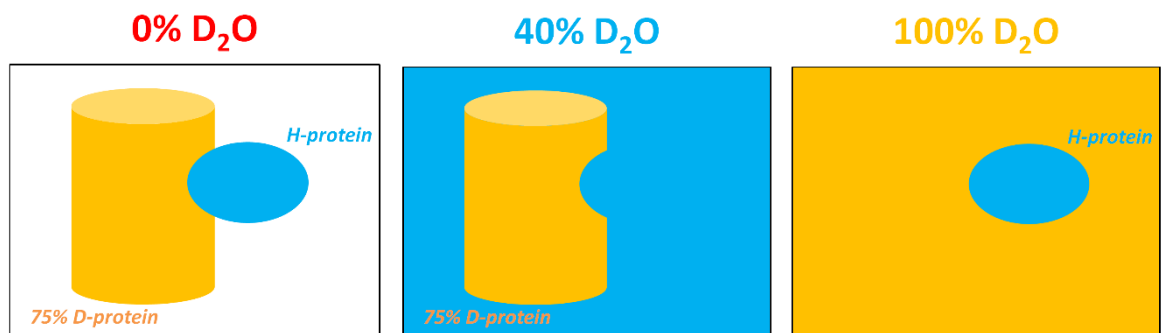


Figure 2.11: Principle of a contrast variation experiment, example of a two-protein complex involving an unlabelled protein and a 75 % deuterated protein.

2.3.3 The ESRF synchrotron radiation source

The European Radiation Synchrotron Facility (ESRF, Grenoble, France) operates the most intense synchrotron X-ray source in the world (Figure 2.12). The linear accelerator (linac) accelerates the electrons previously emitted by an electron gun which are then transferred to a booster synchrotron. After reaching the energy of 6 GeV, they are transferred to the storage ring. Bending magnets direct the beam from one straight section of the storage ring to the next. Synchrotron radiation, emitted from insertion devices (ID) or from the bending magnets (BM) (Figure 2.13), is captured and tailored (energy selection, focussing) for use in experiments on the beamlines.



Figure 2.12: Aerial picture of the ESRF located on the EPN campus in Grenoble. Photograph from www.esrf.eu.

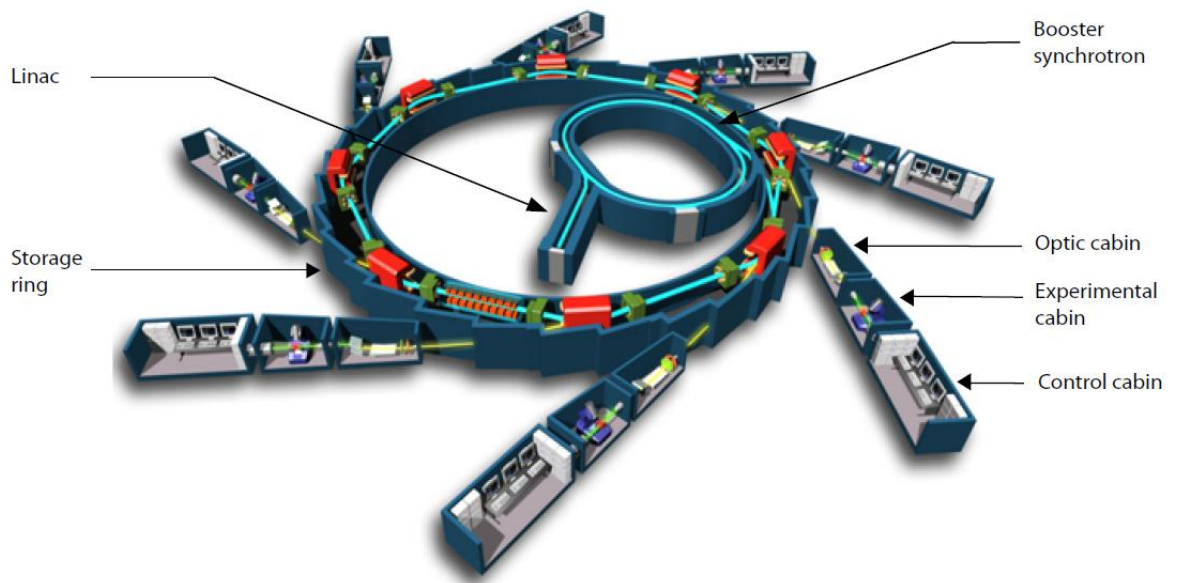


Figure 2.13: Schematic of the inside of a synchrotron showing the linac and booster synchrotron as well as bending magnets, insertion devices and the beamlines of the main storage ring. Picture from www.esrf.eu.

Small-angle X-ray scattering measurements were performed at the bioSAXS beamlines of the ESRF (Grenoble, France, *Pernot et al., 2013*), dedicated to measuring biological macromolecules in solution. During this work beamline ID14-3 was transferred to BM29. Both the ID14-3 beamline and now the BM29 beamline are equipped with an optimised optical layout for bioSAXS experiments with a multilayer monochromator for maximising the available flux, and a toroidal focusing mirror for the generation of a $100 \mu\text{m}^2$ beam. The working energies are between 7 to 15 keV with an achievable q-range of 0.01 to 5 nm^{-1} . The bioSAXS facility is equipped with a Pilatus 1M detector and a liquid handling robot (Figures 2.14 and 2.16). The minimum sample size is $5 \mu\text{l}$, but $30 \mu\text{l}$ is needed for the 30 frame measurement. A schematic of the bioSAXS beamline is presented in Figure 2.15.

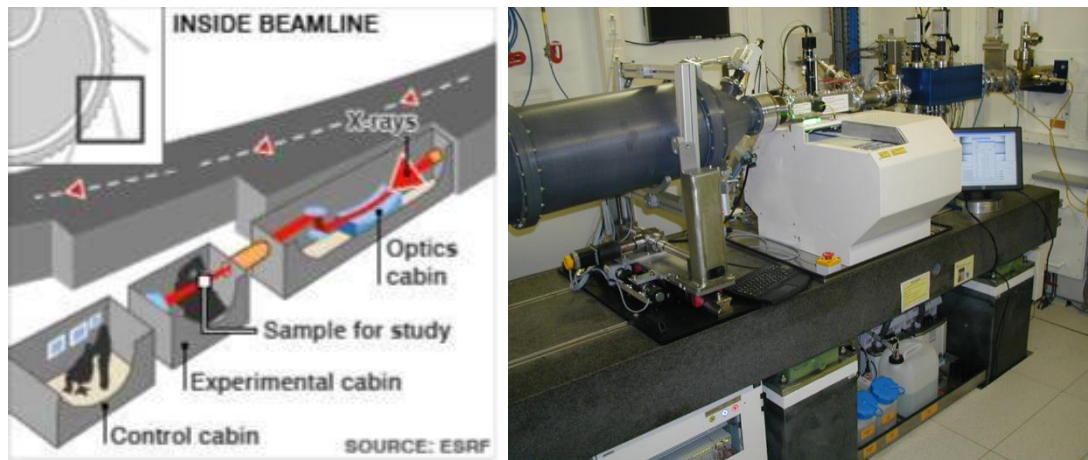


Figure 2.14: Left: General organisation of an ESRF beamline; right: Photo of the beamline experimental cabin. Pictures from <http://www.esrf.eu>.

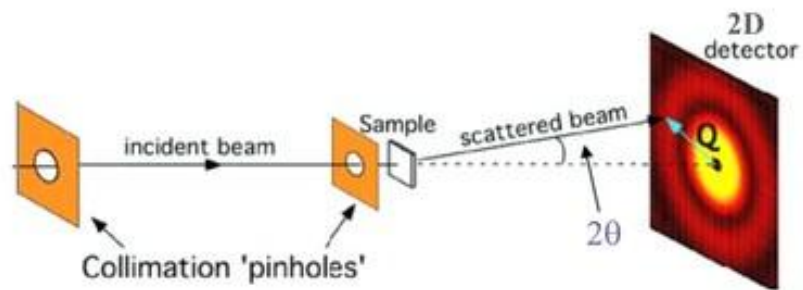


Figure 2.15: Schematic of the BioSAXS beamline. Picture from <http://www.esrf.eu>.

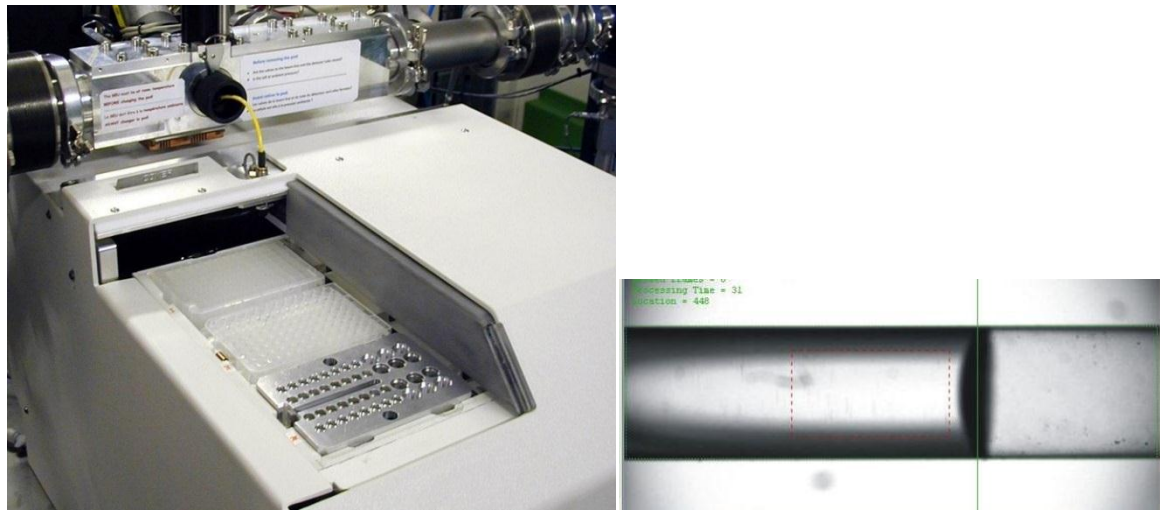


Figure 2.16: Left: Photo of the Sample charger robot; right: photo of the sample loaded in the measurement capillary. Pictures from <http://www.esrf.eu>.

2.3.4 The Institut Laue Langevin neutron source

The high-flux research reactor of the Institut Laue-Langevin (ILL, Grenoble, France) produces neutrons through fission to be then directed at a suite of instruments and used to probe the structure and behaviour of many forms of matter by elastic and inelastic neutron scattering (general photo presented in Figure 2.17).



Figure 2.17: The Institut Laue-Langevin and its nuclear reactor (blue dome) producing the neutrons used for the SANS data collection. Photograph from Patrick Blanc.

The small-angle neutron scattering measurements presented in Chapter 4 were performed on the D22 instrument of the Institut Laue-Langevin (ILL, Grenoble, France) as presented in Figure 2.18 and Figure 2.19. D22 has a resolution range varying from 10-15 Å to approximately 1000 Å. The maximum flux at sample is equal to 1.2×10^8 neutrons/cm²/s. Neutrons are firstly selected in function of their speed via the velocity selector (used as a monochromator) in order to define the wavelength (from 4.5 to 40 Å). Then, a guide is used for their collimation to reduce the divergence of the flux. Samples in their quartz cuvettes are set in the sample holder which allows temperature control. After scattering through the samples, neutrons are finally detected in a surface detector which can be moved along a 20 metre long and 2.5 metre width evacuated tube under vacuum, providing

sample-to-detector distances of 1.1 m to 17.6 m. Thus, in standard conditions D22 covers a total q range of 4×10^{-4} to 0.44 \AA^{-1} (no detector offset) or 0.85 \AA^{-1} (with detector offset). The data are collected via a two dimensional ^3He multidetector of 1 m^2 size, before being integrated to get the final intensity vs diffusion angle curve.



Figure 2.18: Picture of the small-angle diffractometer D22 instrument with its 20 metre long detector tube. Photograph from <http://www.ill.eu>.

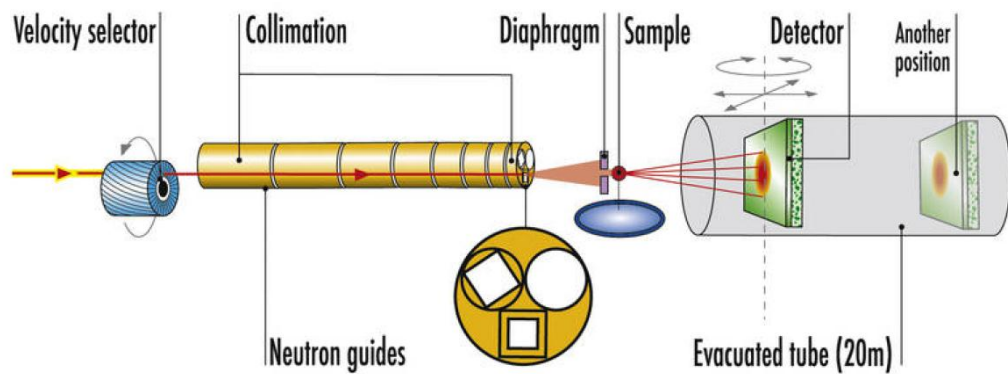


Figure 2.19: Schematic of the D22 instrument of the ILL. Picture from <http://www.ill.eu>.

2.3.5 Data treatment

The software used for the SAXS and SANS data analysis are summarised in Table 2.1 (see Chapter 4 for further details).

| | |
|---|--|
| Software employed for SAXS analysis | |
| Primary data reduction | Primus (<i>Konarev et al., 2003</i>) |
| Data processing | Gnom (<i>Svergun, 1992</i>) |
| <i>Ab initio</i> modelling | DAMMIF (<i>Franke et al., 2009</i>) DAMAVER (<i>Volkov and Svergun, 2003</i>) |
| 3D model display | PyMol (Schrödinger company) |
| Software employed for SANS analysis | |
| Primary data reduction | RNILP, DETCI, RMSK (ILL / R. Ghosh) |
| Data processing | SPOLLY, Gnom (ILL / R. Ghosh) |
| <i>Ab initio</i> modelling | DAMMIF (<i>Franke et al., 2009</i>) DAMAVER (<i>Volkov and Svergun, 2003</i>) |
| 3D model display | PyMol (Schrödinger company) |
| Software employed for analysis of the predicted models | Crysol (<i>Svergun et al., 1995</i>) |

Table 2.1: Summary of the main programs employed for SAXS and SANS data treatment.

2.4 Protein crystallisation

2.4.1 Fundamental principles:

The first step necessary for obtaining a structure by X-ray or neutron diffraction is the preparation of good quality crystals. Various methods are available to crystallise a protein (Figure 2.20, left), amongst which are microbatch (A) to vapour diffusion (B), dialysis (C), and free interface diffusion (D).

In the vapour diffusion method small drops (tens of nanolitres to few microlitres) are made containing the target protein in buffered solution with a low concentration of precipitate (e.g. PEG, MPD, salts). These are set in ‘wells’ containing the same buffer, without protein but a high concentration of precipitant, and the whole system is then sealed. The equilibrium of the chemical potentials in the well is responsible for the evaporation of water in the drop towards the well, increasing gradually the protein concentration. If this balance crosses the nucleation zone, crystalline seeds will form. The concentration of soluble protein will then decrease, allowing crystalline growth phase. The two techniques used here are referred to as ‘hanging drop vapour diffusion’ and ‘sitting drop vapour diffusion’.

In hanging drop vapour diffusion (Figure 2.20, right side), the protein solution is contained in a hanging drop held downward over the well containing a higher concentration of the precipitating agent(s). In sitting drop style, the protein solution is held in small ‘bridges’ above the well.

In this work, initial screening for crystallisation conditions was performed automatically using the High Throughput Crystallisation (HTX) platform of EMBL Grenoble. The *Cartesian*® (Genomics Solution) robot of the HTX platform realises nano-drops (0.1 µl) in the sitting drop configuration allowing automatic screening of 576 standard conditions. Promising conditions were then reproduced manually with the technique of hanging drop vapour diffusion to optimise the conditions and obtain the best possible quality and size of crystals.

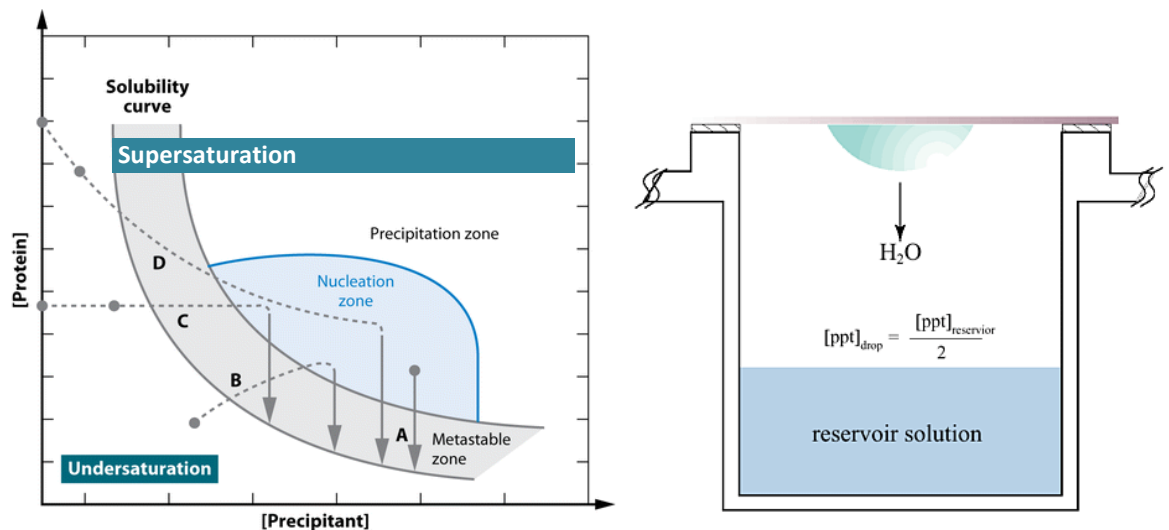


Figure 2.20: **Left:** simplified phase diagram for protein crystallisation. Paths for microbatch (A), vapour diffusion (B), dialysis (C), and free interface diffusion (D) are shown. Crystals only form in the zone of supersaturation (right of the solubility curve), where the concentration of the protein in the precipitant solution is above its solubility. In the zone of undersaturation (left of the solubility curve), crystals dissolve and cannot form (Figure modified from *Chayen et al., 1998*). **Right:** hanging drop vapour diffusion method.

2.4.2 Preparation of crystallography experiments

To avoid radiation damage (review *Garman et al, 2002*) resulting from the high intensity synchrotron beams, protein crystals are routinely cooled to liquid nitrogen temperatures (*Garman and Mitchell, 1996*). However, protein crystals typically contain a high percentage of water. To therefore avoid the formation of ice in loops during their cooling, crystals are usually transferred in to a solution containing the crystallisation mother liquor with addition of a cryo-protectant - typically glycerol, PEG or MPD, before being cryo-cooled.

3. THE EBA181-4.1R PROTEIN COMPLEX: EXPRESSION AND CHARACTERISATION BY BIOCHEMICAL AND BIOPHYSICAL TECHNIQUES

Abstract

This chapter focuses on biochemical and biophysical characterisation of the interaction between the EBA181 and 4.1R proteins that have been associated with the Plasmodium falciparum infection of red blood cells. This work was essential preparation for the structural work carried out for the SAXS, SANS and NMR solution studies that are described in later chapters of this thesis. Initial work from collaborators in the group of Prof. Theresa Coetzer at Witwatersrand University Medical School identified the interaction between these proteins using phage display. The subsequent development and optimisation of the expression systems involved are described in detail in this chapter. Following this, work that characterises the interaction between the two proteins is described. This includes detailed pull-down assays as well as a wide ranging biochemical and biophysical investigation that was designed to yield information on the conformation and behaviour of individual proteins as well as of the complex.

The pull-down assays have demonstrated the interaction in a tag- or -resin-independent way and allowed the characterisation of the interaction between the two proteins by testing complex stability in different conditions (pH variation, salt strength) as well as competition

assays using small peptides. Light scattering has provided the size of individual proteins and the complex, showing that both proteins behave as monomers in solution, and that the complex has a 1:1 protein stoichiometry, with a K_d estimated at 0.3 μM using micro-scale thermophoresis (MST). Structural characterisation by $^1\text{H-NMR}$ and circular dichroism has shown the intrinsically disordered state of EBA-181 and the low level of compactness of 4.1R.

3.1 Introduction

This chapter focuses on the structural and functional characterisation of the interaction between the 10 kDa domain of the human erythrocyte 4.1R protein (the 4.1R_{10kDa} sequence is well-known in the literature and defined as from residue 405 to 471, *Leto et al., 1984*) and the PfJ fragment of the malarial *Plasmodium falciparum* protein EBA181 (EBA181₉₄₅₋₁₀₉₇). The background to this work has been described in Chapters 1 and 2, and exploits the availability of the expression systems developed by the Coetzer Group as well as facilities within the Partnership for Structural Biology (PSB) in Grenoble.

The overall strategy was to modify and develop the expression systems to yield suitable quantities of proteins to allow complementary biochemical and biophysical techniques to get insights and allow a consensus on key aspects of the structure, interactions and functional relationships of these proteins. Section 3.2 describes the expression system and the way in which the two proteins were produced and purified. A considerable amount of

time was put into this given the need for high efficiency and large quantities of protein. The plasmid DNA transposition, bacterial transformation and protein solubility tests are described. The way in which the large scale bacterial cultures were carried out is also described, as are the purification methods used for both proteins. Additionally, information on the preparation of the complex, pull-down, and inhibition assays is given in Section 3.3. Section 3.4 describes the biochemical characterisation carried out, mainly focused on pull-down assays that were designed to yield information on the complex by testing its stability. Section 3.5 describes an extensive set of biophysical studies including one dimensional nuclear magnetic resonance ($^1\text{H-NMR}$), circular dichroism (CD), thermal shift assays (TSA), size exclusion chromatography coupled to multi-angle laser light scattering and refractive index detection (SEC-MALLS-RI), dynamic light scattering (DLS), micro-scale thermophoresis (MST) and many crystallisation trials. The chapter finishes with a discussion/conclusion section that summarises the overall results and their implications for the structural studies on this interaction which are described later in this thesis.

3.2 Production of individual EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins

3.2.1 Plasmid DNA: extraction and quality control

Both plasmids, pGEX-4T-2 containing the 4.1R_{10kDa} coding sequence, and pET15b containing the EBA181₉₄₅₋₁₀₉₇ coding sequence, were extracted from *E.coli* DH5 α bacteria via mini-preps based on the alkaline lysis method. The quality of the mini-preps was checked on agarose gels, and UV spectra over the range from 220 nm to 350 nm

(Figure 3.1) were recorded in order to measure the concentrations. The two mini-prep samples with the highest concentration are shown with black and red arrows in Figure 3.1: pGEX-4T-2 (GST-4.1R_{10kDa}) concentrated at 135 ng/μl and pET15b (His-EBA181₉₄₅₋₁₀₉₇) concentrated at 95 ng/μl. Prior to use the plasmids were stored at -20°C.

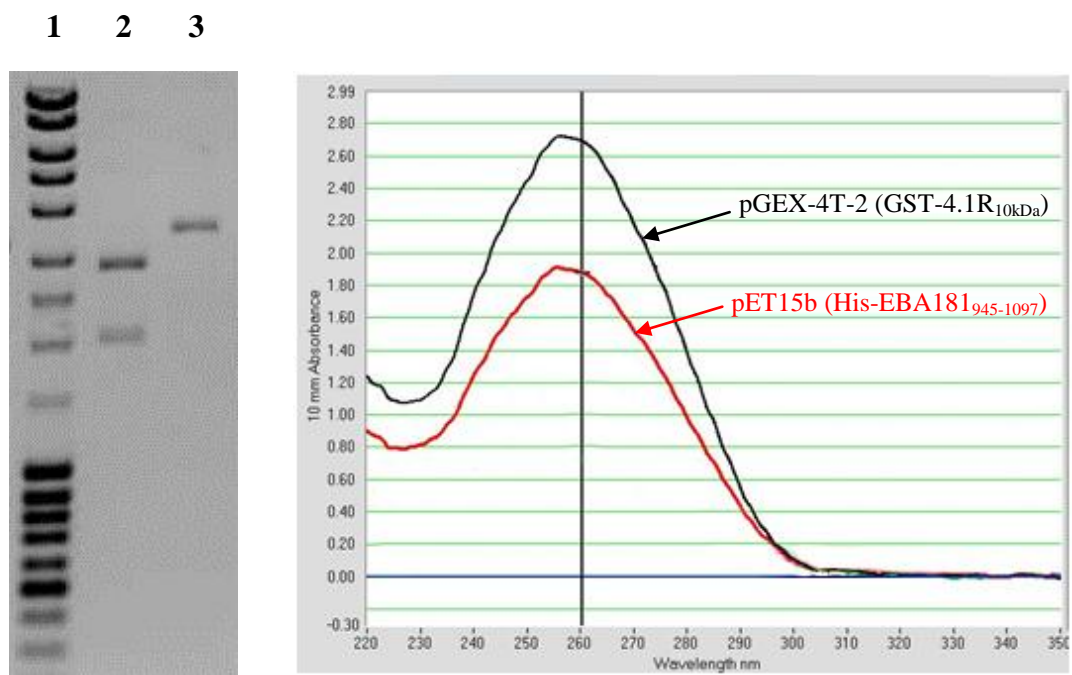


Figure 3.1: 1.2% Agarose gel (left) and UV spectra (right) of the different plasmids after the mini-prep. Lanes of the agarose gel are:

- 1: DNA marker used as a calibrant.
- 2: pGEX-4T-2 containing the GST-4.1R_{10kDa} coding sequence.
- 3: pET15b containing the His-EBA181₉₄₅₋₁₀₉₇ coding sequence.

3.2.2 Change of resistance marker

In vitro transposition was used in order to change the antibiotic resistance from ampicillin to kanamycin for both plasmids. This was necessary because in high cell density cultures and/or cultures involving deuterated media the expressed beta-lactamase enzymes (from the *bla* gene carried by the plasmid), have a tendency to leak outside the cell, where they inactivate the ampicillin antibiotic, leading to a loss of selection pressure and overgrowth of plasmid-less cells. Hence recombinant plasmid expression is drastically reduced or eliminated. In the case of kanamycin resistant cells the expression of the protein involved occurs in the cytosol and hence leakage to the medium is much less likely to occur since the protein would have to cross two bacterial membranes.

The kit used for *in vitro* transposition was the *EZ-Tn5*TM <KAN> *Insertion Kit* from Epicentre Technologies Corporation®. A total reaction volume of 10 µl was prepared with 0.2 µg of each plasmid of interest containing one of the two sequences (pGEX-4T-2 or pET15b) mixed with diluted transposon at a molar equivalent concentration to the plasmid (in 10 mM Tris-HCl pH 7.5 with 1 mM EDTA), 1 U of transposase (in 50% glycerol containing 50 mM Tris-HCl pH 7.5, 0.1 M NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, and 0.1% Triton® X-100), 1 µl of 10x reaction buffer (0.50 M Tris-acetate pH 7.5, 1.5 M potassium acetate, 100 mM magnesium acetate, and 40 mM spermidine) and finally completed with ddH₂O.

The reaction was incubated for 2 hours at 37°C, and spun in a centrifuge for 30 seconds in case of evaporation on the lid. The reaction was terminated by the addition of 1 µl of 10x

stop solution (1% SDS solution). After being mixed, the solution was heated for 10 minutes at 70°C. Top10 and JM109 bacteria cells were used for transformation. 3 µl and 7 µl of the previous reaction were respectively used with 30 µl of Top10 and the JM109 bacteria cells, and maintained on ice for 15 minutes. Heat shock was performed at 42°C for 30 seconds and the tubes were immediately put back on ice for 5 minutes. After the addition of 1 ml of super-optimal broth with catabolite repression medium (SOC) (see composition detailed in Appendix A.1), the tubes were incubated at 37°C for 1 hour to facilitate cell growth before being spread on kanamycin-containing plates to select transformants which had the kanamycin transposon inserted into their plasmid DNA.

The plates were grown overnight at 37°C. In order to screen for clones having the kanamycin transposon inserted into the ampicillin resistance gene, each kanamycin resistant colony was tested on two new plates: one plate containing kanamycin to double check, and one containing ampicillin to screen for insertional inactivation of the ampicillin gene. The clones that grew on the kanamycin plate and that did not grow on the ampicillin plate were picked and subsequently grown overnight in LB medium to perform mini-preps in order to extract DNA from each of the kanamycin resistant pGEX-4T-2 or pET15b plasmids.

3.2.3 Bacterial transformation and long-term storage of clones

To express the proteins, each plasmid was transformed into competent bacterial cells suitable for protein expression. *E. coli* Rosetta™ 2 (DE3) was chosen for its particularity to improve expression levels of gene/cDNAs containing rare codons (codons rarely used in

E. coli). These strains supply tRNAs for seven rare codons (AGA, AGG, AUA, CCC, CGG, CUA and GGA) on a compatible chloramphenicol-resistant plasmid.

1 µl of each different plasmid DNA concentrated to 50 ng/µl was mixed with 50 µl of *E. coli* (DE3) Rosetta™ 2 cells and incubated on ice for 15 minutes. To increase transformation efficiency and allow the DNA to pass through the bacterial cell wall/membrane, a heat-shock was performed: 30 seconds at 42°C in a water-bath, then 15 minutes on ice. 1 ml of SOC medium (see medium composition in Appendix A.1) was added to the mixture and incubated for 1 h at 37°C under rapid agitation (200 rpm). Finally, 200 µl of each previous culture was plated onto Petri dishes containing LB-agar medium and the appropriate antibiotics (Appendix A.1).

E. coli Rosetta™ 2 cells for each of the expression constructs were stored at -80°C. A fresh and dense culture (having reached the stationary phase) of each individual clone was used to make long-term storage stocks using sterile beads in a cryo-preserved solution (*Technical Service Consultants Ltd*).

3.2.4 Expression and solubility tests

For the expression and solubility tests of GST-4.1R_{10kDa}, separate cultures of LB medium supplemented with 30 µg/ml of kanamycin and with 34 µg/ml of chloramphenicol (resistance given by the *E. coli* Rosetta™ 2 cells) were started using colonies from Petri dishes. These were then incubated overnight at 37°C under agitation. The overnight culture was used to start three different LB cultures (with appropriate antibiotics) in order to test

three different induction conditions: 18°C overnight, 30°C for 3 hours and 37°C for 3 hours. The bacteria were grown at 37°C until the optical density at 600 nm reached 0.7 and protein expression was then induced with 1 mM IPTG for the three different temperature tested (see Figure 3.2, top).

For His-EBA181₉₄₅₋₁₀₉₇, the optimal expression conditions were known from our collaborators: induction with 1mM IPTG at 37°C for 3 hours (see Figure 3.2, bottom). Similarly to the expression of GST-4.1R_{10kDa}, cultures of LB medium containing 30 µg/ml of kanamycin and 34 µg/ml of chloramphenicol were started using colonies from Petri dishes. These were incubated overnight at 37°C under agitation and this overnight culture was used to start the final LB culture. The bacteria were grown at 37°C until the optical density at 600 nm reached 0.7 and protein expression was then induced with 1 mM IPTG for 3 hours at 37°C.

3.2.5 Large-scale bacterial cultures

Following the change of the antibiotic resistance, eight kanamycin resistant clones expressing His-EBA181₉₄₅₋₁₀₉₇ or GST-4.1R_{10kDa} were tested in order to compare the solubility and expression levels of each protein (data not shown). The expression levels as judged by SDS-PAGE of the two kanamycin resistant clones that showed the highest expression yield of soluble His-EBA181₉₄₅₋₁₀₉₇ or GST-4.1R_{10kDa} after 3 hours of induction at 37°C are shown in Figure 3.3.

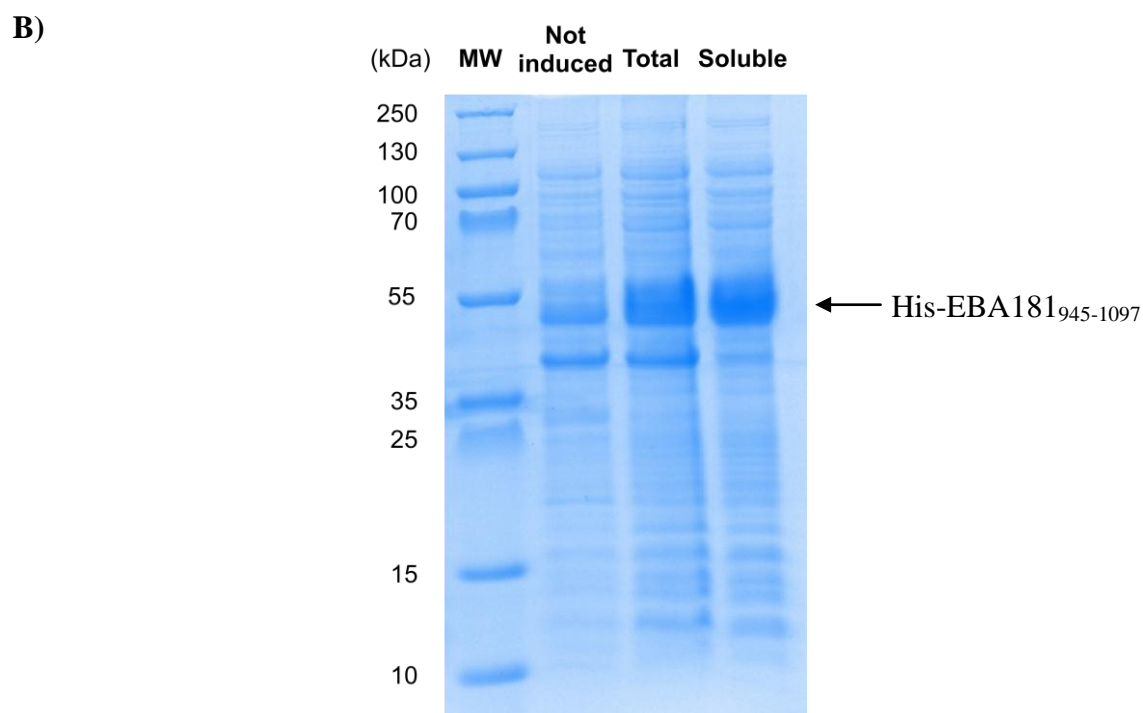
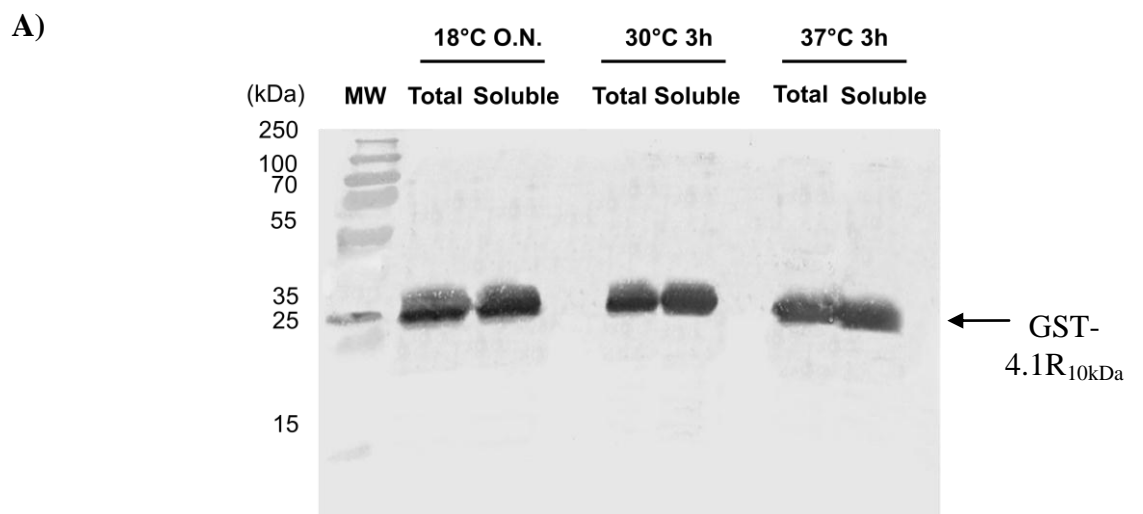


Figure 3.2: A: Western blot analysis after expression of GST-4.1R_{10kDa} at three different temperatures. **B:** SDS-PAGE (12 % Tris-Tricine gel) showing the expression and solubility test of the His-EBA181₉₄₅₋₁₀₉₇ protein after induction at 37°C.

MW: Molecular weight marker (*PageRuler™ Plus Prestained SM181* from *Fermentas*).

Not induced: Sample before induction with IPTG.

Total: Insoluble and soluble protein after IPTG induction.

Soluble: Soluble proteins after IPTG induction and centrifugation.

For the large-scale protein expression, the selected medium was LB medium supplemented with 30 µg/ml of kanamycin and with 34 µg/ml of chloramphenicol. Firstly, 5 ml precultures were inoculated from Petri dish and incubated until reaching the stationary phase of growth (37°C, overnight under agitation). Further precultures were subsequently started from the 5 ml ones, in three steps with 20 ml, 100 ml to finally 500 ml. After this, each 1 litre culture was inoculated using the starter 500 ml precultures (1:50 dilution). The bacteria were grown at 37°C to an optical density of around 0.7 at 600 nm. Protein expression was induced with 1 mM IPTG over 3 hours as defined from the previous expression and solubility tests. The bacteria were lysed by sonication to release the proteins; the suspension was then centrifuged at 40 000 g for 1 h and at 4°C to separate the soluble material.

It should be noted that EBA181₉₄₅₋₁₀₉₇ is very negatively charged at pH 7, which leads to an unusual migration pattern on SDS-PAGE (*Iakoucheva et al., 2001*). Instead of having a migration at around 19 kDa which is its real molecular weight, the migration appears close to 50 kDa, due to decreased binding of the SDS which is also negatively charged. N-terminal sequencing and mass spectrometry experiments were performed on EBA181₉₄₅₋₁₀₉₇ which confirmed the molecular weight and N-terminal sequence (first eight amino-acids).

3.2.6 Protein purification

Both proteins were purified using two different techniques (Table 3.1).

- A) Immobilised metal affinity chromatography (IMAC, using an Äkta Prime system or a peristaltic pump), following three different steps (binding, washing and elution) and the buffers listed in Appendix A.2, Table A.1.
- B) Size exclusion chromatography, carried out using the Äkta Purifier system and Superdex columns from GE Healthcare® (S75 or S200).

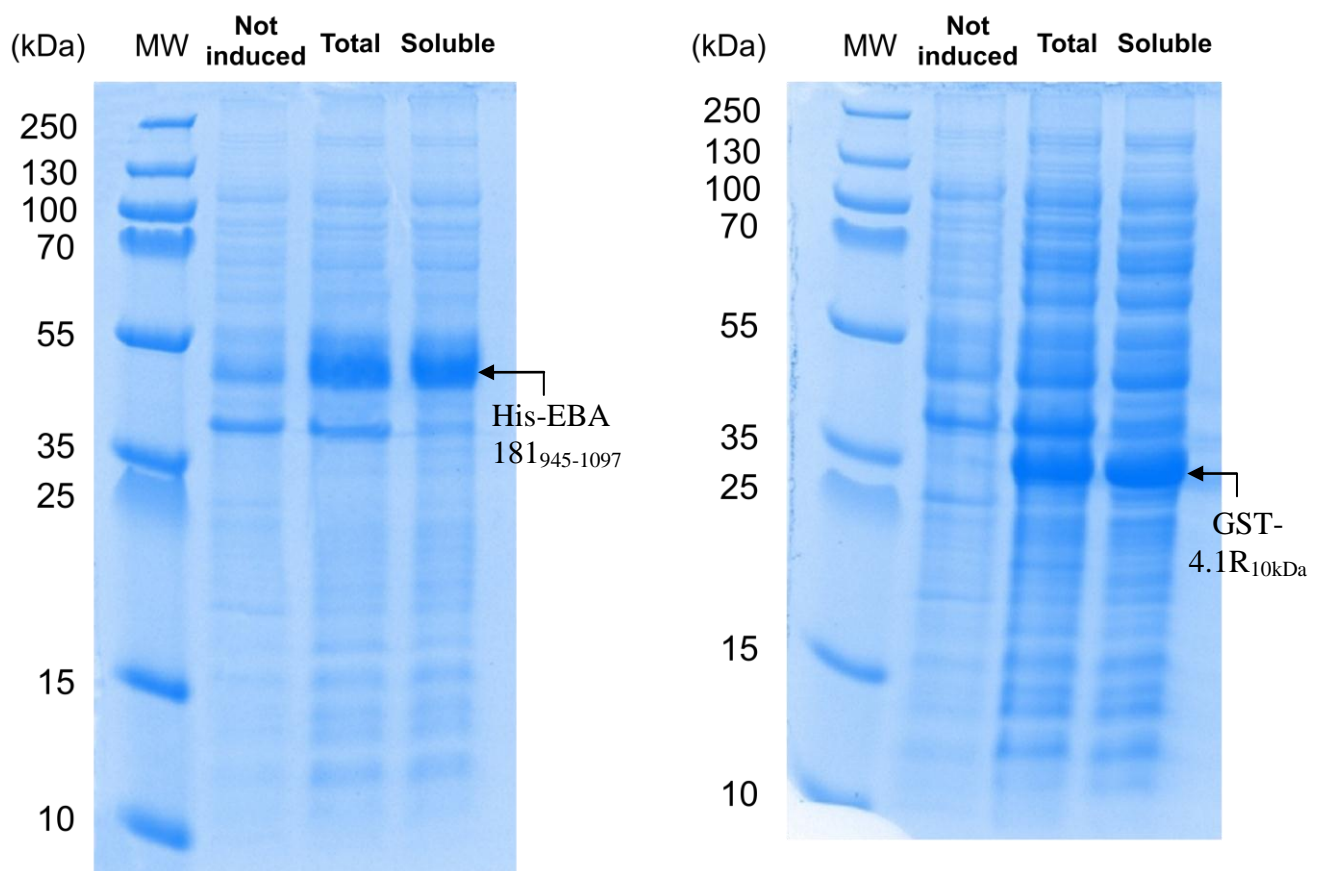


Figure 3.3: SDS-PAGE showing expression of His-EBA_{181⁹⁴⁵⁻¹⁰⁹⁷} (**left**) and GST-4.1R_{10kDa} (**right**) after the change of the antibiotic resistance.

MW: Molecular weight marker (*PageRulerTM Plus Prestained SM181* from *Fermentas*).

Not induced: Sample before induction with IPTG.

Total: Insoluble and soluble protein after IPTG induction.

Soluble: Soluble proteins after IPTG induction and centrifugation.

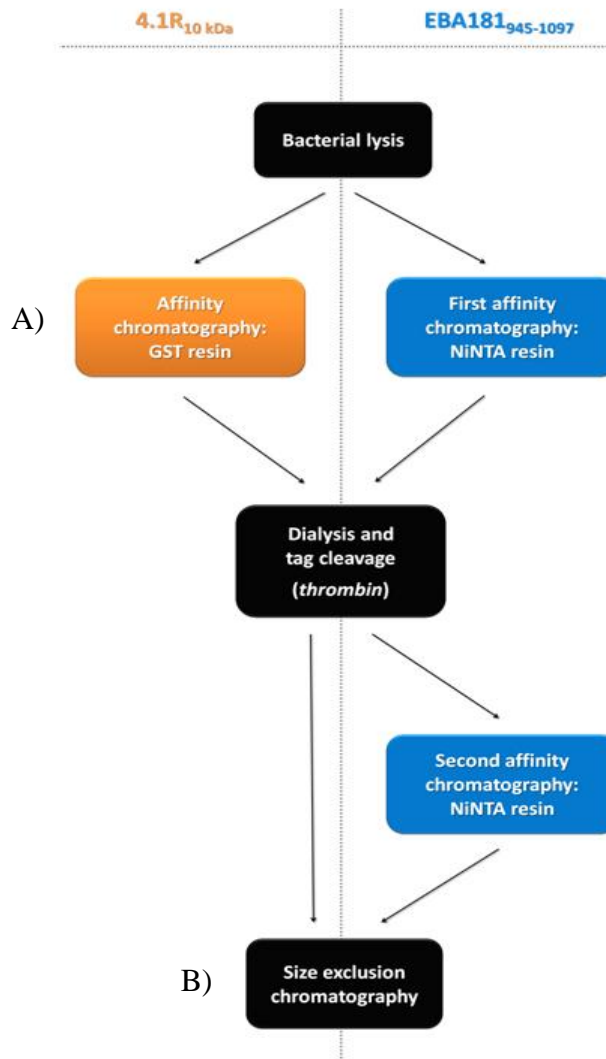


Diagram 3.1: Summary and comparison of the main steps of EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} purification protocol.

3.2.6.1 Purification of EBA181₉₄₅₋₁₀₉₇

His-tagged EBA181₉₄₅₋₁₀₉₇ can be purified by affinity chromatography (Figure 3.4) on a nickel column (packed with NiNTA Superflow resin from Qiagen®), equilibrated in lysis

and binding buffer (Appendix A.2, Table A.1). Soluble proteins extracted from *E. coli* culture were loaded onto the NiNTA resin column. Washing buffer was then used to remove the non-specifically-bound proteins, and His-EBA181₉₄₅₋₁₀₉₇ was finally eluted using the elution buffer containing 150 mM imidazole (Appendix A.2, Table A.1). After elution, His-EBA181₉₄₅₋₁₀₉₇ was dialysed overnight at 8°C against the lysis buffer with thrombin protease (3U / mg of protein), in order to cleave the histidine tag.

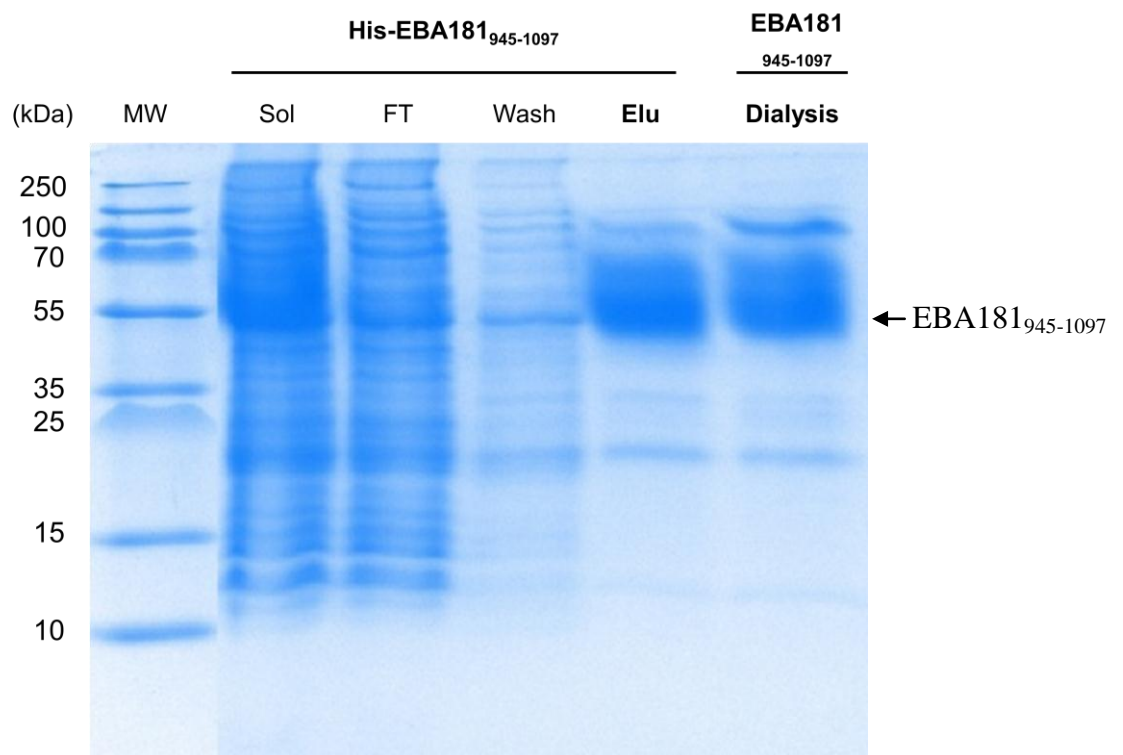


Figure 3.4: SDS-PAGE of the first NiNTA purification of EBA181₉₄₅₋₁₀₉₇.

MW: Molecular weight marker (*PageRuler™ Plus Prestained SM181* from *Fermentas*).

Sol: Soluble proteins after centrifugation.

FT: Flow through the NiNTA resin column.

Wash: Wash of the column with the washing buffer.

Elu: Elution of the proteins with the elution buffer.

Dialysis: Sample after dialysis and tag cleavage.

After dialysis, EBA181₉₄₅₋₁₀₉₇ was again purified on a NiNTA column in order to retain the uncleaved (intact) fraction of His-EBA181₉₄₅₋₁₀₉₇ as well as the His-tag itself, leaving the untagged EBA181₉₄₅₋₁₀₉₇ protein in the flow-through fraction (Figure 3.5, 'FT'). According to the SDS-PAGE, the tag was successfully removed from EBA181₉₄₅₋₁₀₉₇ since the whole sample was not retained by the NiNTA column and no His-EBA181₉₄₅₋₁₀₉₇ was detectable in the elution fraction (Figure 3.5, 'Elu'). The His-tag alone resulting from the cleavage is too small to be seen on the gel.

The untagged (FT) fraction was concentrated to a volume of approximately 1 ml (by a factor of 15-20) using an Amicon Ultra centricon tube with a cut-off of 10 kDa and then filtered (0.2 µm filter). The resulting solution was then injected onto a size exclusion chromatography column (Figure 3.6, column reference: *Superdex S200 10/300 GL, GE Healthcare*®) that had previously been equilibrated in filtered and degassed buffer for two column volumes (50 ml) at a flow of 0.5 ml/min. 500 µl of protein was injected per run in a 1 ml loop using the same flow of 0.5 ml/min.

Because of the absence of aromatic residues in EBA181₉₄₅₋₁₀₉₇, the absorbance during the gel filtration was monitored at 205 nm, corresponding to the peptide bond absorbance. This leads to a noisier spectrum, though still sufficiently good to identify the fractions containing the protein. The size exclusion chromatography led to good separation of EBA181₉₄₅₋₁₀₉₇ from the other contaminants (Figure 3.6). Approximately 5 mg of pure EBA181₉₄₅₋₁₀₉₇ was obtained per litre of bacterial culture. The pooled fractions containing EBA181₉₄₅₋₁₀₉₇ (Figure 3.6, fractions C1 to C9) were concentrated (from 5 to 90 mg/ml) and finally centrifuged (to eliminate any aggregates) for further analysis. The resulting

EBA181₉₄₅₋₁₀₉₇ was considered highly pure for subsequent biochemical and biophysical analysis, as judged by SDS-PAGE (Figure 3.7).

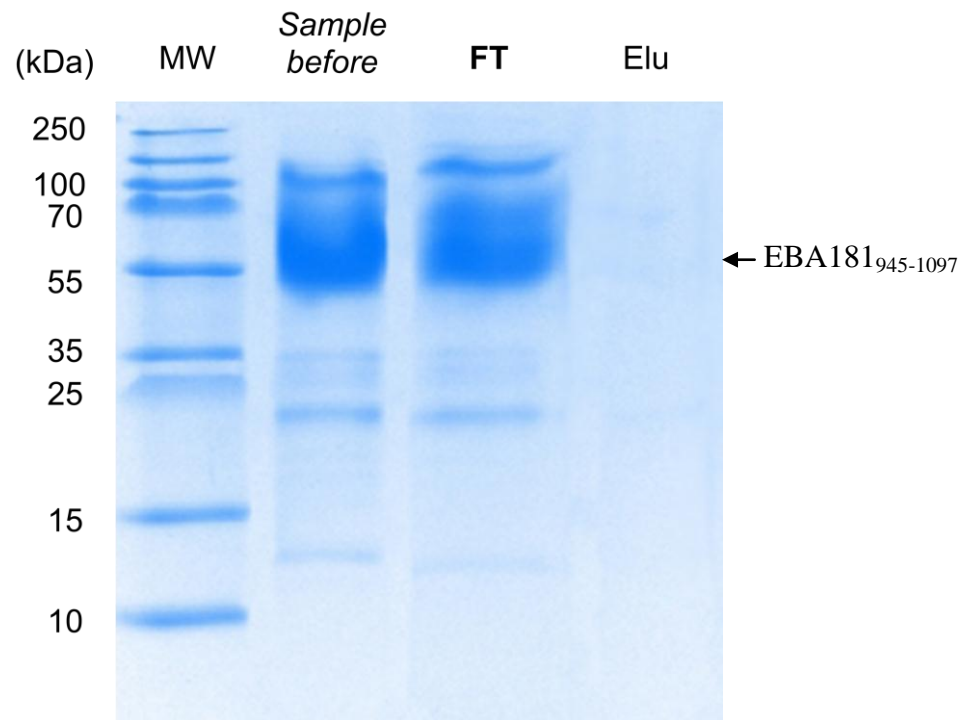


Figure 3.5: SDS-PAGE of the second NiNTA purification after tag cleavage of EBA181₉₄₅₋₁₀₉₇.

MW: Molecular weight marker (*PageRulerTM Plus Prestained SM181* from *Fermentas*).

Sample before: Sample kept after the first NiNTA purification (control).

FT: Flow through the NiNTA resin column.

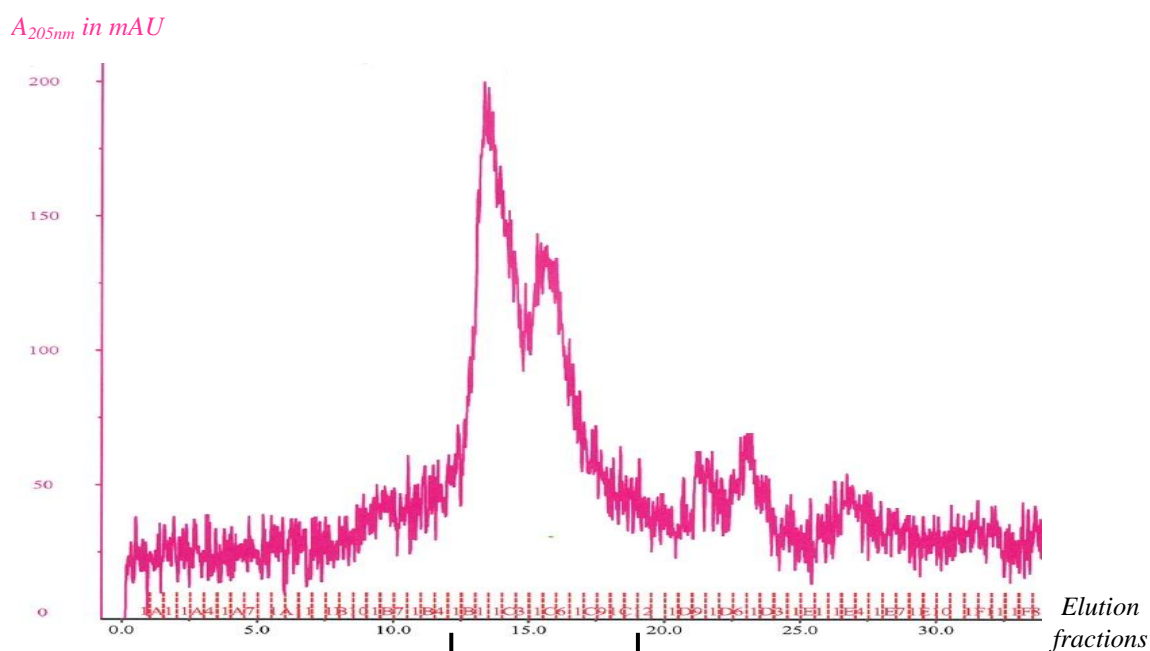
Elu: Elution with the elution buffer.

3.2.6.2 Purification of 4.1R_{10kDa}

Due to the fused GST tag on its N-terminus, the GST-tagged 4.1R_{10kDa} protein (GST-4.1R_{10kDa}) can be purified by affinity chromatography, on a glutathione Sepharose column (packed with *Glutathione SepharoseTM High Performance* resin from *GE Healthcare*®).

Following *E. coli* cell lysis and centrifugation, the supernatant (15 ml) was loaded on the GST resin column (Figure 3.8, A) at a flow of 0.1 ml/min, previously equilibrated with the lysis buffer (Appendix A.2, Table A.1). The column was then washed using 7 to 8 column volumes of the washing buffer (Appendix A.2, Table A.1) in order to remove the contaminant proteins (Figure 3.8, top, 'Wash 1'). During the optimisation of the large scale purification of GST-4.1R_{10kDa}, a significant amount of nucleic acids was found to be co-eluted with GST-4.1R_{10kDa}. Therefore an extended overnight washing step with a least 75 column volumes of washing buffer containing 1 M NaCl was used to eliminate these nucleic acid contaminants (Figure 3.8, 'Wash 2'). After the overnight wash GST-4.1R_{10kDa} was eluted from the column using 5 ml of the elution buffer containing reduced state glutathione (buffer composition in Appendix A.2, Table A.1).

A)



B)

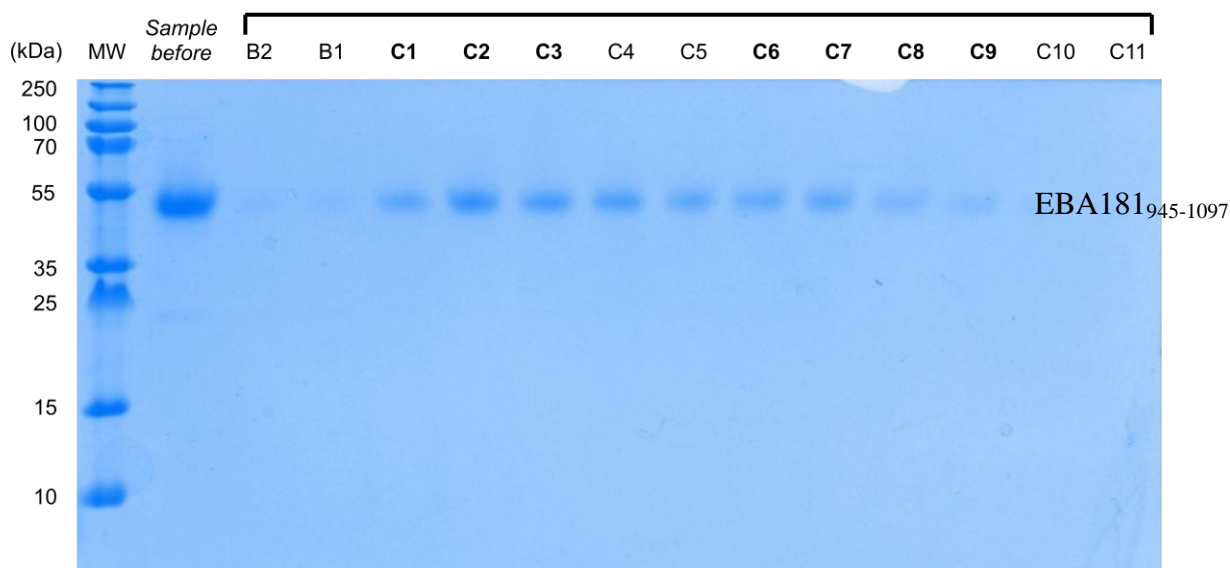


Figure 3.6: A: Size exclusion chromatography purification of EBA181₉₄₅₋₁₀₉₇ using *Superdex S200 10/300 GL* column, *GE Healthcare*[®]. Pink curve: absorption curve ($A_{205\text{nm}}$ in mAU). **B:** SDS-PAGE analysis of collected fractions. The two black lines show the corresponding chromatography fractions on the SDS PAGE.

MW: Molecular weight marker (*PageRulerTM Plus Prestained SM181* from *Fermentas*).

Sample before: Sample kept after the second NiNTA purification step (control).

B2 to C11: Elution fractions.

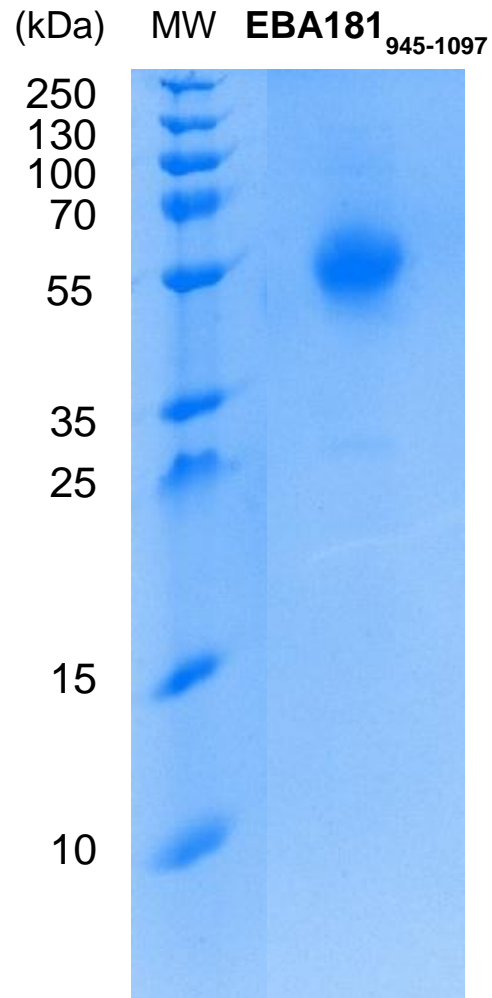


Figure 3.7: SDS-PAGE of EBA181₉₄₅₋₁₀₉₇ after the final size-exclusion step of purification.

MW: Molecular weight marker (*PageRuler™ Plus Prestained SM181* from *Fermentas*).

EBA181₉₄₅₋₁₀₉₇: Purified EBA181₉₄₅₋₁₀₉₇ protein.

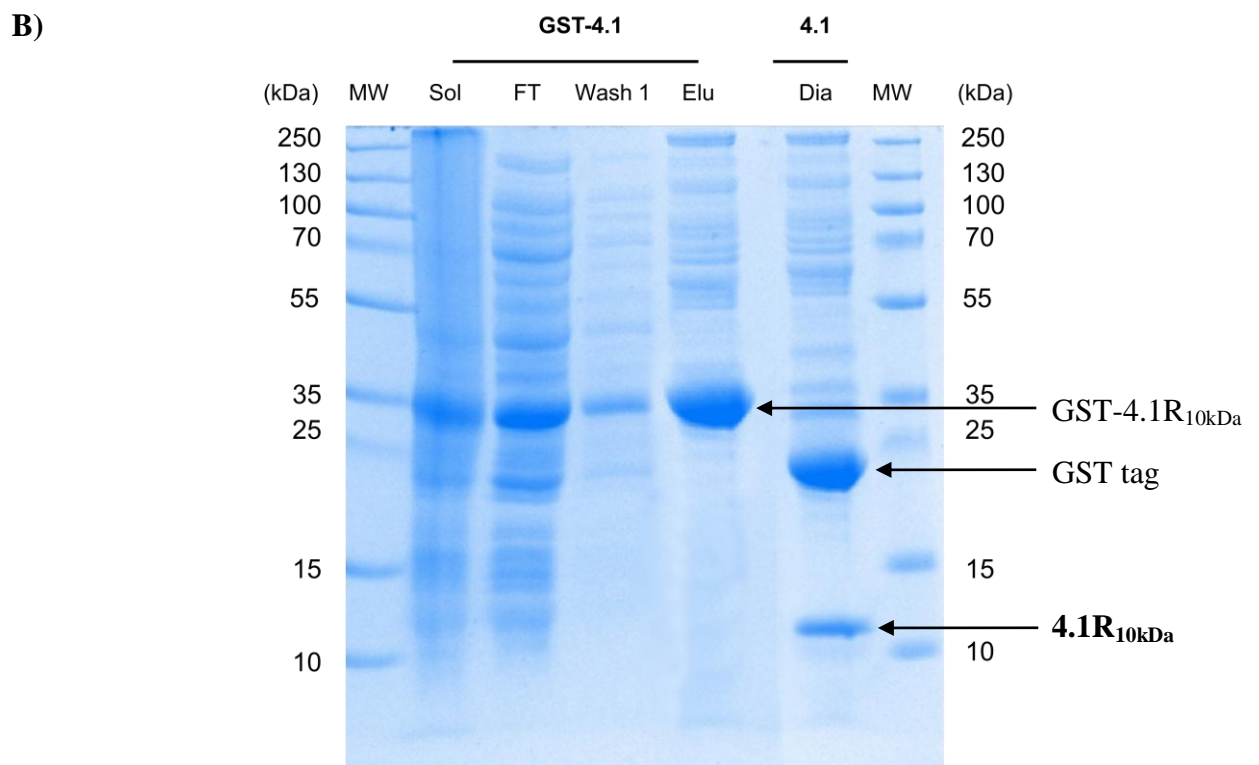


Figure 3.8: Purification of GST-4.1R_{10kDa} on the GST resin column and tag cleavage.

A: Chromatogram of the different steps for the GST column purification (blue curve: A_{280nm} in mAU, red curve: conductivity in mS). **B:** SDS-PAGE of the different steps of the GST column purification.

MW: Molecular weight marker (*PageRuler™ Plus Prestained SM181* from *Fermentas*).

Sol: Soluble proteins after centrifugation.

FT: Flow through the GST resin column.

Wash 1: Wash of the column with the washing buffer.

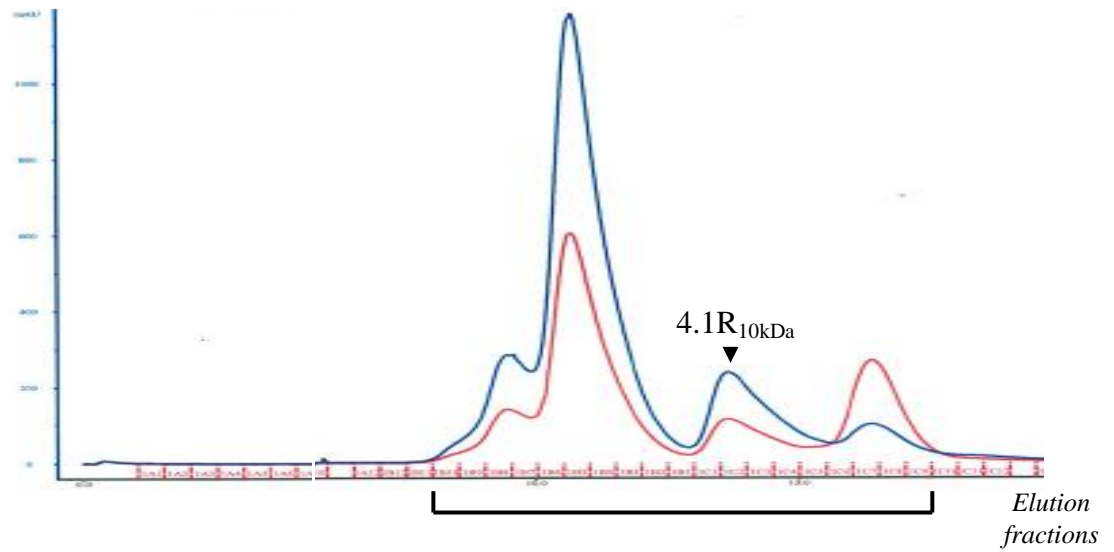
Elu: Elution of the proteins with the elution buffer.

Dia: Sample after dialysis and GST tag cleavage.

After elution, GST-4.1R_{10kDa} was dialysed overnight at 8°C in the same buffer as for EBA181₉₄₅₋₁₀₉₇, with thrombin protease (3U / mg of protein) in order to cleave the GST tag. After the dialysis, the solution containing 4.1R_{10kDa} was concentrated to around 1 ml volume (by a factor of 15) using an Amicon Ultra centricon tube with the right weight cut-off (3.5 kDa for 4.1R_{10kDa}), filtered through a 0.2 µm filter and injected in a size exclusion chromatography column (column reference: *Superdex S75 10/300 GL, GE Healthcare*[®]), previously equilibrated in filtered and degassed dialysis buffer for two column volumes (50 ml) at a flow of 0.5 ml/min. 500 µl of protein is injected per run in a 1 ml loop using the same flow of 0.5 ml/min (Figure 3.9). In this case, the absorbance of 4.1R_{10kDa} could be monitored as usual at 280 nm due to the presence of aromatic residues (Figure 3.9, A, blue curve) with the contaminant nucleic acids absorption monitored at 254 nm (red curve). Using the size exclusion chromatography, 4.1R_{10kDa} was successfully separated from contaminants (mostly nucleic acids, cleaved GST tag and uncleaved GST-4.1R_{10kDa}, Figure 3.9, B). Approximately 0.5 mg of pure 4.1R_{10kDa} was obtained per litre of bacterial culture. The fractions containing purified 4.1R_{10kDa} (Figure 3.9, C3 to C5 fractions, red arrows) were centrifuged to eliminate any aggregates present and finally concentrated to around 5 mg/ml (Figure 3.10).

A)

A_{280nm} and A_{254nm} in mAU



B)

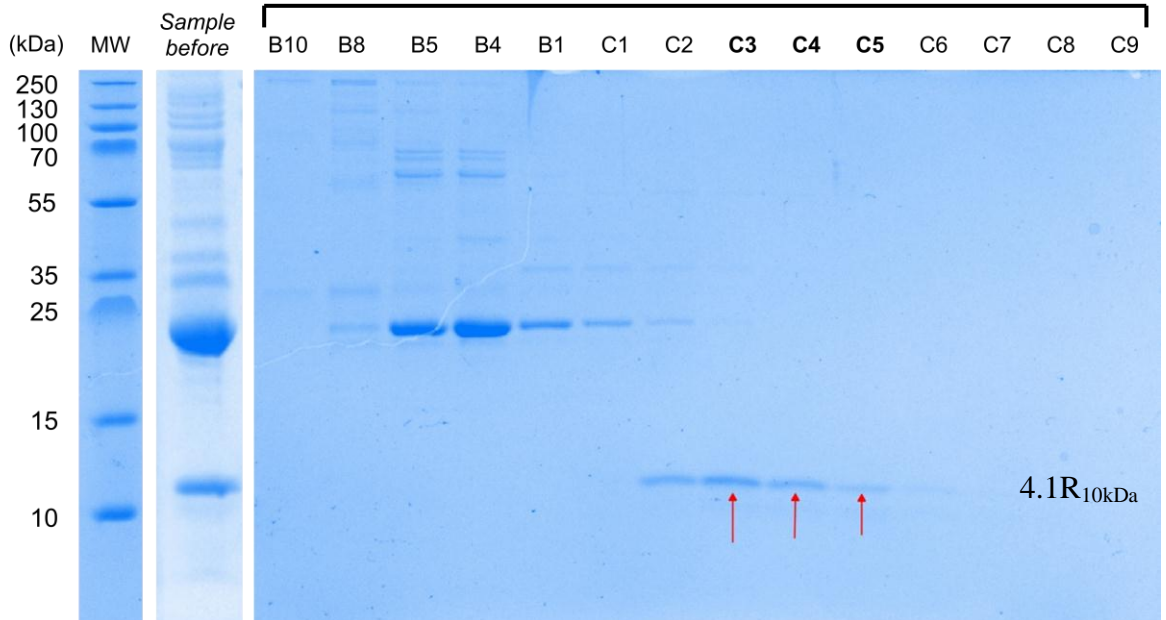


Figure 3.9: A: Size exclusion chromatography purification of 4.1R_{10kDa} using *Superdex S75 10/300 GL* column, *GE Healthcare*[®]. Blue curve: A_{280nm} in mAU, red curve: A_{254nm} in mAU. **B:** Corresponding SDS-PAGE. The two black lines show the corresponding chromatography fractions on the SDS PAGE and the thin red arrows show the fractions pulled together for concentration.

MW: Molecular weight marker (*PageRulerTM Plus Prestained SM181* from *Fermentas*).

Sample before: 4.1R_{10kDa} after dialysis and tag-cleavage (control).

B10 to C9: Elution fractions.

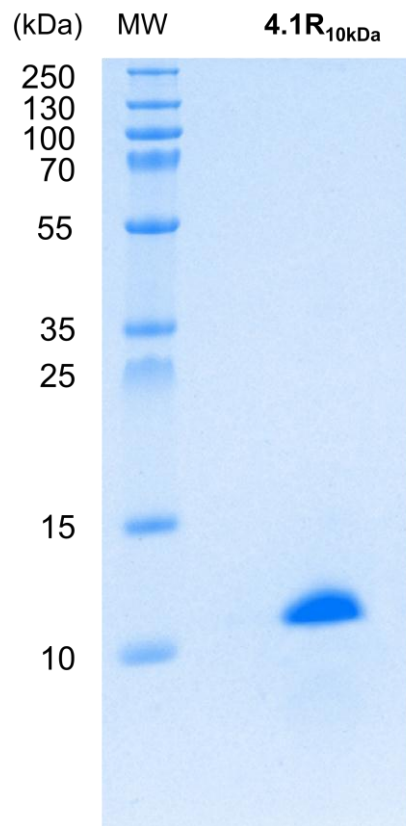


Figure 3.10: SDS-PAGE of 4.1R_{10kDa} after final size-exclusion step of purification.

MW: Molecular weight marker (*PageRuler™ Plus Prestained SM181* from *Fermentas*).

4.1R_{10kDa}: Purified 4.1R_{10kDa} protein.

3.3 Preparation and purification of the complex

The complex formed by EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} was formed either by mixing both previously freshly purified individual proteins in a 1:1 molar ratio, or by using a pull-down assay to purify the complex from bacterial crude extract (this latter method is presented in this section).

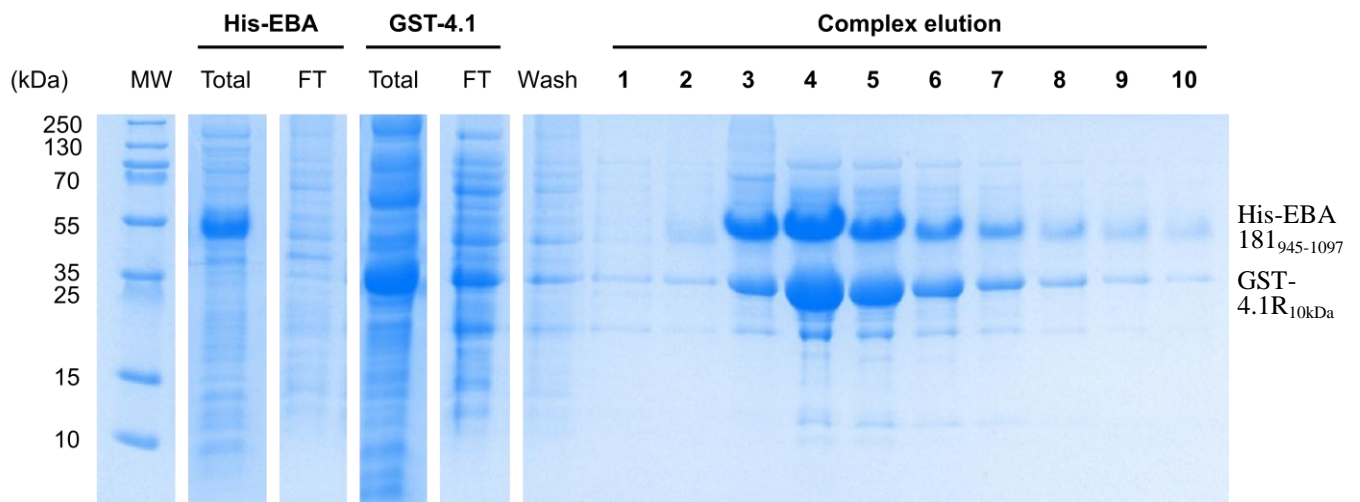


Figure 3.11: SDS-PAGE of the pull-down of the complex His-EBA181₉₄₅₋₁₀₉₇ / GST-4.1R_{10kDa} using NiNTA resin. EBA181₉₄₅₋₁₀₉₇ is used as the bait for 4.1R_{10kDa}.

MW: Molecular weight marker (*PageRulerTM Plus Prestained SM181* from *Fermentas*).

FT: Flow through the NiNTA resin column.

Wash: Washing steps.

1 to 10: Elution fractions.

In the first step, bacterial crude extract from a 1 litre culture expressing His-EBA181₉₄₅₋₁₀₉₇ was loaded onto the IMAC column (Figure 3.11, 'His-EBA Total' and 'His-EBA FT'). Due to the His-tag and the resin used (Nickel NiNTA, as previously) the His-EBA181₉₄₅₋₁₀₉₇ protein bound to the column. In the second step, bacterial crude extract from a 1 litre culture expressing GST-4.1R_{10kDa} was added and bound to the previously loaded His-EBA181₉₄₅₋₁₀₉₇ (Figure 3.11, lanes 'GST-4.1 Total' and 'GST-4.1 FT'). In the third step, the contaminant proteins were removed by washing the column with the His-EBA181₉₄₅₋₁₀₉₇ washing buffer (composition in Appendix A.2, Table A.1). The complex was then eluted (Figure 3.11, elution fractions from 1 to 10) using the His-EBA181₉₄₅₋₁₀₉₇ elution buffer and loaded on a size exclusion chromatography column (*Superdex S200, GE Healthcare*[®]) to further increase its purity (Figure 3.12).

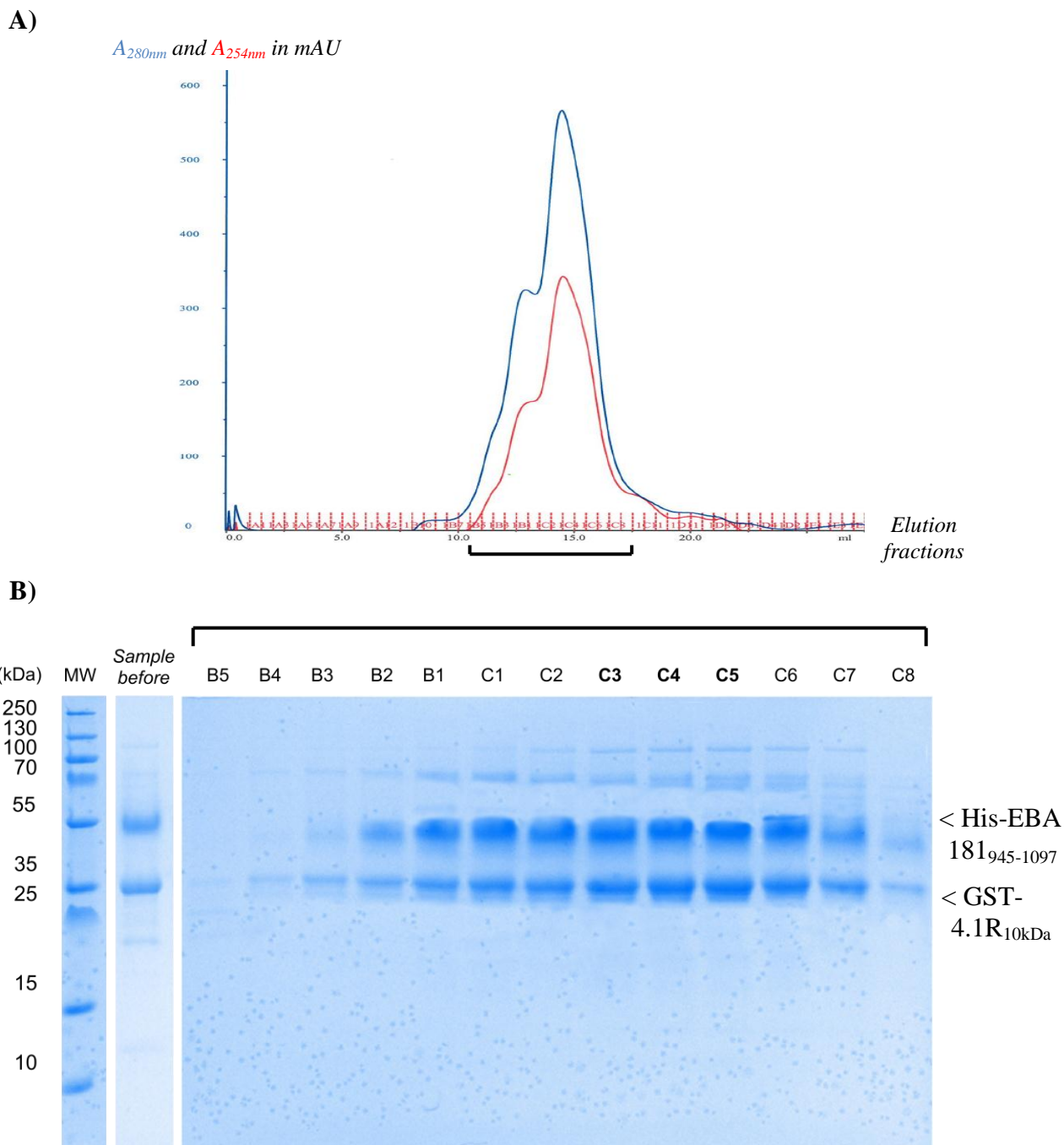


Figure 3.12: A: Size exclusion chromatography purification using *Superdex S200 10/300 GL* column, *GE Healthcare*[®] (blue curve: A_{280nm} in mAU, red curve: A_{254nm} in mAU) of the complex between His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} and **B:** Corresponding SDS-PAGE. The two black lines show the corresponding chromatography fractions on the SDS-PAGE.

3.4 Biochemical characterisation of individual EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins and the complex

3.4.1 Quantification of nucleic acid contamination

The purity of the isolated proteins was confirmed by SDS-PAGE gel. However, these protein samples might still be contaminated by nucleic acids through non-specific binding to the protein. In order to check if 4.1R_{10kDa} preparations were largely nucleic-acid free, the ratio protein/nucleic acids was determined by UV spectroscopy (Figure 3.13). For EBA181₉₄₅₋₁₀₉₇, it was not possible to measure this ratio since there is no aromatic residue in its sequence. Nevertheless, since EBA181₉₄₅₋₁₀₉₇ is highly negatively charged at pH between 6 and 8, as DNA or RNA are, binding is therefore, in this case, very unlikely to occur.

The ratio of absorptions at 260 nm vs. 280 nm of 4.1R_{10kDa} (Figure 3.13) is equal to 0.61, which means that the sample is composed of essentially 100% protein (Table 3.2, *Sambrook and Russell, 2001*), demonstrating the effectiveness of the overnight washing step during the GST purification which targeted to remove any nucleic-acid contamination.

| % Protein | % Nucleic acid | A_{260}/A_{280} |
|-----------|----------------|-------------------|
| 100 | 0 | 0.57 |
| 95 | 5 | 1.06 |
| 90 | 10 | 1.32 |
| 85 | 15 | 1.48 |
| 80 | 20 | 1.59 |
| 75 | 25 | 1.67 |
| 70 | 30 | 1.73 |
| 65 | 35 | 1.78 |
| 60 | 40 | 1.81 |
| 55 | 45 | 1.84 |

| % Protein | % Nucleic acid | A_{260}/A_{280} |
|-----------|----------------|-------------------|
| 50 | 50 | 1.87 |
| 45 | 55 | 1.89 |
| 40 | 60 | 1.91 |
| 35 | 65 | 1.93 |
| 30 | 70 | 1.94 |
| 25 | 75 | 1.95 |
| 20 | 80 | 1.97 |
| 15 | 85 | 1.98 |
| 10 | 90 | 1.98 |
| 5 | 95 | 1.99 |
| 0 | 100 | 2.00 |

Table 3.2: Determination of DNA/RNA contamination in a protein sample using the absorbance ratio $A_{260\text{nm}}$ (nucleic acids) / $A_{280\text{nm}}$ (protein, tryptophan residues mainly).

Source: *Sambrook et al., 2001.*

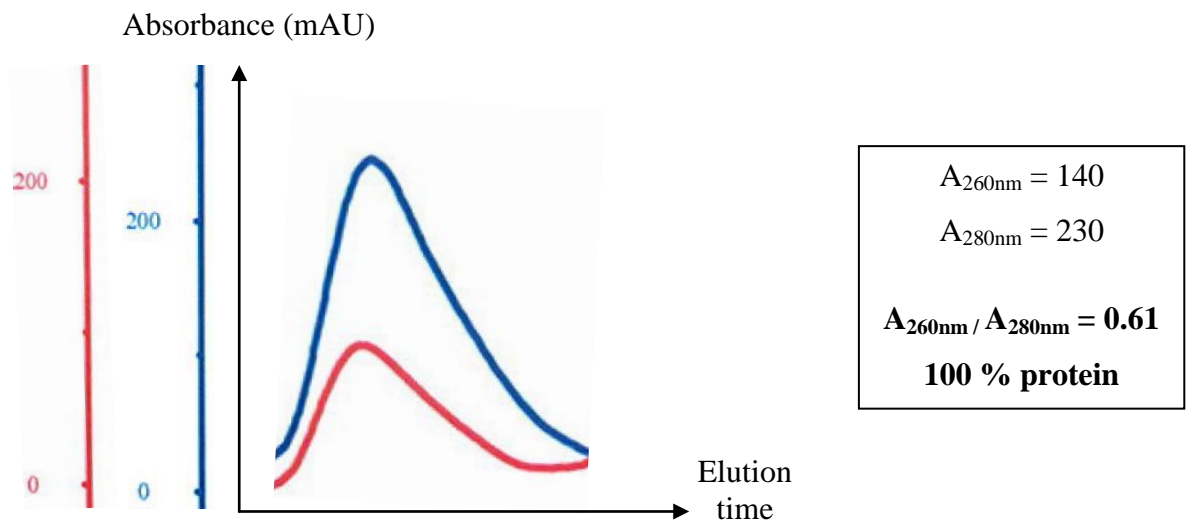


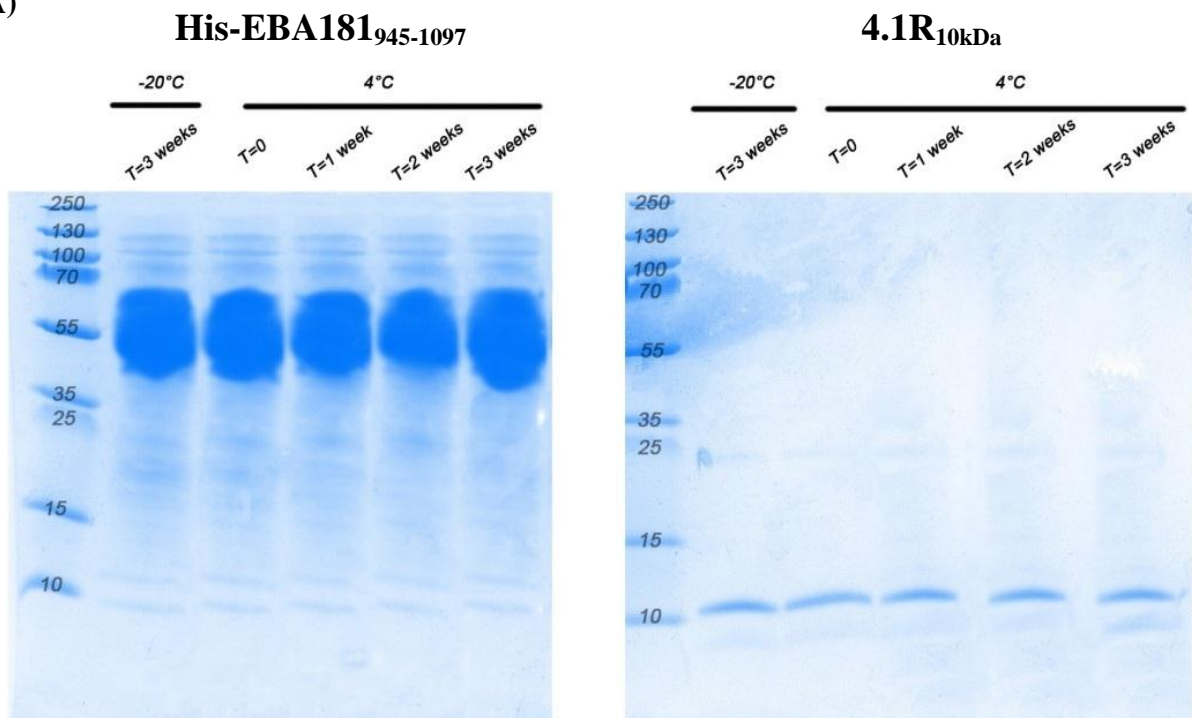
Figure 3.13: UV measurements of 4.1R_{10kDa} at 260 nm (red) and 280 nm (blue) during the size exclusion chromatography purification step.

3.4.2 Storage and degradation tests

To be able to store either the tagged His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} or the untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins, anti-proteases were added after final step of purification and concentration (1 mM PMSF final or commercial *Complete* from ROCHE®). Storage/degradation tests were performed. Each protein was stored at 4°C and aliquots were taken at T=0, after 1 week, 2 weeks and finally 3 weeks. These samples were mixed with gel loading buffer and heated at 100°C for 5 minutes to stop any degradation process. In parallel, a ‘freezing test’ was performed in order to test the feasibility to store the proteins at -20°C. Fresh samples were put at -20°C for three weeks, then defrosted to be analysed and allow the comparison with the 4°C test results (Figure 3.14). His-EBA181₉₄₅₋₁₀₉₇ samples were intentionally overloaded on the SDS-PAGE in order to be able to detect even weak degradation. It was not possible to load as much for 4.1R_{10kDa} because of the weak expression yield.

According to the SDS-PAGE (Figure 3.14, A), EBA181₉₄₅₋₁₀₉₇ can be stored at 4°C for at least three weeks without significant degradation. Moreover, the freezing process does not induce major proteolytic cleavage of the protein. Nevertheless it is impossible to see on a SDS-PAGE if freezing the protein has an impact on secondary structures. CD measurements were done as quality checks of the secondary structure on purified EBA181₉₄₅₋₁₀₉₇ before and after freezing (not shown). Both spectra were identical, indicating no significant change in secondary structure upon freeze/thaw cycles.

A)



B)

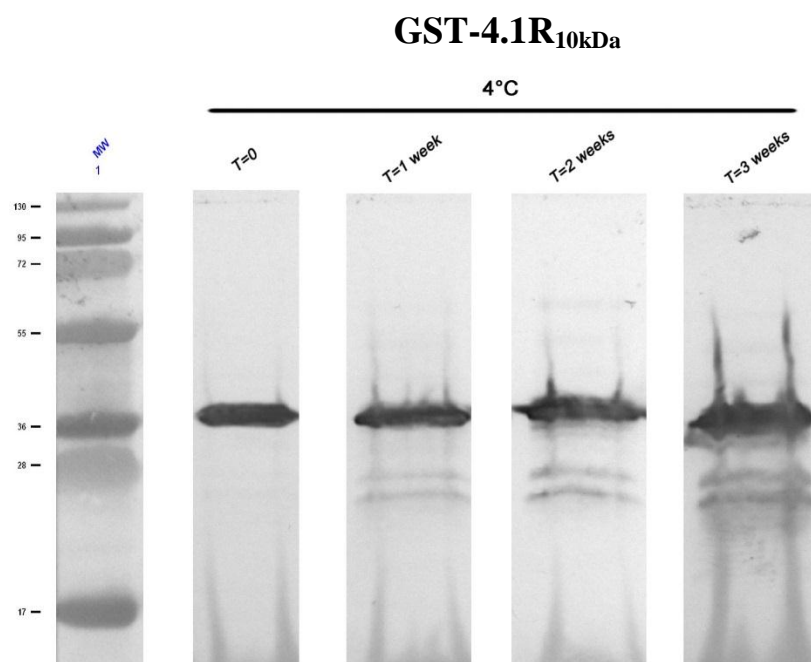


Figure 3.14: Storage and degradation tests.

A: SDS-PAGE of purified His-EBA181₉₄₅₋₁₀₉₇ (left) and 4.1R_{10kDa} (right) proteins.

B: Western blot anti GST tag on purified GST-4.1R_{10kDa} protein.

For 4.1R_{10kDa} the SDS-PAGE from the degradation tests have shown that the protein is stable for approximately two weeks at 4°C, with some degradation visible in three week old aliquots. As for EBA181₉₄₅₋₁₀₉₇, no difference is visible on gel for 4.1R_{10kDa} after three weeks at -20°C and being defrosted. Similarly to EBA181₉₄₅₋₁₀₉₇, CD measurements were performed on purified 4.1R_{10kDa} before and after freezing (not shown). The two spectra were also identical, showing no significant structural change upon a freeze-thaw cycle.

On the contrary, GST-4.1R_{10kDa} is not stable and rapidly degraded (few days), as noticed during the purification steps. This result shows the crucial importance of working at low temperature and the usage of anti-protease to minimise the activity of trace proteases.

3.4.3 Development of a 'micro pull-down' protocol

To study the complex and characterise the interaction between the EBA-181 and 4.1R proteins, a micro pull-down protocol was developed. The use of a 96 mini column system and a 96 well micro-plate permits the use of only a small amount of protein whilst testing different conditions in parallel, as well as using the same sample and protocol to allow a good and reliable comparison between each tested condition.

After testing different volumes of samples, resins and buffers (data not shown), the following optimised protocol was used for the pull-down characterisation studies (Figure 3.15). Firstly, 50 µl of the chosen resin (NiNTA or GST) was applied onto each mini column (the resin was retained by the filter, see photo of the receiver plate in Figure 3.15). Then, 500 µl of ddH₂O was added on top of the resin and drawn down into it

using a pipette in order to remove the ethanol storage solution of the resin. Depending upon the chosen characterisation test, 500 µl of the buffer of interest (see Appendix A.2 for buffer composition and 3.4.3 for the results and the different characterisation performed) were added on top of the resin to equilibrate the resin and again drawn down to the level of the top of the resin to remove any excess. In the meantime, the two freshly purified proteins were incubated together for 30 minutes in a tube on ice to form the complex. Starting from this step, the study was done using gravity flow only, in order to minimise the flow speed. 50 µl of the protein complex (concentration between 0.1 and 0.5 mg.ml⁻¹) was applied on top of the resin. Once the protein sample had passed through the resin, 50 µl of the protein buffer was added to compensate the volume of the resin and recover the entire flow-through (total 100 µl). To remove the non-specifically bound proteins and to test the complex under different conditions and stresses (pH, salt concentration, competition with peptides), the resin was washed in several steps with the chosen washing buffer (two resin volumes, 100 µl each time) until no further protein was detected in the washing fractions. Finally, the proteins bound to the resin were eluted with 100 µl of elution buffer. All the fractions were collected to be analysed on SDS-PAGE.

3.4.4 Inhibition assays

Two 30 amino-acid peptides were used for micro pull-down experiments as competitors of the EBA-181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex. The first peptide, called here Pf30, contains the amino acid sequence from EBA-181 binding to human 4.1R 10kDa domain (as found by phage display technique by our collaborators, *Lanzilotti et al., 2005*, see Figure 3.16).

Each different peptide was pre-incubated with the two proteins, with a 10 or 100 fold molar excess of peptide (on ice for 30 minutes). The previously described micro pull-down protocol was then used to perform the inhibition experiment.

3.4.5 Application of the micro pull-down technique

Further characterisation of the interaction between EBA-181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} was performed using the developed micro pull-down technique. The four possible constructs were used: tagged His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa}, and untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}. The following micro pull-down experiments were performed:

- Test of the complex formation in a tag-or-resin independent way,
- Negative controls to confirm the specificity of the tag-to-resin binding and the absence of false positives,
- High-salt (positive and negative charges competition),
- pH variation,
- Inhibition using competitor peptides.

3.4.5.1 Test of the complex formation in a tag-or-resin independent way

These first pull-down experiments involved the four different constructs, in order to confirm the formation of the complex independently of them being tagged/untagged and of the resin used (NiNTA or GST resin). When using the NiNTA resin, His-EBA181₉₄₅₋₁₀₉₇ is retained first. GST-4.1R_{10kDa} or 4.1R_{10kDa} can similarly only be retained by the column if held in complex with His-EBA181₉₄₅₋₁₀₉₇ (Figure 3.17, A1 & A2). In the case of the GST

resin, GST-4.1R_{10kDa} is retained first and His-EBA181₉₄₅₋₁₀₉₇ or EBA181₉₄₅₋₁₀₉₇ can only be retained if they bind to GST-4.1R_{10kDa} (Figure 3.17, B1 & B2).

To study the interaction between His-EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} the chosen buffer was 50 mM sodium phosphate at pH 6.0 with 50 mM NaCl and 5 mM βME. However, because the theoretical iso-electric point of GST-4.1R_{10kDa} is equal to 6.57, the buffer to study the interaction between EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} has to be sufficiently different in pH to avoid the protein to be not charged; PBS at pH 7.4 was therefore used, with 5 mM βME.

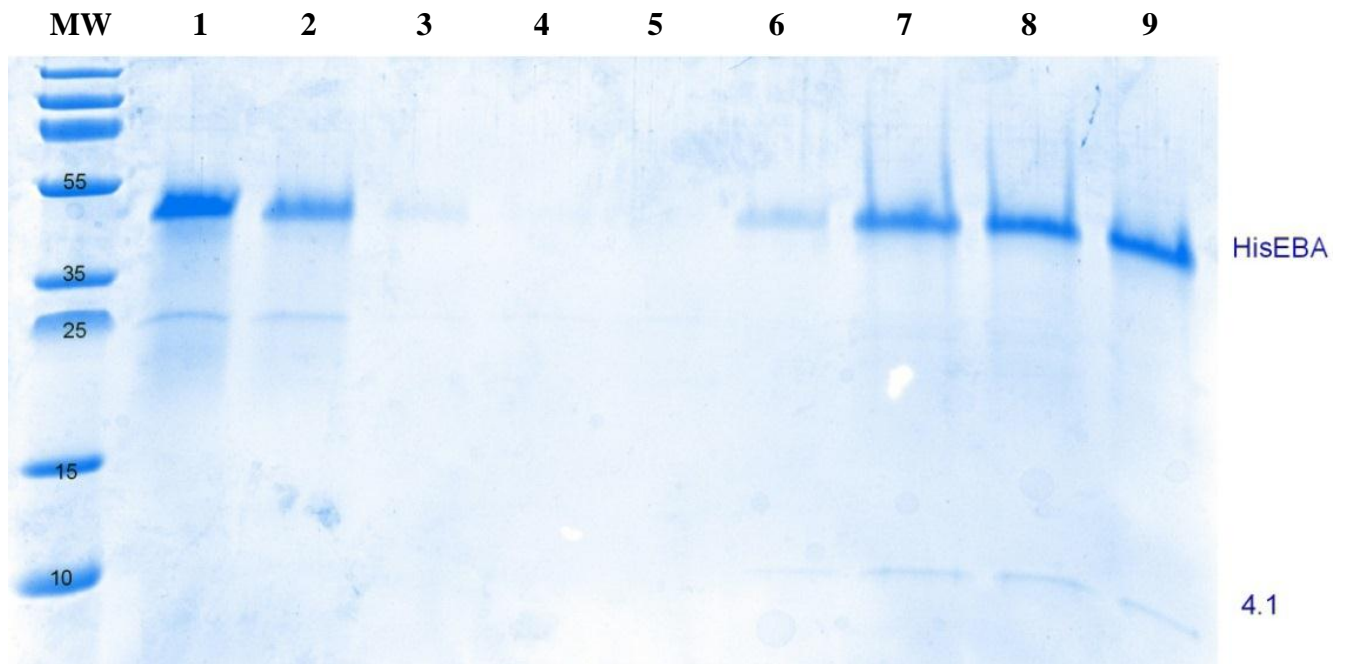
A1) For the pull-down experiment of 4.1R_{10kDa} by His-EBA181₉₄₅₋₁₀₉₇ using the NiNTA resin, some of the His-EBA181₉₄₅₋₁₀₉₇ was visible in the flow-through (lane 2) and the first washing step (lane 3), meaning not 100 % of the protein was bound to the resin. No 4.1R_{10kDa} was detectable in either flow-through or washing fractions. After ten column volumes of wash, no remaining protein was detectable (lane 5), allowing a careful analysis of the following elution fractions. Finally, His-EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} were co-eluted as a complex (lanes 8, 9 and 10).

A2) For the pull-down experiment of GST-4.1R_{10kDa} by His-EBA181₉₄₅₋₁₀₉₇ with the NiNTA resin, some of the His-EBA181₉₄₅₋₁₀₉₇ (not bound to the resin) and GST-4.1R_{10kDa} (not bound to the EBA protein) were visible in the flow-through (lane 2) and the two first washing steps (lanes 3 and 4). After washing with ten column volumes of washing buffer, no remaining protein was detectable (lane 5), allowing the analysis of the following elution fractions. Finally, His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} were co-eluted as a complex (lanes 6 to 10).

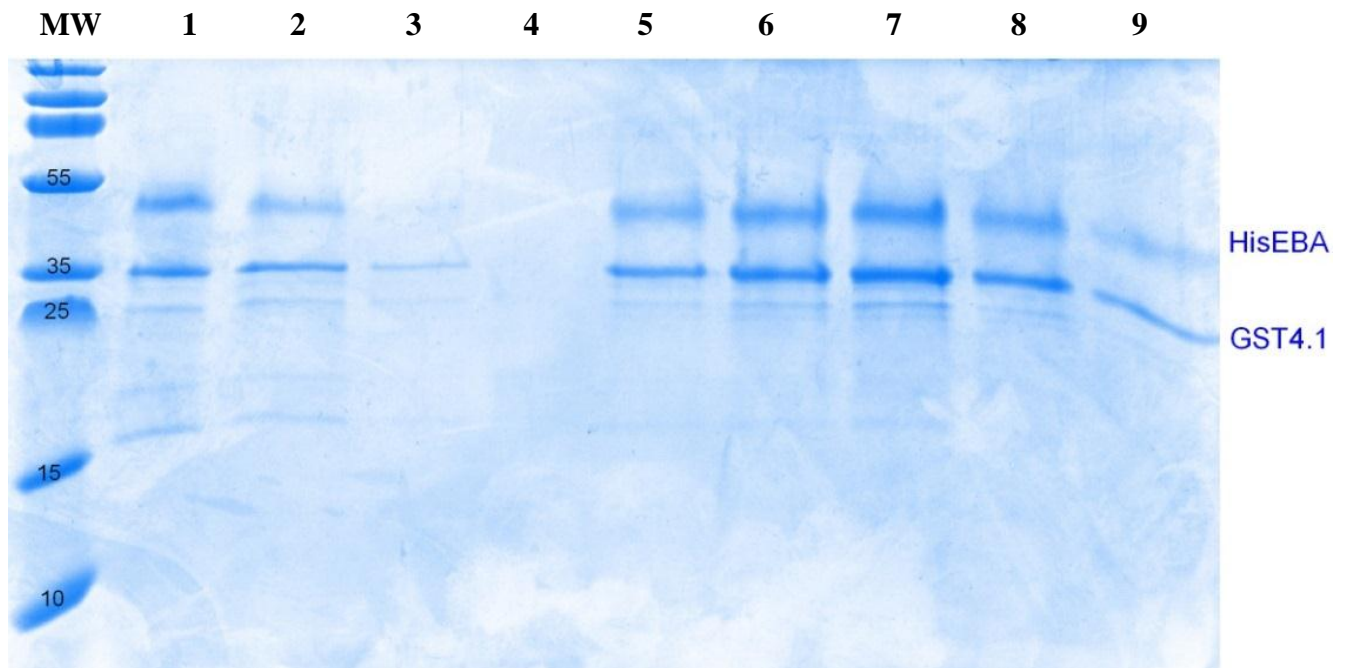
B1) For the pull-down experiment of His-EBA181₉₄₅₋₁₀₉₇ by GST-4.1R_{10kDa} using the GST resin, some of the His-EBA181₉₄₅₋₁₀₉₇ was visible in the flow-through (lane 2) and the first washing step (lane 3), so as a ‘free form’ not in complex with GST-4.1R_{10kDa}. No GST-4.1R_{10kDa} was detectable in either flow-through or washing fractions. After ten column volumes of wash, no remaining protein was detectable (lane 5). As previously, this allows the analysis of the following elution fractions. Finally, GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇ were co-eluted as a complex (lanes 6 to 9).

B2) For the pull-down experiment of untagged EBA181₉₄₅₋₁₀₉₇ by GST-4.1R_{10kDa} with the GST resin, EBA181₉₄₅₋₁₀₉₇ proteins were visible in the flow-through (lane 2) and the first washing step (lane 3) as a ‘free form’, like in the previous pull-down (B1). But no GST-4.1R_{10kDa} was detectable in either flow-through or washing fractions. No remaining protein was detectable after ten column volumes of wash (lane 5). Finally, GST-4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ were co-eluted as a complex (lanes 8 to 10).

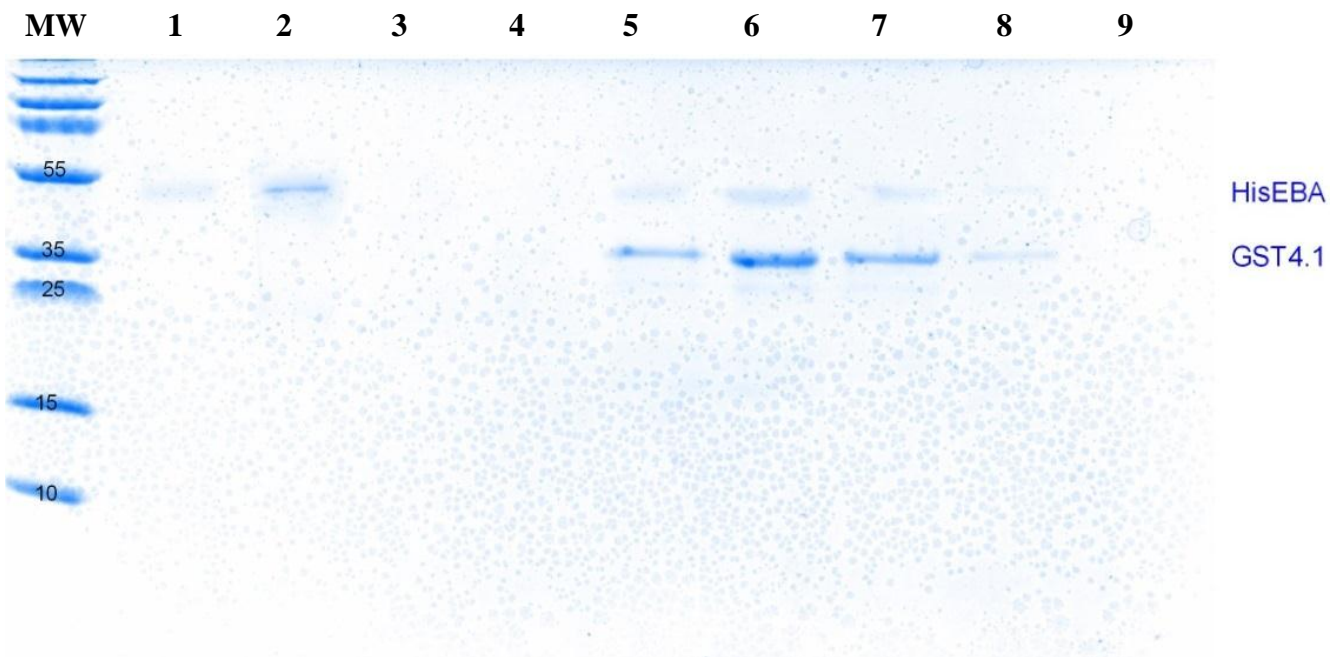
A1) NiNTA resin: Pull-down His-EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa}



A2) NiNTA resin: Pull-down His-EBA181₉₄₅₋₁₀₉₇ / GST-4.1R_{10kDa}



B1) GST resin: Pull-down GST-4.1R_{10kDa} / His-EBA181₉₄₅₋₁₀₉₇



B2) GST resin: Pull-down GST-4.1R_{10kDa} / EBA181₉₄₅₋₁₀₉₇

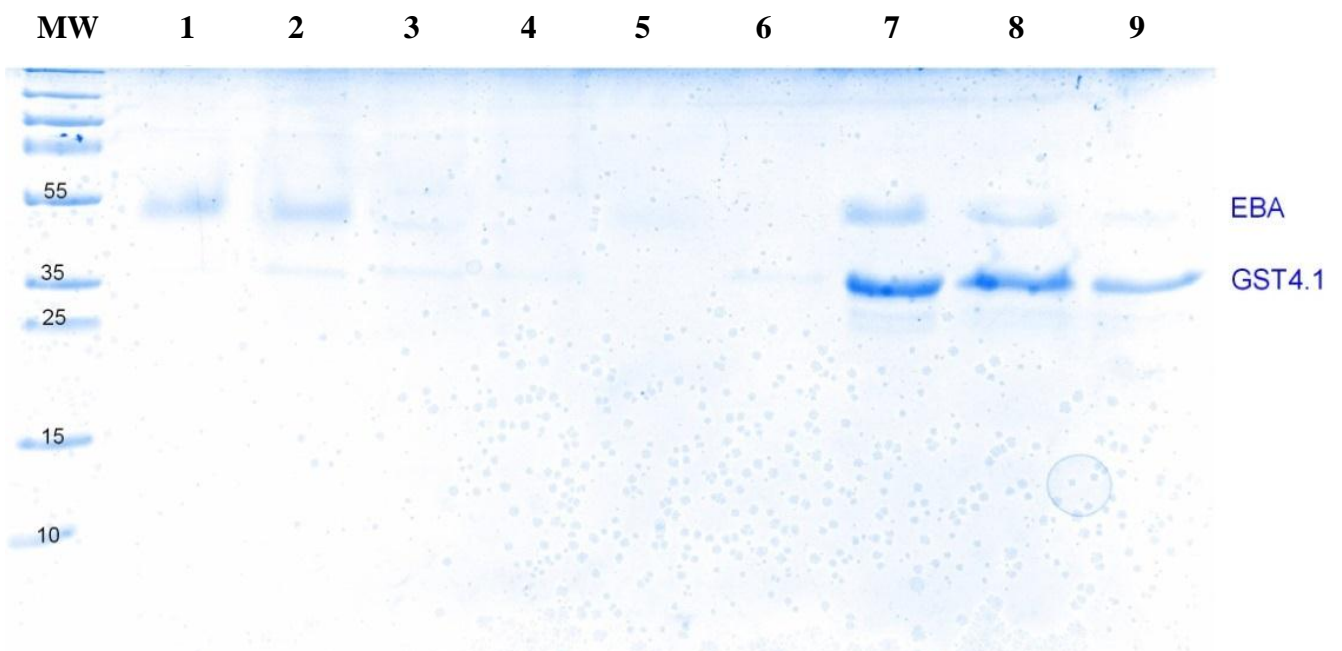


Figure 3.17: SDS-PAGE of the pull-down experiments with (A) NiNTA and (B) GST resins; using (A1) His-EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}, (A2) His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa}; or (B1) GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇, (B2) GST-4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇.

MW: Molecular marker. **1:** Flow-through. **2 to 4:** Washes. **5 to 9:** Elution fractions.

3.4.5.2 Negative controls

Negative controls were also performed using both GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇ with the NiNTA and GST resins, in order to confirm that neither His-EBA181₉₄₅₋₁₀₉₇ nor GST 4.1R_{10kDa} could bind non-specifically to the GST and NiNTA resins respectively (Figure 3.18).

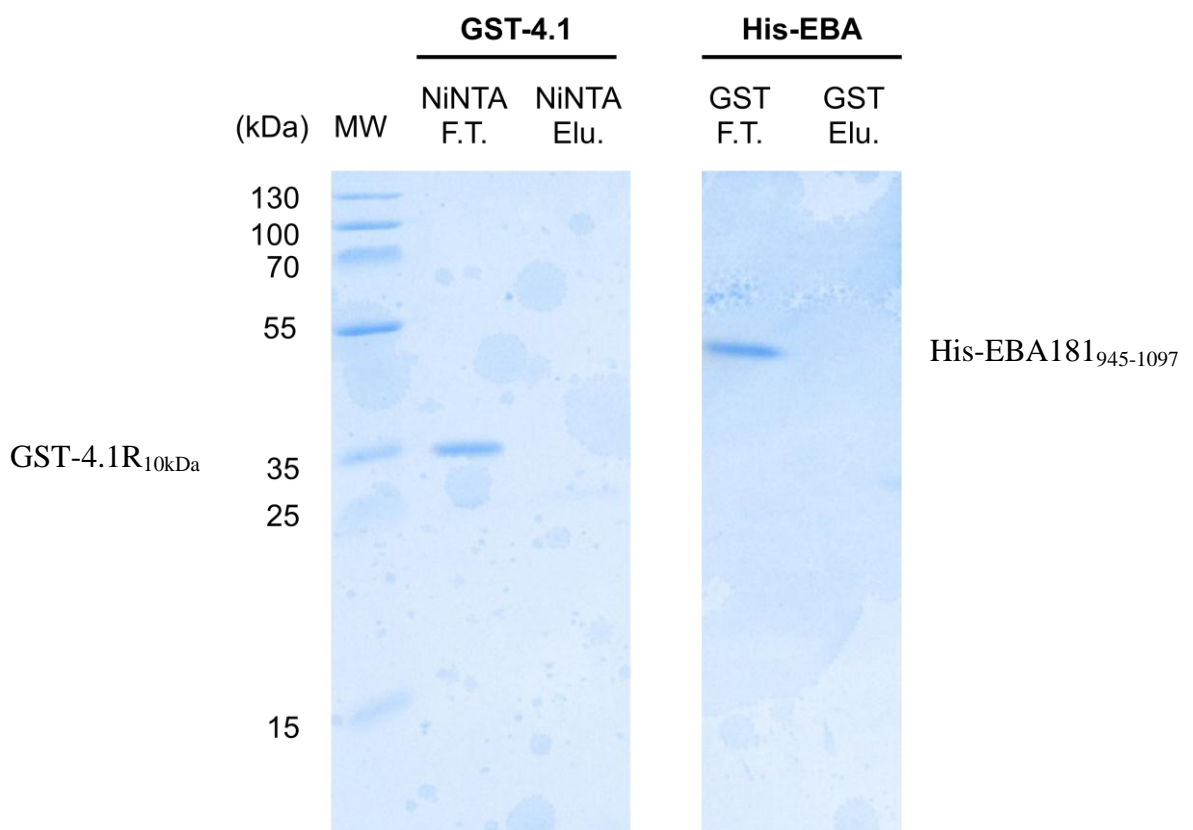


Figure 3.18: Left: SDS-PAGE of negative controls with GST-4.1R_{10kDa} on NiNTA and **right:** His-EBA181₉₄₅₋₁₀₉₇ on GST resin.

3.4.5.3 High salt characterisation

The following characterisation experiment was to apply two high salt stresses in order to test if the interaction is sensitive to charge competition given by the Na^+ and the Cl^- , and mimicking the cell local concentrations (Figure 3.19). The three NaCl concentrations used were 250 mM (lanes from 2 to 4), 1 M (lanes from 5 to 7) and 2 M (lanes from 8 to 10). As shown by the SDS-PAGE from the three different micro pull-down experiments (Figure 3.19), the complex is stable in competition with 250 mM or 1 M (elution fractions 4 and 7); and when the salt concentration was increased to 2 M NaCl, no complex dissociation was neither noticeable in both flow-through and washing fractions. The complex was therefore present in the elution fraction (fraction 10).

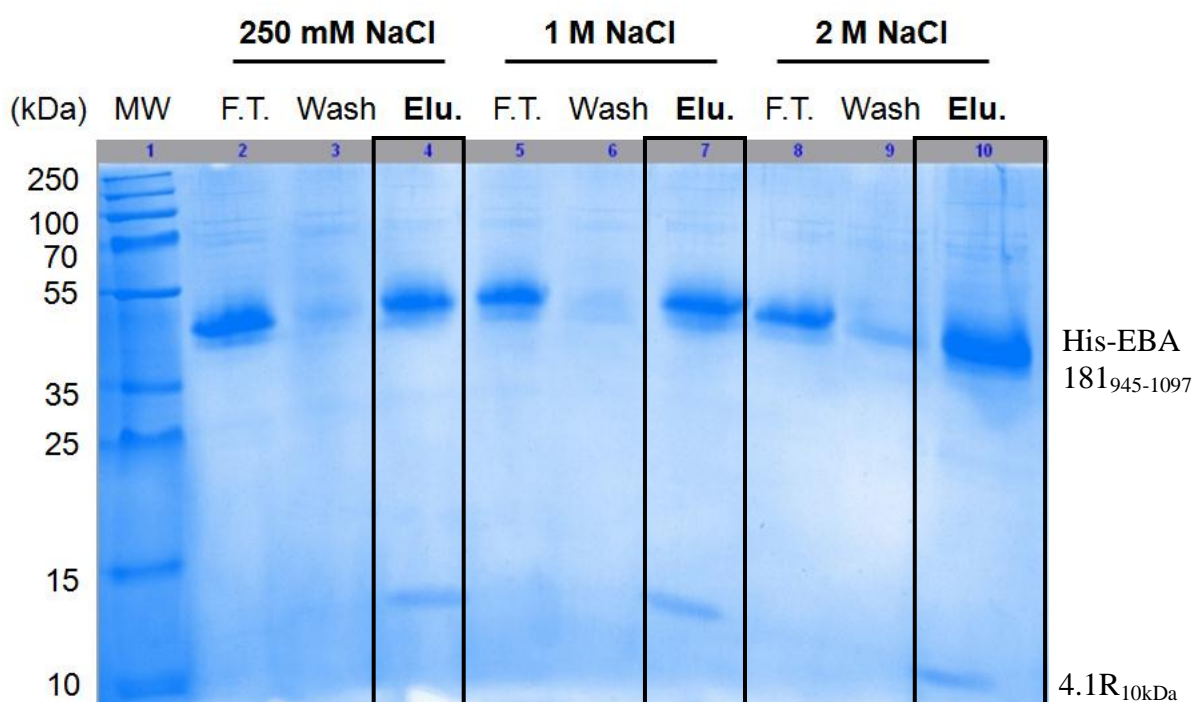


Figure 3.19: SDS-PAGE of pull-down experiment on His-EBA_{181⁹⁴⁵⁻¹⁰⁹⁷} and 4.1R_{10kDa} complex, resistance in high salt conditions. The rectangles show the eluted complex after each of the three different experiments.

‘F.T.’: flow through the NiNTA resin column.

‘Wash’: wash with the washing buffer containing: **3**: 250mM NaCl, **6**: 1M NaCl, **9**: 2M NaCl.

‘Elu.’: elution of the proteins with the elution buffer.

3.4.5.4 pH variation

The next characterisation was a pull-down test carried out to study the complex stability against pH, from pH 5.0 to 9.0 (Figure 3.20). The nickel resin used (NiNTA) has a pH binding capacity between 5.0 and 9.0; it is therefore compatible with the range of pH targeted. The micro pull-down protocol was used as described earlier (see 3.3.1). The complex between His-EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} was pre-formed on ice and incubated 30 minutes in each of the different pH solutions before being loaded onto the columns. Washing and elution steps were also performed in the corresponding pH values.

The complex between His-EBA and 4.1R was detected at each pH point tested (pH 5.0, 6.0, 7.0, 8.0 and 9.0; Figures 3.20 and 3.21), as shown on the SDS-PAGE elution lanes (rectangle). Nevertheless, it is noticeable that the complex seems to be slightly less stable at the extreme pH 9.0, since the flow-through fraction at pH 9.0 showed a slightly higher quantity of the 4.1R partner (compared to pH 8.0 on the same gel). These results provide insights into the robustness of the complex to these high negative or positive charges (given by the Na⁺ and Cl⁻ ions); and so upon local pH variations in cells.

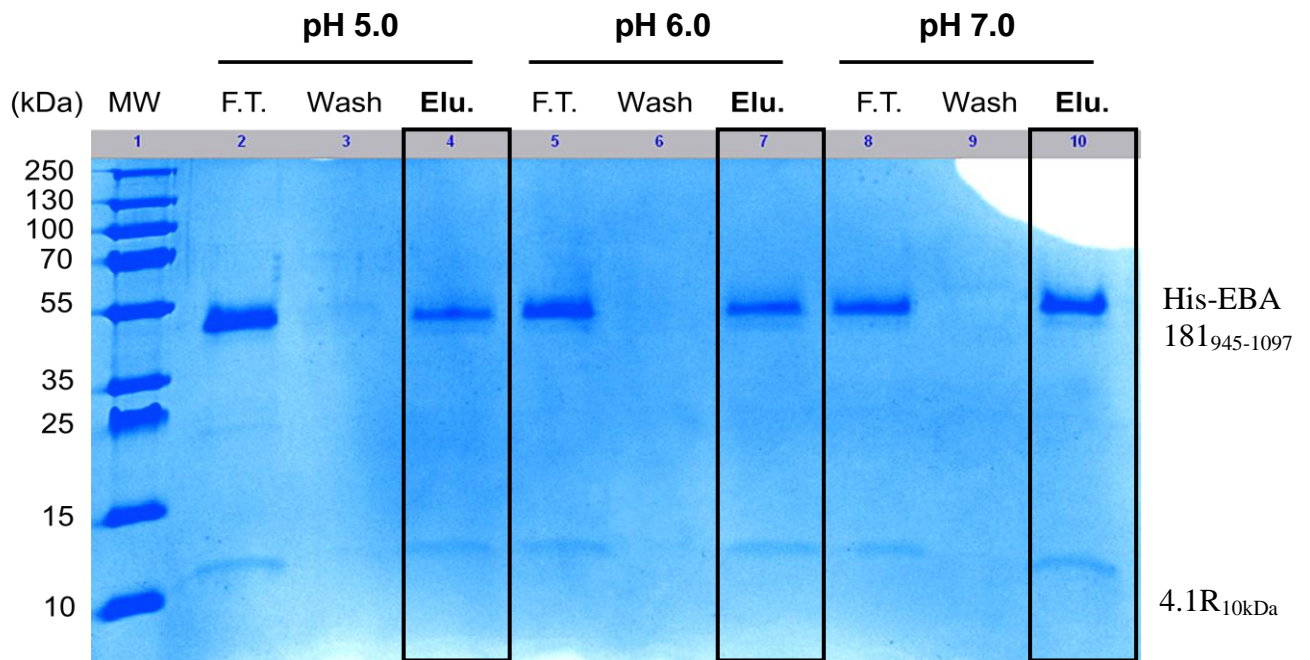


Figure 3.20: SDS-PAGE of pull-down experiments on the His-EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} complex, resistance of the complex to pH variations (pH from 5.0 to 7.0). The rectangles show the eluted complex at each different pH.

‘F.T.’: flow through the NiNTA resin column.

‘Wash’: wash with the washing buffer.

‘Elu.’: elution of the proteins with the elution buffer.

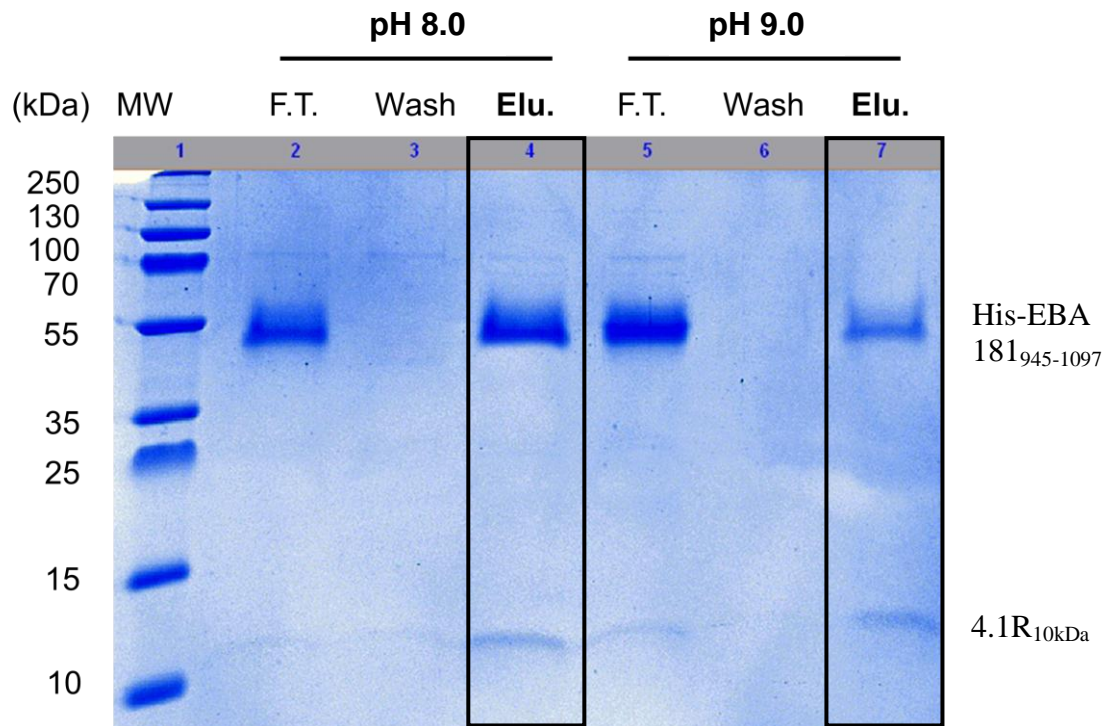


Figure 3.21: SDS-PAGE of pull-down experiments on the His-EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} complex, resistance of the complex to pH variations (pH 8.0 and 9.0). The rectangles show the eluted complex at each different pH.

‘FT’: flow through the NiNTA resin column.

‘Wash’: wash with the washing buffer.

‘Elution’: elution of the proteins with the elution buffer.

3.4.5.5 Binding competition assays

The last pull-down experiment was to perform a binding competition assay by the use of two different 30 amino-acid peptides (Pf30 and Pr30) with the complex. Pf30 contains the EBA binding sequence to 4.1R, so is therefore likely to bind to 4.1R and will act as a competitor to EBA. On the other hand, Pr30 contains the homologue sequence of Pf30 but from *Plasmodium*

reichenowi. As this parasite does not infect humans, the Pr30 peptide might not bind to 4.1R. Pr30 was therefore used for comparison and as a likely negative control. In order to facilitate and improve the protein detection on gel, only tagged proteins were used for this pull-down experiment (His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa}, Figure 3.22).

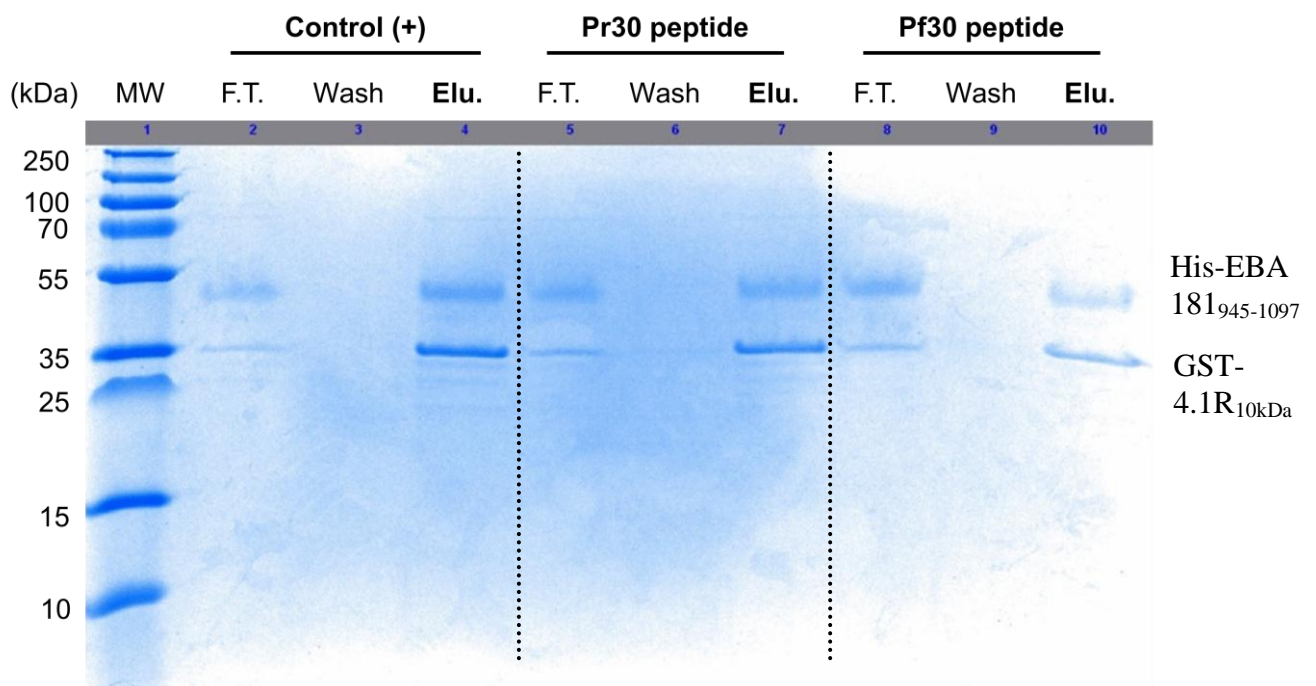


Figure 3.22: SDS-PAGE of pull-down experiments on His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} complex, binding competition by the two Pf30 and Pr30 peptides.

‘F.T.’: flow through the NiNTA resin column.

‘Wash’: final washing fraction.

‘Elu.’: elution of the proteins with the elution buffer.

‘+’: positive control, pull-down without any peptide.

‘Pr30 peptide’: complex binding competition with Pr30 peptide (100 fold excess).

‘Pf30 peptide’: complex binding competition with Pf30 peptide (100 fold excess).

The same competition pull-down experiment was performed twice, with 10 times (data not shown) or 100 times molar excess peptide (Figure 3.22). In both cases there was no inhibition of complex formation. Neither of the 30 amino-acid peptides disrupted the interaction between EBA and 4.1, even with a 100 fold excess. This suggests that the whole EBA181₉₄₅₋₁₀₉₇ sequence is important for binding to 4.1R. Thus, even if the 30 amino-acids are sufficient to bind to 4.1R, the rest of the sequence might increase the affinity and/or help to stabilise the complex.

3.4.6 Native polyacrylamide gel electrophoresis

Native polyacrylamide gels were used to analyse and demonstrate the complex formation, as well as to detect any charge and/or conformation changes of the two proteins upon binding. Native gel electrophoresis does not use a charged denaturing agent; therefore the molecules being separated differ not only in molecular mass and intrinsic charge, but also in their cross-sectional area, and thus experience different electrophoretic forces dependent on the shape of the overall structure. These analyses were carried out using 5 µg of each purified protein and gels were run during 4 hours at 15 mA and at 8°C (Figure 3.23).

Native gel electrophoresis of 4.1R_{10kDa} showed that 4.1R_{10kDa} did not enter the gel: since the protein was almost uncharged at the required pH used (pH=8.9, theoretical iso-electric point of 4.1R_{10kDa} = 8.25, 13 residues with negatively charged side chain and 14 residues with positively charged side chain, see Appendix B.4 for more details on the sequence properties). This result was used to compare the change of overall charges when bound to EBA181.

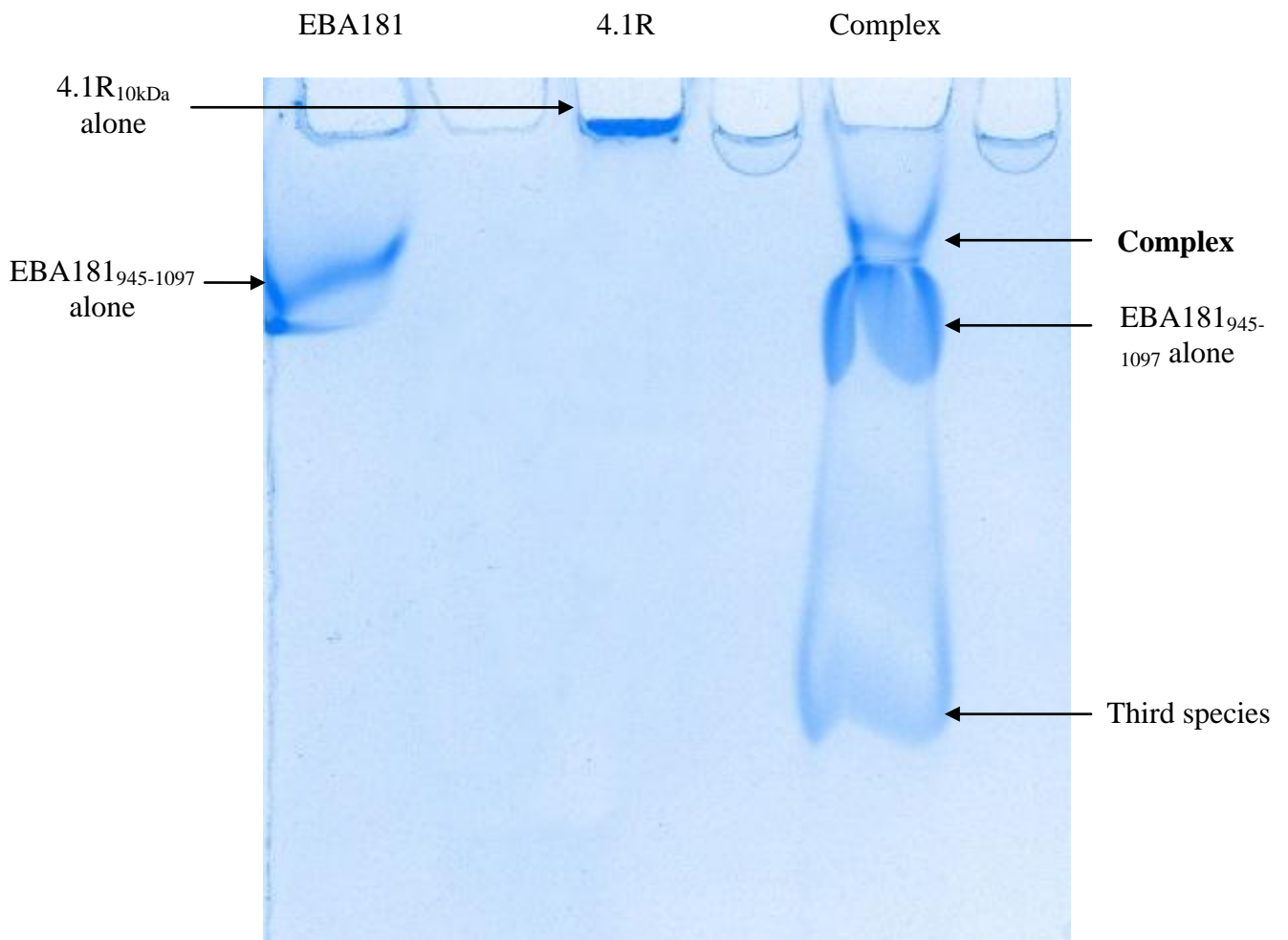


Figure 3.23: Native polyacrylamide gel of EBA181₉₄₅₋₁₀₉₇, 4.1R_{10kDa} and the complex.

In contrast to 4.1R_{10kDa}, electrophoresis of EBA181₉₄₅₋₁₀₉₇ showed that the protein entered in the gel (theoretical iso-electric point of EBA181₉₄₅₋₁₀₉₇ = 4.20, 37 negatively-charged side chain residues for only 13 positively-charged, see Appendix B.3). The migration pattern showed a single band, reflecting the presence of a single species. Nevertheless, probably due to the disordered character and its very negatively charged property, the migration pattern of the band was smeared.

When 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ were in complex, the electrophoresis migration pattern was different. The result was also another smeary migration pattern, with the appearance of two new bands in addition to the 'EBA181₉₄₅₋₁₀₉₇ alone' band: one above it, and one which migrated further down, close to the end of the gel. This is evidence for a change in apparent charges upon binding of the two proteins. A further analysis of this first native gel electrophoresis permitted the migration band of the complex to be identified. Thus, since 4.1R_{10kDa} alone (not in complex) is not charged at the pH used here, it might decrease the average negative charges carried by EBA181₉₄₅₋₁₀₉₇ when both proteins are in complex, and so slowing its migration in the gel. Moreover, the weight of the complex is expected to be higher than the individual proteins. This intrinsic property will also lead to a slower migration, and so to a higher band on the gel. Based on these assumptions, the higher migration band was assigned to the complex (Figure 3.23, lane 'complex', band 'complex'). In addition to the migration band of EBA181₉₄₅₋₁₀₉₇ alone, the third band showed the presence of a third species, which is hard to interpret. The migration is further along the gel, reflecting the presence of a small species, or an even more negatively charged than the EBA alone migration band.

In order to cross-check these results, a further native gel electrophoresis experiment was performed. In this, the complex was loaded on a native gel at different times during a 6 hour migration (Figure 3.24). Here, the complex was carefully pre-formed in a tube at a 1 to 1 molar ratio. In this case and as shown by the native gel, the 'EBA181₉₄₅₋₁₀₉₇ alone' and '4.1R_{10kDa} alone' bands are almost invisible (Figure 3.24, lanes 'complex'), showing that almost all the proteins are effectively in complex. The results are similar to the first native gel, showing the

appearance of a higher band upon binding of the two proteins; as another evidence of the formation of the EBA/4.1R complex.

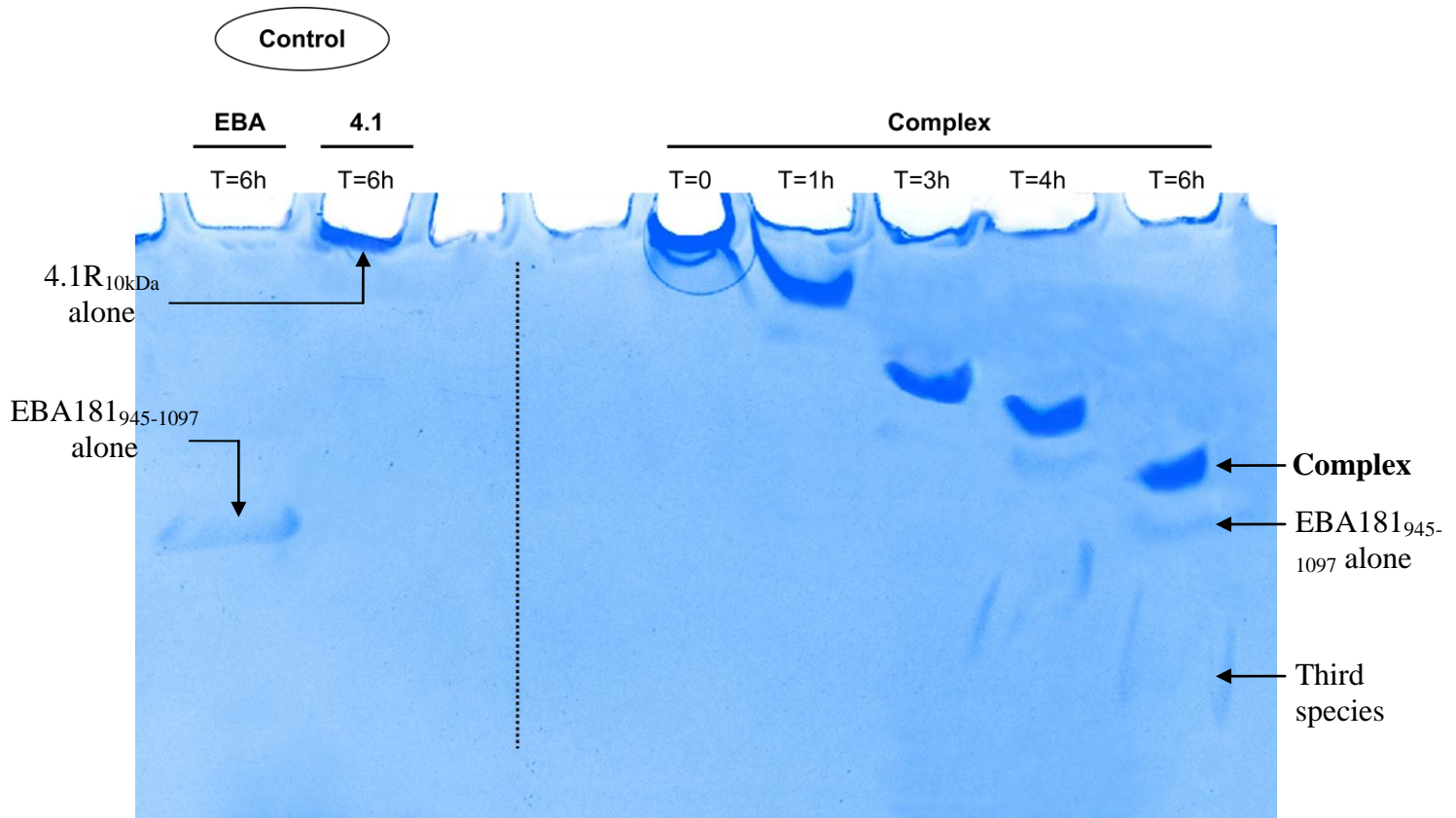


Figure 3.24: Native polyacrylamide gel of: EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} alone, migrated for 6 hours (control); and the complex migrated for 0h, 1h, 3h, 4h and 6h.

3.4.7 Protein quantification using SDS-PAGE

Quantification of protein bands on SDS-PAGE was made to allow an estimation of the molar ratio of EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} in the protein complex.

Firstly, the staining power of Coomassie blue was measured on each of the tagged and untagged proteins (Figure 3.25) to give a baseline/calibration for the further analysis of gels run on the complex. For this, respectively 2 μg and 5 μg of EBA181₉₄₅₋₁₀₉₇, His-EBA181₉₄₅₋₁₀₉₇, 4.1R_{10kDa} and GST-4.1R_{10kDa} were loaded onto a SDS-PAGE and the luminosity of each band was measured with the following protocol. After scanning the gel (1200 dpi resolution TIFF file, Figure 3.25), the resulting image was colour inverted with Photoshop software. A box covering the biggest band (in this case lane GST-4.1R_{10kDa}, 5 μg) was drawn using the Photoshop marquee tool. The same sized box was then used for all the other bands. Using the Photoshop histogram tool, the total luminosity was measured for each individual band. The background value was obtained in the same way, using the same sized box but on a 'blank' area of the gel containing no protein bands. After this, for each individual luminosity value the corresponding background value was subtracted. It was decided to only keep the values from the 5 μg lanes (lane 'Luminosity' in Tables 3.3, 3.4 and 3.5) since those from the 2 μg samples were weaker and at the limit of the Coomassie binding capacity. The GST-4.1R_{10kDa} value was set to 1 as the reference; therefore all other protein values were divided by the GST-4.1R_{10kDa} value (lane 'Ratio of luminosity normalised to GST-4.1R_{10kDa}' in Tables 3.3, 3.4 and 3.5). These divided values were then converted according to the ratio GST-4.1R_{10kDa} / each protein weights, in order to have the molar ratio.

The final results from the Coomassie staining quantification are listed in Table 3.3. Figures for 4.1R_{10kDa} could not be extracted since the protein was poorly stained. Thus GST-4.1R_{10kDa} was used and analysed afterward.

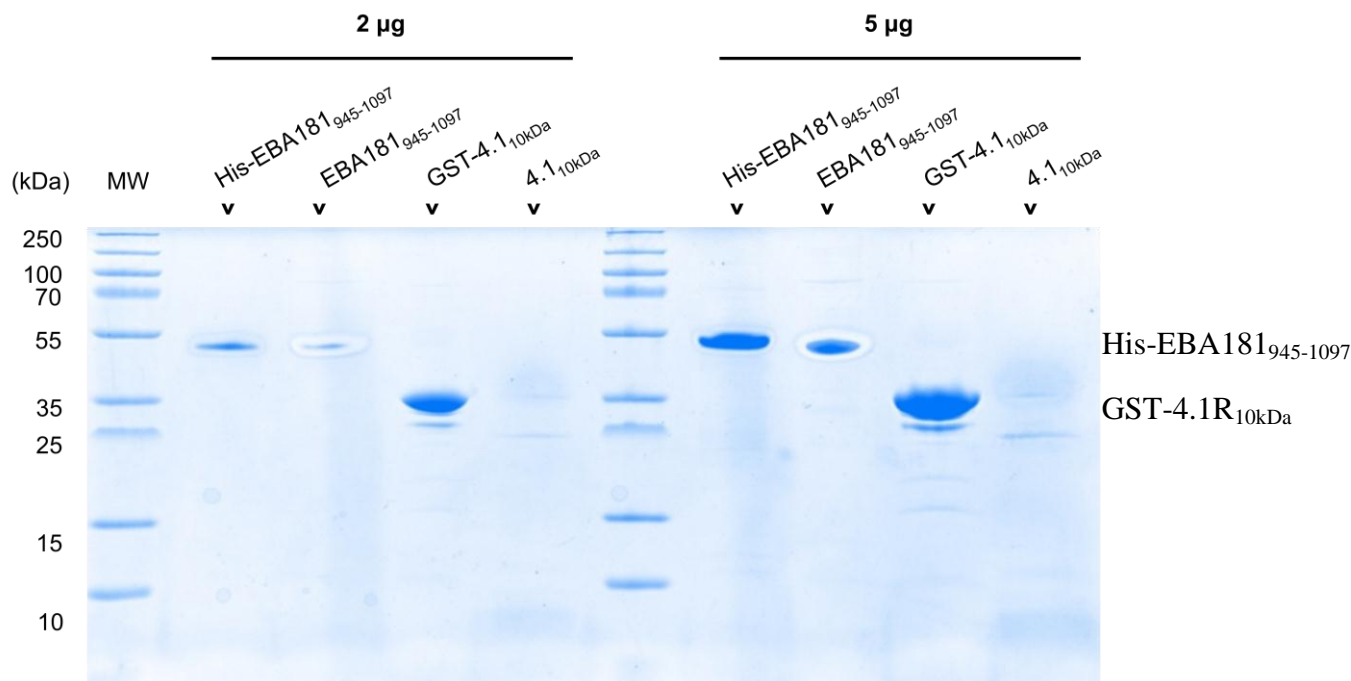


Figure 3.25: Original SDS-PAGE gel used to quantify the Coomassie blue staining of tagged and untagged 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ proteins. **Left:** 2 µg of each protein. **Right:** 5 µg of each protein.

Following this first calibration step, the same method was then applied to the pull-down gels (see Section 3.4.3.1 for the full size gels) in order to determine the molar ratio of the complex. Since SDS-PAGE gels of the eluted pull-down fractions displayed several bands for each protein, the only difference with the protocol previously described and used for calibration was the calculation of the ‘mean luminosity’ coming from each single band luminosity, by averaging each individual band value (and subtraction of the background). Since the 4.1R_{10kDa} staining was not sufficient, it was chosen to only analyse the gels with tagged GST-4.1R_{10kDa}. Both GST (Table 3.4) and NiNTA (Table 3.5) resins were used to cross-check the final result.

| Protein | Calibration gel | | |
|---|---------------------------|--------------------------------|----------------------------|
| | GST-4.1R _{10kDa} | His-EBA181 ₉₄₅₋₁₀₉₇ | EBA181 ₉₄₅₋₁₀₉₇ |
| Luminosity (<i>Photoshop</i>) 5 µg | 96 | 61 | 35 |
| Ratio of luminosity (<i>luminosity normalised to GST-4.1R_{10kDa}</i>) | 1 (=96/96) | 0.64 (=61/96) | 0.36 (=35/96) |
| Molar ratio (<i>weight corrected to GST-4.1R_{10kDa}</i>) | 1 (36 kDa) | 0.32 (18 kDa) | 0.18 (17 kDa) |

Table 3.3: Coomassie blue staining capability of His-EBA181₉₄₅₋₁₀₉₇, EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} from the calibration gel (5 µg).

3.4.7.1 GST resin pull-down gels

The first stoichiometry calculation was performed on the elution fractions coming from the GST resin pull-downs (Section 3.4.3.1; B1 and B2).

The data in Table 3.4 were extracted from the pull-down gels in Figure 3.26. After measurement of the mean luminosity and subtraction of the background, each mean luminosity value was normalised to the GST-4.1R_{10kDa} one (as previously described). Finally, the stoichiometry is defined by applying the coefficients from the calibration gel.

For both tagged and untagged EBA181₉₄₅₋₁₀₉₇ the final result was similar and gave a stoichiometry of close to 1 to 1 when binding to GST-4.1R_{10kDa}.

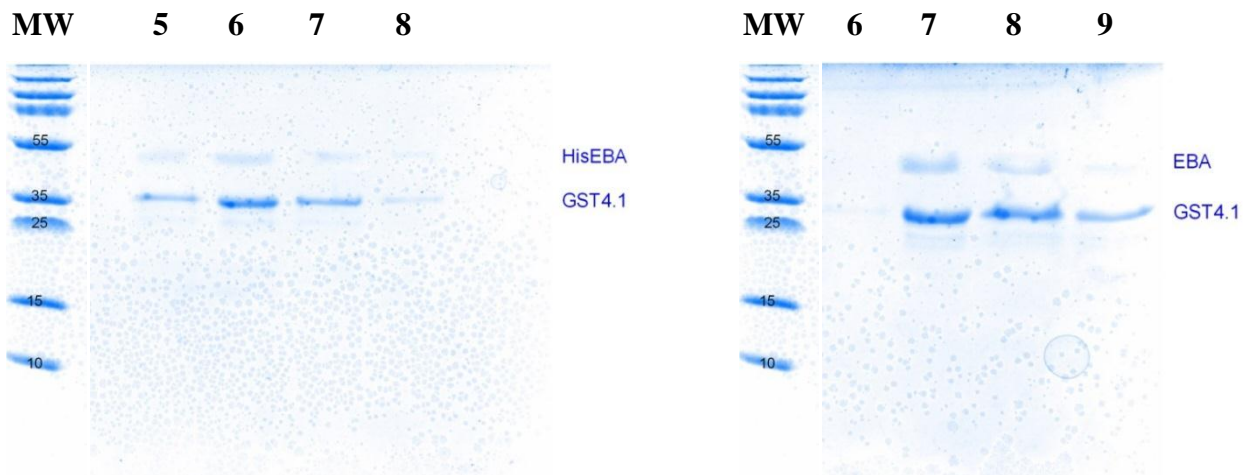


Figure 3.26: SDS-PAGE from elution fractions from GST resin pull-down used to determine the $4.1R_{10kDa}$ / $EBA181_{945-1097}$ protein ratio when in complex.

Left: $GST-4.1R_{10kDa}$ in complex with tagged His- $EBA181_{945-1097}$. **Right:** $GST-4.1R_{10kDa}$ in complex with untagged $EBA181_{945-1097}$.

MW: Molecular weight markers.

5 to 9: Elution fractions.

| Protein | GST resin pull-down gels | | |
|--|--------------------------|--------------------------------|--------------------------------|
| | $GST-4.1R_{10kDa}$ | His- $EBA181_{945-1097}$ | $EBA181_{945-1097}$ |
| Mean luminosity | 81 | 29 | 16 |
| Ratio of luminosity (<i>luminosity normalised to $GST-4.1R_{10kDa}$</i>) | 1 (=81/81) | 0.36 (=29/81) | 0.20 (=16/81) |
| Stoichiometry (<i>molar ratio calculated from calibration gel</i>) | / | 1 : 1.12 (0.32/0.36) | 1 : 1.06 (0.18/0.20) |

Table 3.4: Luminosity and ratio calculation on the SDS-PAGE from GST resin pull-downs of the $GST-4.1R_{10kDa}$ / His- $EBA181_{945-1097}$ complex and $GST-4.1R_{10kDa}$ / $EBA181_{945-1097}$ complex.

3.4.7.2 NiNTA resin pull-down gel

The second and last stoichiometry calculation was performed on the elution fractions coming from the NiNTA resin pull-downs (Section 3.4.3.1; A2).

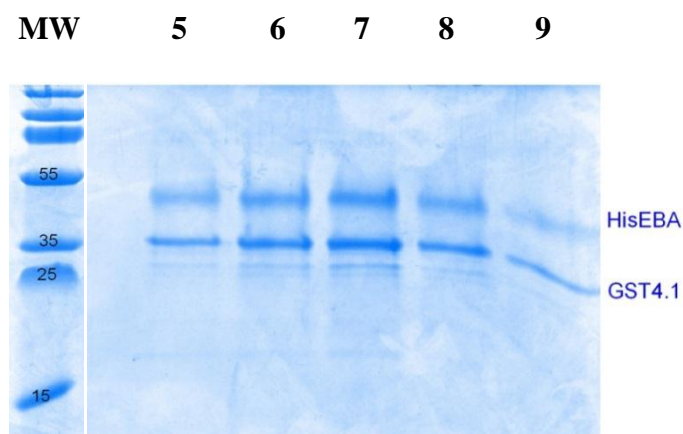


Figure 3.27: SDS-PAGE from elution fractions from NiNTA resin pull-down used to determine the $4.1R_{10kDa}$ / $EBA181_{945-1097}$ protein ratio when in complex. Tagged His-EBA181₉₄₅₋₁₀₉₇ in complex with GST-4.1R_{10kDa}.

MW: Molecular weight markers. **5 to 9:** Elution fractions.

| Protein | NiNTA resin pull-down gel | |
|---|---------------------------|--------------------------------|
| | GST-4.1R _{10kDa} | His-EBA181 ₉₄₅₋₁₀₉₇ |
| Mean luminosity | 122 | 43 |
| Ratio of luminosity (<i>luminosity normalised to GST-4.1R_{10kDa}</i>) | 1 | 0.35 |
| Stoichiometry (<i>molar ratio calculated from calibration gel</i>) | / | 1 : 1.1 |

Table 3.5: Luminosity and ratio calculation on SDS-PAGE from NiNTA resin pull-downs of the His-EBA181₉₄₅₋₁₀₉₇ / GST-4.1R_{10kDa} complex.

The data in Table 3.5 were extracted from the pull-down gels in Figure 3.27. The final result was similar to the previous result using GST resin and gave a close to 1 to 1 stoichiometry for the GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇ binding.

In combining both GST and NiNTA resin pull-downs, it was possible to determine that the interaction between EBA-181 and 4.1R has a 1 to 1 stoichiometry.

This relative coarse method was the first attempt to determine the binding stoichiometry of untagged EBA-181 and 4.1R proteins and was used to complement the set of pull-down characterisation. This method is indirect and requires preliminary calibration steps, which could lead to loss of accuracy and accumulation of errors coming from the different measurements (errors coming from the initial pull-down experiments, the SDS-PAGE, the scan of the gels, the Photoshop software for measuring each different luminosity, the calibration process,...). In order to confirm this 1 to 1 binding stoichiometry, SEC-MALLS-RI experiments were performed (see Section 3.5.5).

3.5 Biophysical characterisation of individual EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins and the complex

3.5.1 Structural bioinformatics analysis

Bioinformatics analysis is a powerful tool for the prediction and analysis of the structure of proteins and nucleic acids. The specific rules of protein folding are poorly understood, and thus bioinformatics predictions need to be interpreted with care. Nonetheless, as part of an attempt to provide additional information in the context of biochemical and biophysical results, and to provide insight for preparing effective constructs and crystallisation strategies, secondary structure and disorder predictions were performed for the two entire 4.1R and EBA-181 proteins as-well-as for both the 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ expressed domains. Since it is difficult to estimate with certitude the quality of the predictions, it was thought important to deploy different bioinformatics approaches.

Secondary structure predictions were carried out using the web-based PSIPRED and I-TASSER servers (*Ambrish et al., 2010*). PSIPRED gave an overview of the protein secondary structure, while I-TASSER provided models with tertiary structure. The I-TASSER server provides a confidence score for estimating the quality of predicted models called the C-score. The C-score is typically in the range of [-5, 2], with a C-score of higher value signifying a model with a high confidence. The I-TASSER server also calculates a ‘TM-score’ and RMSD of the predicted models by comparison to the native structures based on the C-score. The TM-score weights smaller RMSD pair distances greater than longer ones to avoid local structural differences in a small part of the

prediction from overly biasing a standard RMSD calculation. A TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means a random similarity.

In addition to the structural predictions, disorder predictions were carried out for the proteins, using the two web-based server programs DISOPRED (*Ward et al., 2004*) and IUPRED (*Dosztányi et al, 2005*). Linked to IUPRED, potential binding sequence in the disordered regions were mapped with the ANCHOR prediction server program. Since various disordered proteins function via binding to a structured partner and undergo a disorder-to-order transition, the ANCHOR server predicts disordered domains which cannot form enough favourable intra-chain interactions to fold on their own but can be stabilised to a more ordered structure by interacting with a globular protein partner.

This section is divided into four parts:

- Modelling of full-length EBA-181.
- Modelling of the EBA181₉₄₅₋₁₀₉₇ domain.
- Modelling of full-length 4.1R.
- Modelling of the 4.1R_{10kDa} domain.

Each of these parts is then subdivided into (1) the results of the secondary and tertiary structure prediction results from PSIPRED and I-TASSER, and (2) the disorder predictions from DISOPRED, IUPRED and potential binding sequences from ANCHOR.

3.5.1.1 Secondary structure and disorder predictions of full-length EBA-181

According to PSIPRED and I-TASSER, the EBA-181 protein seems from a structural point of view to be clearly divided into three well-defined domains. The secondary structure map of full-length EBA-181 by PSIPRED is presented in Figure 3.29 and the overall three-dimensional structure prediction of I-TASSER is shown in Figure 3.30.

The first half of the sequence from the N-terminus to approximately residue 900 is predicted to be structurally folded. This prediction is largely based on the existing crystal structure of the homologue EBA-175 RII domain (see Section 2.1.1) which is composed of only alpha-helices (Figure 3.29, residues in pink colour). This domain contains the signal peptide required for correct microneme trafficking as well as the two F1 and F2 Duffy binding-like (DBL) domains (Figure 3.28), known to play a crucial role in the binding of merozoites to the host cell and in defining erythrocyte specificity.

Following this structured domain, the second domain is predicted to be disordered, spanning approximately residues 900 to 1350, as shown in Figure 3.29. This part includes the 4.1R_{10kDa} binding sequence (971-1001).

Finally, the C-terminus is predicted to be largely folded as alpha-helices. The C-terminus end is known to host the cysteine-rich motif essential for correct microneme trafficking and a transmembrane domain (Figure 3.28); this could explain the required fold of the C-terminus region.

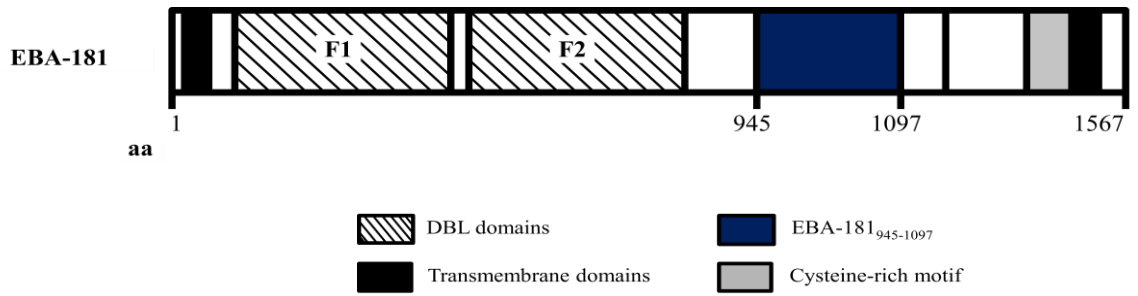


Figure 3.28: Schematic of the domain organisation of the full-length *P. falciparum* invasion EBA-181 protein (accession number: PFA0125c).

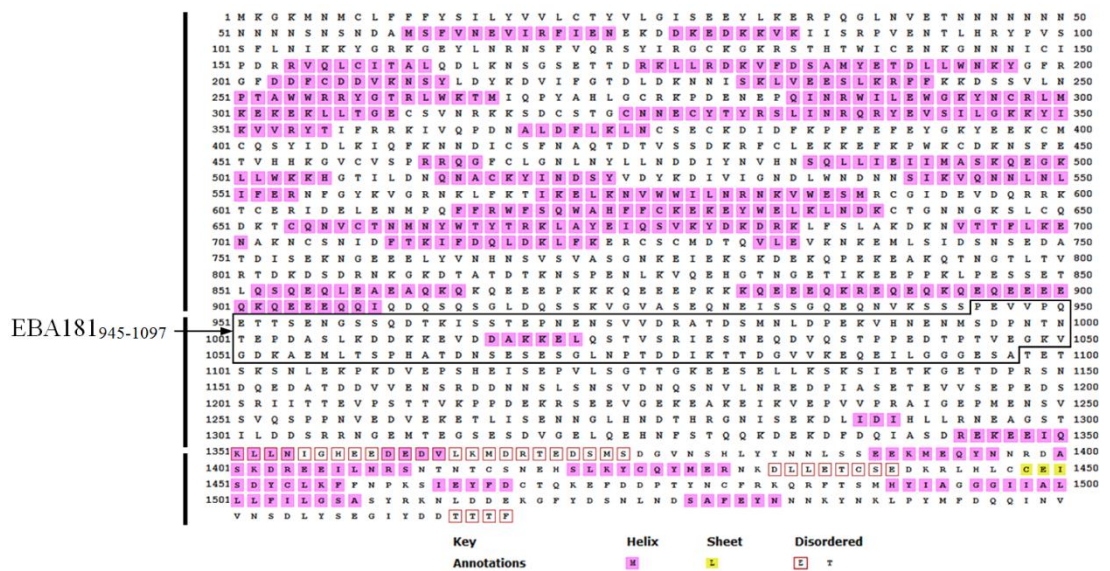


Figure 3.29: Secondary structure map of full-length EBA-181 given by PSIPRED web-based server.

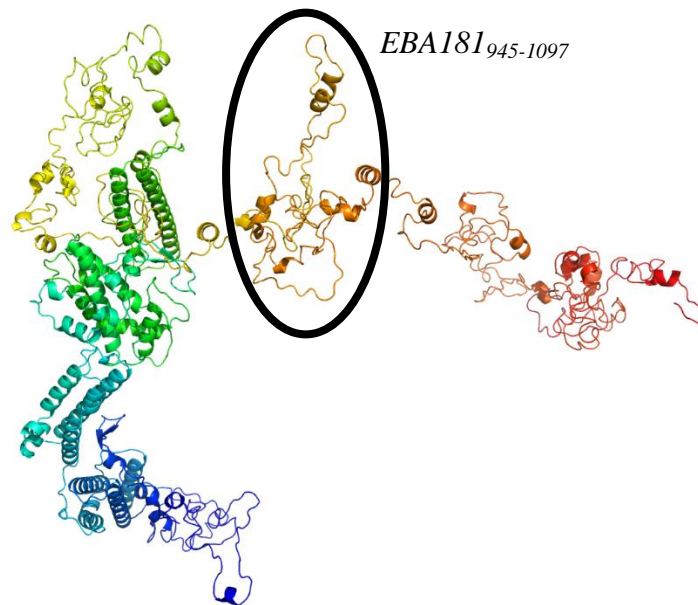


Figure 3.30: The highest scoring predicted 3D model of the full-length EBA-181 protein using the I-TASSER server. Estimated accuracy: C-score= -1.23, TM-score= 0.56 ± 0.15 and RMSD= $13.0 \pm 4.2\text{\AA}$. The corresponding EBA181₉₄₅₋₁₀₉₇ domain is circled in black.

DISOPRED and IUPRED were used to study the likelihood of disorder of EBA-181. According to the disorder prediction shown in Figures 3.31 and 3.32, the EBA-181 protein shows two principal regions. The N-terminus part to around residue 650/700 is predicted as being ordered, the subsequent 250 residues becomes less ordered, and from roughly residue 900 to almost the end of the sequence, it becomes disordered once again. The EBA181₉₄₅₋₁₀₉₇ domain is located in this latter region. The ANCHOR server predicts the disordered regions which might fold and thus interact with a globular partner, as also shown in Figure 3.32. Interestingly, despite its disordered nature, the EBA181₉₄₅₋₁₀₉₇ domain is predicted as being a putative binding region.

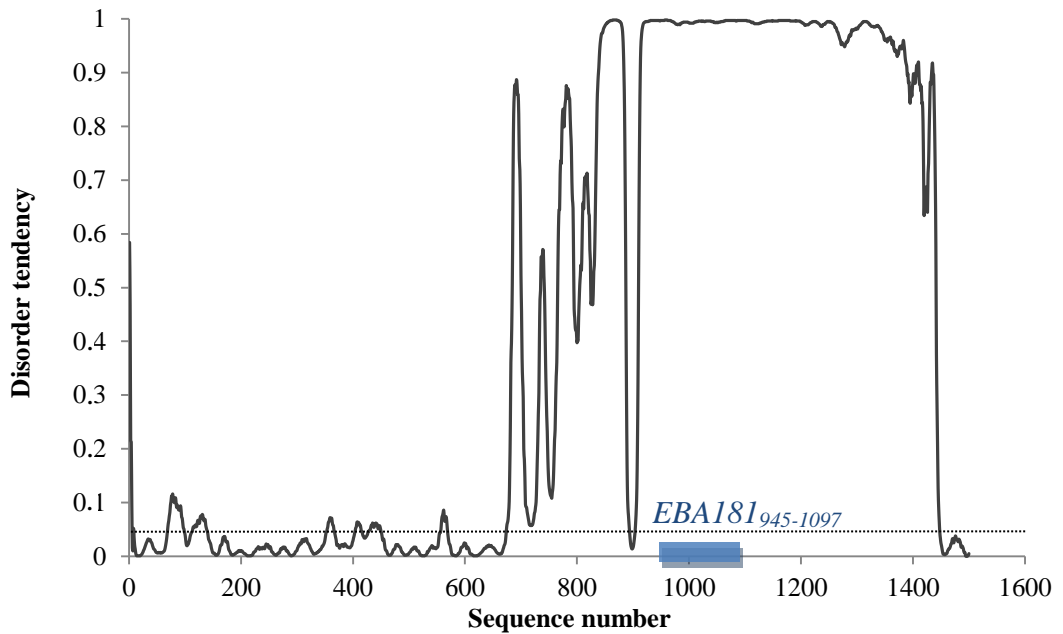


Figure 3.31: Disorder prediction of the full-length EBA-181 protein by DISOPRED server. The stippled line represents the disorder threshold to distinguish between ordered and disordered regions set by the program to 0.05.

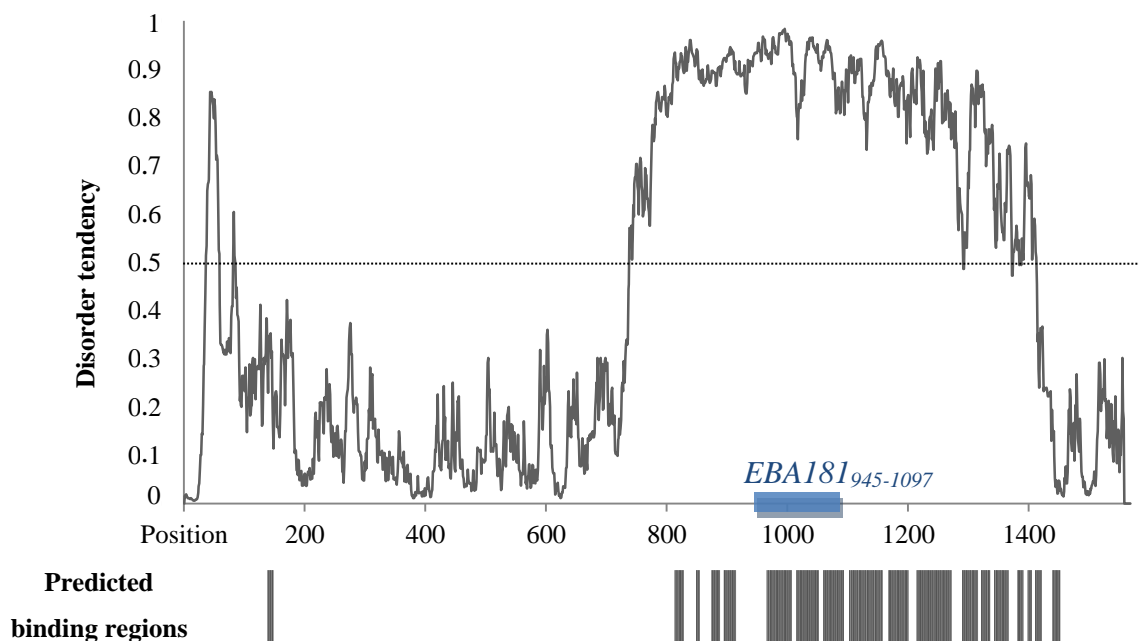


Figure 3.32: Disorder prediction of the full-length EBA-181 protein by IUPRED server and binding regions prediction using ANCHOR. The stippled line represents the disorder threshold set to 0.5 by the program.

3.5.1.2 Secondary structure and disorder predictions of EBA181₉₄₅₋₁₀₉₇

According to the PSIPRED and I-TASSER secondary structure predictions, the EBA181₉₄₅₋₁₀₉₇ domain is predicted as being mainly disordered, but with some residual alpha-helices (residues in pink, Figure 3.33 for the PSIPRED server; Figure 3.34 for the I-TASSER prediction). In Figure 3.33 it is possible to have some doubt about the actual existence of the first helix with only four residues. Nevertheless, the predictions result in the fact that EBA181₉₄₅₋₁₀₉₇ is probably globally very weakly structured.

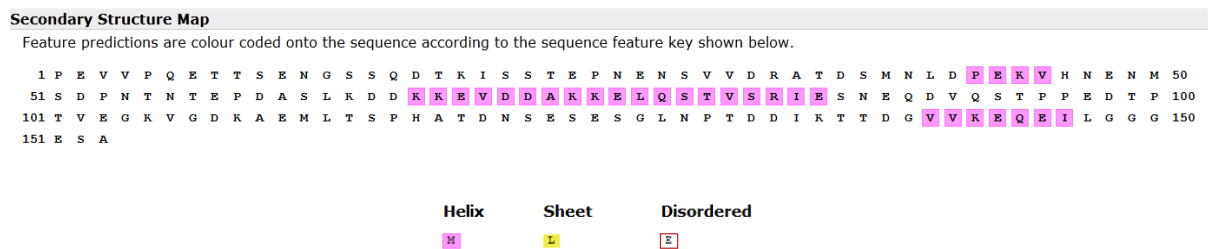


Figure 3.33: Secondary structure map of the EBA181₉₄₅₋₁₀₉₇ domain given by the PSIPRED web-based server.

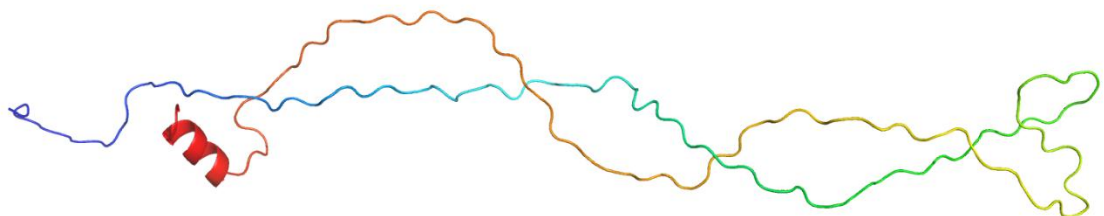


Figure 3.34: The highest scoring predicted 3D model of the EBA181₉₄₅₋₁₀₉₇ domain using the I-TASSER server. Estimated accuracy: C-score= -3.95.

The DISOPRED and IUPRED disorder prediction servers gave a very clear result about the disordered state of EBA181₉₄₅₋₁₀₉₇ (Figure 3.35 and Figure 3.36), with the disorder tendency over the threshold for the whole sequence. Nevertheless, in the DISOPRED prediction, the probability fluctuations and in particular the four fragments with a disorder tendency close to the threshold could mean the presence of few residual structures (Figure 3.35). It is clearer in the IUPRED prediction where the disorder tendency is almost equal to 1 all along the sequence. Here there is no probability fluctuation as seen with the DISOPRED server and therefore no secondary structure predicted in the EBA181₉₄₅₋₁₀₉₇ domain (Figure 3.36). The ANCHOR server predicts four disordered regions to interact with partners, as shown in Figure 3.36 and table 3.6; with almost the whole EBA181₉₄₅₋₁₀₉₇ domain being predicted as a putative binding region. Moreover, the 4.1R_{10kDa} binding region located between the residue 34 and 63 is situated in one of these predicted binding regions.

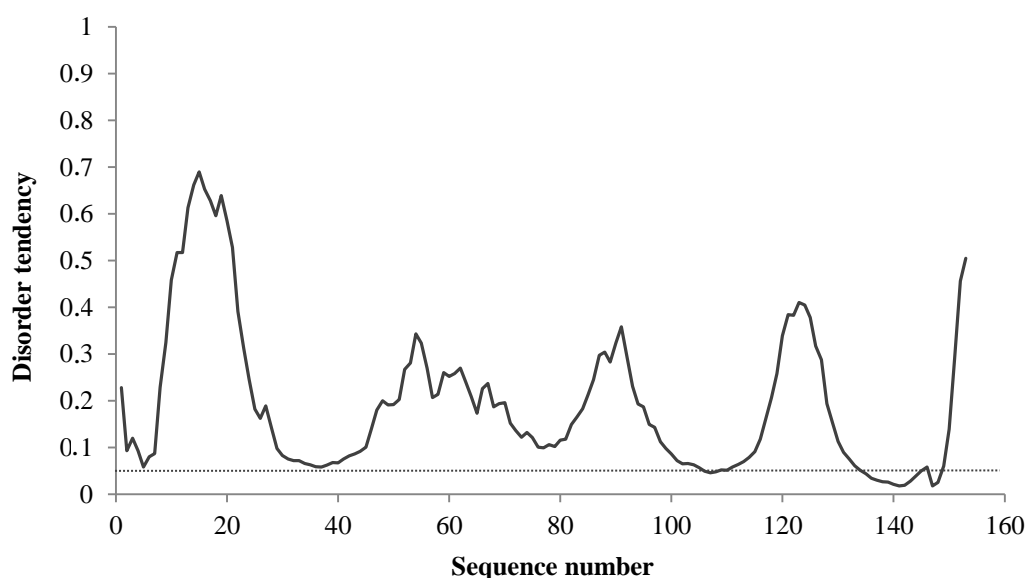


Figure 3.35: Disorder prediction of the EBA181₉₄₅₋₁₀₉₇ domain by DISOPRED server. The stippled line represents the disorder threshold (equal to 0.05).

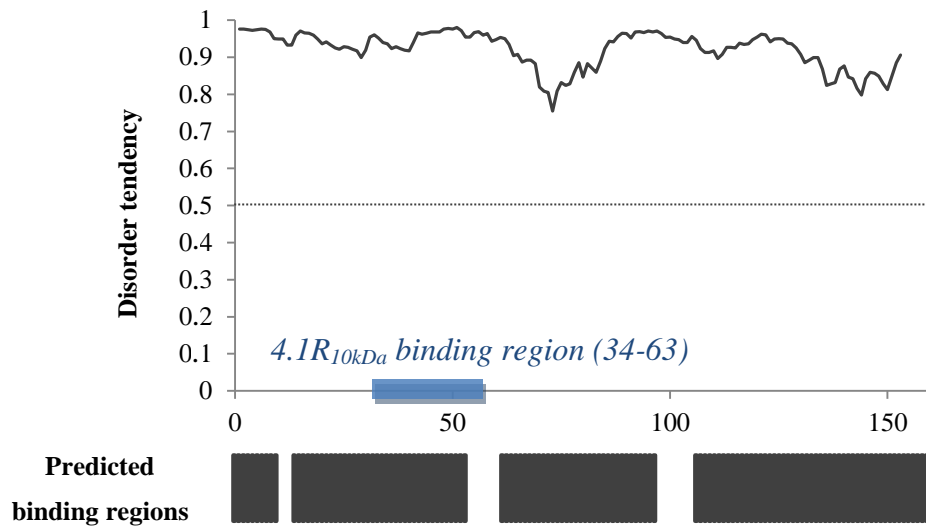


Figure 3.36: Disorder prediction of the EBA181₉₄₅₋₁₀₉₇ domain by IUPRED server and binding regions prediction using ANCHOR. The stippled line represents the disorder threshold (equal to 0.5). The binding region to 4.1R_{10kDa} is represented by the blue box.

Predicted disordered binding regions

| | From | To | Length |
|---|------|-----|--------|
| 1 | 1 | 10 | 10 |
| 2 | 14 | 51 | 38 |
| 3 | 59 | 92 | 34 |
| 4 | 101 | 153 | 53 |

Table 3.6: Predicted disordered binding regions of the EBA181₉₄₅₋₁₀₉₇ domain by ANCHOR server.

3.5.1.3 Secondary structure and disorder predictions of full-length 4.1R

According to the secondary structure prediction servers PSIPRED and I-TASSER (respectively Figures 3.38 and 3.39), the full-length 4.1R protein is folded for the first half of its sequence from its N-terminus part until roughly residue 310 (corresponding to the 30 kDa domain, see Figure 3.37, alpha helices and beta sheets based on the available X-ray structure, see Section 2.1.2). Then, the sequence becomes predicted as being mainly disordered, from residue 310 to 500 (corresponding to both the 16 kDa and the 10 kDa domains). As a reminder, the 4.1R_{10kDa} domain includes residues from 404 to 471. In the I-TASSER prediction, it can be noticed that 4.1R_{10kDa} is envisaged to have some residual alpha helices. Finally, the C-terminus part of the sequence is predicted as being folded again, with alpha helices and beta sheets. This region corresponds to the C-terminus 22 kDa domain.

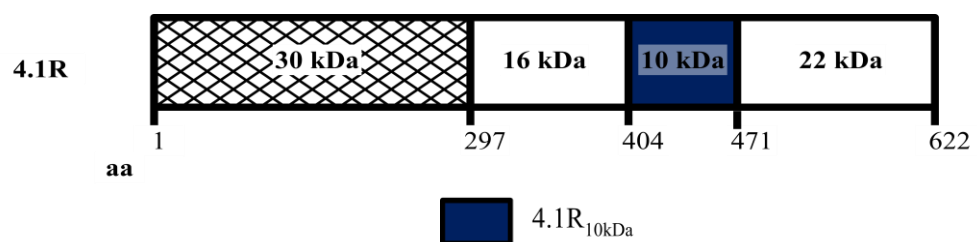


Figure 3.37: Schematic of the domain organisation of the full-length 80 kDa 4.1R protein.

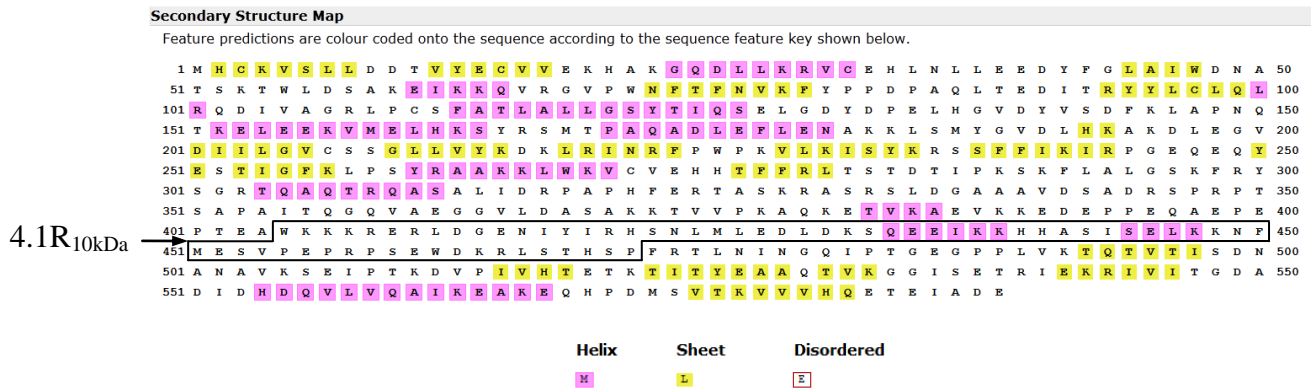


Figure 3.38: Secondary structure map of the full-length 4.1R protein given by PSIPRED server.

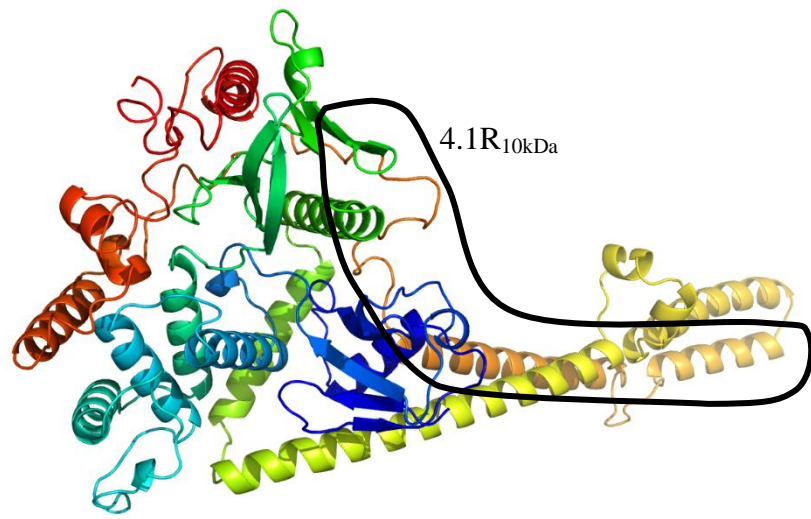


Figure 3.39: The highest scoring predicted 3D model of the full-length 4.1R protein using the I-TASSER server. Estimated accuracy: C-score= -0.86, TM-score= 0.61 ± 0.14 Å and RMSD= 9.7 ± 4.2 Å. The corresponding 4.1R_{10kDa} domain is encircled in black colour.

The DISOPRED and IUPRED disorder prediction servers (respectively Figures 3.40 and 4.41) gave a very clear result about the folded state of the N-terminus 30 kDa domain of full-length 4.1R, since the X-ray structure of this domain is really compact. From roughly the residue 300 and so the 16 kDa domain, the disorder prediction rose and is above the 0.05 threshold until the C-terminus end. Thus, the three 16 kDa, 10 kDa and 22 kDa domains are here predicted as being unfolded. Nevertheless, the presence of several oscillations of the disorder tendency and its range around the threshold for some regions might suggest the presence of residual secondary structures. This is particularly true for the IUPRED prediction of the 4.1R_{10kDa} domain showed to be unfolded but with a disorder probability rather close to the threshold which might also suggest the possible presence of secondary structures. The ANCHOR server predicts seven disorder regions to interact with partners, as shown in Figure 3.41 and table 3.7. Three of these seven predicted binding regions are located in the 4.1R_{10kDa} domain.

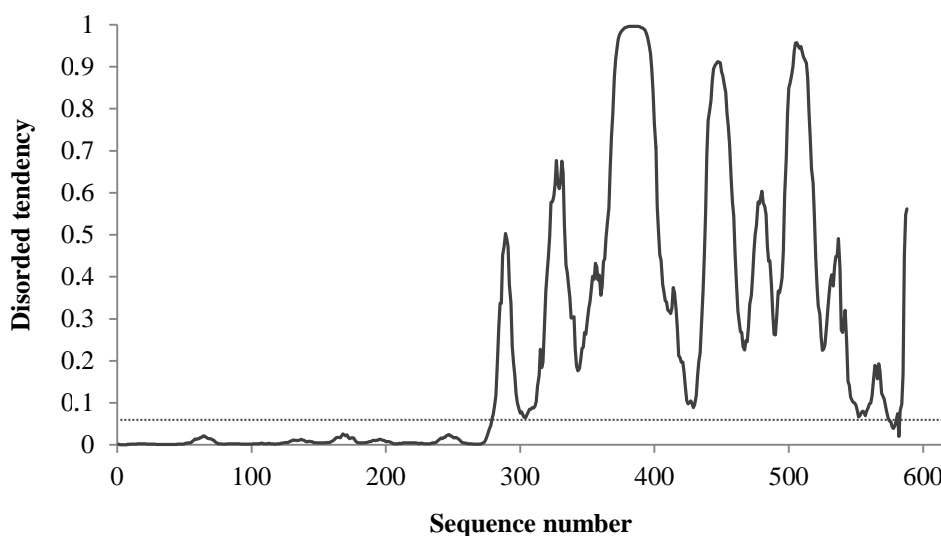


Figure 3.40: Disorder prediction of the full-length 4.1R protein by DISOPRED server. The stippled line represents the disorder threshold (equal to 0.05).

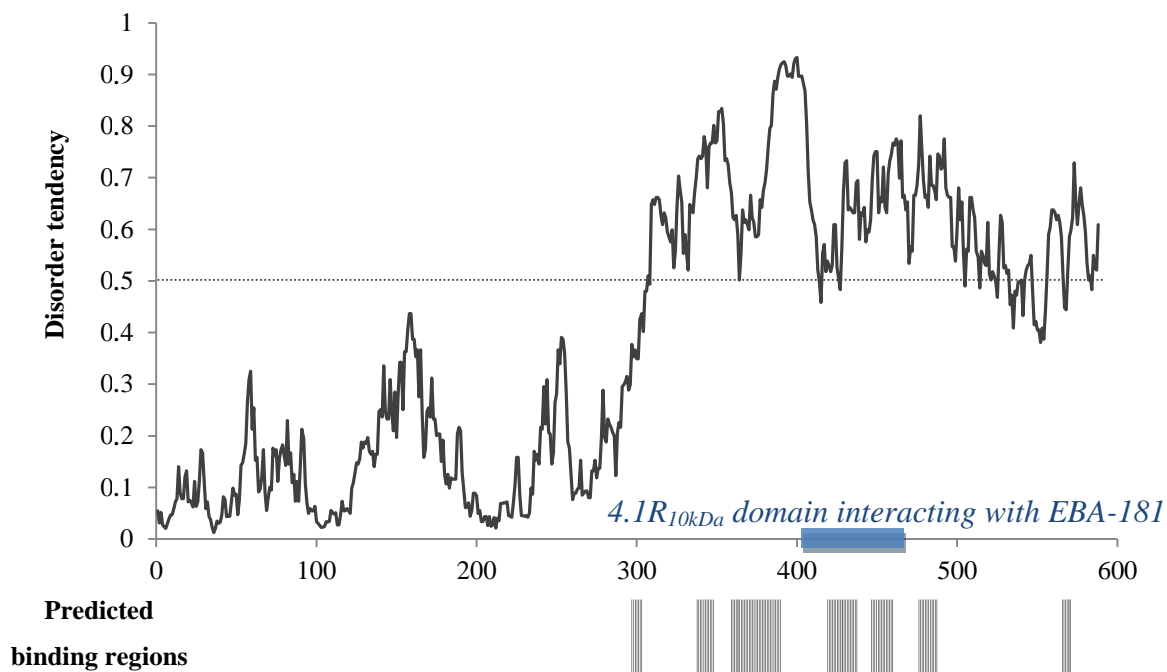


Figure 3.41: Disorder prediction of the full-length 4.1R protein by IUPRED server and binding regions prediction using ANCHOR. The stippled line represents the disorder threshold (equal to 0.5). The 4.1R_{10kDa} domain binding to EBA181₉₄₅₋₁₀₉₇ is represented by the blue box.

Predicted disordered binding regions

| | From | To | Length |
|---|------|-----|--------|
| 1 | 292 | 298 | 7 |
| 2 | 332 | 342 | 11 |
| 3 | 353 | 383 | 31 |
| 4 | 412 | 430 | 19 |
| 5 | 439 | 452 | 14 |
| 6 | 468 | 479 | 12 |
| 7 | 556 | 561 | 6 |

Table 3.7: Predicted disordered binding regions of the full-length 4.1R protein by ANCHOR server.

3.5.1.4 Secondary structure and disorder predictions of 4.1R_{10kDa}

The PSIPRED and I-TASSER servers (Figures 3.42 and 3.43) predicts 4.1R_{10kDa} to have three dominant alpha-helix domains, separated by coiled regions. In addition, the PSIPRED server predicted five residues to be in a beta sheet conformation.

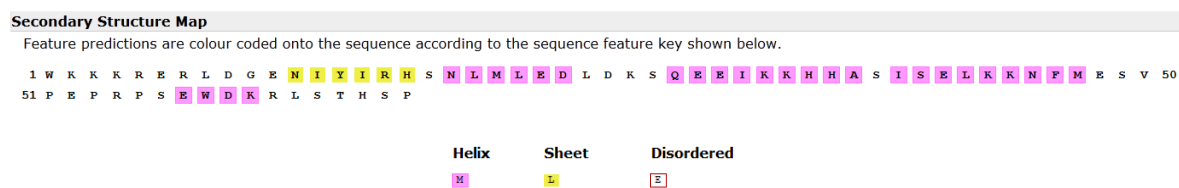


Figure 3.42: Secondary structure map of the 4.1R_{10kDa} domain given by PSIPRED server.

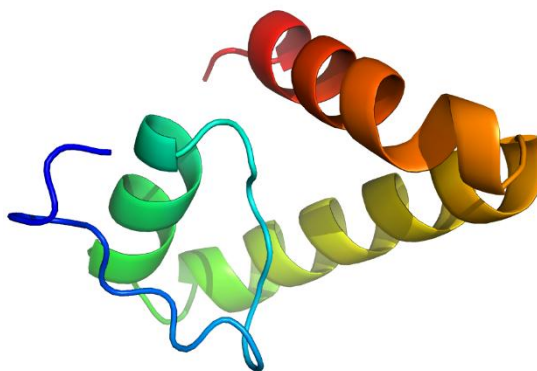


Figure 3.43: The highest scoring predicted 3D model of the 4.1R_{10kDa} domain using the I-TASSER server. Estimated accuracy: C-score= -2.99, TM-score= 0.38 ± 0.13 Å and RMSD= 9.6 ± 4.6 Å.

The DISOPRED and IUPRED servers (Figures 3.44 and 3.45) predicted 4.1R_{10kDa} to be disordered but with a disorder tendency around the threshold. This might mean that the domain is not fully disordered and might have some structure. Nevertheless, in the DISOPRED prediction, the disorder tendency for the N-terminus and the C-terminus ends, with respectively the 10 first and the last 5 amino acids is really high and might signify that these regions are clearly disordered. In conclusion and from these two disorder prediction servers, 4.1R_{10kDa} is neither totally unfolded nor fully ordered; and the protein could for example be very flexible. Finally, the ANCHOR server (Figure 3.45 and Table 3.12) predicts two putative binding sites, respectively from residues 12 to 22, and 37 to 46, corresponding to the two folded and helical domains predicted by DISOPRED and I-TASSER.

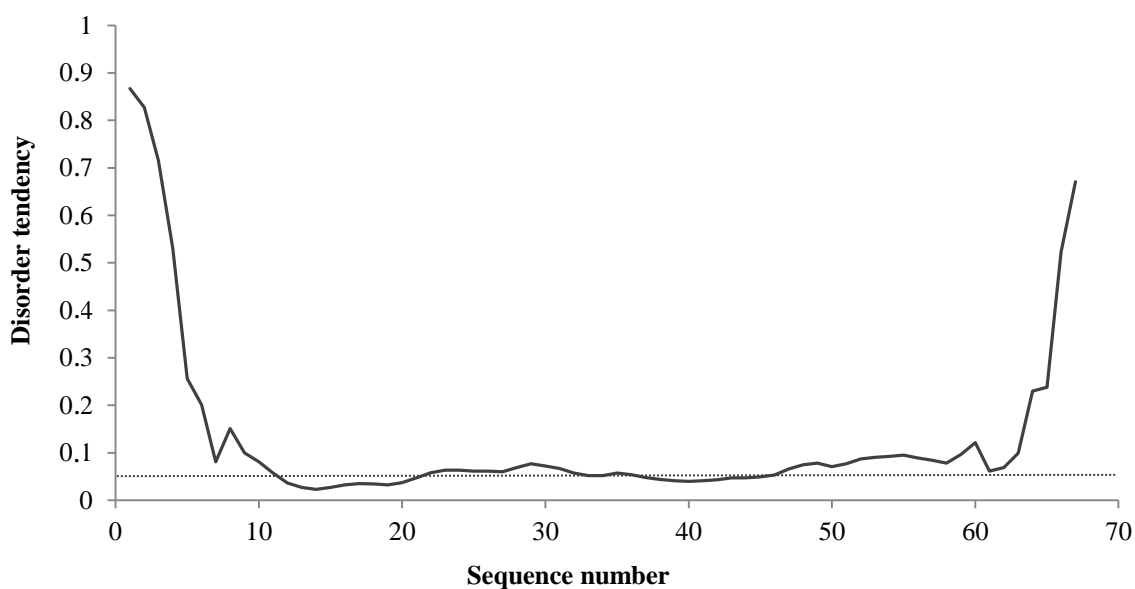


Figure 3.44: Disorder prediction of the 4.1R_{10kDa} domain by DISOPRED server. The stippled line represents the disorder threshold (equal to 0.05).

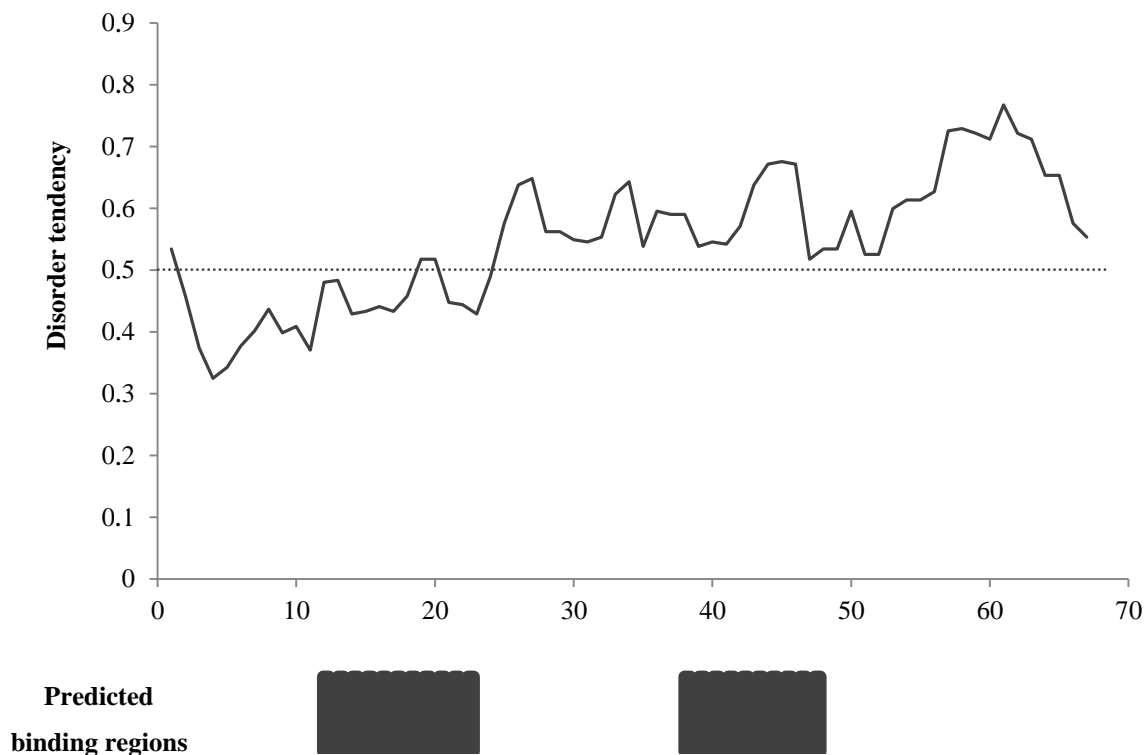


Figure 3.45: Disorder prediction of the 4.1R_{10kDa} domain by IUPRED server and binding regions prediction using ANCHOR. The stippled line represents the disorder threshold (equal to 0.5).

Predicted disordered binding regions

| | From | To | Length |
|---|------|----|--------|
| 1 | 12 | 22 | 11 |
| 2 | 37 | 46 | 10 |

Table 3.12: Predicted disordered binding regions of the 4.1R_{10kDa} domain by ANCHOR server.

3.5.2 1D NMR assessment of EBA181₉₄₅₋₁₀₉₇, 4.1R_{10kDa} and the complex

One dimensional NMR (1D NMR) was carried out to check the conformational state of the two individual untagged proteins and of the complex. 200 μ l of protein was used at a minimum concentration of 100 μ M on a 600 MHz NMR spectrometer equipped with a cryo-probe. 10 % of D₂O was added to the sample in order to allow the calibration of the magnetic field. Experiments were finally carried out at different temperatures, from 4°C to 40°C.

3.5.2.1 Protein EBA181₉₄₅₋₁₀₉₇

In contrast to folded proteins that show a very broad spectrum and signals above 8.5 ppm and below 0.5 ppm, no signal could be seen in this folded region. This indicates that EBA181₉₄₅₋₁₀₉₇ is unfolded at 5°C, 25°C and 40°C (Figure 3.46).

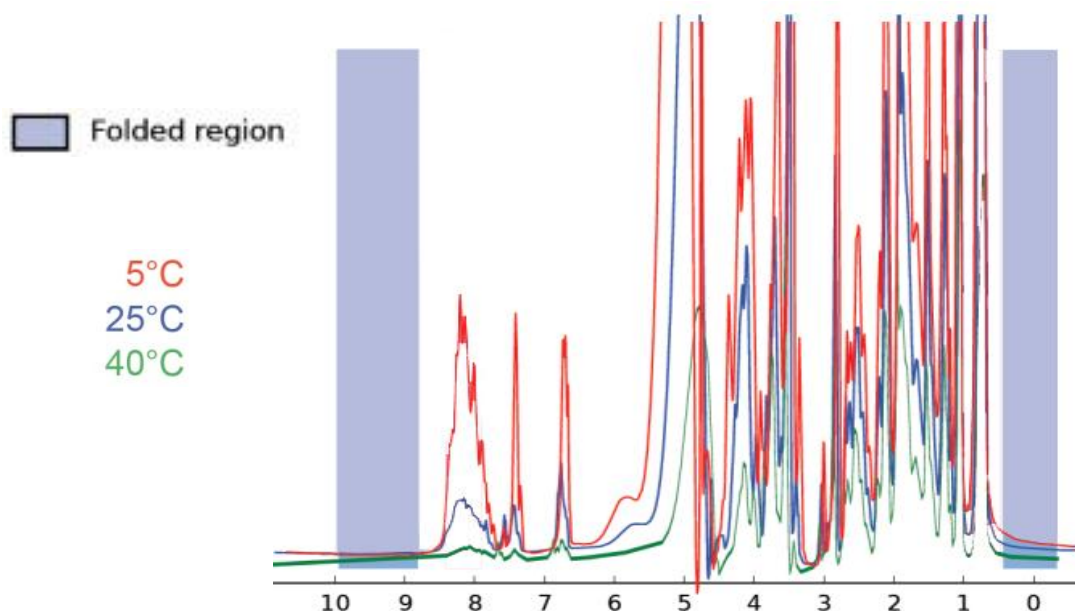


Figure 3.46: ¹H-NMR experiments on EBA181₉₄₅₋₁₀₉₇. Temperatures: 5°C (red curve), 25°C (blue curve) and 40°C (green curve).

3.5.2.2 Protein 4.1R_{10kDa}

Similarly, the ¹H-NMR spectrum of 4.1R_{10kDa} (Figure 3.47) showed no signal in the chemical shift ranges of structured region. However, in contrast to EBA181₉₄₅₋₁₀₉₇ more intense and sharper signals were observed around 7 p.p.m. These signals are attributable to phenylalanine and would indicate a mobility of side chains compared to the protein backbone. Thus, it is possible that amide protons from the backbone were in an exchange state and so invisible to NMR. Therefore aromatic signals appear more intense because of the loss of the N-H signals from the backbone. Another possibility is the combination of the phenomenon explained earlier with the low compactness/fold of the protein. It was impossible to conclude with certitude the degree of folding of 4.1R_{10kDa}, it was therefore decided to continue the studies using other techniques.

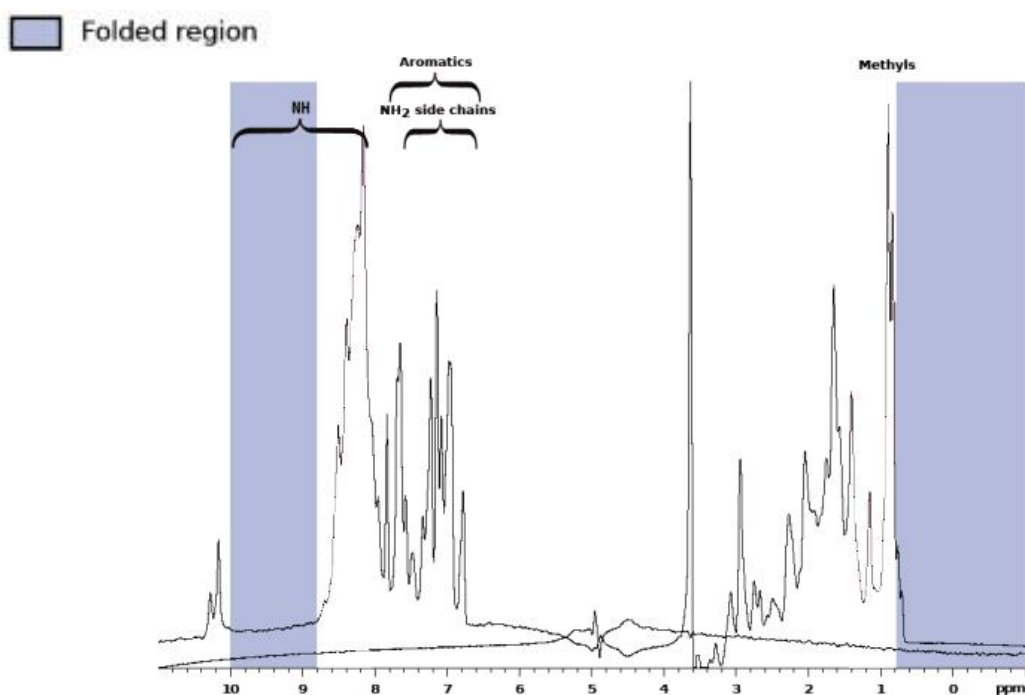


Figure 3.47: ¹H-NMR experiment on 4.1R_{10kDa} at 4°C.

3.5.2.3 EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex

In order to check for conformational changes when the two EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins are in complex, ¹H-NMR experiments were also carried out on the complex. However, no signal is observed in the folded region of the NMR spectra when 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ are in complex (Figure 3.48). This suggests that there is not a significant change in folding state of the proteins upon binding.

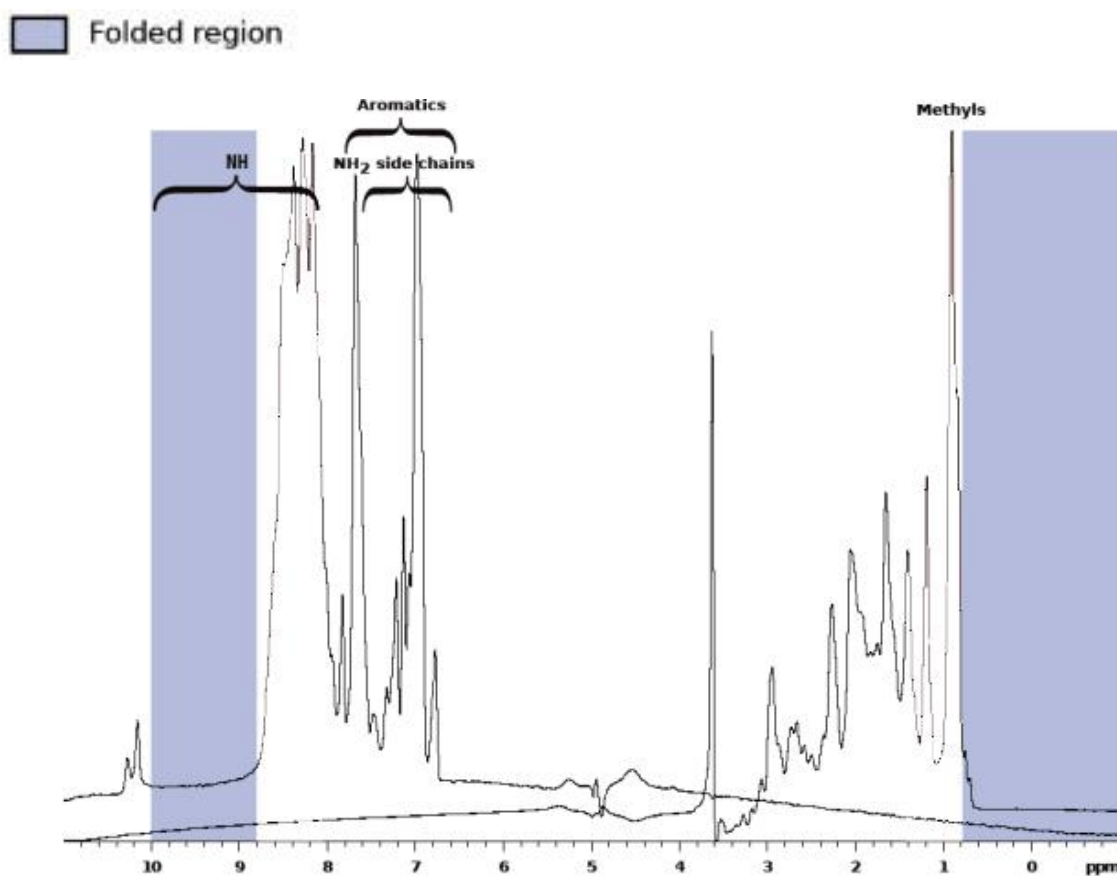


Figure 3.48: ¹H-NMR experiment on the 4.1R_{10kDa} / EBA181₉₄₅₋₁₀₉₇ complex at 4°C.

3.5.3 Circular dichroism (CD) measurements

After cleavage of their respective tags, both protein domains were analysed using circular dichroism. Circular dichroism measurements were carried out by recording a spectrum in far UV (from 190 to 260 nm). Each experiment was recorded three times at both 4°C and 20°C. The blank sample was made with dialysis buffer of each individual protein.

3.5.3.1 Protein EBA181₉₄₅₋₁₀₉₇

Circular dichroism measurements on EBA181₉₄₅₋₁₀₉₇ (Figure 3.49) show a slightly dipped plateau between 250 nm and 215 nm and only one minimum around 200-205 nm. This curve is interpretable as a curve typical of random coil conformation and disordered protein (Figure 3.49, blue calculated curve as estimation of secondary structures from the experimental curve), implying that EBA181₉₄₅₋₁₀₉₇ is in an unfolded state.

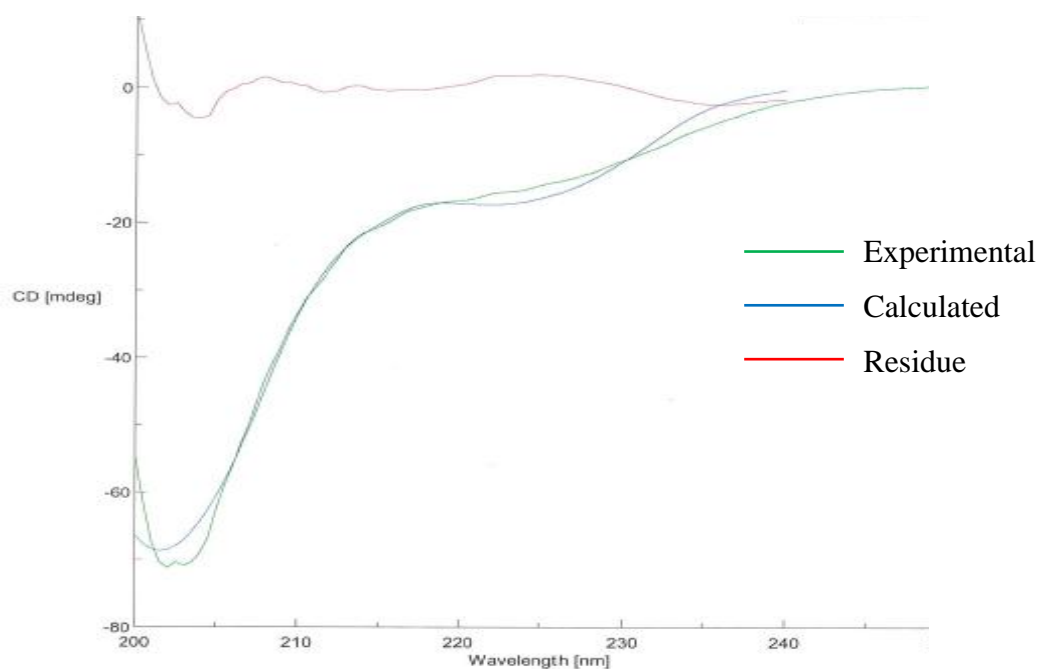


Figure 3.49: Comparison between experimental curve and estimated content in secondary structure (calculated curve) using the far-UV CD spectrum of EBA181₉₄₅₋₁₀₉₇.

Green curve: experimental curve.

Blue curve: calculated curve used to determine the secondary structure content.

Red curve: residual curve after subtraction of experimental and calculated curves.

3.5.3.2 Protein 4.1R_{10kDa}

In contrast to the circular dichroism spectra of EBA181₉₄₅₋₁₀₉₇, the same experiment performed on 4.1R_{10kDa} (Figure 3.50) shows two very specific minima at 208 nm and 222 nm. This curve is typical of alpha-helix content in molecules. Modelling by the JASCO software Spectra Manager II (Figure 3.50, blue calculated curve) estimated 4.1R_{10kDa} to be composed of 35 % of alpha helices, 20 % of turn and 45 % of random coil.

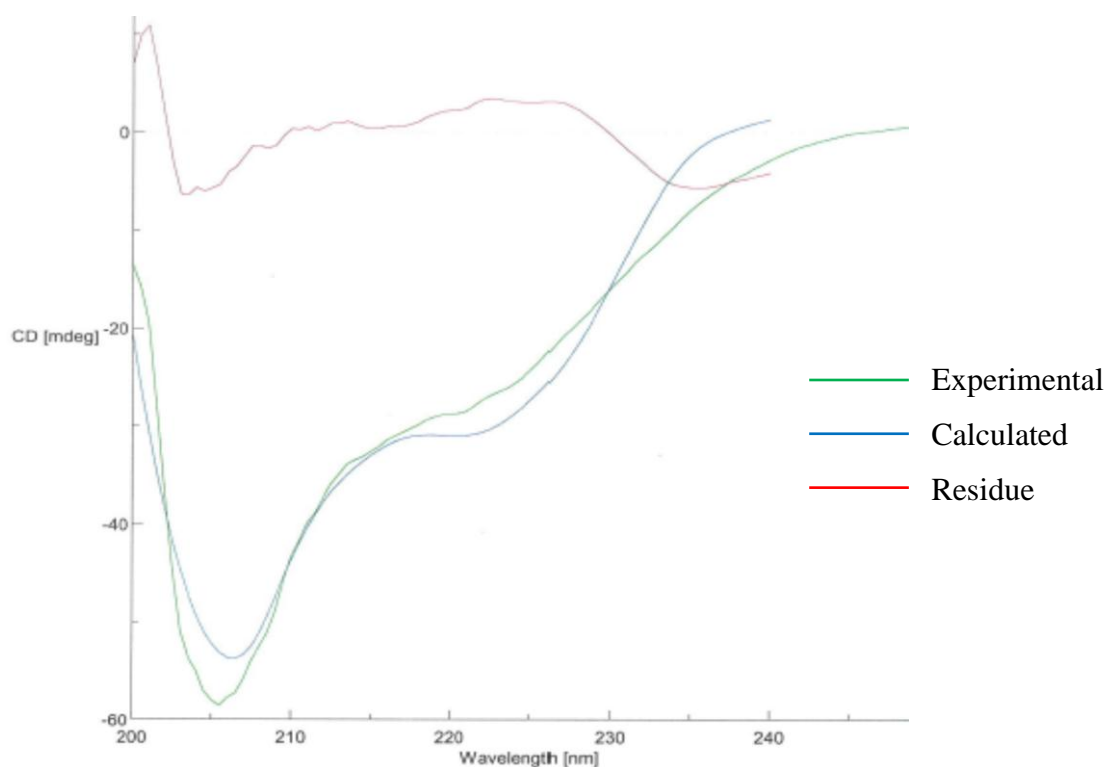


Figure 3.50: Comparison between experimental curve and estimated content in secondary structure (calculated curve) using the far-UV CD spectrum of 4.1R_{10kDa}.

Green curve: experimental curve.

Blue curve: calculated curve used to determine the secondary structure content.

Red curve: residual curve after subtraction of experimental and calculated curves.

3.5.4 Thermal shift assay (TSA) measurements

Thermal shift assays were carried out to obtain stability information on both proteins and in particular to define the best buffer conditions for protein purification and storage. Each assay experiment was replicated three times. The assays were realised in 96 well plates

(Stratagene) using a pre-defined range of 48 buffers with pH varying from 5.0 to 9.0, and with two different concentrations of salt (100 and 500 mM NaCl).

3.5.4.1 **Protein EBA181₉₄₅₋₁₀₉₇**

From the assay result performed on EBA181₉₄₅₋₁₀₉₇, it was impossible to determine the T_m for any of the buffer and salt conditions for EBA181₉₄₅₋₁₀₉₇. There was no significant shift of the fluorescence of the probe whilst increasing the temperature, meaning that the protein was already unfolded at 25°C (Figure 3.51). The fluorescence shifts only when the probe binds in exposed hydrophobic regions following the protein melting.

Moreover, the measured fluorescence started at very high levels (between 10 000 and 40 000 depending the buffers) even at 25°C, meaning that the probe was already bound to hydrophobic regions even at low temperatures.

Even if the results of the TSA on EBA181₉₄₅₋₁₀₉₇ did not permit the best buffer conditions to be defined for protein stability, the absence of a hydrophobic core is further evidence for the unfolded nature of the protein.

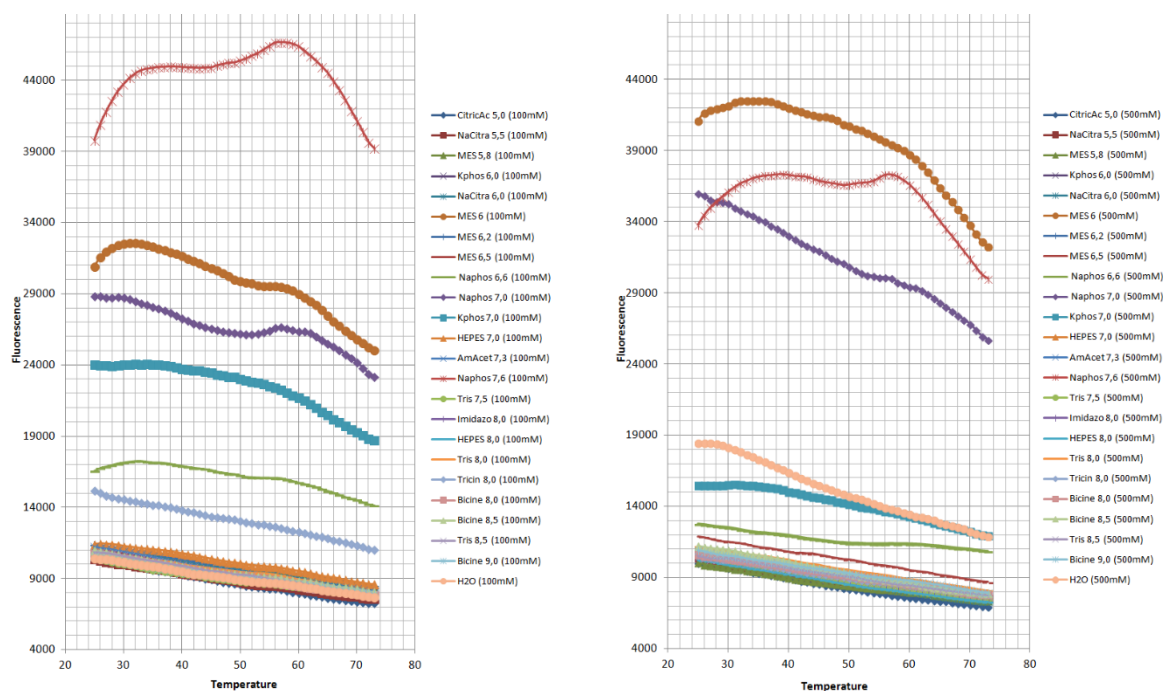


Figure 3.51: Thermal shift assay on EBA181₉₄₅₋₁₀₉₇ from pH 5.0 to pH 9.0. **Left:** experiment at 100 mM NaCl. **Right:** experiment at 500 mM NaCl.

3.5.4.2 Protein 4.1R_{10kDa}

For 4.1R_{10kDa} protein, the highest T_m was determined to be 57°C at pH=6.0 and 100 mM NaCl concentration (Figure 3.52, left, orange curve). Nevertheless, the TSA results had similar overall fluorescence behaviour to EBA181₉₄₅₋₁₀₉₇: a high starting fluorescence point and no shift of the probe fluorescence with increasing the temperature. These results meant that the protein was at least partially unfolded or very flexible.

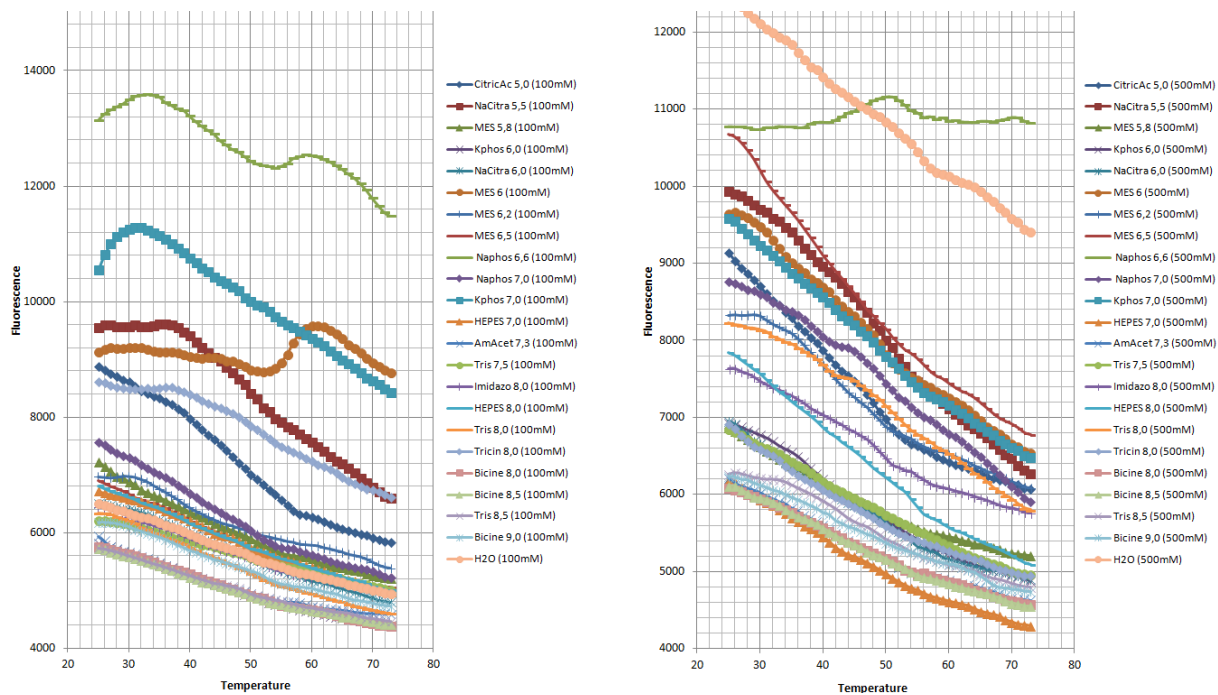


Figure 3.52: Thermal shift assay on 4.1R_{10kDa} from pH 5.0 to pH 9.0. **Left:** experiment at 100 mM NaCl. **Right:** experiment at 500 mM NaCl.

3.5.4.3 Protein GST-4.1R_{10kDa}

For GST-4.1R_{10kDa} protein, a clear T_m was obtained, demonstrating protein folding (most likely due to the GST tag). The highest T_m was determined to be 53°C between pH 7.0 and 8.5, and 100 mM NaCl concentration (Figure 3.53, left).

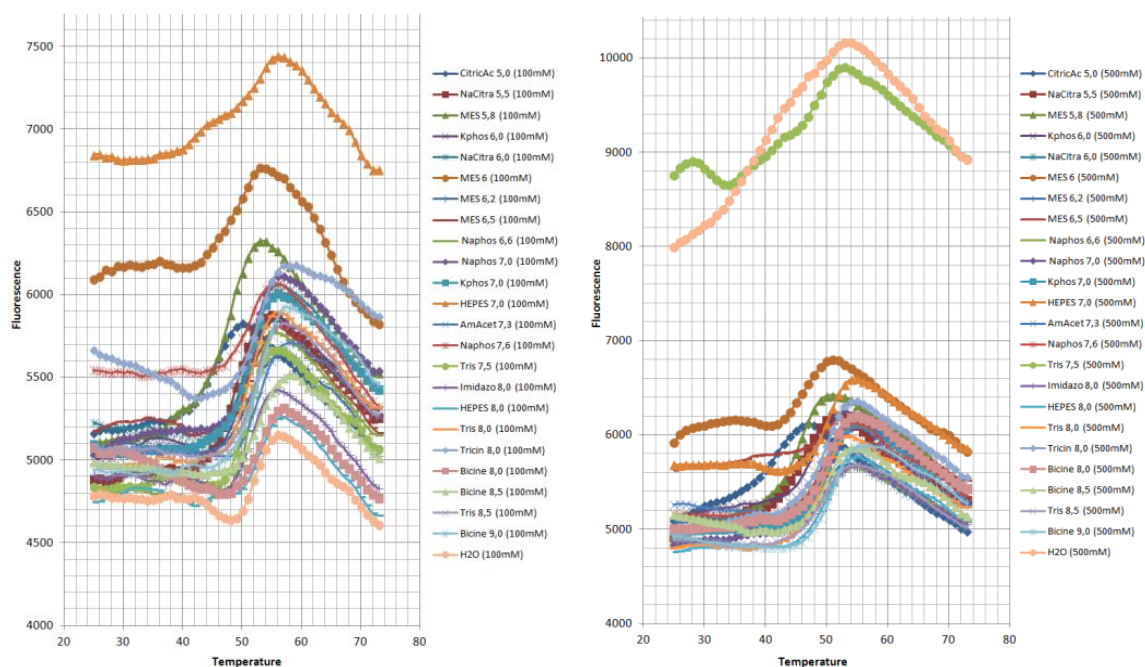


Figure 3.53: Thermal shift assay on GST-4.1R_{10kDa} from pH 5.0 to pH 9.0. **Left:** experiment at 100 mM NaCl. **Right:** experiment at 500 mM NaCl.

3.5.5 Multi-angle laser light scattering coupled to size exclusion chromatography and refractive index (SEC-MALLS-RI) for determination of oligomeric state and molar mass calculation

SEC-MALLS-RI experiments were performed to analyse the oligomeric state of each individual protein in solution and to check the molarity of the complex. The column and device were equilibrated overnight in the corresponding dialysis buffer at a flow of 0.5 ml/min, and 50 μ l of each purified and concentrated protein were then injected. Finally the data were analysed using the ASTRA software (WYATT Technology).

3.5.5.1 Protein EBA181₉₄₅₋₁₀₉₇

According to the SEC-MALLS-RI data (Figures 3.54 and 3.55), EBA181₉₄₅₋₁₀₉₇ is monomeric in solution, with a measured size of 17.4 kDa \pm 3 % (0.5 kDa), which corresponds to the expected size of 17 kDa given by the amino-acid sequence.

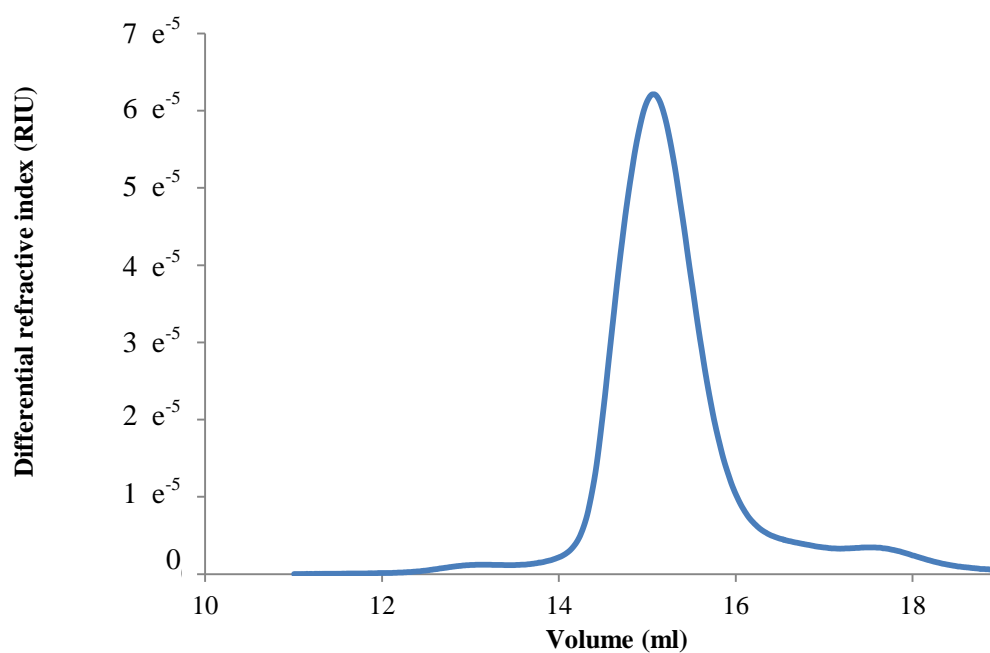


Figure 3.54: SEC-MALLS-RI analysis of EBA181₉₄₅₋₁₀₉₇ using a size-exclusion chromatography Superdex™ 200 column.

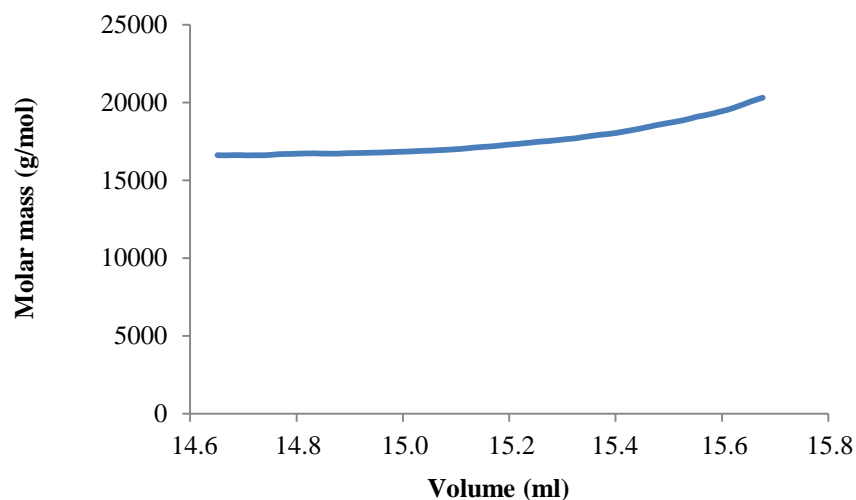


Figure 3.55: Molar mass calculation using the SEC-MALLS-RI analysis of EBA181₉₄₅₋₁₀₉₇.

3.5.5.2 Protein 4.1R_{10kDa}

Regarding its relative small size, two different size-chromatography columns were used to analyse the 4.1R_{10kDa} protein by MALLS (SuperdexTM 200 and SuperdexTM 75 columns).

According to the MALLS result using the S200 column (Figure 3.56 and Figure 3.57), 4.1R_{10kDa} is monomeric in solution, with a measured size of 10.8 kDa \pm 36 % (3.9 kDa), which corresponds to the expected size of 8.5 kDa, given by the amino-acid sequence. Nevertheless, the accuracy of the measure was low. This result could be caused by the S200 size-exclusion chromatography resolution which was not sufficient for a low molecular weight protein.

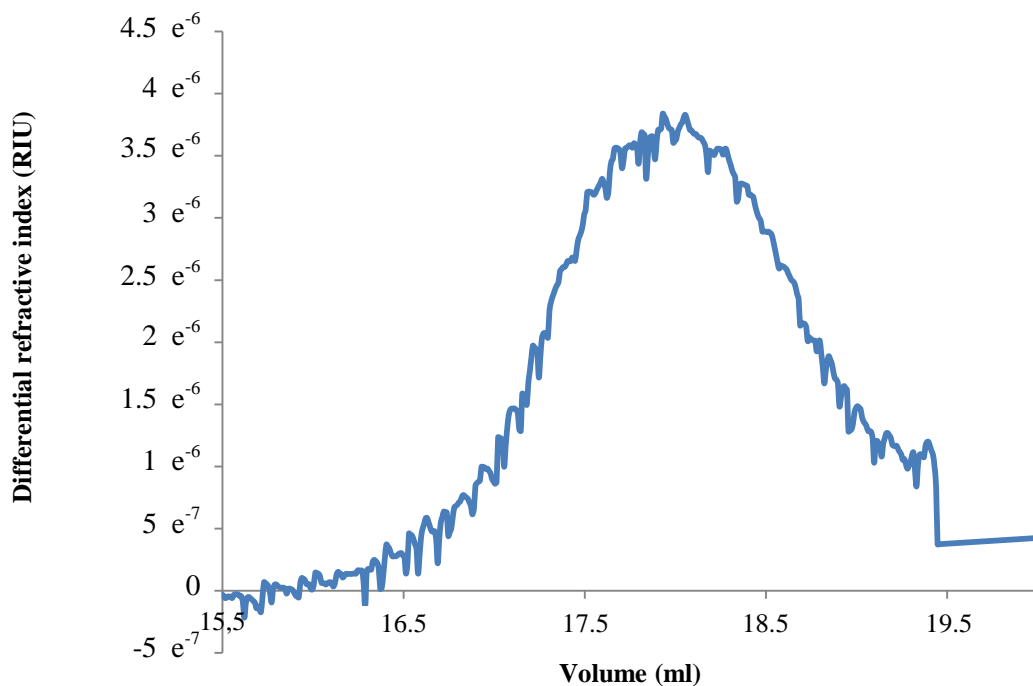


Figure 3.56: SEC-MALLS-RI analysis of 4.1R_{10kDa} on the size-exclusion chromatography Superdex™ 200 column.

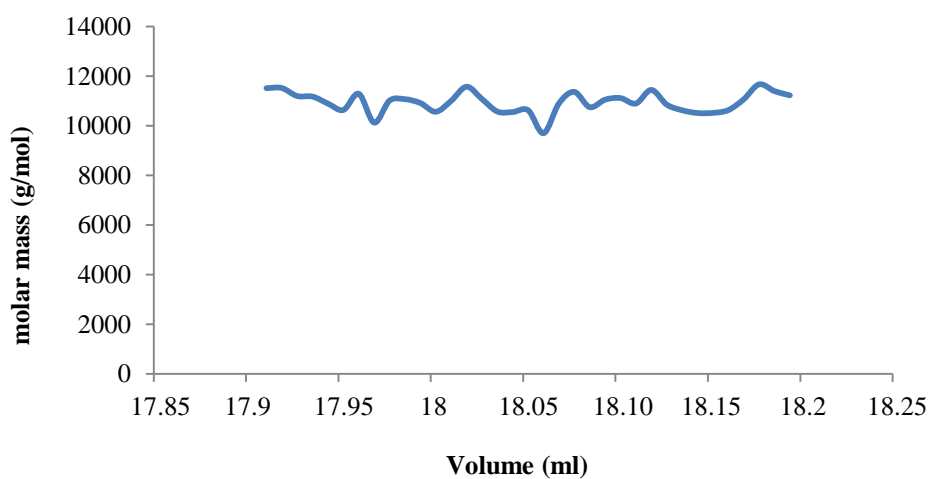


Figure 3.57: Molar mass calculation using the SEC-MALLS-RI analysis of 4.1R_{10kDa} on the S200 column.

Regarding the high error range given by the S200 size-exclusion column, it was decided to use a S75 column to cross-check and improve the previous MALLS result. Using the S75 size-exclusion chromatography column, it was possible to improve the quality of the results (Figure 3.58) and thus, the accuracy of the measured size was higher (Figure 3.59). The resulting data also showed 4.1R_{10kDa} to be a monomer in solution, with a measured size of 8.8 kDa \pm 9% (0.8 kDa).

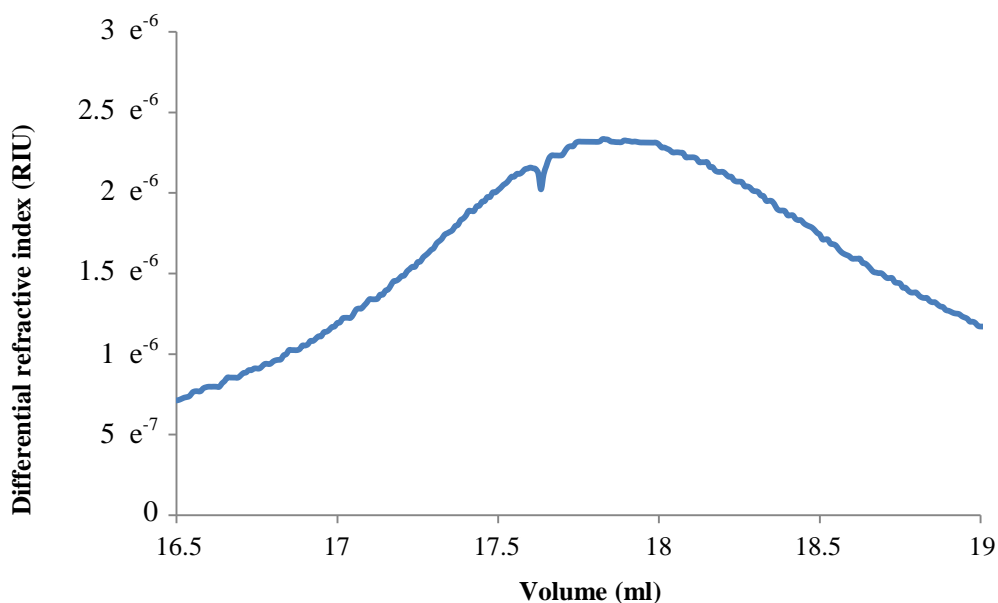


Figure 3.58: SEC-MALLS-RI analysis of 4.1R_{10kDa} on the size-exclusion chromatography SuperdexTM 75 column.

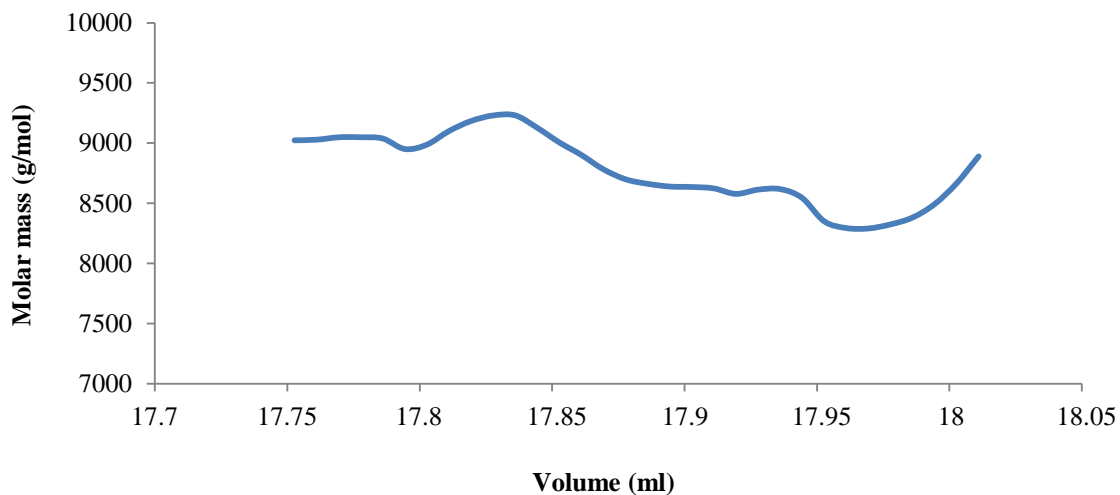


Figure 3.59: Molar mass calculation using the SEC-MALLS-RI analysis of 4.1R_{10kDa} on the S75 column.

3.5.5.3 EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex

Using the SEC-MALLS-RI method it was possible to determine the molarity of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex (Figure 3.60). The measured size was equal to 24 kDa \pm 0.9 % (0.2 kDa). The theoretical sizes of EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} are respectively 17 kDa and 8.5 kDa for a complex of 24.5 kDa, close to the resulting weight from the SEC-MALLS-RI and indicating a 1:1 molar ratio in the complex.

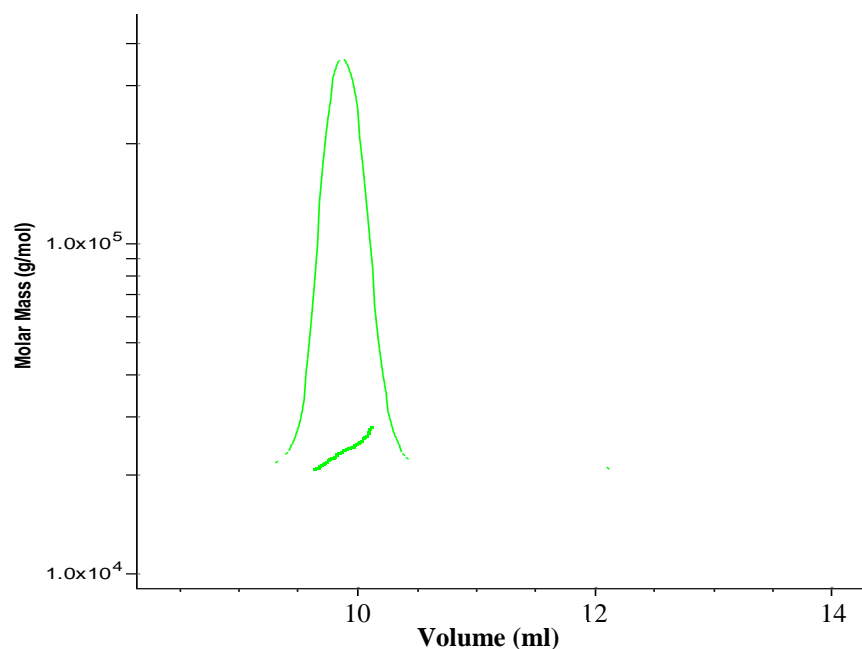


Figure 3.60: SEC-MALLS-RI analysis of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex on TSKgel G3000SWxl size-exclusion chromatography column. Green: refractometry.

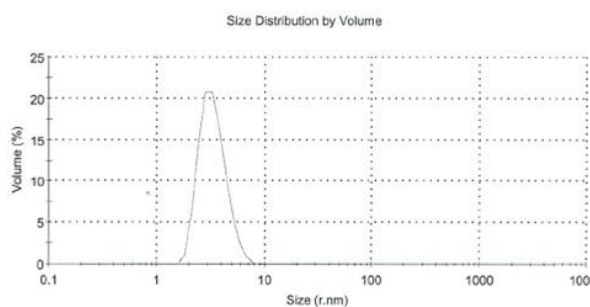
3.5.6 Dynamic light scattering (DLS) measurements

Dynamic light scattering was used to check the quality of the samples after the final step of purification and to determine the hydrodynamic radius (R_h) of 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ (see Figure 3.61). Dynamic light scattering measurements were performed on 25 to 45 μ l of protein sample in a concentration range between 0.5 and 5 mg/ml in a three windows quartz cuvette (*Hellma*) on a Zetasizer (*Malvern*) device set at 4°C. In order to characterise the effect of buffer composition and pH on their respective sizes, further DLS hydrodynamic radius measurements were carried out on the two proteins purified in four different buffers: Tris pH 8.0, PBS pH 7.4, sodium phosphate pH 7.0 and sodium phosphate pH 6.0 (Figures 3.61, 3.62 and 3.63; buffer compositions shown in Appendix A.2).

3.5.6.1 Protein EBA181₉₄₅₋₁₀₉₇

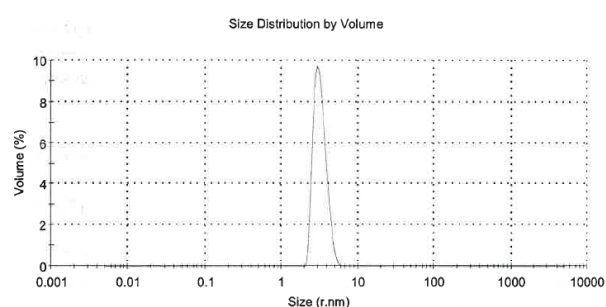
The hydrodynamic radius of EBA181₉₄₅₋₁₀₉₇ was measured for each of the four pH conditions (Figure 3.61) with the smallest R_h equal to 2.95 nm at pH=6.0. The results were similar for the tagged His-EBA181₉₄₅₋₁₀₉₇ construct (data not shown). This experiment defined the best condition (buffer/pH) for EBA181₉₄₅₋₁₀₉₇: sodium phosphate (NaP) buffer at pH 6.0; and gave its physical size (R_h) of 2.95 nm.

Tris pH=8.0:



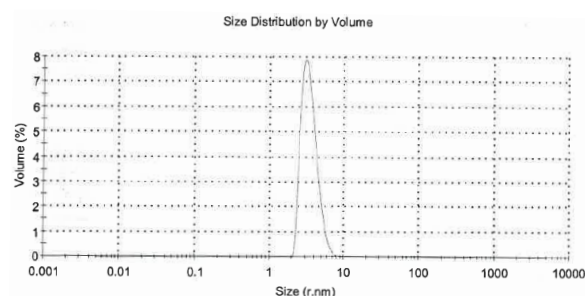
$$R_h = 3.4 \text{ nm}$$

PBS pH=7.4:



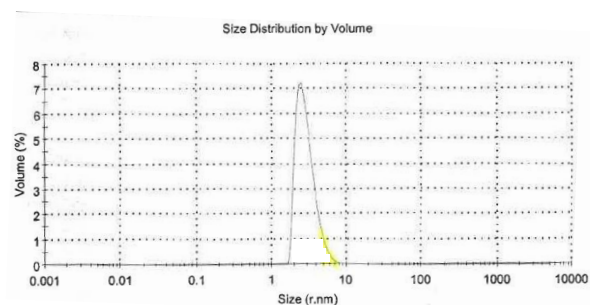
$$R_h = 3.2 \text{ nm}$$

Sodium phosphate pH=7.0:



$$R_h = 3.5 \text{ nm}$$

Sodium phosphate pH=6.0:



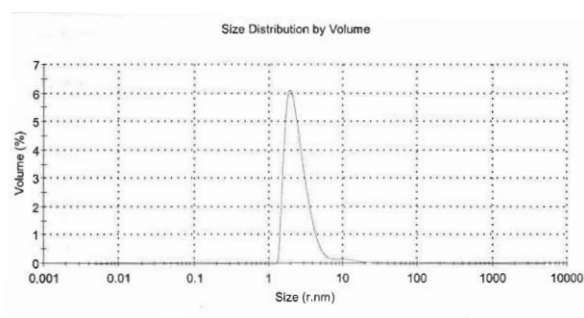
$$\underline{R_h = 2.95 \text{ nm}}$$

Figure 3.61: DLS measurements of hydrodynamic radius (R_h) of EBA181₉₄₅₋₁₀₉₇ in four different pH buffers (Tris pH=8.0, PBS pH=7.4, sodium phosphate pH=7.0 and sodium phosphate pH=6.0). The smallest R_h is underlined.

3.5.6.2 Protein 4.1R_{10kDa}

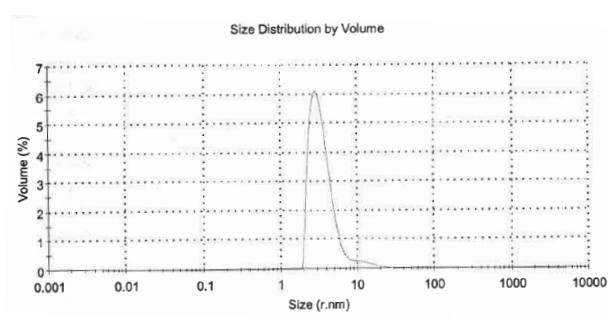
Hydrodynamic radius of 4.1R_{10kDa} was measured for each of the four pH conditions (Figure 3.62) with the minimum R_h equal to 2.0 nm at pH=6.0. This experiment defined the best buffer/pH for 4.1R_{10kDa}: sodium phosphate buffer at pH 6.0; and gave its physical size (R_h) of 2.0 nm.

Tris pH=8.0:



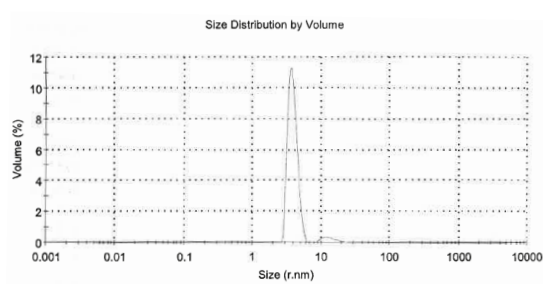
$$R_h = 2.5 \text{ nm}$$

PBS pH=7.4:



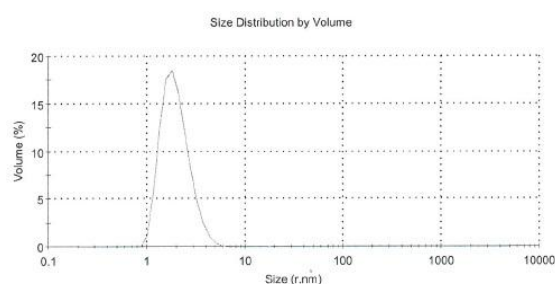
$$R_h = 3.9 \text{ nm}$$

Sodium phosphate pH=7.0:



$$R_h = 3.9 \text{ nm}$$

Sodium phosphate pH=6.0:



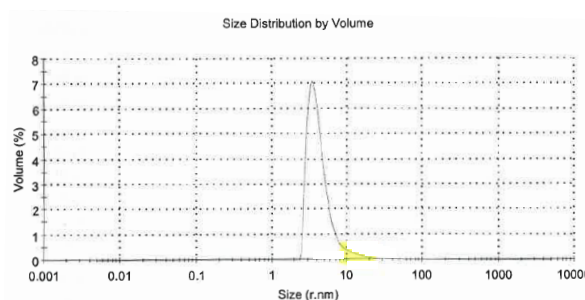
$$\underline{R_h = 2.0 \text{ nm}}$$

Figure 3.62: DLS measurements of hydrodynamic radius (R_h) of 4.1R_{10kDa} in four different pH buffers (Tris pH=8.0, PBS pH=7.4, sodium phosphate pH=7.0 and sodium phosphate pH=6.0). The smallest R_h is underlined.

3.5.6.3 EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex

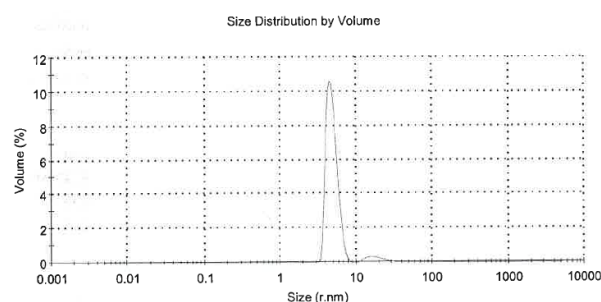
The hydrodynamic radius of the complex was measured for three pH conditions (Figure 3.63) with the best (minimum) R_h equal to 4.4 nm at pH=6. In addition to defining the best buffer/pH for the complex (sodium phosphate buffer at pH 6.0) and the calculation of its physical size ($R_h=4.4$ nm), the DLS measurement added evidence for the complex interaction by showing an increase of the observed size over the individual proteins.

Tris pH=8.0:



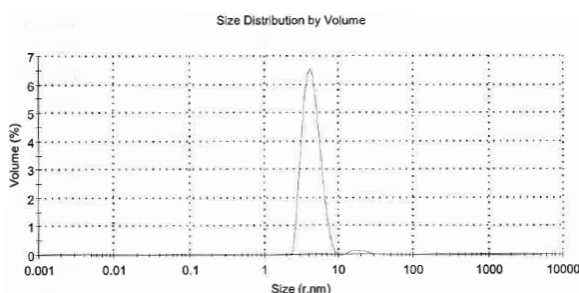
$$R_h = 4.6 \text{ nm}$$

PBS pH=7.4:



$$R_h = 4.9 \text{ nm} \\ + 18.5 \text{ nm (slight aggregation)}$$

Sodium phosphate pH=6.0:



$$\underline{R_h = 4.4 \text{ nm}}$$

Figure 3.63: DLS measurements of hydrodynamic radius (R_h) of the complex formed between EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} in three different pH buffers (Tris pH=8.0, PBS pH=7.4 and NaP pH=6.0). The smallest R_h is underlined.

These different R_h measurements from the DLS experiments (Tables 3.13) could also provide additional information when compared to the R_h of globular proteins, so called R_{min} (Table 3.14). R_{min} is therefore the minimal radius of a sphere that could contain the given mass of protein. With an R_h of 2.0 nm, 4.1R_{10kDa} and its 10 kDa size (calculated from amino acid composition) is comparable to a globular protein of at least 30 kDa. This shows that 4.1R_{10kDa} is neither globular nor even compact. EBA181₉₄₅₋₁₀₉₇ has a R_h equal to 2.95 nm and is therefore comparable to a globular protein of approximately 90 kDa. Regarding its size of 18 kDa, this shows that EBA181₉₄₅₋₁₀₉₇ is very likely to be unfolded. The same behaviour is noticeable with the complex between EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}, and its R_h of 4.4 nm. This would be the size of a 350 kDa globular protein, thus showing the disordered/unfolded character of this complex.

| | 4.1R_{10kDa} | EBA181₉₄₅₋₁₀₉₇ | | Complex |
|-------------------------|---|--|--|---|
| R_h | 2.0 nm <i>Sodium phosphate</i> <i>pH=6.0</i> | 2.95 nm <i>Sodium phosphate</i> <i>pH=6.0</i> | | 4.4 nm <i>Sodium phosphate</i> <i>pH=6.0</i> |

Table 3.13: Summary of the smallest hydrodynamic radius measured by DLS for EBA181₉₄₅₋₁₀₉₇, 4.1R_{10kDa} and the complex.

EBA181₉₄₅₋₁₀₉₇ injected in the PBS buffer. For all three controls, as would be expected with no interaction (i.e no heat released) between the reservoir and injected solutions, the calorimetry curve stayed flat.

Unfortunately, it was not possible to obtain an interpretable ITC curve for the complex formation from any of the subsequent tests performed. Aggregation of GST-4.1R_{10kDa} in the reservoir was noticed (Figure 3.64), even in varying the temperature from 4°C to 40°C or using different concentration of both proteins. This behaviour masked any changes of heat during the titration of GST-4.1R_{10kDa} with EBA181₉₄₅₋₁₀₉₇. It was therefore decided to use another method, micro-scale thermophoresis, to define the K_d of the interaction between the two proteins.

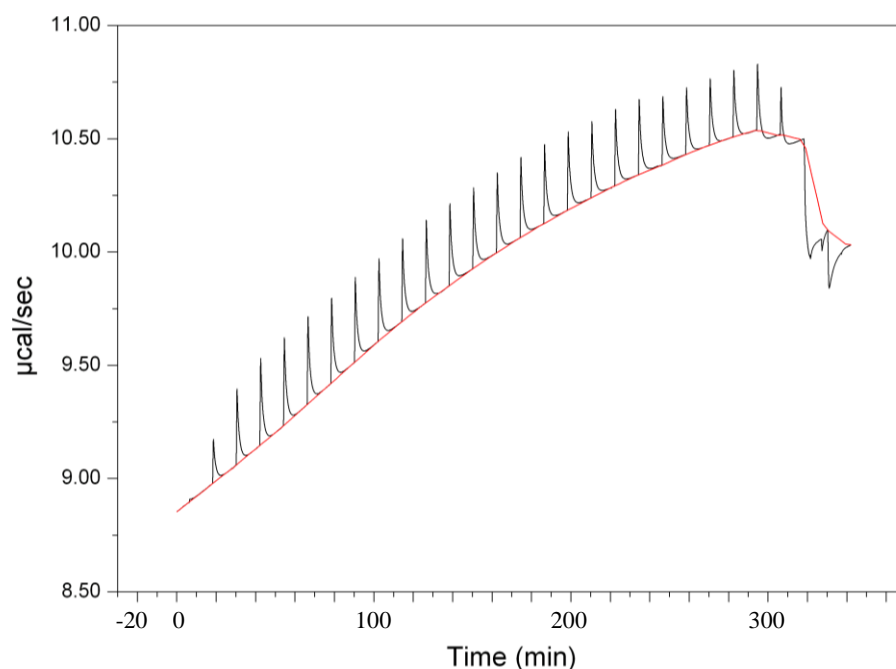


Figure 3.64: Uninterpretable ITC measurement recorded by titration of GST-4.1R_{10kDa} with EBA181₉₄₅₋₁₀₉₇.

3.5.8 Micro-scale thermophoresis (MST) studies of the complex

Since the ITC failed, micro-scale thermophoresis was used in order to determine the K_d of the complex between EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}. In contrast to ITC, the amount of material required for MST is low; it was therefore possible to use untagged 4.1R_{10kDa}. Moreover, since the GST-tag is cleaved, the K_d obtained could be considered to be more relevant.

The *Monolith NTTM Protein Labelling Kit RED-NHS* from *NanoTemper Technologies* was used to label the EBA181₉₄₅₋₁₀₉₇ protein, and 4.1R_{10kDa} was kept as the non-labelled partner. The *NT-647-NHS* reactive dye contains N-hydroxysuccinimide esters and forms highly stable crosslinks with proteins by reacting with protein primary amines. Thus since buffers that contain primary amines (Tris buffer for example) would significantly reduce protein labelling, a first step of buffer exchange was required using the labelling buffer from the kit. In a second step, the solid *NT-647-HNS* fluorescent dye was solubilised using DMSO, mixed with the EBA181₉₄₅₋₁₀₉₇ protein in a 1 to 1 molar ratio and incubated for 30 minutes in the dark. The main interest of this kit comes from the optimisation for elimination of unreacted free dye. To achieve this, the last step was the purification of the dye using a small desalting column equilibrated in the buffer chosen for the MST analysis (50 mM sodium phosphate pH 6.0, 50 mM NaCl and 5 mM β ME). Finally, the ratio protein to dye was checked (both were approximately equal to 10 μ M) and aliquots of labelled EBA181₉₄₅₋₁₀₉₇ were kept at -80°C.

To prepare the MST experiment, the concentration of EBA181₉₄₅₋₁₀₉₇ was kept constant (2 μM) while its partner 4.1R_{10kDa} was diluted 16 times (from 100 μM to 2 nM) by serial dilution assay method. 10 μl of each diluted 4.1R_{10kDa} was mixed with 10 μl of labelled EBA181₉₄₅₋₁₀₉₇ for a total of 16 capillaries. The experiment was carried out in triplicate at room temperature on the *NanoTemper Monolith NT.115 Instrument* and finally gave a K_d close to **0.3 μM** (Figure 3.65).

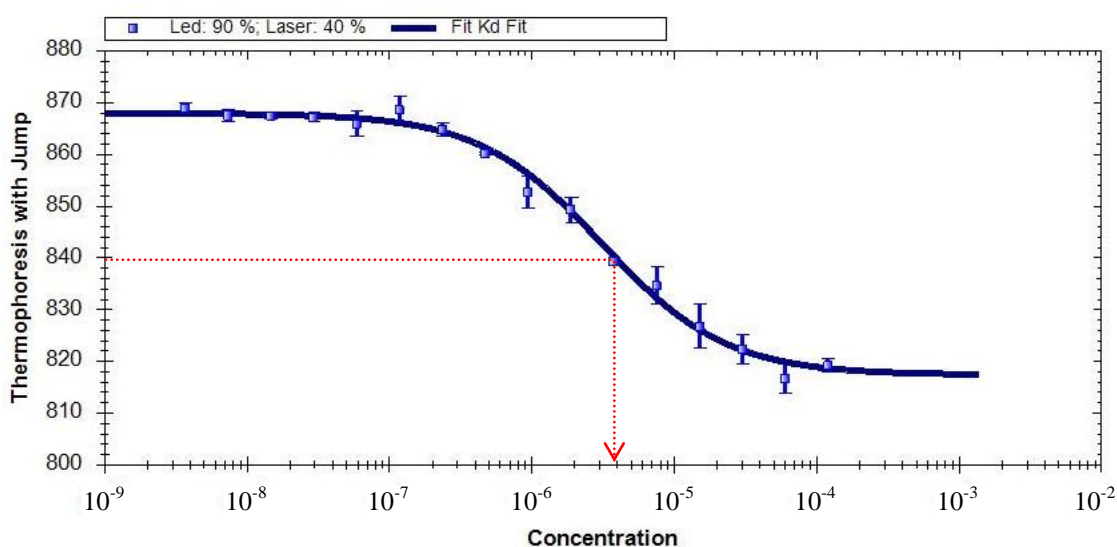


Figure 3.65: MST measurement of 4.1R_{10kDa} with labelled EBA181₉₄₅₋₁₀₉₇.

3.5.9 Crystallisation trials

As a matter of course, crystallisation trials were carried out on all of the protein systems studied in this thesis work. These included both the 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ proteins with and without their respective tags. Complexes between 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇, and between GST-4.1R_{10kDa} and the Pf30 peptide were also submitted to crystallisation

trials. The high-throughput crystallisation robot from the ‘High Throughput Crystallisation Platform’ of the EMBL was used, and manual crystallisation attempts were also carried out. Unfortunately these trials were unsuccessful.

3.5.9.1 Pre-crystallisation

The ‘PCT Pre-Crystallisation Test kit’ (*Hampton Research*) was used in order to establish the protein concentration range suitable for first crystallisation screenings. The results were 5 mg/ml for 4.1R_{10kDa} and 20 mg/ml for EBA181₉₄₅₋₁₀₉₇.

3.5.9.2 High-throughput and manual screenings

Screening for initial crystallisation was initially made with the use of the EMBL High Throughput Crystallisation robot. Sitting drop vapour diffusion at 4°C was made across 576 different standard conditions of various buffers, pH, salts and precipitants. Then, manual drops (gradients pH vs precipitant concentration) were set according to the best conditions from the crystallisation robot. The following trials were performed:

Individual proteins:

For His-EBA181₉₄₅₋₁₀₉₇ and EBA181₉₄₅₋₁₀₉₇, crystallisation attempts were carried out at protein concentrations of 20 mg/ml, 50 mg/ml and 90 mg/ml (the protein solubility limit of the protein).

For the GST-4.1R_{10kDa} construct, trials were carried out using a variety of buffer conditions, with protein concentrations of 4 mg/ml and 8 mg/ml (solubility limit). Some

crystals were obtained using Tris and HEPES buffer conditions, but diffraction tests at the ESRF showed them to be salt. For the 4.1R_{10kDa} domain, crystallisation trials were also performed, using two concentrations: 2.5 mg/ml and 5 mg/ml (solubility limit).

| Construct | Protein concentration |
|--------------------------------|---|
| His-EBA181 ₉₄₅₋₁₀₉₇ | <ul style="list-style-type: none"> ▪ 20 mg/ml ▪ 50 mg/ml ▪ 90 mg/ml (solubility limit) |
| EBA181 ₉₄₅₋₁₀₉₇ | <ul style="list-style-type: none"> ▪ 20 mg/ml ▪ 50 mg/ml ▪ 90 mg/ml (solubility limit) |
| GST-4.1R _{10kDa} | <ul style="list-style-type: none"> ▪ 4 mg/ml ▪ 8 mg/ml (solubility limit) |
| 4.1R _{10kDa} | <ul style="list-style-type: none"> ▪ 2.5 mg/ml ▪ 5 mg/ml (solubility limit) |

Table 3.15: Crystallisation trials of the individual tagged or untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins.

Complex:

Crystallisation trials of various forms of the complex were also carried out: EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} (at 3 mg/ml and 4.5 mg/ml, the latter being the solubility limit)

and EBA181₉₄₅₋₁₀₉₇ / GST-4.1R_{10kDa} (at 3 mg/ml, 5 mg/ml, 7 mg/ml). No crystals were observed.

For Pf₃₀ / GST-4.1R_{10kDa} crystallisation trials at 4 mg/ml and 8 mg/ml were carried out. (8 mg/ml was the solubility limit of the protein). Crystals were obtained from the 8 mg/ml trial of the Pf₃₀ / GST-4.1R_{10kDa} system in a variety of conditions (phosphate buffer and HEPES buffers). The crystals were reproduced by manual screening but X-ray diffraction tests at ESRF also showed them to be salt crystals.

| Construct | Protein concentration |
|--|---|
| EBA181 ₉₄₅₋₁₀₉₇ / 4.1R _{10kDa} | <ul style="list-style-type: none"> ▪ 3 mg/ml ▪ 4.5 mg/ml (solubility limit) |
| EBA181 ₉₄₅₋₁₀₉₇ / GST-4.1R _{10kDa} | <ul style="list-style-type: none"> ▪ 3 mg/ml ▪ 5 mg/ml ▪ 7 mg/ml |
| Pf ₃₀ / GST-4.1R _{10kDa} | <ul style="list-style-type: none"> ▪ 4 mg/ml ▪ 8 mg/ml (solubility limit) |

Table 3.16: Crystallisation trials of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex.

3.6 Discussion and conclusions

The EBA-181 protein sequence from the residue 945 to the residue 1097 and encompassing the binding region to the 4.1R protein was expressed, purified and intensively studied using biophysical and biochemical techniques. From the purification procedure, which includes size-exclusion chromatography, it was found that the EBA181₉₄₅₋₁₀₉₇ protein eluted in a long series of fractions and not as a sharp peak, with the beginning of the elution appearing earlier than for a globular protein of this weight (18 kDa), although MALLS data showed the protein to be monomeric in solution. This characterisation was the first of a series of observations that suggested the protein to be unfolded. 1D NMR experiments yielded a narrow ¹H spectrum, typical of unfolded protein. CD analysis also resulted in spectra that are typically associated with a random coil conformation. Moreover, TSA analysis showed the protein to not have a defined hydrophobic core, something that is characteristic of unfolded proteins. DLS measurements showed the protein to have a size (R_h) of at least 2.95 nm, corresponding to a 90 kDa globular protein. All other aspects of the protein characterisation were consistent with EBA181₉₄₅₋₁₀₉₇ existing in a disordered state – one that was very unlikely to crystallise. Hence it seems that the EBA181₉₄₅₋₁₀₉₇ domain is an intrinsically disordered protein, consistent with the information given by the structural and disorder predictions on the full-length protein and to all the biophysical characterisation performed on the EBA181₉₄₅₋₁₀₉₇ domain. According to the bioinformatics predictions on the full-length protein, the N-terminus part (which includes the transmembrane domain) appears to be structurally folded. Nevertheless, at least half of its sequence, after the N-terminus and almost until the C-terminus, is predicted be disordered. Moreover, and very interestingly,

the binding region to the 4.1R protein as well as all other predicted and putative binding regions can only be found in this unfolded region.

For the 4.1R_{10kDa} protein, the results were rather different. Following expression and purification, CD analysis showed a typical curve of helical content, which could be estimated as followed: 35 % of alpha helices, 20 % of turns and 45 % of random coil. DLS measurements showed a minimum size (R_h) of 2.0 nm, comparable to a globular protein of at least 30 kDa and showing the relatively low compactness of 4.1R_{10kDa}. 1D NMR showed a narrow spectrum, typical of disordered proteins or dynamic alpha helices. This last possibility of a dynamic behaviour including the appearance of alpha helices, is likely to be the biological relevant conformation, when comparing to the other characterisation. The consensus of all the biophysical characterisation performed shows that 4.1R_{10kDa} is neither globular nor even compact, but has some 3D structure with an estimation of 35 % of alpha helices. The protein as a whole is well known to be digested by proteases (*Leto et al., 1984*) as four well-defined domains of 30 kDa, 16 kDa, 10 kDa and 22 kDa; the 10 kDa one interacting with the EBA-181 protein. The 30 kDa domain is the only one with a known 3D structure, but the three other domains are very likely to have some secondary structures too, as shown by the trypsin digestion experiment and the structure and disorder predictions carried out.

The complex between EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} was also thoroughly characterised. The micro pull-down procedure, allowing the use of very small amounts of sample and the characterisation of a range of different conditions in parallel, confirmed the complex formation and showed that the complex formed independently of the presence or the

absence of the His or GST tag, carried respectively by the EBA181₉₄₅₋₁₀₉₇ or 4.1R_{10kDa} proteins. Moreover, the complex also formed during size exclusion chromatography (tagged proteins); or using native polyacrylamide gel electrophoresis (untagged proteins) with the appearance of a novel band after co-migration of the two proteins. Further pull-down analysis demonstrated the complex to be stable in various pH (from at least pH 5.0 to 9.0), in the presence of high salt concentration (up to 2 M NaCl), or even when in competition with shorter EBA peptides encompassing the 4.1R binding sequence. Quantification in gel filtration studies and in MALLS experiments provided new insights on the stoichiometry of the interaction, demonstrating that the EBA-4.1R complex is formed as a 1 to 1 complex. Interestingly, 1D NMR spectra did not show any folding upon binding, as it was also suggested by the DLS data (R_h of minimum 4.4 nm, corresponding to the size of a 350 kDa globular protein). Finally, MST analysis gave the K_d of the biological relevant complex between untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} to be close to 0.3 μ M, showing medium affinity of the two partners.

| | His-EBA181₉₄₅₋₁₀₉₇ | EBA181₉₄₅₋₁₀₉₇ | GST-4.1R_{10kDa} | 4.1R_{10kDa} |
|---------------------------------|--|---|---|---|
| Plasmid and tag | pET15b (Kan^R) with His tag | pET15b (Kan^R) with His tag | pGEX-4T-2 (Kan^R) with GST tag | pGEX-4T-2 (Kan^R) with GST tag |
| Expression yield | <i>~ 5 mg/l of culture in E.coli Rosetta2 (DE3) Cam^R</i> | <i>~ 5 mg/l of culture in E.coli Rosetta2 (DE3) Cam^R</i> | <i>~ 2 mg/l of culture in E.coli Rosetta2 (DE3) Cam^R</i> | <i>~ 0.5 mg/l of culture in E.coli Rosetta2 (DE3) Cam^R</i> |
| 1D NMR | Data suggest protein is not folded | Protein not folded |  | Protein not folded (or alpha helices invisible) |
| Circular Dichroism | Protein not folded (4°C and 25°C) | Protein not folded (4°C and 25°C) | Protein ~70% folded (30 % helices and 40 % beta) | Protein weakly folded (~35% of helices, 20% of turn and 45% of random coil) |
| Thermal Shift Assay | Protein not folded | Protein not folded | Protein folded, success in defining the best stability condition | Very flexible |
| Gel filtration and MALLS | Protein not globular => <i>Monomer</i> | Protein not globular => <i>Monomer</i> |  | Protein not globular => <i>Monomer</i> |
| DLS | Protein solution monodisperse, not globular (3.0 nm, size of a 100 kDa globular protein) | Protein solution monodisperse, not globular (2.95 nm, size of a 100 kDa globular protein) |  | Protein solution monodisperse, not globular (2.0 nm, size of a 30 kDa globular protein) |
| Crystallisation trials | No crystal 20 mg/ml 50 mg/ml 90 mg/ml | No crystal 20 mg/ml 50 mg/ml 90 mg/ml | Salt crystals 4 mg/ml 8 mg/ml | No crystal 2.5 mg/ml 5 mg/ml |

Table 3.17: Summary of the main results on individual tagged or untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins.

| | Complex |
|---------------------------------|---|
| Gel filtration and MALLS | Evidence of 1 to 1 stoichiometry EBA-181 / 4.1R: measured MW of 24 kDa |
| Quantification on gel | Evidence of 1 to 1 stoichiometry EBA-181 / 4.1R |
| Native gel | Appearance of a band consistent with complex formation |
| DLS | Hydrodynamic radius $R_h = 4.4$ nm |
| 1D NMR | No evidence of folding upon binding |
| ITC | Failed |
| MST | $K_d = 0.3 \mu\text{M}$ |
| Crystallisation trials | <p>EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} (3 mg/ml and 4.5 mg/ml) => no crystal</p> <p>EBA181₉₄₅₋₁₀₉₇ / GST-4.1R_{10kDa} (3 mg/ml, 5 mg/ml, 7 mg/ml) => no crystal</p> <p>Pf₃₀ / GST-4.1R_{10kDa} (4 mg/ml and 8 mg/ml) => salt crystals</p> |

Table 3.18: Summary of the main results on the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex.

4. SAXS AND SANS CHARACTERISATION OF EBA181₉₄₅₋₁₀₉₇, 4.1R_{10kDa} AND THEIR COMPLEX.

Abstract

The main goal of the work described in this chapter was to carry out solution structural studies of the individual proteins EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}, (tagged and untagged) as well as of the complex formed between them. The overall strategy was to apply complementary small-angle X-ray and neutron scattering techniques (SAXS and SANS respectively) in a way that provided low resolution information on the solution structures of the two proteins and of the complex.

The SAXS experiments demonstrated that both proteins are elongated and non-globular, consistent with the disordered structures as described in Chapter 3. SANS measurements on the complex were also carried out. Here full use of deuteration was exploited by making samples of the complex in which the EBA-181 protein was deuterated, and the 4.1R protein hydrogenated. This allowed H₂O/D₂O-based solvent contrast variation to be used in distinguishing and imaging the two parts of the intact complex. This also permitted a direct comparison of the complex structure with that derived from SAXS experiments on both labelled and unlabelled analogues of the EBA-181 / 4.1R system. These results show that (I) the SAXS and the SANS analyses provide good agreement in terms of the solution state modelling of the complex, (II) the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex adopts a 'T-shaped'

structure in which the 4.1R_{10kDa} molecule protrudes from the central region of the elongated EBA181₉₄₅₋₁₀₉₇ molecule, and (III) the EBA181₉₄₅₋₁₀₉₇ and the 4.1R_{10kDa} protein structures, at the resolution of the SAS analysis, appear to be relatively unchanged following complex formation, with their respective radius of gyration and D_{max} parameters remaining very similar between the uncomplexed and complexed forms of both proteins.

4.1 Introduction

As described in previous chapters, a clear interaction has been identified between the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} domains. The studies described in Chapter 3 have demonstrated that the EBA181₉₄₅₋₁₀₉₇ protein is intrinsically disordered and that the 4.1R_{10kDa} protein is also disordered to a large degree. One major consequence of this characterisation work is that neither the components of the complex nor the complex itself are likely to be amenable to traditional crystallography. In the circumstances, the only options remaining for structural information are solution state studies using small-angle X-ray scattering (SAXS), small-angle neutron scattering (SANS) or nuclear magnetic resonance (NMR).

This chapter focuses on solution scattering work carried out on this system using both SAXS and SANS. The project exploited the optimised expression system described in Chapter 3 as well as facilities of the PSB in Grenoble, including SAXS beamlines of the ESRF and the SANS instrument of the ILL. Both approaches are strongly complementary

in the information they provide, and in both cases *ab-initio* modelling can be applied to image the proteins or protein components under study. In the case of the SANS work, it was necessary to label the EBA-181 protein with deuterium so that solvent contrast variation could be used to distinguish between the two components of the complex.

Section 4.2 describes the material and methods used to perform the scattering experiments, including sample preparation for SAXS/SANS and the unique deuteration methods used in SANS for contrast variation (see Chapter 2 for further technical details). Section 4.3 describes the structural characterisation results using small-angle X-ray scattering on the individual proteins as well as the complex. Section 4.4 presents the small-angle neutron scattering structural characterisation on the complex using deuteration and contrast variation. Finally, Section 4.5 summarises the combined SAXS and SANS analyses in the context of the other information that has been gathered on the EBA-181/4.1R interaction.

4.2 Material and methods

This study required the large scale production of both EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins (tagged and untagged) for SAXS. In the case of SANS, an additional requirement was the production of deuterium-labelled EBA181₉₄₅₋₁₀₉₇. The following sections describe how these proteins were produced and characterised prior to the small-angle scattering experiments.

4.2.1 Sample production for small-angle X-ray scattering

The large-scale expression and the different purification steps of the individual proteins were performed as previously described in Chapter 3. The tagged complex (His-EBA181₉₄₅₋₁₀₉₇ / GST-4.1R_{10kDa}) was formed as described in Chapter 3, whereas the untagged complex (EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa}) was formed by mixing the individual untagged proteins in a 1 to 1 molar ratio, as shown by the MALLS analysis (Chapter 3). In order to optimise the quality of the final SAXS experiment a variety of buffers were evaluated following the final purification steps using gel filtration, DLS and preliminary SAXS measurements. These measurements were designed to detect specific issues of importance for the analysis of the small-angle scattering data, in particular aggregation issues. The buffers used in this evaluation were: Tris buffer at pH 8.0, PBS buffer at pH 7.4, sodium phosphate (NaP) at pH 7.0 and NaP buffer at pH 6.0. The NaCl concentration was another variable parameter, as well as the final protein concentration in each sample (Figure 4.2). Typical DLS measurements on the proteins and their complex are shown in Figure 4.1.

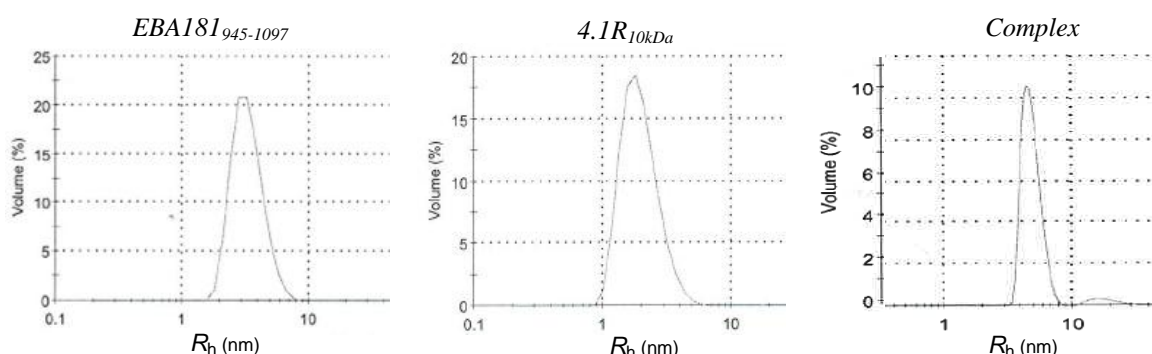


Figure 4.1: DLS measurements showing the individual 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ proteins and the complex to be monodisperse after the final step of purification (in NaP buffer at pH 6.0 + 50 mM NaCl).

4.2.2 Sample production for small-angle neutron scattering

As described in Chapter 2, small-angle neutron scattering is a powerful technique whereby solution contrast can be used to distinguish between, and model, specific components of a biological macromolecular complex, especially if deuteration methods are deployed. In these SANS studies of the EBA181 / 4.1R protein complex, the EBA181₉₄₅₋₁₀₉₇ protein was deuterated at approximately 75 % in ‘non-exchangeable’ positions (as noted in Chapter 2, this level of deuteration was chosen so as to produce material that would be optimally matched out in pure D₂O), and the 4.1R_{10kDa} protein was kept hydrogenated. In principle either protein could have been deuterated, but in this case EBA-181 was chosen because of its high expression level, as opposed to the low expression of 4.1R_{10kDa} (see Chapter 3).

One of the important issues that arise in connection with the use of deuterated proteins and D₂O based solvents is the fact that there is anecdotal evidence that deuteration increases aggregation problems when handling protein samples. Hence some care was needed in evaluating this potential problem prior to the SANS experiments.

4.2.2.1 Bacterial adaptation to deuterium

In order to be able to grow in deuterated medium and thus express the protein of interest to a sufficient level, bacteria need an adaptation process. To achieve this and express the deuterated EBA181₉₄₅₋₁₀₉₇ protein, 30 cycles of 10 ml bacterial cultures in 85 % deuterated Enfors medium with Kan + Cam antibiotics were performed as follows: the initial 10 ml culture was initiated using 1 ml of stationary phase LB culture of *E. coli* RosettaTM 2 cells carrying the plasmid for the expression of the His-EBA181₉₄₅₋₁₀₉₇ protein (initiated from

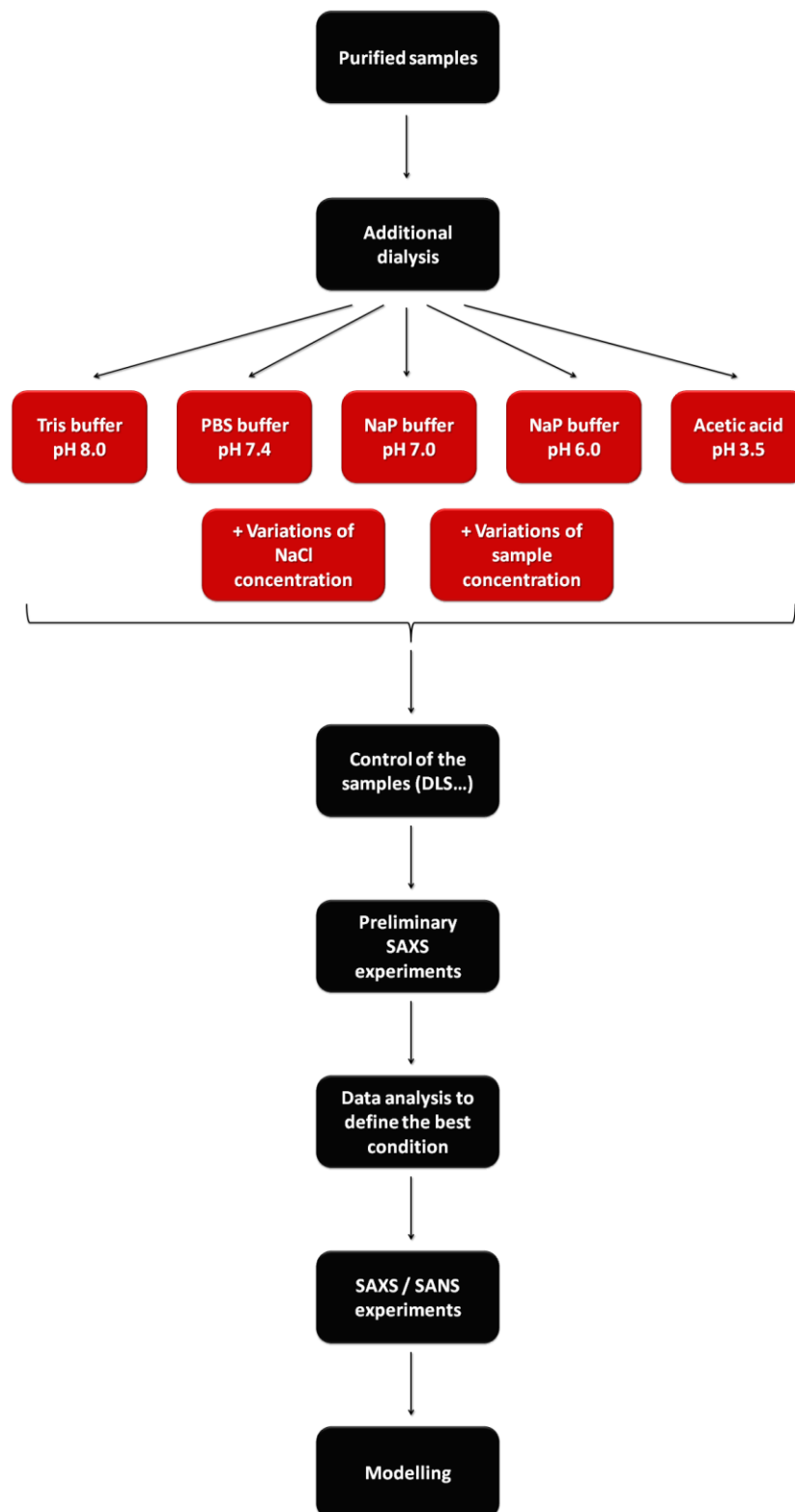


Figure 4.2: Summary of the main steps of EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} purification protocol and sample preparation for the SAXS experiments.

the kanamycin resistant clone stored at -80°C, see Chapter 3). The 29 subsequent cycles were initiated using 1 ml of the previous stationary phase culture and 9 ml of fresh Enfors medium. Each step required an O.N. growth at 30°C under agitation.

The deuterium-adapted *E. coli* Rosetta™ 2 expressing the His-EBA181₉₄₅₋₁₀₉₇ protein was stored at -80°C. For this, the 30th adaptation cycle (having reached the stationary phase) was used to make long-term storage stocks using sterile beads in a cryo-preservative solution (*Technical Service Consultants Ltd*).

Table 4.1 A and B presents the composition of the bacterial growth media used for the expression of deuterated EBA181₉₄₅₋₁₀₉₇ protein (modified from *Enfors et al., 2000*). The powders were resuspended in 85% D₂O, and the medium was sterilised by filtering through a 0.2 µm filter.

| Enfors medium | Concentration |
|---|----------------------|
| (NH ₄) ₂ SO ₄ anhydrous | 6.86 g/l |
| KH ₂ PO ₄ anhydrous | 1.56 g/l |
| Na ₂ HPO ₄ anhydrous | 6.48 g/l |
| (NH ₄) ₂ -H-citrate anhydrous | 0.49 g/l |
| 100% hydrogenated glycerol | 5 g/l |
| MgSO ₄ 1M | 1 ml/l |
| Metal salts* | 1 ml/l |

Table 4.1 A: Enfors growth medium used for EBA181₉₄₅₋₁₀₉₇ protein deuteration.

| *Metal salts (x1000) | Concentration |
|--------------------------------------|----------------------|
| CaCl ₂ .2H ₂ O | 0.5 g/l |
| FeCl ₃ .6H ₂ O | 16.7 g/l |
| ZnSO ₄ .7H ₂ O | 0.18 g/l |
| CuSO ₄ .5H ₂ O | 0.16 g/l |
| MnSO ₄ .4H ₂ O | 0.15 g/l |
| CoCl ₂ .6H ₂ O | 0.18 g/l |
| Na-EDTA anhydrous | 20.1 g/l |

Table 4.1 B: Metal salts used to prepare the Enfors growth medium.

4.2.2.2 Expression and solubility tests

Since deuteration might have an impact on expression and solubility of proteins, it was necessary to perform new expression and solubility tests to determine the condition where the expression and solubility of the EBA181₉₄₅₋₁₀₉₇ protein was the highest (Figure 4.3).

The results show that an overnight expression at 30°C gave the best expression and solubility when compared to the three other conditions tested (37°C for 3h, 37°C for 7h and 20°C for 48h).

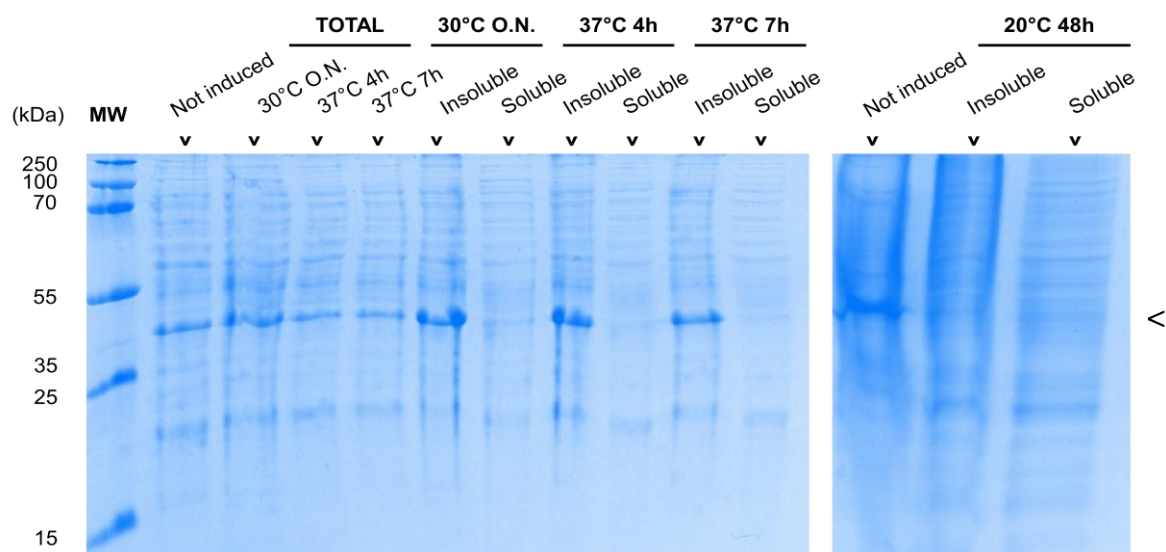


Figure 4.3: SDS-PAGE of expression tests of His-EBA181₉₄₅₋₁₀₉₇ in 85 % D-Enfors medium. The sign “<” indicates the migration position of His-EBA181₉₄₅₋₁₀₉₇.

4.2.2.3 Large-scale expression of deuterated EBA181₉₄₅₋₁₀₉₇ and hydrogenated

4.1R_{10kDa} proteins

In order to express the hydrogenated 4.1R_{10kDa} protein, *E. coli* Rosetta™ 2 (DE3) cells were transformed with the pGEX-4T-2 plasmid (protocol similar to Section 3.2.3) and plated on LB-plates (+ Kan and Cam antibiotics). The selected medium for the large-scale hydrogenated protein expression was the LB medium supplemented with 30 µg/ml of kanamycin and with 34 µg/ml of chloramphenicol. Firstly, precultures were inoculated from the LB plates (LB-Agar, Kan + Cam) from the transformation and incubated until reaching the stationary phase of growth (37°C, overnight under agitation). Further preculture steps were required to inoculate each of the 1 litre culture (1:50 dilution). The

bacteria were grown at 37°C to an optical density of around 0.7 at 600 nm. Protein expression was induced with 1 mM IPTG over 3 hours as described in Chapter 3; and the bacteria were then lysed by sonication to release the proteins. The suspension was then centrifuged at 40 000 g for 1 h and at 4°C to separate the soluble material as showed in Figure 4.4.

The selected medium for the large-scale expression of the 75 % deuterated EBA181₉₄₅₋₁₀₉₇ was the Enfors medium in 85 % D₂O, supplemented with 30 µg/ml of kanamycin and 34 µg/ml of chloramphenicol. By using unlabelled glycerol as carbon source, the protein produced was approximately 75% deuterated in 85 % D₂O medium. In order to express the deuterated EBA181₉₄₅₋₁₀₉₇ protein, the deuterium-adapted *E. coli* Rosetta™ 2 (DE3) cells expressing the His-EBA181₉₄₅₋₁₀₉₇ protein (pET15b) were used (from the -80°C long-term storage clone) and plated on a LB plate (LB-Agar, Kan + Cam). Precultures were inoculated from this Petri dish and incubated until reaching the stationary phase of growth (37°C, overnight under agitation). Finally, 1 litre of 85 % D-Enfors was inoculated with the last preculture (1:50 dilution) and split into two 3 litre flasks in order to maximise the aeration. The bacteria were grown at 30°C to an optical density of around 0.7 at 600 nm. Protein expression was induced with 1 mM IPTG overnight as previously defined in Figure 4.2, and the bacteria were lysed by sonication to release the proteins. The suspension was then centrifuged at 40 000 g for 1 h and at 4°C to separate the soluble material as showed in Figure 4.4.

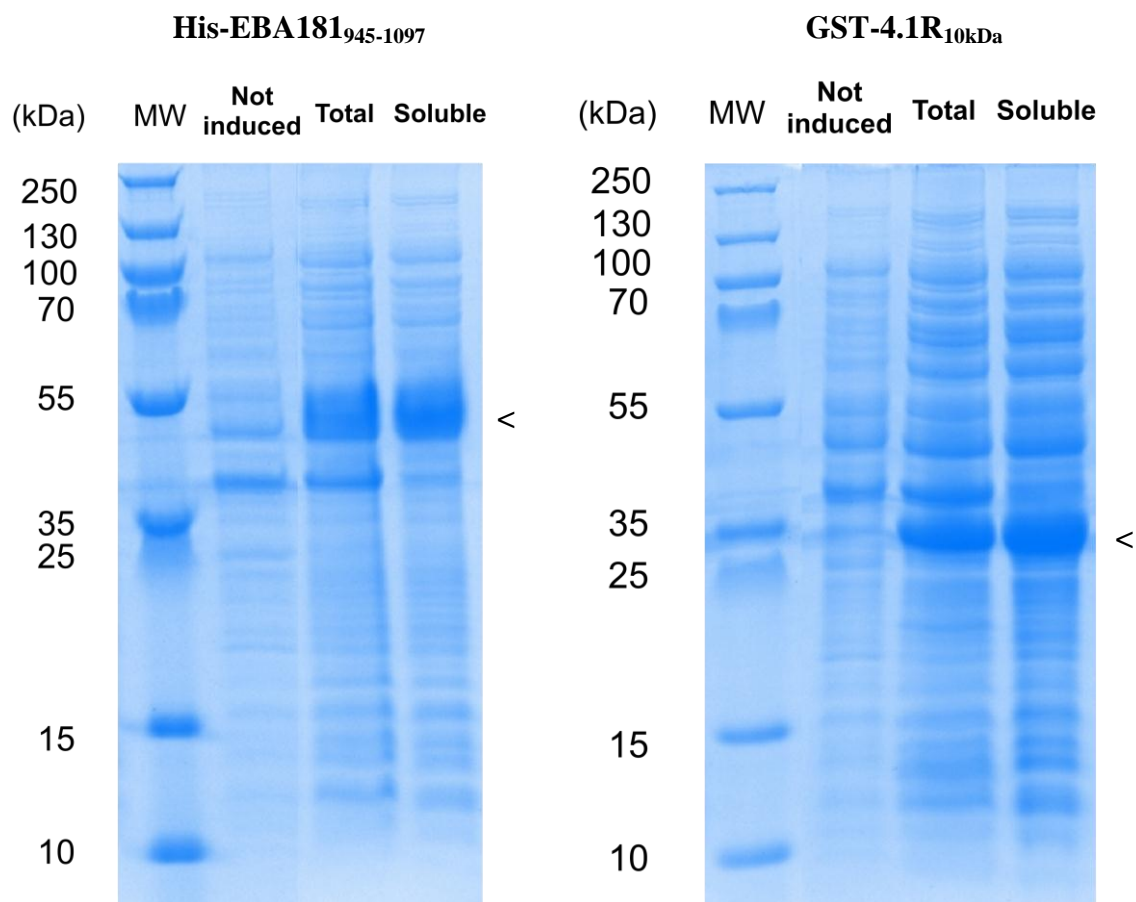


Figure 4.4: **Left:** SDS-PAGE showing the expression of deuterated His-EBA181₉₄₅₋₁₀₉₇ after adaptation and **right:** hydrogenated GST-4.1R_{10kDa}. The sign “<” indicates the migration position of the respective proteins.

MW: Molecular weight marker (*PageRuler™ Plus Prestained SM181* from *Fermentas*).

Not induced: Sample before induction with IPTG.

Total: Insoluble and soluble protein after IPTG induction.

Soluble: Soluble proteins after IPTG induction and centrifugation.

Both deuterated EBA181₉₄₅₋₁₀₉₇ and hydrogenated 4.1R_{10kDa} proteins were subsequently purified as described in Chapter 3 and the complex was formed by mixing the individual proteins in a 1 to 1 molar ratio. In order to allow the contrast variation experiment, the final

buffer (50 mM NaP buffer + 50 mM NaCl) was used either as 100 % hydrogenated (the solvent is pure H₂O), or as respectively 40 % and 100 % deuterated. To achieve this, the deuteration of the inorganic materials was achieved by flash evaporation: the hydrogenated NaP buffer was evaporated at 60°C with the use of a rotary evaporator system and the resulting powder was resuspended in the minimal required volume of D₂O. This process was performed three times in order to remove traces of H₂O. Finally, the powder was resuspended in 40 % or 100 % D₂O to prepare the two buffers required for the contrast variation experiment. SDS-PAGE and DLS measurements were performed on the purified deuterated EBA181₉₄₅₋₁₀₉₇ and hydrogenated 4.1R_{10kDa} proteins to control the quality of the purification and check for aggregation. The deuterated EBA181₉₄₅₋₁₀₉₇ samples as well as the hydrogenated 4.1R_{10kDa} samples were monodisperse in each of the 0 %, 40 % or 100 % D₂O based NaP pH 6.0 buffers. Typical DLS measurements in 100 % D₂O are shown in Figure 4.5.

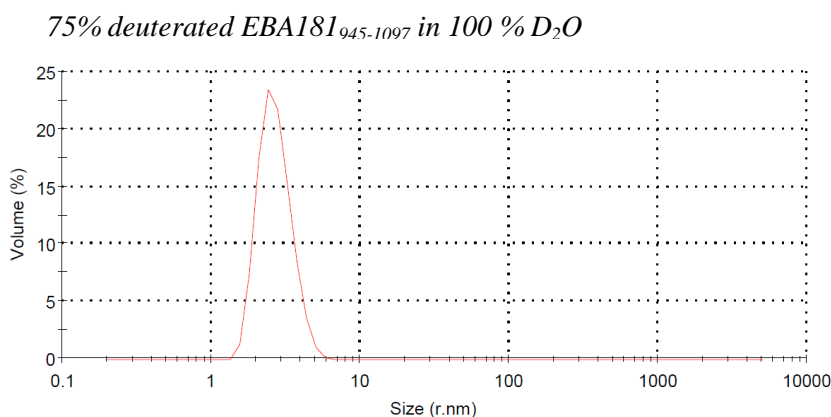
4.2.3 SAXS and SANS data collection

4.2.3.1 BioSAXS

Small-angle X-ray scattering measurements were performed at the BioSAXS beamlines of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Two similar SAXS beamlines were used: ID14-3 and BM29 (*Pernot et al., 2013*). Prior to the measurements each sample was centrifuged and filtered to remove aggregates. The scattering measurements were performed using the capillary pre-installed on the beamline (1 mm width) under temperature control (4°C and 20°C) using the sample changer robot (see Chapter 2) and recorded using a 2D sensitive pixel detector (Pilatus). The

sample-to-detector distance was 2.841 metres with an incident radiation wavelength (λ) of 0.99 Å, providing an accessible q -range from 0.22 to 4 nm⁻¹. 10 frames were recorded for each sample (1 second per frame under constant flow). Each single measurement was manually inspected for radiation damage and then averaged in order to maximise the signal/noise ratio. The different dialysis buffers corresponding to the samples were also measured (before and after every sample) to allow their background scattering to be subtracted from the sample measurements.

A)



B)

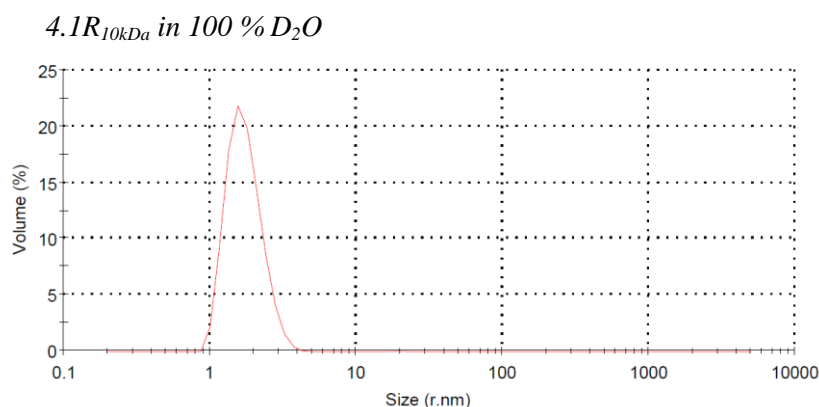


Figure 4.5: DLS measurements showing the deuterated EBA181₉₄₅₋₁₀₉₇ (**A**) and hydrogenated 4.1R_{10kDa} (**B**) proteins to be monodisperse after the final step of purification and in the 100 % D₂O NaP pH 6.0 buffer.

| | |
|---------------------------------|---|
| Sample-detector distance | 2.841 metres |
| Wavelength | 0.99 Å |
| q-range | 0.22 to 4 nm ⁻¹ |
| Exposure time | 10 x 1 second frames |
| Sample temperature | 4°C and 20°C |
| Sample concentration | 4.1R _{10kDa} : 0.4 to 3 mg/ml GST-4.1R _{10kDa} : 3 mg/ml EBA181 ₉₄₅₋₁₀₉₇ : 5 to 30 mg/ml His-EBA181 ₉₄₅₋₁₀₉₇ : 10 to 30 mg/ml 4.1R _{10kDa} / EBA181 ₉₄₅₋₁₀₉₇ complex: 5 mg/ml |

Table 4.2: Summary of the main parameters of the SAXS experiments.

4.2.3.2 SANS

The small-angle neutron scattering measurements presented here were performed on the D22 instrument of the Institut Laue Langevin (ILL, Grenoble, France) as described in Chapter 2.

The samples were placed in quartz cuvettes and set in the sample holder which allows temperature control. During the experiments carried out on D22, the instrument was set up to work at a wavelength of 8 Å with sample-detector distances of either 2.0 metres, 5.6 metres or 17 metres; and the temperature control set to 4°C (as a reminder, in standard

conditions D22 covers a total q -range of 4×10^{-4} to 0.44 \AA^{-1} no detector offset, or 0.85 \AA^{-1} with detector offset). The data were collected via the two dimensional ^3He multidetector, before being integrated to obtain the final intensity vs diffusion angle curve.

| | |
|---------------------------------|---|
| Sample-detector distance | 2.0 metres / 5.6 metres / 17 metres |
| Wavelength | 8 \AA |
| q-range | 4×10^{-4} to 0.44 \AA^{-1} or 4×10^{-4} to 0.85 \AA^{-1} |
| Exposure time | 30 minutes |
| Sample temperature | 4°C |
| Sample concentration | 4.1R _{10kDa} / EBA181 ₉₄₅₋₁₀₉₇ complex: 2 mg/ml (for the three solvent contrasts chosen: 0 %, 40 % and 100 % D ₂ O) |

Table 4.3: Summary of the main parameters of the SANS experiment.

4.2.4 Data treatment

For each SAXS sample the 10 recorded frames and the two buffer measurements were automatically averaged using the beamline software AUTOSUB (beamline software BsxCuBE and ATSAS package – EMBL) in order to obtain one averaged sample scattering curve and one averaged buffer scattering curve. The data were then analysed manually using the PRIMUS (Konarev *et al.*, 2003) software for buffer subtraction, data reduction, as well as curve merging for the different sample concentrations and the Guinier (low q) / Porod (high q) analysis.

The SANS data were analysed using a range of software (from *R. Ghosh, ILL*): RNILP was used to list the transmission and scattering data, RNILS to calculate the radial distribution function, DETCI to adjust the position of the centre of the detector (as a function of its distance) in order to mask it using the RMSK program. SPOLLY was then used in order to analyse further the data by means of three different including cadmium and empty cell scattering measurements as well as the transmission from a H₂O sample. Following this, each sample was corrected for the scattering from their respective buffer.

For both SAXS and SANS data, GNOM (*Svergun et al., 1992*) was then used to define the D_{\max} and the P(r) function. The modelling of each individual analysed curve was carried out using the DAMMIF (*Franke et al., 2009*) software (set in slow mode) in order to create an ensemble of models. The averaging of these different models coming from DAMMIF was carried out using the DAMAVER program (*Volkov et al., 2003*).

4.3 Structural characterisation using small-angle X-ray scattering

In order to optimise the SAXS data and perform a complete characterisation, both tagged and untagged versions of the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins were analysed. For this a wide range of buffers, pH, NaCl and sample concentrations was tested to find the most suitable conditions for SAXS measurements. The main results on the individual proteins are presented below in two sections (Sections 4.3.1 and 4.3.2) and the corresponding tables

(Tables 4.4 and 4.5). The results presented in bold in the tables correspond to the conditions that gave the highest quality SAXS measurements in terms of aggregation and signal/noise ratio and which were chosen thereafter to perform the structural characterisation and the *ab-initio* modelling.

The complex formed between the two EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins was also characterised using a similar approach. The different conditions tested and scattering results are presented in Section 4.3.3 with the summary shown in Table 4.6. Finally, the untagged complex displaying the best scattering quality measurement was chosen and used in order to perform the structural analysis and the *ab-initio* modelling.

4.3.1 SAXS characterisation of tagged GST-4.1R_{10kDa} and

His-EBA181₉₄₅₋₁₀₉₇ proteins

Protein samples of GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇ were analysed by SAXS using several conditions and optimisation in order to reach the highest quality possible and obtain a reliable structural characterisation of both proteins. The results are summarised in Table 4.4.

The GST-4.1R_{10kDa} protein was kept in the PBS buffer at pH 7.4 since its higher iso-electric point strongly constrains the possible range of pH that can be used (the theoretical pI of GST-4.1R_{10kDa} is 6.6) and thus the use of the dialysis conditions presented in Figure 4.2 and especially the buffers with a pH close to its iso-electric point could lead

to its precipitation. This makes the use of buffers with a pH below 7 impossible. Two NaCl concentrations were used, 137 mM and 1 M. The scattering from the centrifuged sample in PBS in 1 M NaCl was nearly identical to the scattering of its buffer indicating protein precipitation and gave unusual results with a nonsensical R_g . Facing the difficulty to concentrate GST-4.1R_{10kDa} to more than 3-4 mg/ml, only two sample concentrations were tested: 1 and 3 mg/ml. In 137 mM NaCl they both show a similar scattering profile (Table 4.4), thus the data from the sample concentrated to 3 mg/ml and presenting a slightly better signal to noise ratio was kept for the *ab-initio* modelling. The radius of gyration of GST-4.1R_{10kDa} (molecular weight = 35 kDa) is close to 4 nm (R_g results are equal to 4.17 nm at very low scattering angles q in the Guinier analysis, and 4.14 nm at higher q in the Porod region), with a calculated Porod volume of 187 nm³. The P(r) function analysis gave a slightly smaller R_g , at 3.71 nm, and a D_{max} equal to 12 nm. These results suggest that the protein is not globular and quite elongated since the theoretical R_g of a 35 kDa globular protein would be close to 2.0 nm (*Bernado et al., 2009*, see calculation below).

For globular proteins: R_g (Å) $\sim 6.25 \times M^{1/3}$ (M in kDa)

For unfolded proteins: R_g (Å) $\sim 8.05 \times M^{0.522}$ (M in kDa)

It is also interesting to compare the results of GST-4.1R_{10kDa} with the SAXS measurements performed on the GST tag alone, giving a R_g of approximately 2.5 nm and a D_{max} of 6.6 nm (molecular weight = 26 kDa). The addition of the 4.1R_{10kDa} protein to the GST tag adds only 9 kDa, but the measured R_g drastically increases, from approximately 2.5 nm to 4.2 nm. The same behaviour can be noticed when comparing the D_{max} between the GST tag

alone and the GST-4.1R_{10kDa} construct, with a rise from 6.6 nm to 12 nm. These results clearly demonstrate the non-globular and elongated state of the 4.1R_{10kDa} protein which dominates the hydrodynamic radius compared to its molecular weight.

For the His-EBA181₉₄₅₋₁₀₉₇ protein two different buffers and pH were chosen according to DLS results displaying the lowest aggregation (data not shown): Tris pH 8.0 and NaP pH 6.0. The scattering results from the sample in sodium phosphate buffer at pH 6.0 gave the highest quality results as expected after the different biophysical characterisations performed (Chapter 3). The scattering quality of this experiment was sufficient to perform *ab-initio* modelling, and thus no further conditions were tested. The measurement performed in Tris buffer at pH 8.0, with His-EBA181₉₄₅₋₁₀₉₇ concentrated at 14 mg/ml, gave a result almost as good as in NaP pH 6.0. The increase of protein concentration to 20 mg/ml resulted in dramatic aggregation during the SAXS irradiation as highlighted by the Guinier analysis, with a R_g almost doubled to 4.21 nm. The radius of gyration of His-EBA181₉₄₅₋₁₀₉₇ (measured in NaP pH 6.0) was determined from the 15 mg/ml experiment in the range from 2.31 nm (Porod region) to 2.91 nm (P(r) function), with the Guinier analysis giving a R_g of 2.40 nm. The D_{max} is equal to 9.1 nm which is very wide for an 18 kDa protein, highlighting the very elongated state of the protein; the theoretical R_g of a globular protein of the same mass would be close to 1.6 nm.

| Sample | Buffer | NaCl concentration | Sample concentration (mg/ml) | Guinier region | | Porod region | | P(r) function | |
|--------------------------------|-------------------|--------------------|------------------------------|----------------|-----------|--------------|-----------------------|---------------|----------------|
| | | | | R_g (nm) | MM (kDa) | R_g (nm) | Vp (nm ³) | R_g (nm) | D_{max} (nm) |
| GST-4.1R _{10kDa} | PBS pH 7.4 | 137 mM | 1 | 3.98 | 45 | 4.02 | 182 | 4.16 | 13.9 |
| | PBS pH 7.4 | 137 mM | 3 | 4.17 | 41 | 4.14 | 187 | 3.71 | 12.0 |
| | PBS pH 7.4 | 1 M | 3 | 0.28 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| GST tag alone | PBS pH 7.4 | 137 mM | 5 | 2.60 | 22 | 2.58 | 99 | 2.41 | 6.6 |
| His-EBA181 ₉₄₅₋₁₀₉₇ | Tris pH 8.0 | 100 mM | 14 | 2.40 | 14 | 2.31 | 11 | 2.90 | 9.3 |
| | Tris pH 8.0 | 100 mM | 20 | 4.21 | 40 | 4.24 | 24 | 2.51 | 9.0 |
| | NaP pH 6.0 | 100 mM | 15 | 2.40 | 21 | 2.31 | 15 | 2.91 | 9.1 |

Table 4.4: Summary of the experimentally determined structural SAXS parameters and results of tagged His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} proteins. The data sets shown in bold were chosen thereafter to perform structural characterisation and *ab-initio* modelling. *N.D.*: non-determined values due to insufficiently good data.

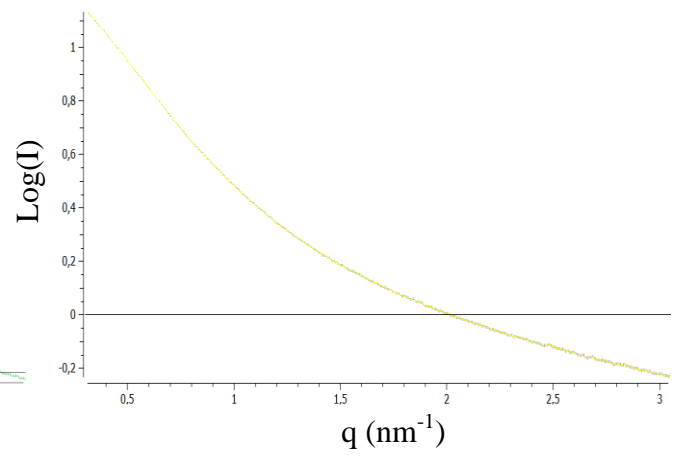
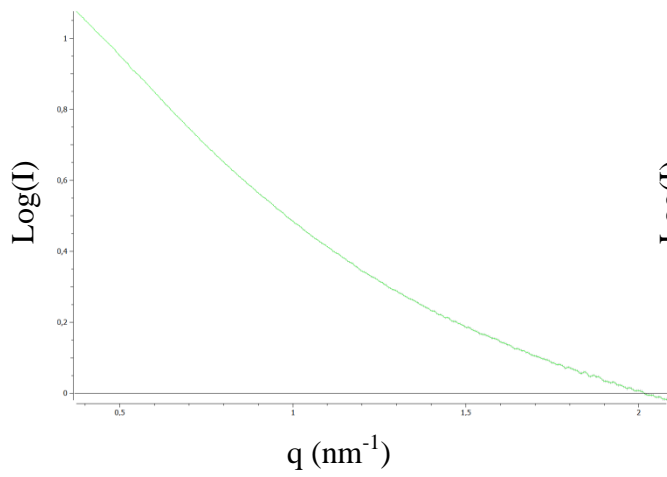
The small-angle X-ray scattering curve of GST-4.1R_{10kDa}, after subtraction of the scattering data from the buffer alone and reduction of the data in the range from 0.4 nm⁻¹ to 2.1 nm⁻¹, is displayed on the left hand side of Figure 4.6 A. The associated paired vector distribution function P(r) which shows the histogram of all inter-atomic distance vectors, was calculated and is displayed in Figure 4.6 B.

The small-angle X-ray scattering curve of His-EBA181₉₄₅₋₁₀₉₇ after subtraction of the data recorded from the buffer alone and reduction of the data in the range from 0.3 nm⁻¹ to 3 nm⁻¹ is displayed on the right hand side of Figure 4.8 A. The associated P(r) was calculated and is displayed on the right hand side of Figure 4.8 A.

GST-4.1R_{10kDa}

His-EBA181₉₄₅₋₁₀₉₇

A)



B)

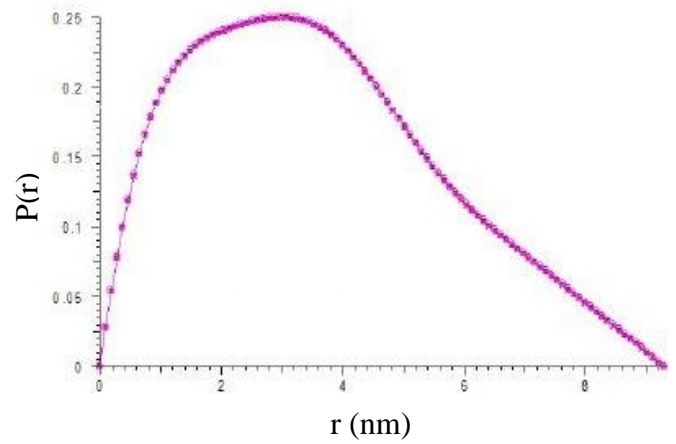
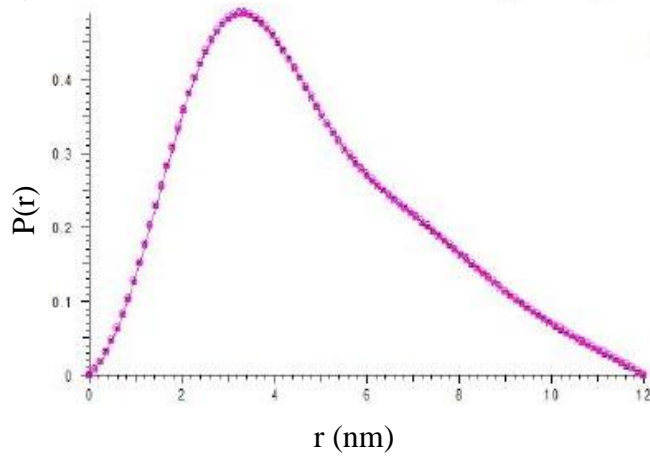
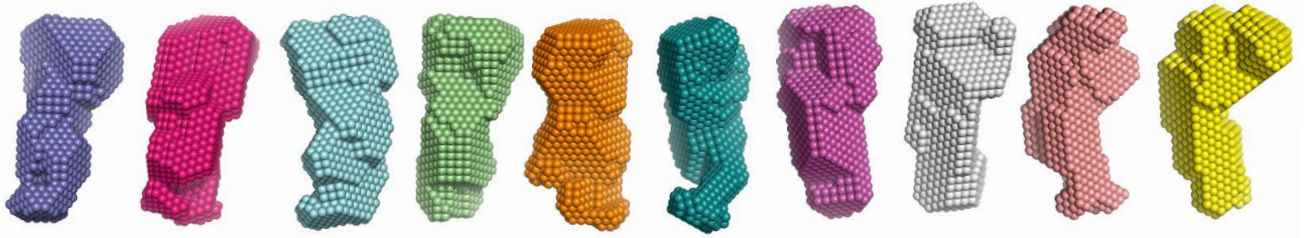


Figure 4.6: Left: GST-4.1R_{10kDa} and Right: His-EBA181₉₄₅₋₁₀₉₇ with (A) buffer corrected SAXS curves and (B) associated calculated $P(r)$.

Ab-initio modelling was performed on both GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇ proteins, and ten models were created as respectively shown in Figures 4.7 A and 4.8 A. The final models after averaging the 10 different 3D models for each of the tagged proteins using DAMAVER are presented in Figures 4.7 B and 4.8 B. The SAXS *ab-initio* models of GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇ are consistent with the biophysical characterisation results (DLS, CD and structural predictions). His-EBA181₉₄₅₋₁₀₉₇ appears to be a very elongated protein in contrast to GST-4.1R_{10kDa} which has a more compact core and structural elements due to the GST tag. In order to validate the models, the back-calculated scattering curves of His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} from the models were calculated and are shown in green on the experimental red scattering curve and show a very good match (Figures 4.8 B and 4.9 B, right hand side).

A)



B)

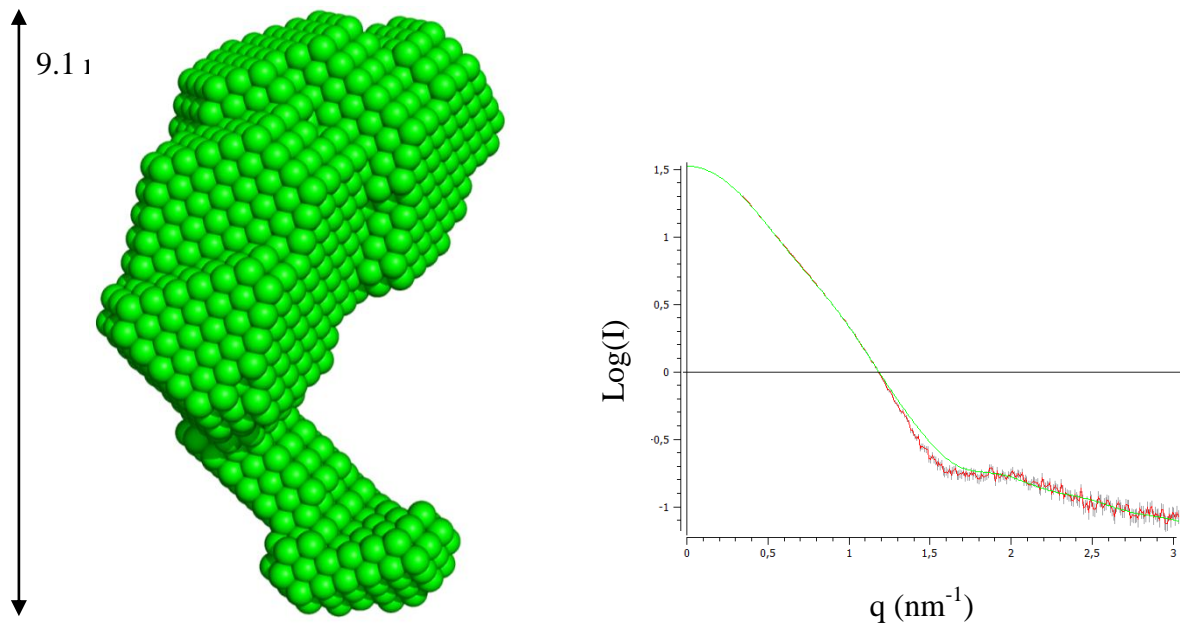
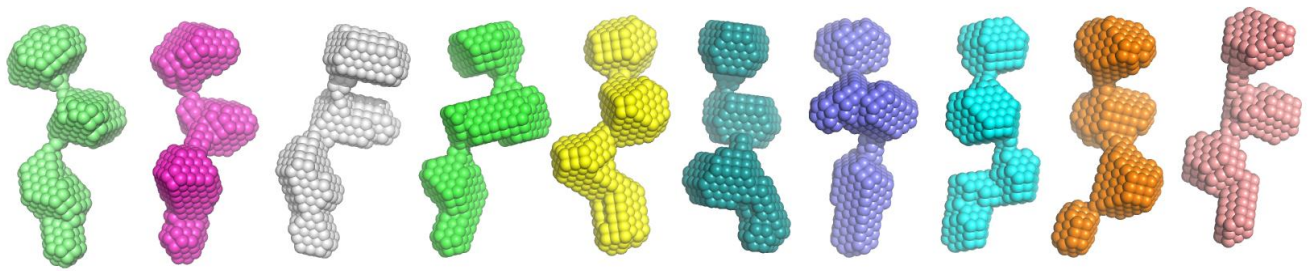


Figure 4.7: Tagged GST-4.1R_{10kDa} with **(A)** *ab-initio* 3D models using DAMMIF (10 models) and **(B)** averaged *ab-initio* 3D model using DAMAVER and displayed as spheres (**left**) and fit of the simulated scattering curve in green versus the experimental data in red (**right**).

A)



B)

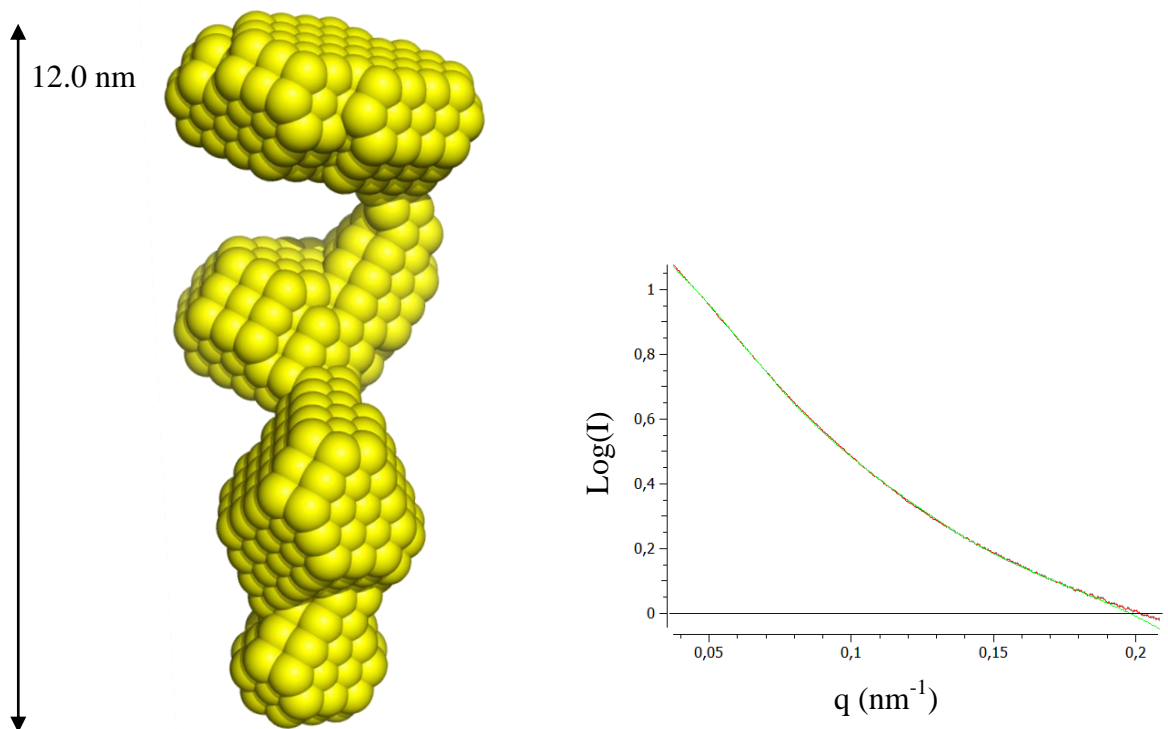


Figure 4.8: Tagged His-EBA181₉₄₅₋₁₀₉₇ with **(A)** *ab-initio* 3D models using DAMMIF (10 models) and **(B)** averaged *ab-initio* 3D model using DAMAVER and displayed as spheres (**left**) and fit of the simulated scattering curve in green versus the experimental data in red (**right**).

The disordered state of a biological macromolecule can be qualitatively assessed by means of a Kratky plot (*Kratky et al., 1955*), when the scattering intensities for macromolecules which behave as random coils would plateau in a $q^2 \times I(q)$ vs q plot. Thus, both His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} as well as the GST tag alone scattering measurements were plotted as a Kratky plot (Figure 4.9). The Kratky plot analysis confirms the disordered state of the His-EBA181₉₄₅₋₁₀₉₇ protein with the characteristic plateau displayed by the blue scattering curve in Figure 4.6. In contrast, the GST tag alone and the GST-4.1R_{10kDa} scattering curves do not display this plateau, demonstrating their more largely non-disordered state (red and green curves in Figure 4.9), as would be expected.

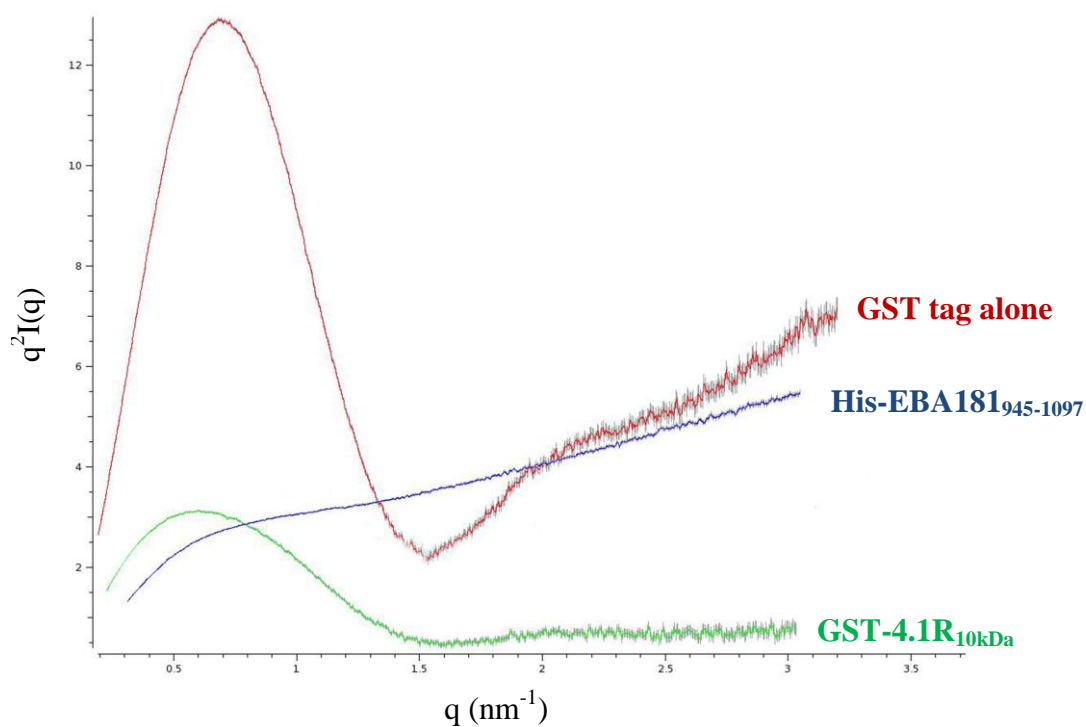


Figure 4.9: Kratky plot of tagged His-EBA181₉₄₅₋₁₀₉₇ (blue), GST-4.1R_{10kDa} (green) and the GST tag alone (red).

Figure 4.10 shows the 3D *ab-initio* model of the GST-4.1R_{10kDa} protein. The GST protein tag *ab-initio* model resulting from the SAXS measurement is also displayed on the right hand side to allow a direct comparison with the *ab-initio* model of tagged GST-4.1R_{10kDa} protein, as well as the X-ray crystal structure of human GST from the PDB (ref. 9GSS, resolution = 1.97 Å). The SAXS measurement of the GST tag alone and with the X-ray crystal structure allowed its probable location to be found in the 3D model of GST-4.1R_{10kDa}, as displayed in Figure 4.10. The GST tag appears to be globular in solution, as would be expected, while the 4.1R_{10kDa} protein is more elongated.

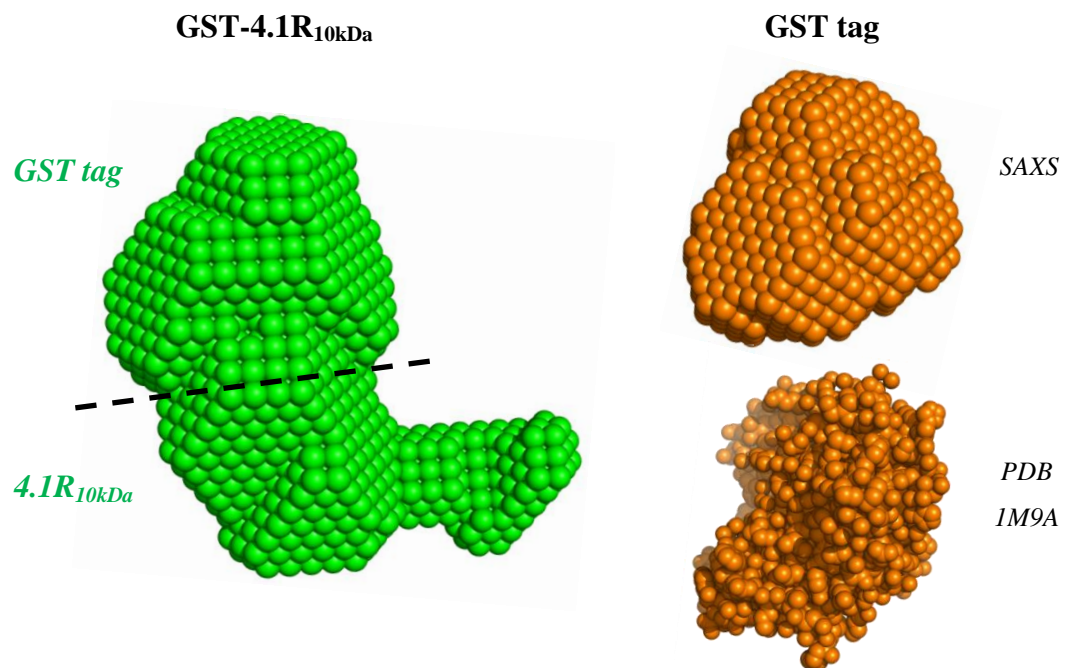


Figure 4.10: *Ab-initio* 3D models displayed as spheres of **left:** the tagged GST-4.1R_{10kDa} protein and **right:** the GST tag alone from SAXS (top) and crystal structure of the GST protein from *Schistosoma japonicum* (bottom, pdb 1M9A, resolution 2.1 Å).

4.3.2 SAXS characterisation of untagged 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ proteins

Protein samples of untagged 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ were analysed by SAXS using a wide range of conditions and optimisation, including buffer type, pH variation or NaCl and sample concentrations (Table 4.5) in order to optimise the SAXS.

The 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ protein samples were dialysed in three different types of buffer: Tris, PBS and NaP; allowing a pH range from 8.0 to 6.0 to be analysed. Several samples were tested and studied for each of these conditions in order to optimise the data quality, particularly at low and high scattering angle values; and to prevent sample aggregation, radiation damage and optimise the signal/noise ratio.

The 4.1R_{10kDa} scattering results in Tris buffer at pH 8.0 display similar values of R_g for all of the conditions tested, with values approximately equal to 2 nm. The D_{max} measured is between 6.2 nm and 9.3 nm. It should be noted that the increase of the NaCl concentration seems to increase the apparent D_{max} of the protein, with the highest values found at 250 mM and 1M NaCl. This means that the addition of ions and charges tend to stretch the protein making it even more elongated, but with no change in the calculated R_g . The highest sample concentration (3 mg/ml) resulted in precipitation of the protein, with a R_g value measured to be almost null. This shows once again the difficulties to obtain concentrated samples of untagged 4.1R_{10kDa}. All of the trials in PBS at pH 7.4 and NaP at pH 7.0 also resulted in precipitation of the protein, with R_g values almost null. These two buffers were thus eliminated in the optimisation process. The two measurements performed at pH 6.0 in the NaP buffer gave the highest quality results; the data from measurements on

the sample with protein concentration of 1.5 mg/ml were used for further analysis and the *ab-initio* modelling process.

The EBA181₉₄₅₋₁₀₉₇ scattering experiments performed at pH 8.0 in the Tris buffer gave all similar results. In 100 mM NaCl the R_g is comprised between 1.7 nm and 3.0 nm (Guinier, Porod and P(r) analysis) and the D_{max} between 6.2 nm and 7.2 nm. When the NaCl concentration is increased to 250 mM or 1 M, the R_g slightly increases and the D_{max} reaches 9.7 nm, showing an effect of the charges on the EBA181₉₄₅₋₁₀₉₇ protein. As observed for 4.1R_{10kDa}, the addition of ions and charges tend to stretch the protein making it more elongated (slight increases of the D_{max} and R_g). The scattering results performed in PBS at pH 7.4 and acetic acid at pH 3.5 were unsuccessful, with precipitation of the protein. The trials in NaP at pH 7.0 gave similar results to those in Tris at pH 8.0. The best scattering results were obtained with a lower pH at 6.0, with a R_g equal to 2.35 nm, 2.30 nm or 1.75 nm when calculated respectively using the Guinier approximation, the Porod law or the P(r) function. The D_{max} was measured to be equal to 7.9 nm which, in addition to the very small calculated value for the Porod volume, indicates a very elongated shape for the EBA181₉₄₅₋₁₀₉₇ protein.

| Sample | Buffer | NaCl concentration | Sample concentration (mg/ml) | Guinier region | | Porod region | | P(r) function | |
|----------------------------------|--------------------|--------------------|--------------------------------------|----------------|-----------|--------------|-----------------------|---------------|----------------|
| | | | | R_g (nm) | MM (kDa) | R_g (nm) | Vp (nm ³) | R_g (nm) | D_{max} (nm) |
| 4.1R_{10kDa} | Tris pH 8.0 | 100 mM | 0.4 | 2.05 | 11 | 1.93 | 1.5 | 1.89 | 6.2 |
| | Tris pH 8.0 | 100 mM | 0.8 | 2.04 | 11 | 1.95 | 1.3 | 2.07 | 7.5 |
| | Tris pH 8.0 | 100 mM | 1 | 2.06 | 16 | 2.03 | 3.2 | 2.18 | 7.4 |
| | Tris pH 8.0 | 250 mM | Combination of 0.4 + 0.8 + 1.6 + 3.2 | 2.02 | 11 | 1.93 | 3.4 | 2.15 | 6.2 |
| | Tris pH 8.0 | 250 mM | 1.5 | 2.36 | 15 | 2.35 | 4.1 | 2.56 | 8.3 |
| | Tris pH 8.0 | 250 mM | 3 | 0.67 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | Tris pH 8.0 | 1 M | 2.5 | 2.02 | 47 | 2.06 | 3.5 | 3.19 | 9.3 |
| | PBS pH 7.4 | 133 mM | 1.5 | 0.76 | 0.6 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | PBS pH 7.4 | 133 mM | 3 | 0.87 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | PBS pH 7.4 | 1 M | 2.5 | 0.89 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | NaP pH 7.0 | 100 mM | 1.5 | 0.39 | 0.5 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | NaP pH 7.0 | 100 mM | 3 | 0.94 | 1 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | NaP pH 6.0 | 100 mM | 1.5 | 2.02 | 10 | 1.90 | 3 | 1.89 | 6.0 |
| | NaP pH 6.0 | 100 mM | 3 | 2.16 | 15 | 2.1 | 5 | 2.03 | 6.3 |
| EBA181₉₄₅₋₁₀₉₇ | Tris pH 8.0 | 100 mM | 10 | 2.33 | 21 | 2.04 | 6 | 1.69 | 6.2 |
| | Tris pH 8.0 | 100 mM | 30 | 2.96 | 26 | 3.03 | 9 | 2.14 | 7.2 |
| | Tris pH 8.0 | 250 Mm | 5 | 3.23 | 26 | 3.22 | 11 | 2.82 | 9.1 |
| | Tris pH 8.0 | 250 Mm | 10 | 3.26 | 29 | 3.15 | 11 | 2.80 | 9.1 |
| | Tris pH 8.0 | 1 M | 8 | 3.25 | 30 | 3.12 | 11 | 2.95 | 9.7 |
| | PBS pH 7.4 | 133 mM | 5 | 0.47 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | PBS pH 7.4 | 133 mM | 10 | 0.62 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | PBS pH 7.4 | 1 M | 10 | 0.65 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | NaP pH 7.0 | 100 mM | 5 | 3.12 | 34 | 3.35 | 12 | 3.08 | 7.6 |
| | NaP pH 7.0 | 100 mM | 10 | 3.15 | 34 | 3.36 | 12 | 3.09 | 7.5 |
| | NaP pH 6.0 | 100 mM | 5 | 2.32 | 20 | 2.32 | 12 | 1.69 | 6.3 |
| | NaP pH 6.0 | 100 mM | 10 | 2.35 | 21 | 2.30 | 15 | 1.75 | 7.9 |
| | Acetic Acid pH 3.5 | 250 Mm | 2 | 0.20 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |

Table 4.5: Summary of the experimentally determined structural SAXS parameters of the untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins. The data sets highlighted in bold were chosen to perform the structural characterisation and the *ab-initio* modelling. *N.D.*: non-determined values due to insufficiently good data.

The small-angle X-ray scattering curve of untagged 4.1R_{10kDa}, after subtraction of the scattering data from the buffer alone and reduction of the data to the range 0.4 nm⁻¹ to 2.8 nm⁻¹, is displayed on the left hand side of Figure 4.11 A. The associated P(r) was calculated and is shown in Figure 4.11 B. After subtraction of the data recorded from the buffer alone and reduction of the data to the range 0.05 nm⁻¹ to 4.2 nm⁻¹ to keep a good signal/noise ratio, the small-angle X-ray scattering curve of untagged EBA181₉₄₅₋₁₀₉₇ is displayed on the right hand side of Figure 4.11 A. The calculated associated P(r) is displayed in Figure 4.11 B.

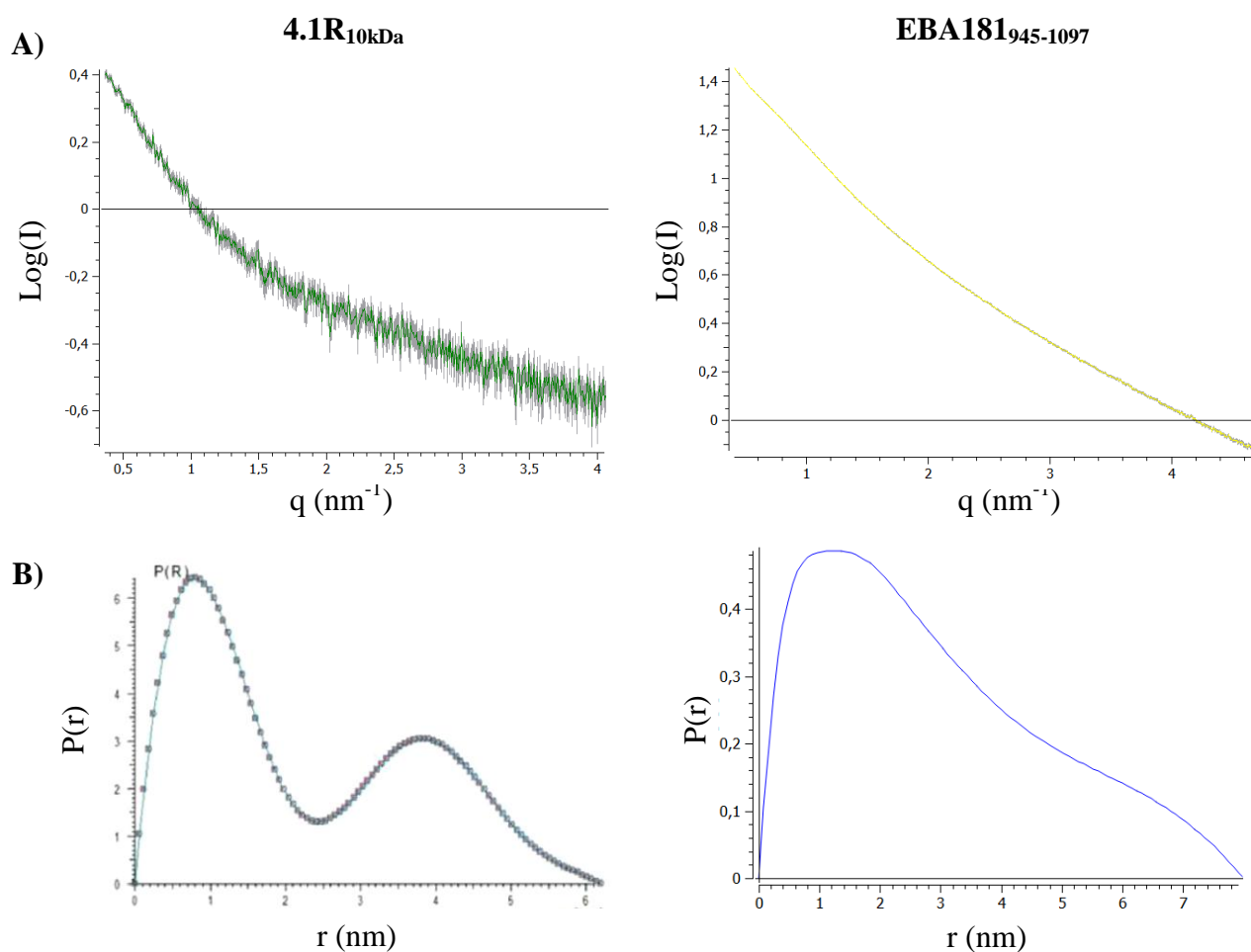


Figure 4.11: Untagged 4.1R_{10kDa} (**left**) and EBA181₉₄₅₋₁₀₉₇ (**right**) with (**A**) buffer corrected SAXS curves and (**B**) associated calculated P(r).

Ab-initio modelling was performed using the data from both untagged 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ proteins and ten models were created as shown in Figures 4.12 A and 4.13 A respectively. For each of the two proteins the ten models were averaged using DAMAVER (Figures 4.12 B and 4.13 B). In order to validate them, the respective back-calculated scattering curves from the different averaged models were also calculated and are shown in green on each experimental red scattering curve. The experimental and calculated curves show a very good match (Figures 4.12 B and 4.13 B, right hand side).

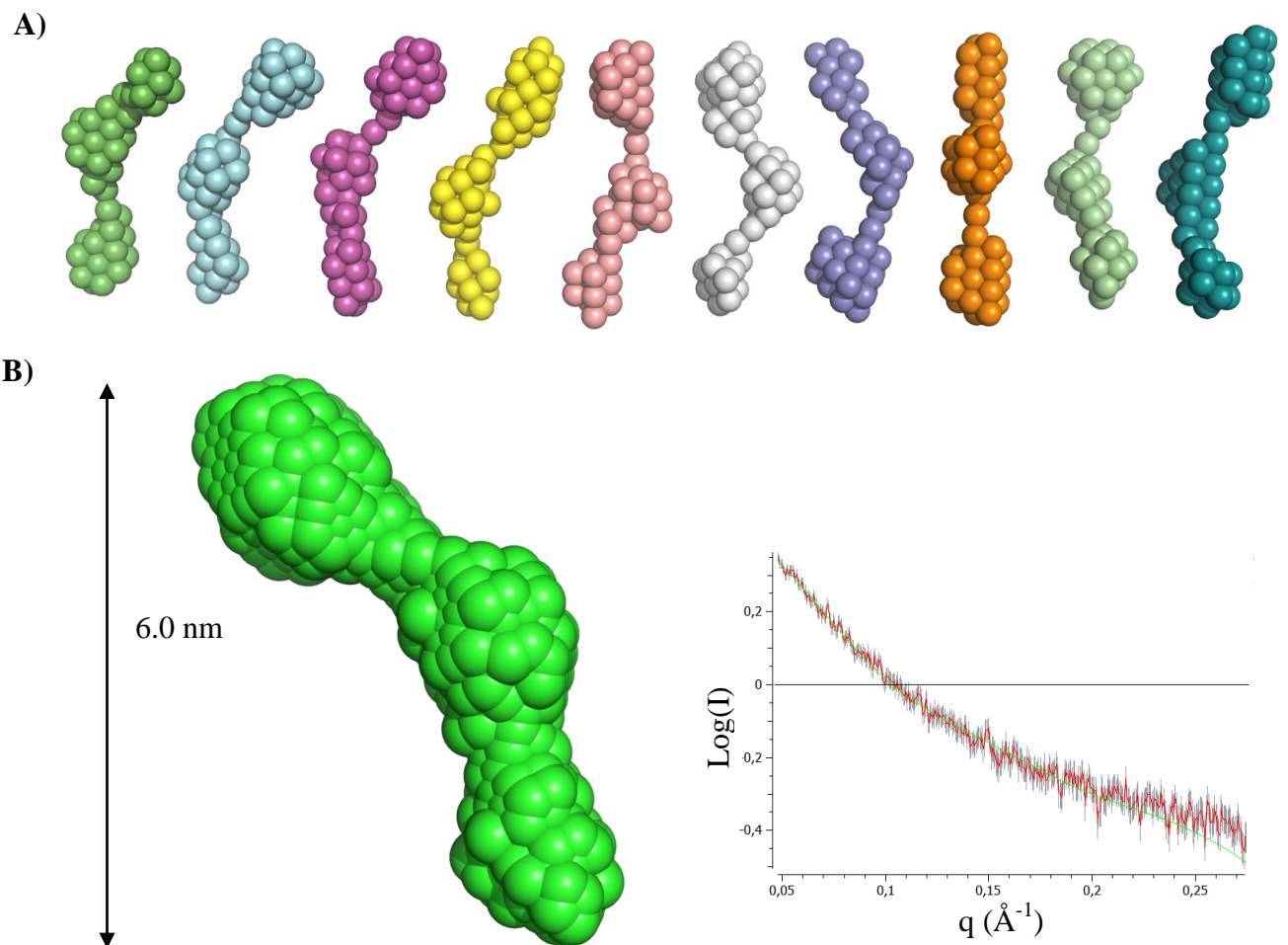


Figure 4.12: Untagged 4.1R_{10kDa} with **(A)** *ab-initio* 3D models using DAMMIF (10 models) and **(B)** averaged *ab-initio* 3D model using DAMAVER and displayed as spheres **(left)** and fit of the simulated scattering curve in green versus the experimental data in red **(right)**.

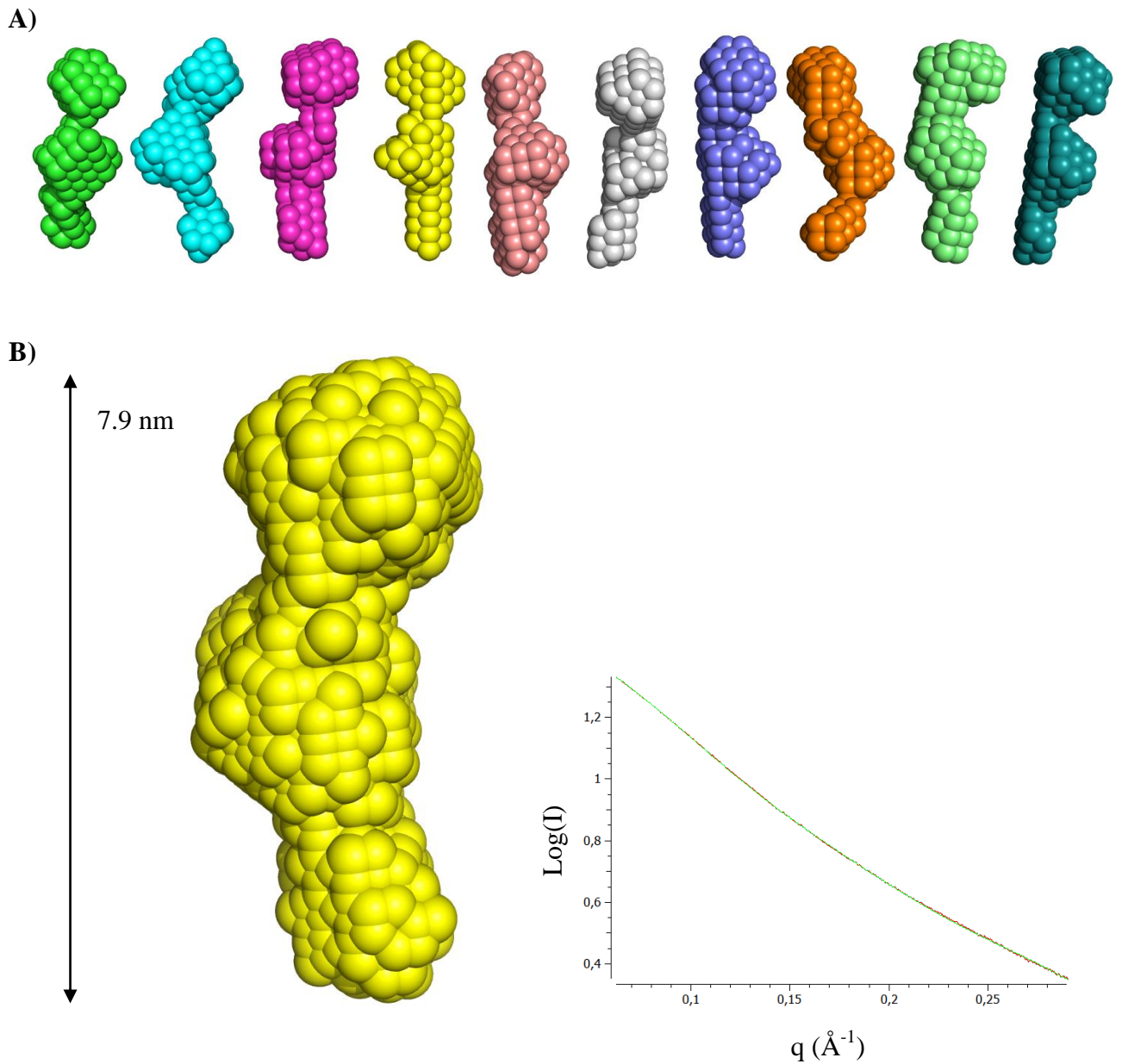


Figure 4.13: Untagged EBA181₉₄₅₋₁₀₉₇ with **(A)** *ab-initio* 3D models using DAMMIF (10 models) and **(B)** averaged *ab-initio* 3D model using DAMAVER and displayed as spheres (**left**) and fit of the simulated scattering curve in green versus the experimental data in red (**right**).

The disordered/ordered state of the untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins have been assessed by the mean of Kratky plot (expressed as $q^2I(q)$ as a function of q).

The Kratky plot analysis of the EBA181₉₄₅₋₁₀₉₇ protein (Figure 4.14, yellow curve) confirmed its disordered state, as already demonstrated with the analysis on tagged His-EBA181₉₄₅₋₁₀₉₇. The Kratky curve on the 4.1R_{10kDa} protein shows the presence of a plateau followed by a slow descent (Figure 4.14, green curve) revealing the protein to be partially unfolded. This is in agreement with the biophysical characterisation performed previously, and particularly the CD analysis showing around 35 % of folding (Chapter 3).

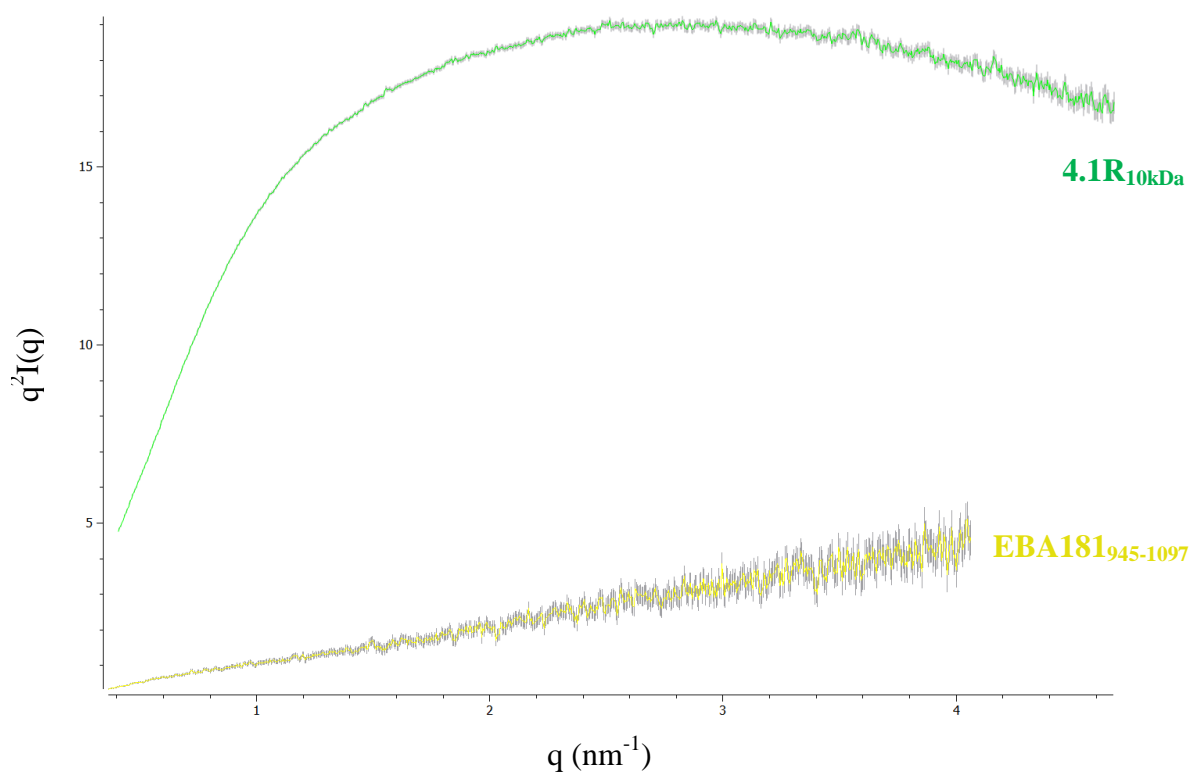


Figure 4.14: Kratky plot of untagged EBA181₉₄₅₋₁₀₉₇ (yellow) and 4.1R_{10kDa} (green).

4.3.3 SAXS characterisation of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex

The protein complex between the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins was analysed by SAXS using a wide set of conditions, optimisation and protein types as listed in Table 4.6. This analysis included the main goal of this study, namely the complex between 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ as untagged proteins. Other constructs were also analysed: the complex between 4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇, between GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇, and between GST-4.1R_{10kDa} or 4.1R_{10kDa} and the two peptides used in Chapter 3 for the biochemical pull-down assays (Pf30, the 30 amino acid peptide encompassing the binding region to 4.1R_{10kDa}; and Pr30, its homologous region from the *Plasmodium reichenowi* parasite).

The 4.1R_{10kDa} / EBA181₉₄₅₋₁₀₉₇ and the 4.1R_{10kDa} / His-EBA181₉₄₅₋₁₀₉₇ complexes were analysed in the different conditions as previously defined for the individual proteins, using a variety of buffers and pH from pH 8.0 to 6.0. The GST-4.1R_{10kDa} / EBA181₉₄₅₋₁₀₉₇ complex was only measured in the PBS buffer, because of the isoelectric point of the GST-4.1R_{10kDa} protein, as already mentioned. Due to the low expression level and the difficulties to concentrate the 4.1R_{10kDa} protein, the concentrations of the different samples were kept in the range 1 to 4 mg/ml. As firstly demonstrated in Chapter 3 and then confirmed with the SAXS measurements on the individual proteins, the measurements performed on the complex gave the best results at pH 6.0, where the least sample aggregation was observed and for which the smallest radius of gyration was calculated. The R_g was equal to 3.53 nm and the D_{max} to 10.4 nm, showing an increase compared to the respective values for the individual 4.1R_{10kDa} or EBA181₉₄₅₋₁₀₉₇ proteins. All of the

different conditions SAXS measurements were analysed but only the most relevant result on the untagged complex is described here-after. The scattering curve of the run at pH 6.0 was used for *ab-initio* modelling.

| Complex | Buffer | NaCl concentration | Sample concentration (mg/ml) | Guinier region | | Porod region | | P(r) function | |
|---|-------------------|--------------------|------------------------------|----------------|-----------|--------------|-----------------------|---------------|----------------|
| | | | | R_g (nm) | MM (kDa) | R_g (nm) | Vp (nm ³) | R_g (nm) | D_{max} (nm) |
| EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} | Tris pH 8.0 | 100 mM | 2 | 3.81 | 29 | 3.84 | 13.1 | 3.70 | 10.8 |
| | Tris pH 8.0 | 1M | 1.5 | 3.80 | 32 | 3.85 | 13.0 | 3.72 | 10.9 |
| | PBS pH 7.4 | 133 mM | 2 | 3.94 | 30 | 3.98 | 14.1 | 3.94 | 11.1 |
| | PBS pH 7.4 | 133 mM | 3 | 3.93 | 29 | 3.88 | 14.2 | 3.93 | 11.2 |
| | NaP pH 7.0 | 100 mM | 2 | 3.82 | 28 | 3.84 | 13.7 | 3.81 | 10.7 |
| | NaP pH 7.0 | 100 mM | 4 | 3.83 | 28 | 3.90 | 13.8 | 3.79 | 10.7 |
| | NaP pH 6.0 | 100 mM | 2 | 3.70 | 26 | 3.72 | 12.9 | 3.56 | 10.4 |
| | NaP pH 6.0 | 50 mM | 2 | 3.54 | 26 | 3.54 | 12.8 | 3.53 | 10.4 |
| His-EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} | Tris pH 8.0 | 100 mM | 2 | 3.98 | 72 | 3.99 | 13.1 | 4.02 | 13.13 |
| | Tris pH 8.0 | 1M | 1.5 | 3.96 | 73 | 3.94 | 13.0 | 4.07 | 13.26 |
| | PBS pH 7.4 | 133 mM | 2 | 0.55 | 2 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | PBS pH 7.4 | 133 mM | 3 | 4.05 | 77 | 4.06 | 13.2 | 4.16 | 13.58 |
| | NaP pH 7.0 | 100 mM | 2 | 0.52 | 1 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | NaP pH 7.0 | 100 mM | 4 | 0.45 | 1 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| | NaP pH 6.0 | 100 mM | 2 | 3.88 | 32 | 3.92 | 13.1 | 3.79 | 10.8 |
| | NaP pH 6.0 | 50 mM | 4 | 3.89 | 32 | 3.81 | 13.1 | 3.69 | 10.9 |
| His-EBA181₉₄₅₋₁₀₉₇ / GST-4.1R_{10kDa} | PBS pH 7.4 | 133 mM | 1 | 4.09 | 89 | 4.12 | | 4.27 | 14.32 |
| Pf30 / GST-4.1R_{10kDa} | PBS pH 7.4 | 133 mM | 1 | 0.57 | 3 | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> | <i>N.D.</i> |
| Pf30 / 4.1R_{10kDa} | Tris pH 8.0 | 250 mM | 2 | 2.62 | 2 | 2.68 | | 2.45 | 7.38 |
| | PBS pH 7.4 | 133 mM | 3 | 4.49 | 42 | 4.57 | | 3.91 | 13.69 |
| Pr30 / GST-4.1R_{10kDa} | PBS pH 7.4 | 133 mM | 1 | 5.13 | 86 | 5.21 | | 4.4 | 14.1 |
| Pr30 / 4.1R_{10kDa} | Tris pH 8.0 | 250 mM | 2 | 2.0 | 8 | 2.21 | | 2.1 | 7 |
| | PBS pH 7.4 | 133 mM | 3 | 2.24 | 10 | 2.32 | | 2.31 | 7.84 |

Table 4.6: Summary of the experimentally determined SAXS parameters obtained on the tagged and untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex, the 4.1R_{10kDa} / Pf30 complex and the 4.1R_{10kDa} / Pr30 complex. The data set highlighted in bold was chosen to perform the structural characterisation and the *ab-initio* modelling. *N.D.*: non-determined values due to insufficiently good data.

The small-angle X-ray scattering curve of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex, after subtraction of the SAXS background of the buffer alone and reduction of the data to the range 0.05 nm⁻¹ to 4.5 nm⁻¹, is shown in Figure 4.15 A. The associated P(r) was then calculated (Figure 4.15 B) and *ab-initio* modelling was performed: ten models were created using DAMMIF, as shown in Figure 4.15 C.

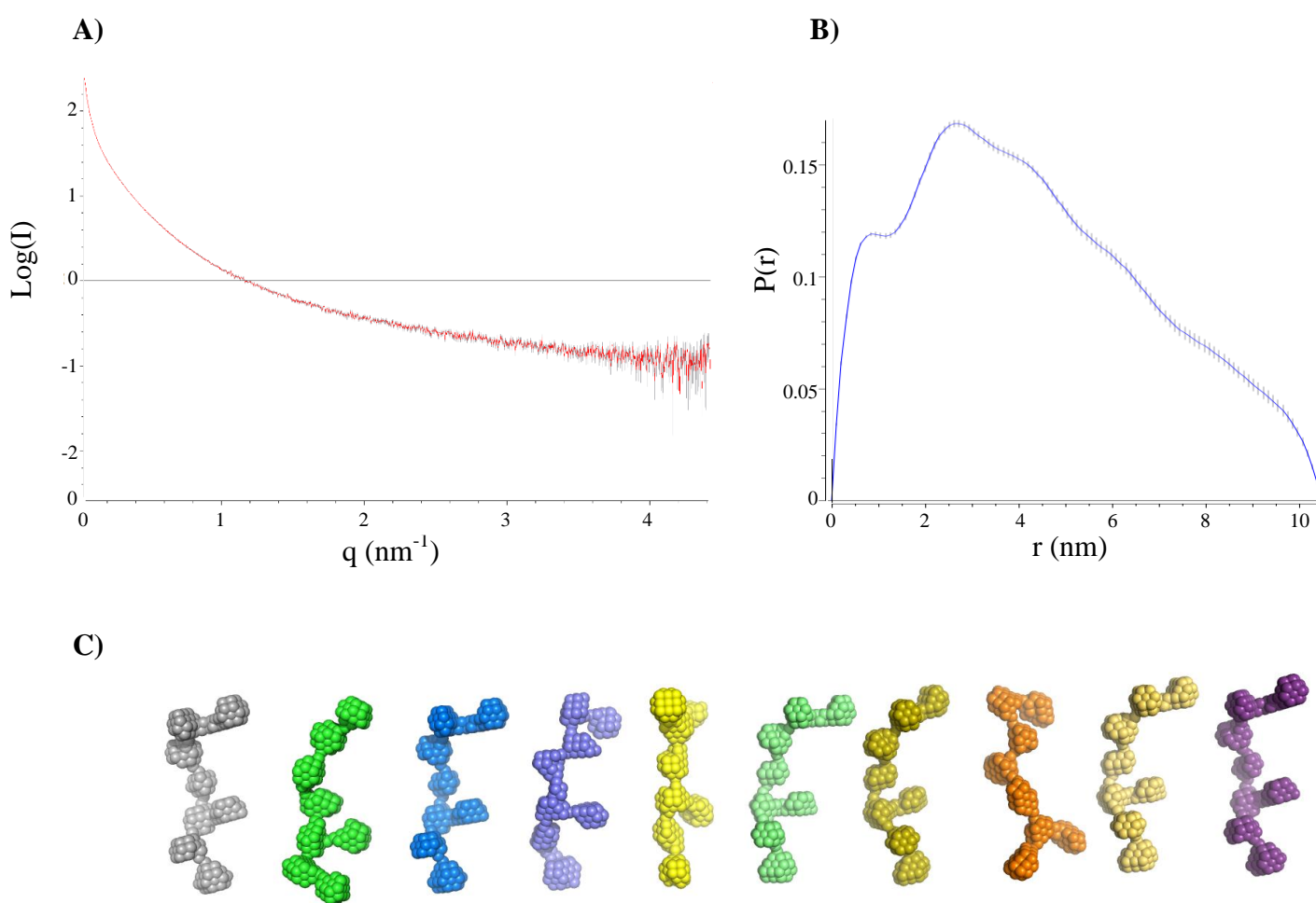


Figure 4.15: (A) Buffer subtracted SAXS curves of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex and (B) associated calculated P(r) and (C) *ab-initio* 3D models using DAMMIF (10 models).

The ten models of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex were averaged using DAMAVER (Figure 4.18). In order to validate them, the back-calculated scattering curve from the modelling was also calculated and is displayed in green on the experimental red scattering curve and shows a good match (right hand side).

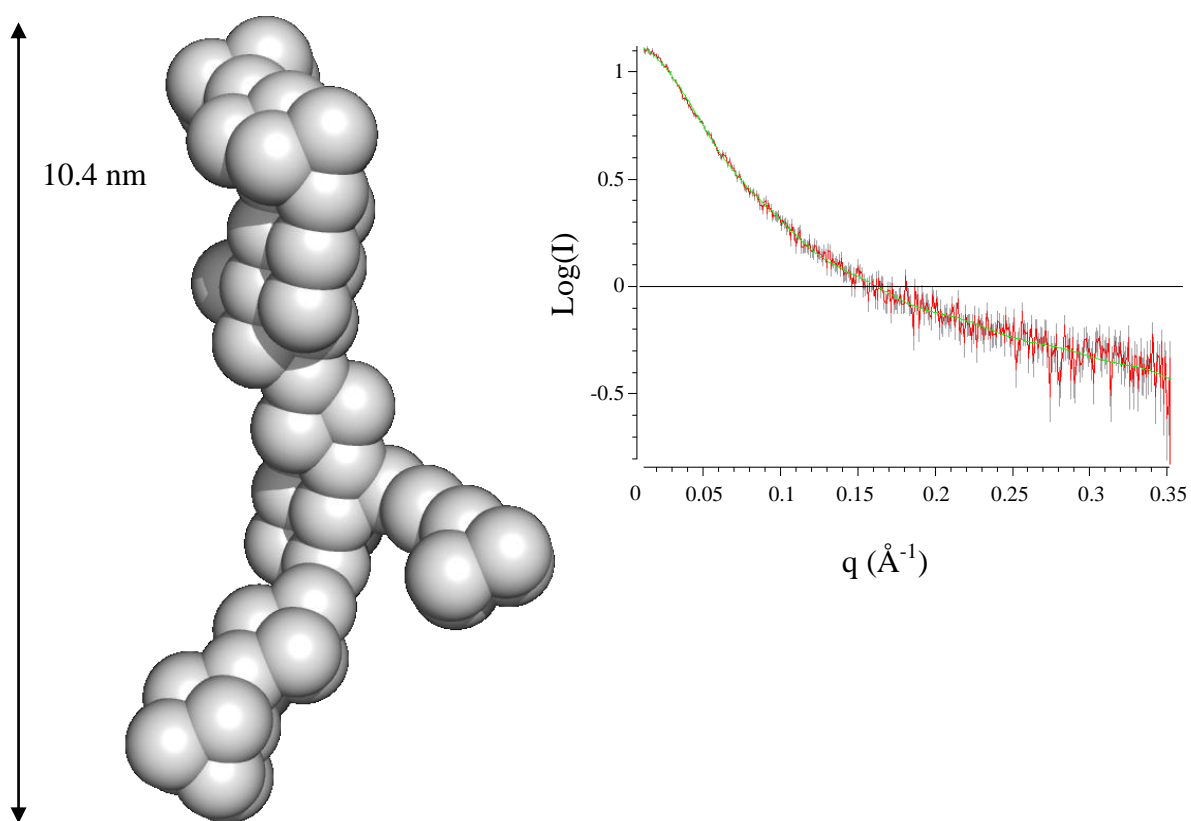


Figure 4.16: Averaged *ab-initio* 3D model of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex using DAMAVER and displayed as spheres (**left**) and fit of the simulated scattering curve in green versus the experimental data in red (**right**).

The disordered state of the untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} protein complex has also been assessed by the means of a Kratky plot. The Kratky plot analysis showing the presence of a typical plateau reveals the complex to be unfolded (Figure 4.17). It is particularly interesting to mention here that, at the SAXS resolution, there is no apparent folding resulting from the binding of the two protein partners, as it is often observed for other intrinsically disordered protein complexes (*Sugase et. al, 2007*).

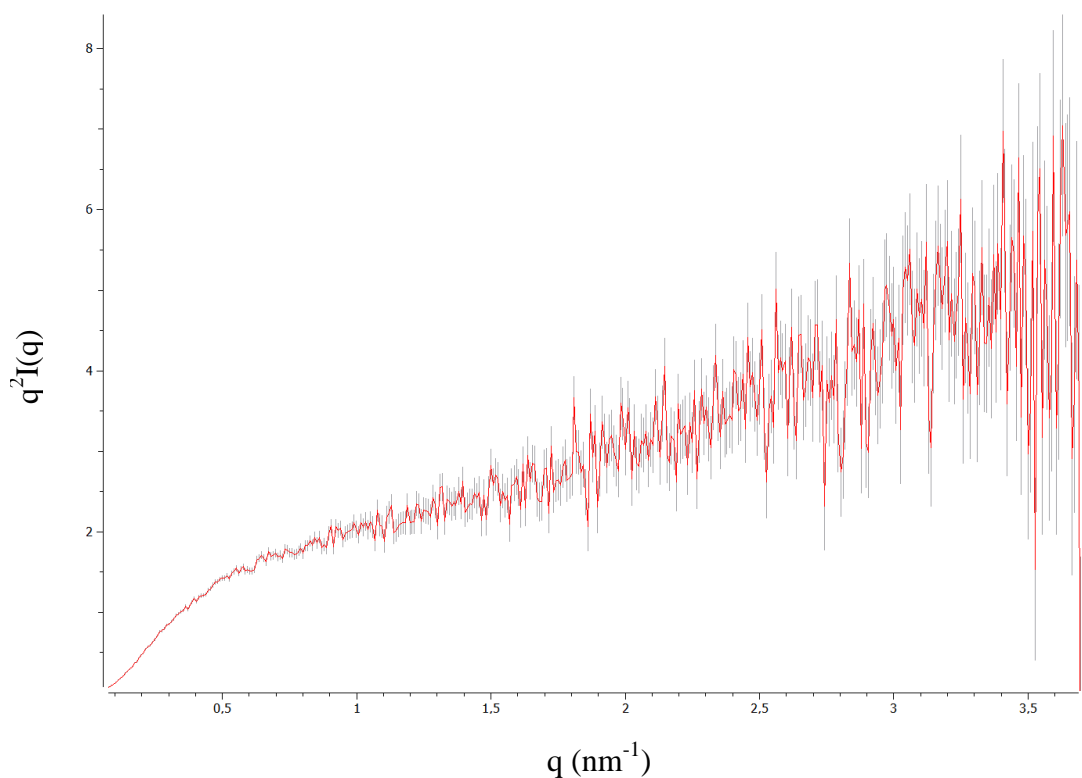


Figure 4.17: Kratky plot of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex.

4.4 Structural characterisation of the EBA-181 / 4.1R complex using small-angle neutron scattering

The complex formed between EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} was analysed by small-angle neutron scattering. The untagged version of the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} protein complex, more biologically relevant than the tagged complex, was chosen and further analysed in this section. To achieve this, the use of protein deuteration was exploited: the EBA181₉₄₅₋₁₀₉₇ protein was expressed as 75% deuterated and the 4.1R_{10kDa} protein as 100% hydrogenated. The H₂O/D₂O based buffers enabled to fully exploit the contrast variation approach by using three different D₂O percentages (Figure 4.19). In the 0% D₂O buffer (NaP pD 6.0), the whole complex is visible (A). In the 40% D₂O buffer, the hydrogenated protein is matched out and only the deuterated one is visible. Thus, the 4.1R_{10kDa} protein is matched out and the EBA181₉₄₅₋₁₀₉₇ protein remains visible (B). In the 100% D₂O buffer, the opposite occurs, with the deuterated EBA181₉₄₅₋₁₀₉₇ protein matched out and the hydrogenated 4.1R_{10kDa} remaining visible (C).

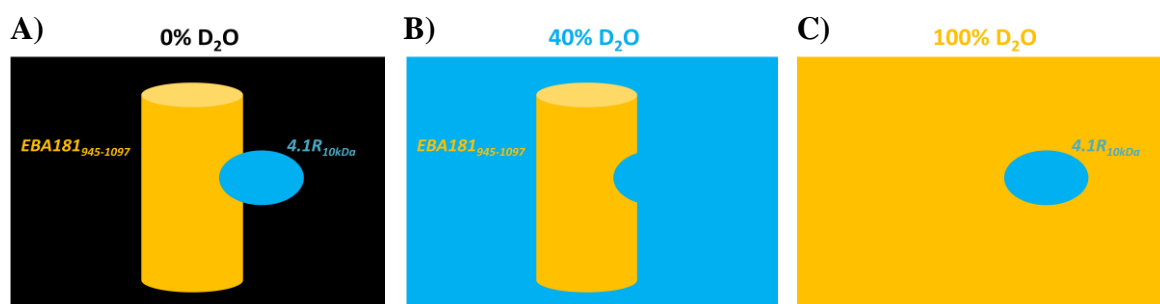


Figure 4.19: Principle of the SANS experiment on the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex using contrast variation (as a reminder). The colours used for the representation of the different protein domains are kept for the next figures and analyses.

The best scattering results were obtained with a distance sample-to-detector of 2 metres and a wavelength of 8Å and are thus presented here-after. The ones performed at 5.6 metres and 17 metres are not shown.

4.4.1 SANS characterisation of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex in 0 % D₂O

The protein complex between EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} was firstly analysed in 0% D₂O, in the NaP buffer at pD 6.0. This condition is the same as that used during the SAXS experiments and previously described in this chapter. As already mentioned, in 0 % D₂O none of the hydrogenated 4.1R_{10kDa} or the deuterated EBA181₉₄₅₋₁₀₉₇ is matched out, therefore the whole complex is visible under these conditions. The scattering of the background was firstly subtracted from the scattering curve (calibration including the adjustment of the position of the centre of the detector and the background correction factors calculated with the scattering measurements of cadmium alone, H₂O alone and one empty cell). The resulting subtracted curve is shown in Figure 4.20.

Following this, this scattering curve of the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} complex sample was subtracted with the scattering curve of its respective buffer (NaP buffer at pH 6.0 in 0 % D₂O), with the result shown in Figure 4.21. For the whole complex the Guinier analysis gives a radius of gyration equal to 3.95 nm (Figure 4.22), a D_{max} of 10.8 nm (Figure 4.23) and a Porod volume of 44.1 nm³.

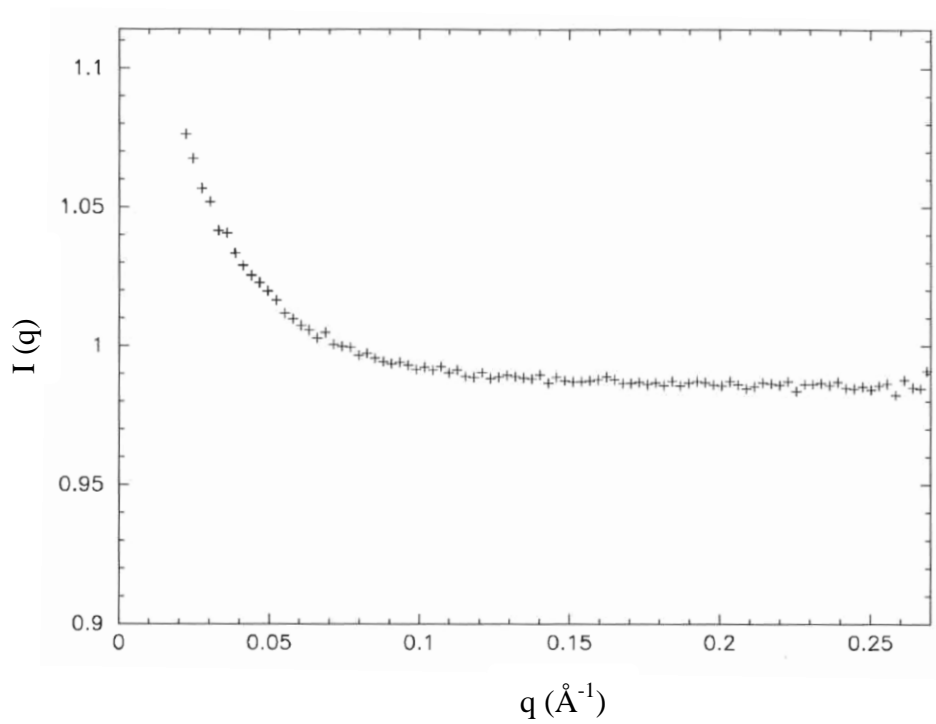


Figure 4.20: Background subtracted SANS curve of the complex in 0 % D₂O buffer.

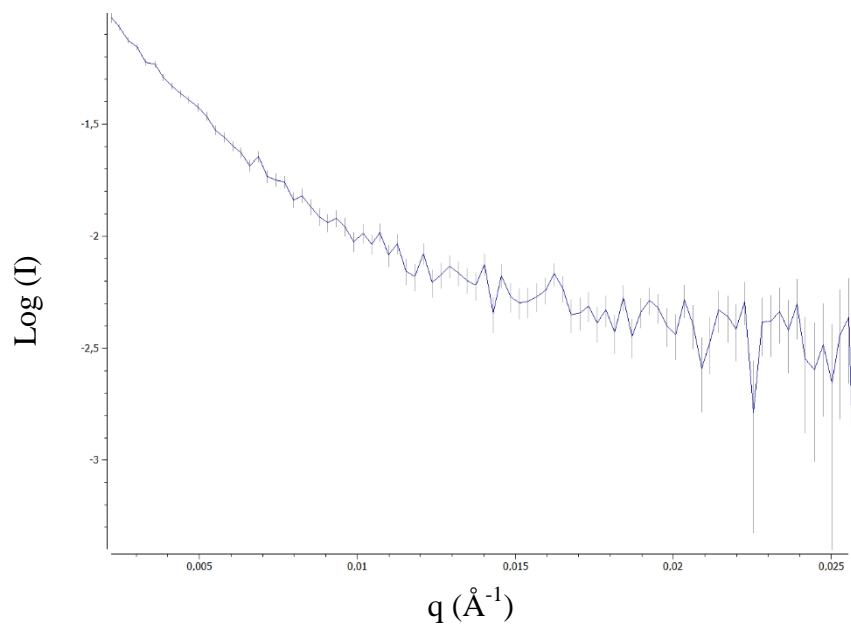


Figure 4.21: Buffer subtracted SANS curve of the complex in 0 % D₂O NaP buffer.

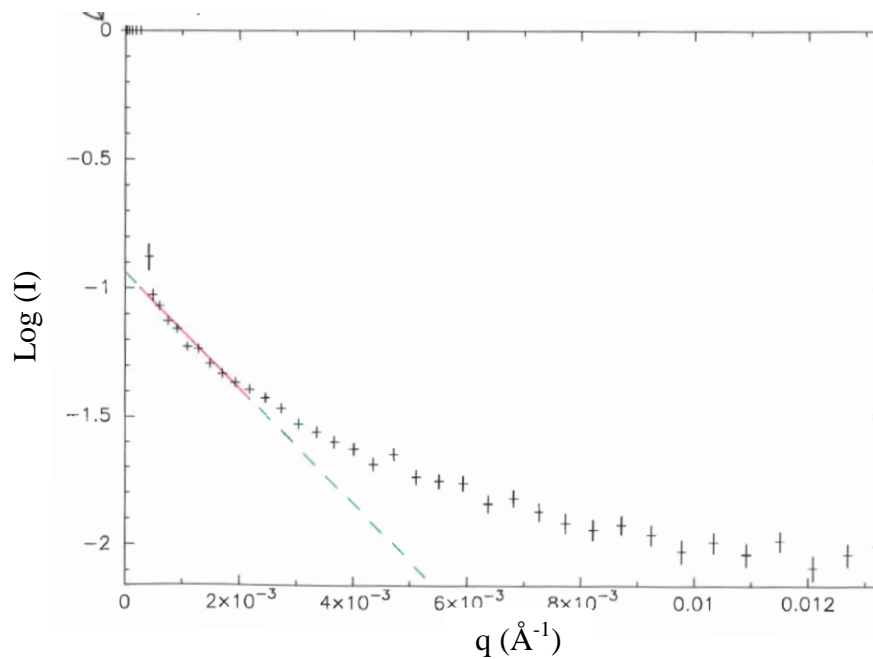
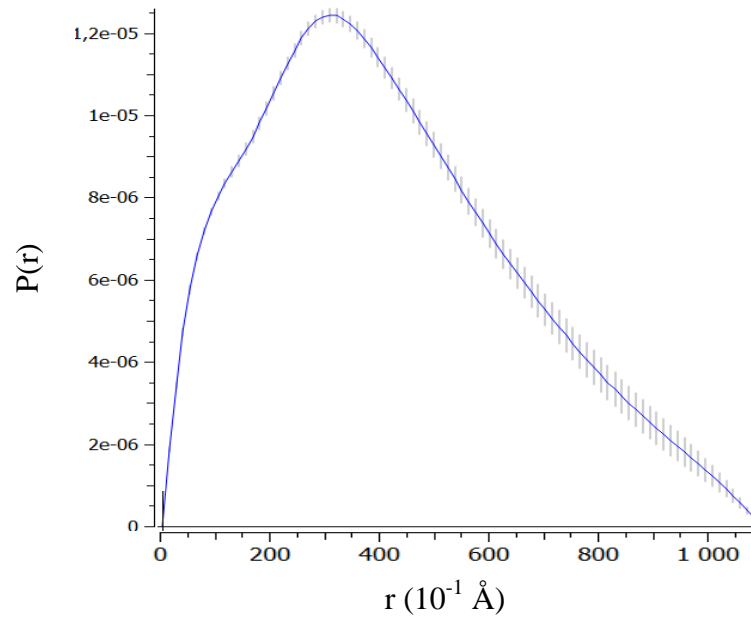


Figure 4.22: Guinier region of the buffer subtracted SANS curve of the complex in 0 % D₂O NaP buffer.

The $P(r)$ function associated to the buffer subtracted SANS curve of the complex in 0 % D₂O was calculated and 10 *ab-initio* 3D models were generated using DAMMIF (Figure 4.23). From these 10 models, an averaged 3D model was generated using DAMAVER as shown in Figure 4.24.

A)



B)

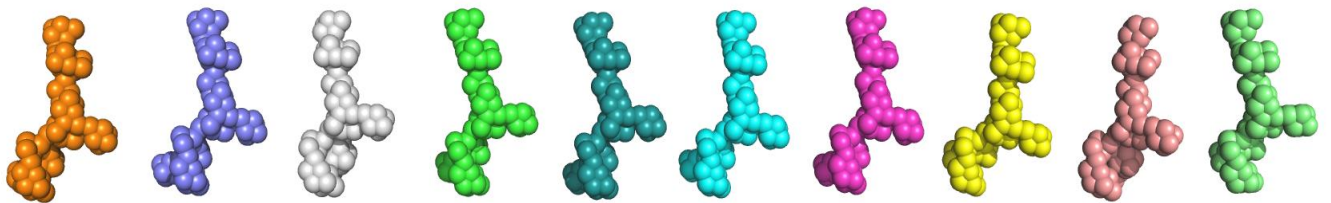


Figure 4.23: (A) Calculated $P(r)$ from the buffer subtracted SANS curves of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex in 0 % D₂O and **(B)** *ab-initio* modelling using DAMMIF (10 models).

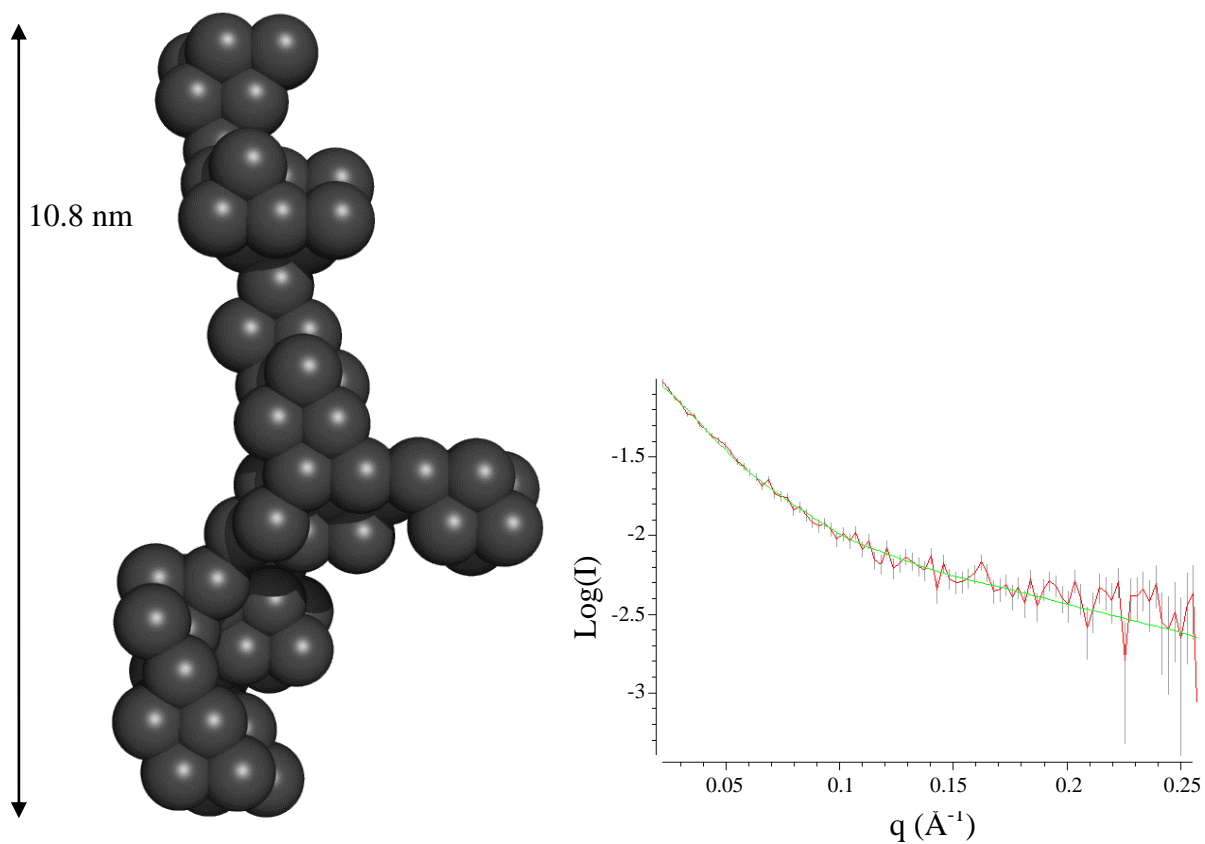


Figure 4.24: Averaged *ab-initio* 3D model of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex in 0 % D₂O using DAMAVER. Displayed respectively as spheres (**left**) and scattering curves (**right**). Red: experimental, green: back calculated from the model.

4.4.2 SANS characterisation of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex in 40 % D₂O

In the second experiment, the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} protein complex was analysed in 40 % D₂O. In this case, the hydrogenated 4.1R_{10kDa} protein is matched out, thus the 75 % deuterated EBA181₉₄₅₋₁₀₉₇ protein should be the only one visible in the complex.

As for the in 0 % D₂O analysis, the scattering of the background was firstly subtracted from the scattering curve. The resulting subtracted curve is shown in Figure 4.25.

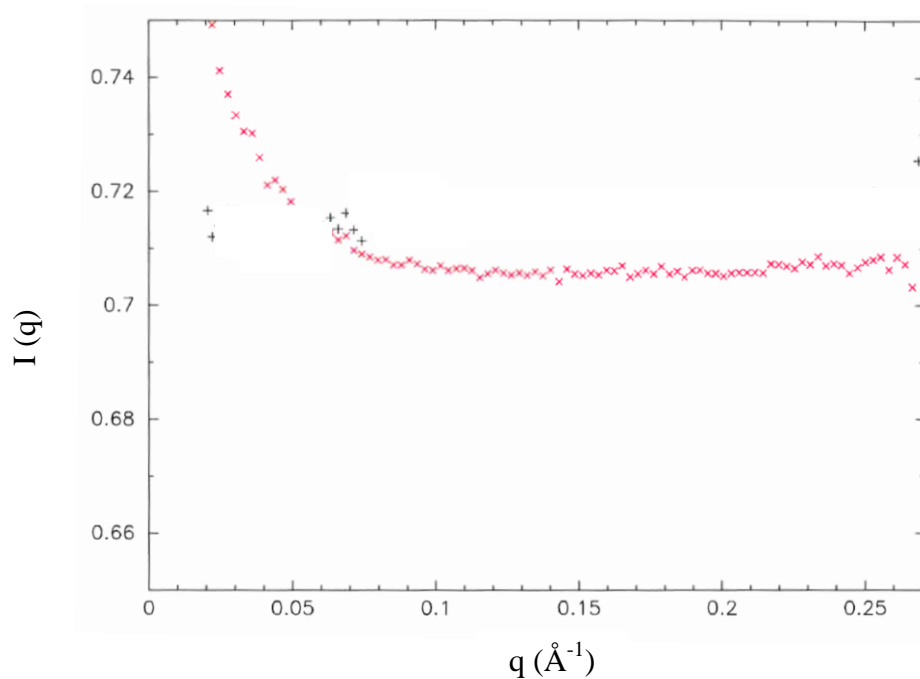


Figure 4.25: Background subtracted SANS curve of the complex in the 40 % D₂O buffer.

Following this step, the scattering curve of the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} complex was subtracted with the scattering curve of its respective buffer (NaP buffer at pH 6.0 in 40 % D₂O), with the result shown in Figure 4.26. The resulting scattering curve shows high noise above 0.12 Å⁻¹. This can be explained by the low protein concentration used here (2 mg/ml) and can be explained by the low intensity of the neutron flux in comparison to the X-ray flux of a synchrotron that allows to study proteins at very low concentrations. The Guinier analysis gives for the EBA181₉₄₅₋₁₀₉₇ molecule in the complex a R_g equal to 4.14 nm (Figure 4.27) and a D_{max} of 11.1 nm (Figure 4.28). These results are close to those obtained in 0% D₂O (10.8 nm) meaning that the EBA181₉₄₅₋₁₀₉₇ protein is, due to its disordered state and with regard to the size, the main object in the complex, with little apparent contribution from 4.1R_{10kDa}.

The P(r) function was calculated using the buffer subtracted SANS curve of the complex in 40 % D₂O and 10 *ab-initio* 3D models were generated using DAMMIF (Figure 4.28). From these 10 models, an averaged 3D model was generated using DAMAVER, as shown in Figure 4.29.

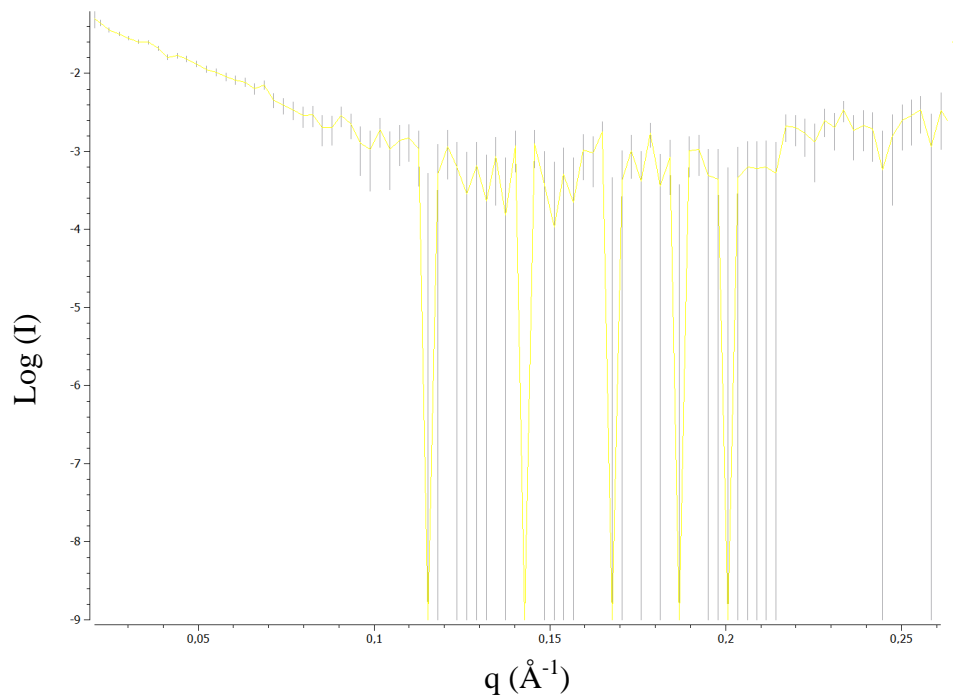


Figure 4.26: Buffer subtracted SANS curve of the complex in the 40 % D_2O buffer.

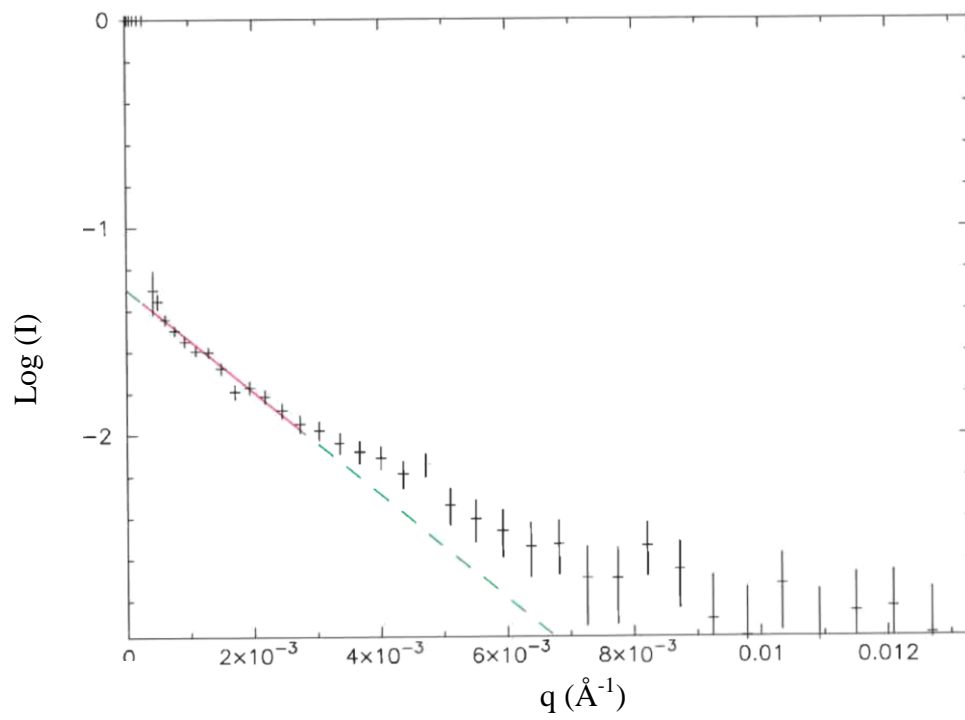
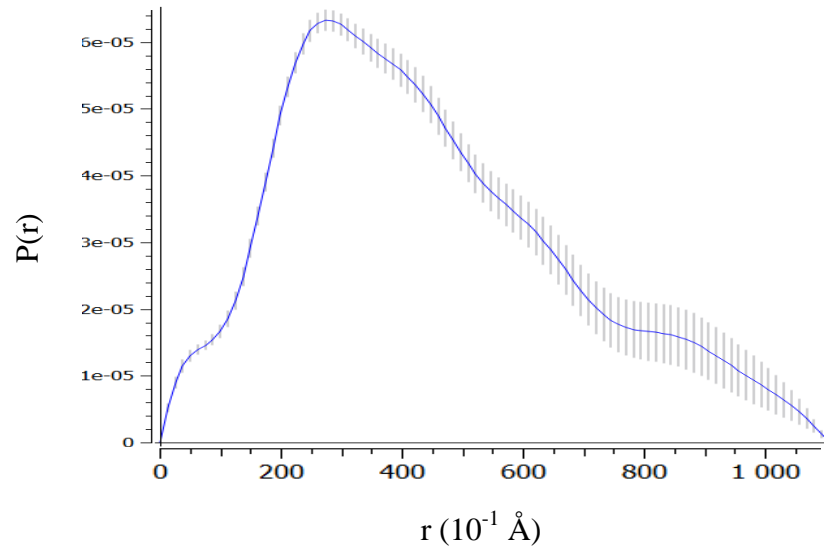


Figure 4.27: Guinier region of the buffer subtracted SANS curve of the complex in the 40 % D_2O buffer.

A)



B)

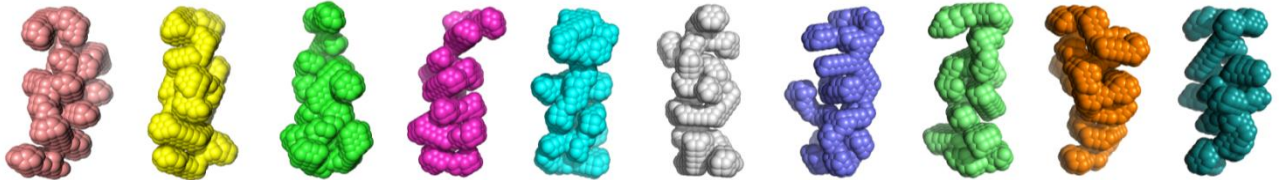


Figure 4.28: **(A)** Calculated $P(r)$ from the buffer subtracted SANS curves of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex in 40 % D₂O and **(B)** *ab-initio* modelling using DAMMIF (10 models).

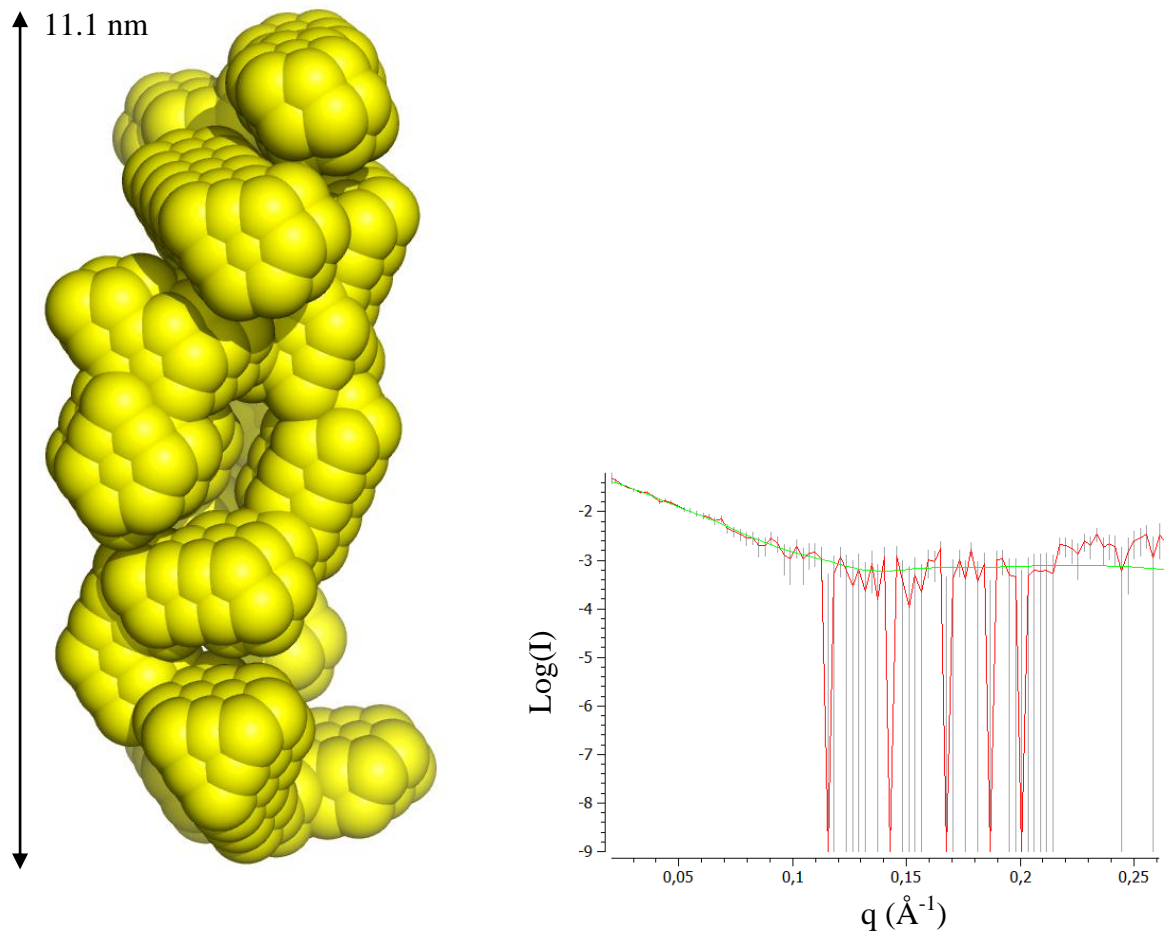


Figure 4.29: Averaged *ab-initio* 3D model of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex in 40 % D₂O using DAMAVER. Displayed respectively as spheres (**left**) and scattering curves (**right**). Red: experimental, green: back calculated from the model.

4.4.3 SANS characterisation of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex in 100 % D₂O

In the final SANS experiment, the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} protein complex was analysed in 100 % D₂O. In this case the 75 % deuterated EBA181₉₄₅₋₁₀₉₇ protein is matched out and the hydrogenated 4.1R_{10kDa} protein remains visible. As performed and presented before, the scattering of the background was firstly subtracted from the scattering curve, with the resulting subtracted curve shown in Figure 4.30.

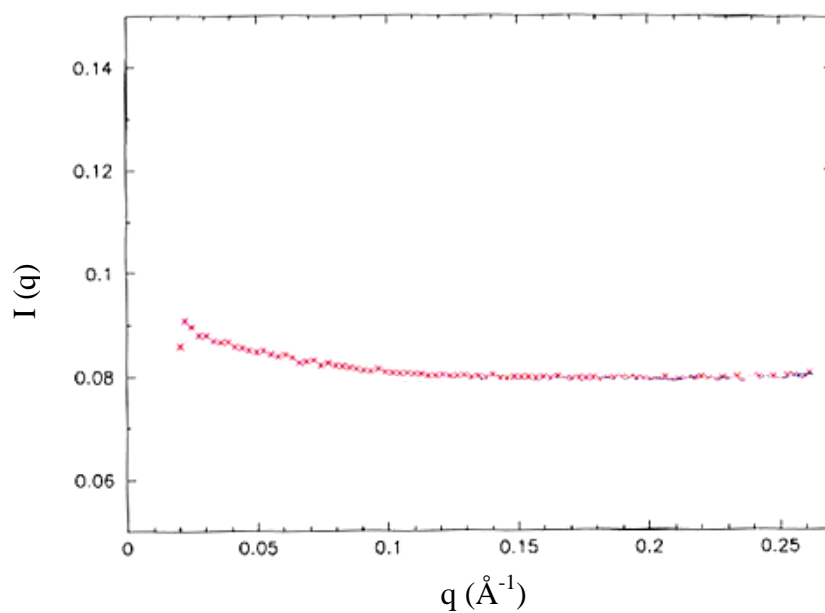


Figure 4.30: Background subtracted SANS curve of the complex in 100 % D₂O NaP buffer.

Following this step, the scattering curve of the EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex was subtracted with the scattering curve of its respective buffer (NaP buffer at pH 6.0 in 100 %

D₂O) as shown in Figure 4.31. The Guinier analysis gives for the 4.1R_{10kDa} molecule in the complex a radius of gyration equal to 1.78 nm (Figure 4.32) and a D_{max} of 6.0 nm (Figure 4.33). These results are consistent with those measured in SAXS for the 4.1R_{10kDa} protein alone, with a R_g of 2.02 nm and a D_{max} of 6.0 nm.

The P(r) function associated with the buffer subtracted SANS curve of the complex in 100 % D₂O was calculated and 10 *ab-initio* 3D models were generated using DAMMIF in slow mode (Figure 4.33). From these 10 models, an averaged 3D model was generated using DAMAVER, as shown in Figure 4.34.

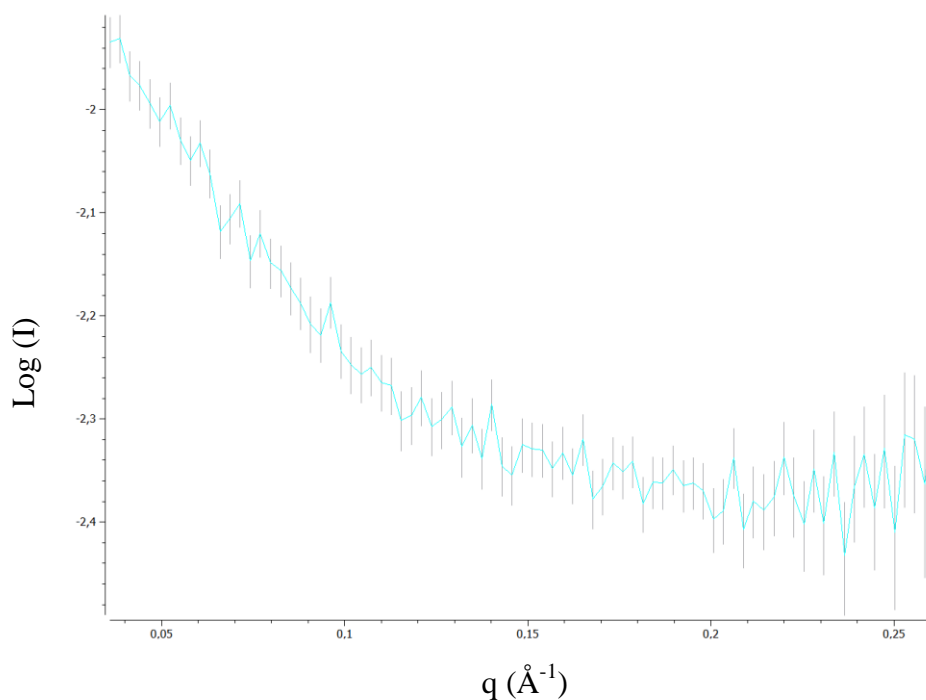


Figure 4.31: Buffer subtracted SANS curve of the complex in 100 % D₂O NaP buffer.

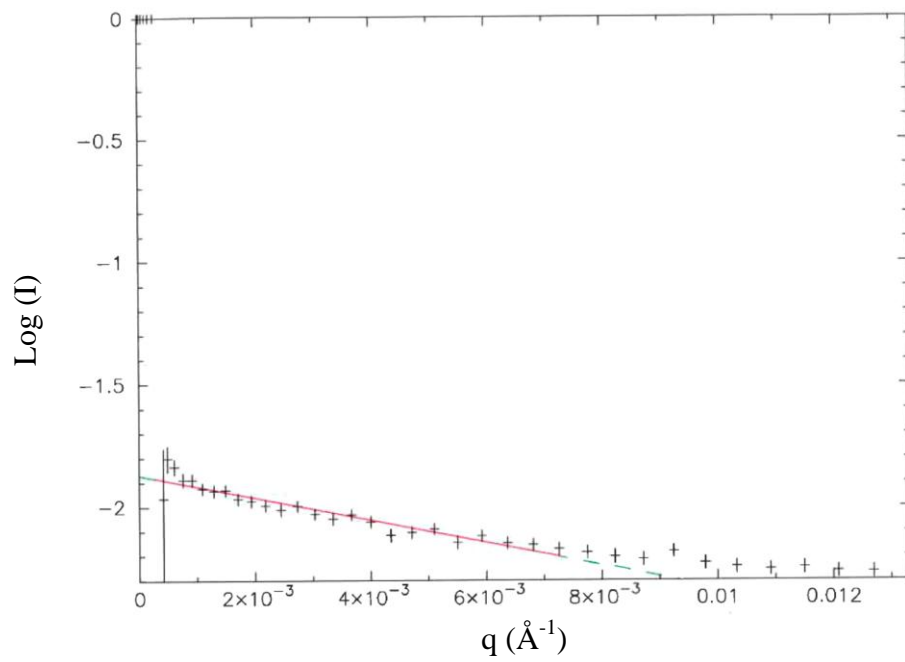
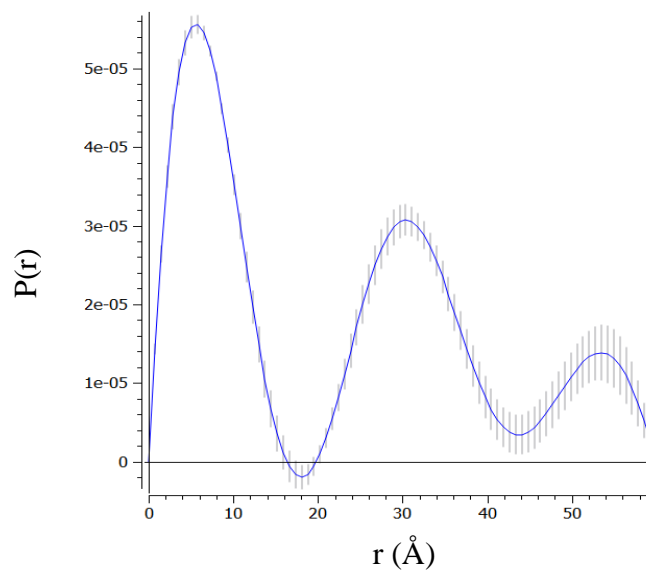


Figure 4.32: Guinier region of the buffer subtracted SANS curve of the complex in 100 % D₂O NaP buffer.

A)



B)

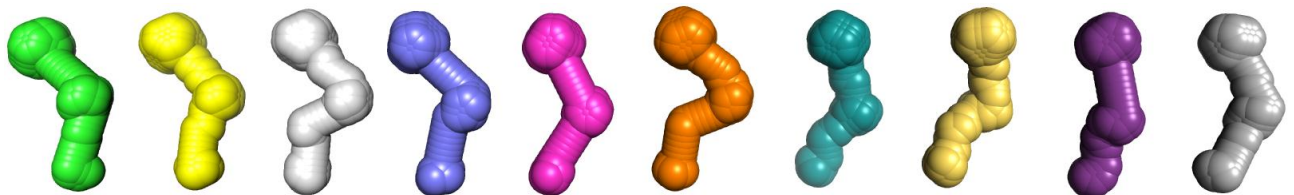


Figure 4.33: **(A)** Buffer subtracted SANS curves of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex in 100 % D₂O; **(B)** associated calculated P(r) and **(C)** *ab-initio* modelling using DAMMIF (10 models).

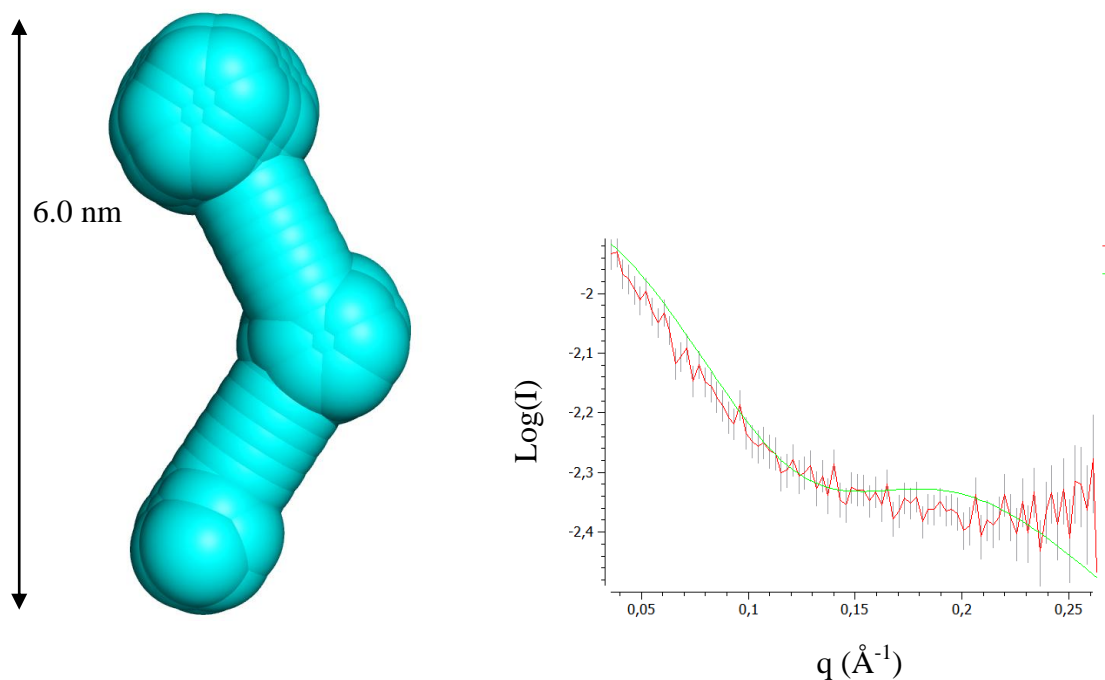


Figure 4.34: Averaged *ab-initio* 3D model of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex in 100 % D₂O using DAMAVER. Displayed respectively as spheres (**left**) and scattering curves (**right**). Red: experimental, green: back calculated from the model.

4.4.4 Validation of the data

It is possible to monitor the quality of a contrast variation experiment by calculating the following parameter $\sqrt{\frac{I_0}{C.T.e}}$ (Ibel *et al.*, 1975) which is proportionally linear with the D₂O percentage in a theoretically perfect experiment. I_0 is the zero intensity, C the sample concentration in mg/ml, T the transmission and e the path-length in cm, and the result can have a positive or a negative sign. The contrast variation results are presented in Table 4.7 and Figure 4.35.

As shown in Figure 4.35, the $\sqrt{\frac{I_0}{C.T.e}}$ parameter results in a linear curve in function of the percentage of D₂O when applied to the three contrasts chosen for the SANS: 0 %, 40 % and 100 % D₂O. This validates the SANS experiment using deuterium-based contrast variation.

| | 0 % D ₂ O | 40 % D ₂ O | 100 % D ₂ O |
|----------------------------|----------------------|-----------------------|------------------------|
| $\sqrt{\frac{I_0}{C.T.e}}$ | (±) 1.08 | (±) 0.64 | (±) 0.28 |

Table 4.7: Contrast variation experiment of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex.

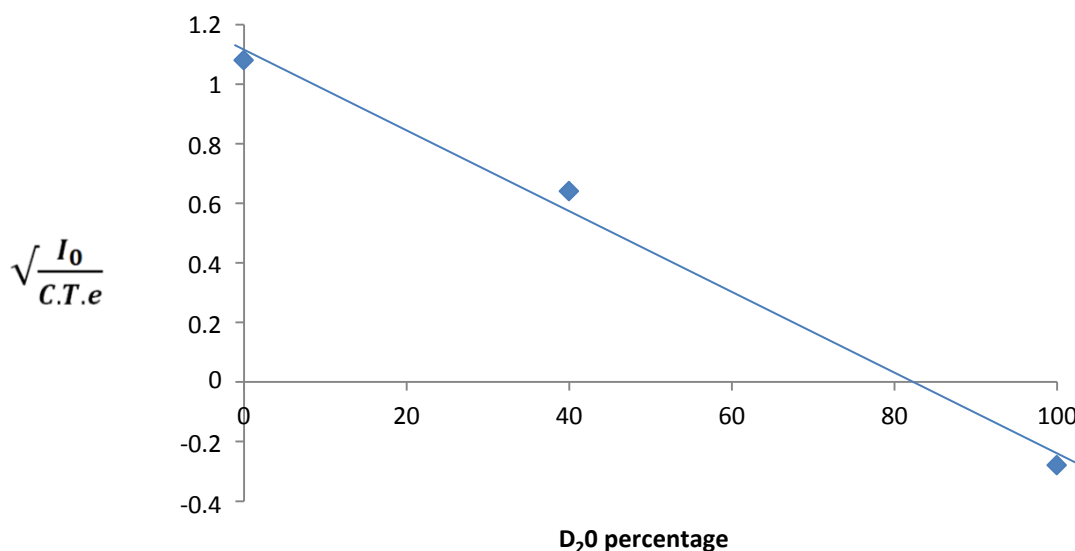


Figure 4.35: Validation of the neutron contrast variation experiment performed on the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex.

4.5 Discussion and conclusions: SAXS and SANS consensus

A summary of the main structural parameters derived from the SAXS and SANS data is given in Table 4.8.

A)

| Individual proteins | | | | |
|---------------------|--------------------------------|----------------------------|---------------------------|-----------------------|
| | His-EBA181 ₉₄₅₋₁₀₉₇ | EBA181 ₉₄₅₋₁₀₉₇ | GST-4.1R _{10kDa} | 4.1R _{10kDa} |
| R_g | 2.40 nm | 2.35 nm | 4.17 nm | 2.02 nm |
| D_{max} | 9.1 nm | 7.9 nm | 12.0 nm | 6.0 nm |

B)

| EBA181 ₉₄₅₋₁₀₉₇ / 4.1R _{10kDa} complex | | | | |
|--|---------|---------------------------|----------------------------|-----------------------------|
| | SAXS | SANS 0 % D ₂ O | SANS 40 % D ₂ O | SANS 100 % D ₂ O |
| R_g | 3.53 nm | 3.95 nm | 4.14 nm | 1.78 nm |
| D_{max} | 10.4 nm | 10.8 nm | 11.1 nm | 6.0 nm |

Table 4.8: Structural parameters (R_g and D_{max}) derived from (A) SAXS measurements on tagged and untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins and (B) SAXS and SANS measurements in the three solvent contrasts on the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex.

Firstly, the analysis of the SAXS data recorded from the tagged constructs of the His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} proteins has provided their structural parameters in solution (Table 4.8, A). In order to further characterise their structures a Kratky plot analysis was performed (Figure 4.9), re-emphasising the disordered state of the tagged His-EBA181₉₄₅₋₁₀₉₇ protein, as described following the biophysical characterisation performed and described in Chapter 3. *Ab-initio* modelling was performed (Figure 4.7 and Figure 4.8) on both tagged His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} proteins. His-EBA181₉₄₅₋₁₀₉₇ is an elongated shape whereas GST-4.1R_{10kDa} appears to be mostly folded, as expected given that approximately two-thirds of the protein constitutes the GST tag. Moreover, a comparison of the 3D models for the GST tag alone, and that derived from the GST-4.1R_{10kDa} protein provides a useful insight into the location and structure of the tag within the protein (Figure 4.10). The GST tag has a globular shape, as expected by comparison with its crystal structure, whilst the 10 kDa domain of the 4.1R protein appears substantially more elongated.

Secondly, the analysis of the SAXS data for the untagged constructs of the proteins (EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}) yielded their structural parameters (Table 4.8, A), with respectively their R_g measured at 2.35 nm and 2.02 nm, and their D_{max} equal to 7.9 nm and 6.0 nm. A Kratky plot analysis for these proteins (Figure 4.17) shows EBA181₉₄₅₋₁₀₉₇ to be disordered and 4.1R_{10kDa} to be partially folded. These results are again in agreement with the biophysical characterisation performed (Chapter 3) which includes CD, 1D NMR and the measurement of their respective sizes (R_h) using DLS, leading to values corresponding to folded proteins with a four times higher molecular mass. In addition, the averaged *ab-initio* models show elongated shapes for both EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}. Finally in

order to validate the data, the simulated scattering curves from the models were calculated and provide a good fit to the experimental data (Figures 4.12 and 4.13).

The final SAXS data analysis was performed on the complex formed by the untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins, with the resulting structural parameters summarised in Table 4.8, B. Interestingly, the Kratky plot analysis suggests that the complex is disordered, implying there is no significant protein structural ordering upon binding and complex formation. *Ab-initio* modelling reveals that the complex adopts a ‘T-shaped’ structure with the 4.1R_{10kDa} protein which protrudes from the central region of the elongated EBA181₉₄₅₋₁₀₉₇ (Figure 4.16).

SANS studies add a powerful complement to the SAXS results in that they allow (i) the identification of the location of the protein partners within the complex, and (ii) by comparison with individual protein forms they allow conformation changes in each protein (if any) before and after binding to be studied and imaged. This SANS study of the untagged EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} protein complex used solvent contrast variation with varying amount of D₂O to match out, in turn, each of the proteins involved in the complex.

At 0 % D₂O (corresponding to schematic A in Figure 4.19) both molecules are visible, providing the structural parameters of the whole complex in solution (Table 4.8) and enabling the complex of the deuterated EBA181₉₄₅₋₁₀₉₇ with the hydrogenated 4.1R_{10kDa} protein to be modelled (Figures 4.24). The *ab-initio* models derived from the SAXS and SANS data give a very similar T-shaped structure (Figure 4.36, C) and close structural parameters. The corresponding R_g and D_{max} values on the complex are respectively

3.95 nm and 10.8 nm for the SAXS measurement, 3.53 nm and 10.4 nm for the SANS measurement in 0 % D₂O.

For the experiment at 40 % D₂O experiment (corresponding to schematic B in Figure 4.19), in principle the hydrogenated 4.1R_{10kDa} protein is matched out and thus only the deuterated EBA181₉₄₅₋₁₀₉₇ is visible in the complex (Figure 4.29). The structural parameters calculated from these SANS data for the deuterated EBA181₉₄₅₋₁₀₉₇ in its complexed form are higher than those obtained from SAXS for the individual EBA181₉₄₅₋₁₀₉₇ protein (Table 4.8), with a measured R_g of 4.14 nm by SANS compared with 2.35 nm by SAXS; and a D_{max} of respectively 11.1 nm by SANS for 7.9 nm by SAXS. This may mean that the EBA181₉₄₅₋₁₀₉₇ is stretched when in complex with 4.1R_{10kDa}. Nevertheless, it should be noted that the SANS data measured in 40 % D₂O have a low signal/noise ratio and thus the derived structural parameters need to be taken with caution. It could also be caused by the 4.1R_{10kDa} protein not being completely matched out and thus contributing to the measured size. Indeed, the 40 % contrast value of D₂O is only theoretical and the experimental contrast to completely match out the 4.1R_{10kDa} protein might thus be slightly different. However, the comparison between the *ab-initio* models from the SAXS and SANS experiments shows a good agreement in their overall elongated shapes, although their relative sizes are different (Figure 4.36, B).

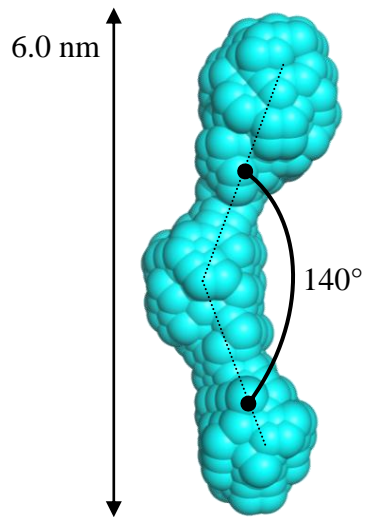
For the SANS experiment performed in 100 % D₂O (corresponding to schematic C in Figure 4.19), the deuterated EBA181₉₄₅₋₁₀₉₇ is matched out, with the hydrogenated 4.1R_{10kDa} staying visible and thus being modelled in its complexed form (Figure 4.34). The R_g and D_{max} obtained were in good agreement with the values obtained by SAXS for the

4.1R_{10kDa} protein in its uncomplexed form (Table 4.8), with a R_g of 1.78 nm (2.02 nm in SAXS) and a D_{max} of 6.0 nm (also 6.0 nm in SAXS). The comparison between the SANS and SAXS data shows similar 3D *ab-initio* models and the only difference comes from the conformation slightly more closed of 4.1R_{10kDa} when in complex, with measurement of its internal angle decreasing to 120°, instead of 140° when in the free form (Figure 4.36, A).

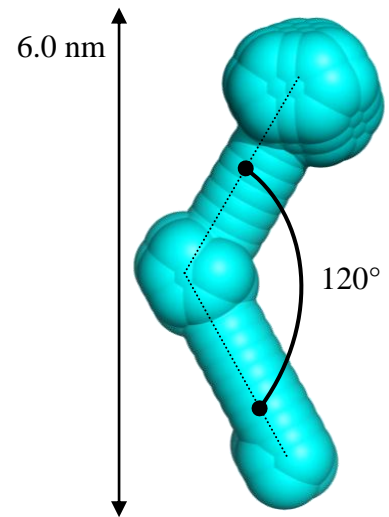
This work highlights key aspects of importance for small-angle solution scattering studies of biological macromolecules. A successful solution scattering study requires monodisperse protein solutions, free of both insoluble and soluble aggregates. This is an essential prerequisite in every scattering experiment and a crucial point for structural modelling. Since the scattering of a particle is related to the square of its scattering length, aggregates contribute disproportionately to the signal, resulting in interpretation difficulties. A mixture of different oligomeric forms also leads to an interpretation problem, highlighting the necessity of performing extensive biochemical and biophysical analysis on the samples as a prerequisite of every SAS experiment, as presented in Chapter 3. In the case of this work, aggregation issues were addressed by working at low temperature during the whole purification process, using freshly purified samples for every scattering experiment, and lastly, by performing a high speed centrifugation of the samples just before the scattering measurements. Hence, the proteins used were pure and monodisperse (as assessed by DLS) after the final step of optimisation and purification.

In conclusion, the SAXS and the SANS analyses provide good agreement in terms of the solution state modelling, including the 3D models of the individual proteins as well as the complex formed between them (Figure 4.36, A, B and C).

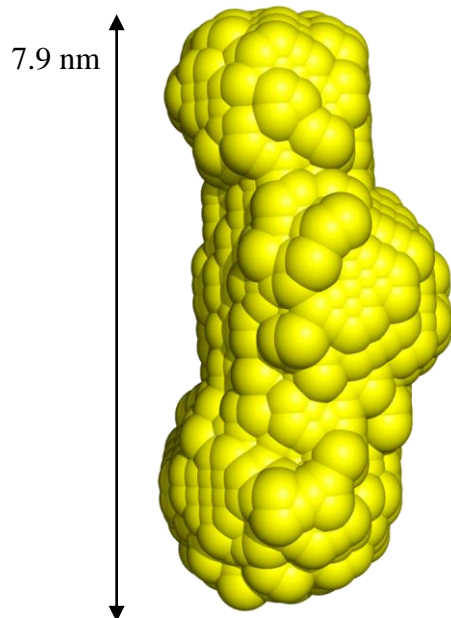
A) 4.1R_{10kDa} alone (SAXS)



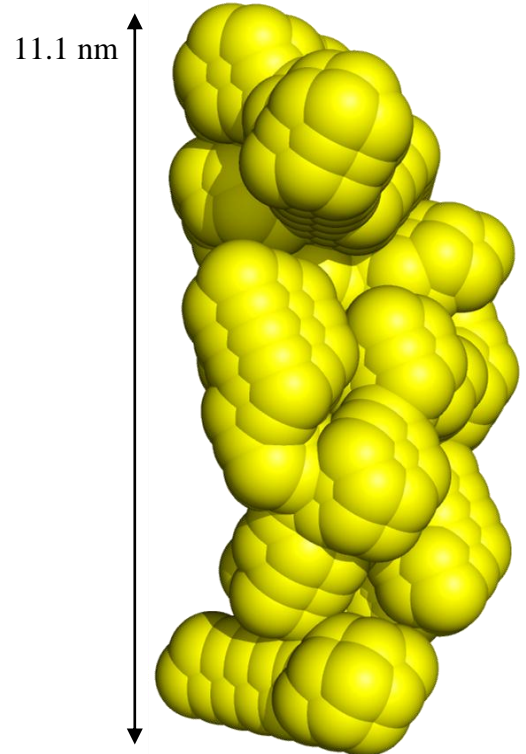
4.1R_{10kDa} in complex (SANS, 100% D₂O)



B) EBA181₉₄₅₋₁₀₉₇ alone (SAXS)



EBA181₉₄₅₋₁₀₉₇ in complex (SANS, 40% D₂O)



C) **EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex**
(SAXS)

EBA181₉₄₅₋₁₀₉₇ / 4.1R_{10kDa} complex
(SANS, 0% D₂O)

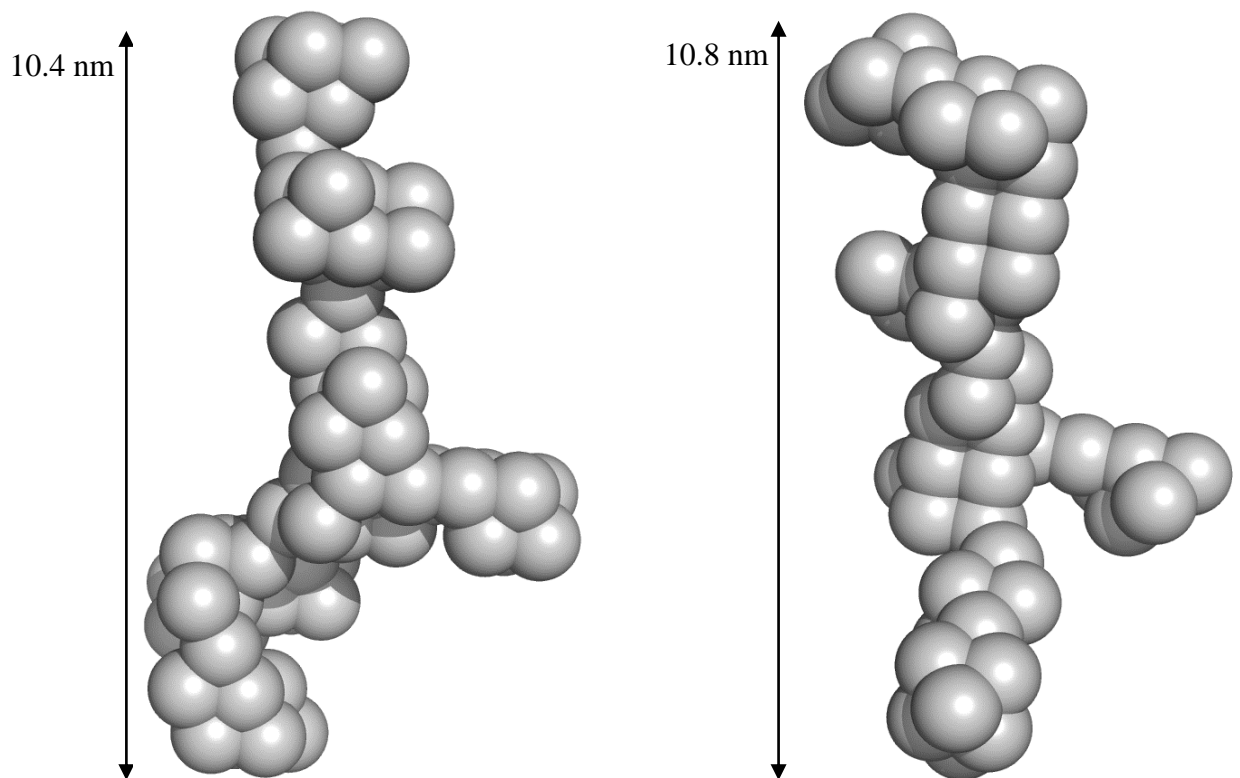


Figure 4.36: Comparison of the final SAXS and SANS 3D models of **(A)** uncomplexed form of 4.1R_{10kDa} (left) and complexed form of 4.1R_{10kDa} (right), **(B)** uncomplexed EBA181₉₄₅₋₁₀₉₇ (left) and complexed EBA181₉₄₅₋₁₀₉₇ (right), and **(C)** the final 3D models of the full complex by SAXS (left) and SANS (right).

5. SOLUTION STATE NMR STUDIES OF EBA181₉₄₅₋₁₀₉₇,

4.1R_{10kDa} AND THEIR COMPLEX

Abstract

This chapter describes NMR solution state studies that were carried out on the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins, and on the complex formed between the two. The motivation for this work derived from the nature of the results described in Chapter 3 and Chapter 4 where a wide body of evidence demonstrated that the EBA181₉₄₅₋₁₀₉₇ protein was itself intrinsically disordered and that the 4.1R_{10kDa} was only partially folded. Despite this, there was clear evidence of complex formation between the two proteins (Chapters 3 and 4). Hence one of the major goals of this NMR work was to further assess the nature of the individual proteins and their complex. This study is also of interest since intrinsically disordered proteins are of emerging importance, given that they are implicated in a wide range of interactions where a specific structure is only formed upon binding with a particular protein partner.

A detailed structural and dynamic characterisation of the protein is presented here, using NMR spectroscopy. The experiments confirmed the unstructured nature of EBA181₉₄₅₋₁₀₉₇, providing detailed insight on its dynamic and structural behaviour. EBA181₉₄₅₋₁₀₉₇ is predominantly a random coil molecule containing several turn motifs. It does not contain transiently populated secondary structures as it occurs in numerous intrinsically

disordered proteins that fold via specific, pre-formed molecular recognition elements. Experiments carried out on the GST tagged 4.1R_{10kDa} demonstrated partial folding. Titration studies of the complex were carried out and showed evidence of complex formation.

5.1 Introduction

Applications of nuclear magnetic resonance can be very broad: characterisation of soluble biomolecules, interfaces and complexes, determination of protein structure or thermodynamic parameters. The samples have to be concentrated (> 0.1 mM) and their molecular weight can be a limiting factor: 30-40 kDa is often taken as a maximum for protein structure determination. NMR spectra of proteins are complex, due to their large molecular weight and number of atoms present. Deuterium labelling is often used to overcome molecular weight limitations by minimising the effect of proton dipolar coupling. ¹⁵N ¹³C isotope labelling are often used to allow backbone and residue assignment respectively.

Results from the previous chapters have clearly demonstrated the EBA181₉₄₅₋₁₀₉₇ protein to belong to the family of intrinsically disordered proteins. The intrinsic disorder of EBA-181 and the lack of structure of 4.1R_{10kDa} clearly limit the applicability of conventional structure determination methods. The data shown in Chapter 3 provide abundant evidence of the unstructured nature of EBA181₉₄₅₋₁₀₉₇ from CD, 1D-NMR and TSA. Similar measurements on 4.1R_{10kDa} suggest approximately 35% alpha-helical content. Despite

clear evidence of complex formation from biochemical and biophysical studies, the complex itself appears to be substantially disordered and the attempts to crystallise the individual proteins and the complex were unsuccessful. The NMR studies were based on the use of selective ^{15}N and ^{13}C labelling such that the individual proteins could be studied and their complex formation assessed in a number of titration experiments, firstly where unlabelled 4.1R_{10kDa} was titrated into double labelled EBA181₉₄₅₋₁₀₉₇, and secondly where unlabelled EBA181₉₄₅₋₁₀₉₇ was titrated into double labelled GST-4.1R_{10kDa}.

In recent years, intrinsically disordered proteins have emerged as an important class of proteins which lack significant amounts of secondary and tertiary structure but play active roles in many biological processes (*Dunker et al., 2001; Dunker et al., 2008; Tompa et al., 2012*). This chapter focuses on the structural and dynamic characterisation of *Plasmodium falciparum* EBA181₉₄₅₋₁₀₉₇ and human 4.1R_{10kDa} domains using solution state NMR to supplement the structural studies presented in the previous chapters. This work was carried out in collaboration with the NMR group of the ‘Institut de Biologie Structural’ (IBS) in Grenoble. Section 5.2 describes the material and methods used to perform the NMR spectroscopy experiments, including sample preparation and ^{15}N / $^{15}\text{N}^{13}\text{C}$ labelling. Section 5.3 describes the spectroscopy results obtained using single- and double-labelled EBA181₉₄₅₋₁₀₉₇ protein, including ^1H - ^{15}N and ^{13}C - ^{15}N heteronuclear single quantum coherence (HSQC) spectra, complete spectral assignment of double labelled EBA181₉₄₅₋₁₀₉₇, chemical shifts, backbone dynamics, structural ensemble descriptions and residual dipolar coupling measurements. Section 5.4 presents the initial results obtained on double labelled GST-4.1R_{10kDa} protein as well as when in complex with EBA181₉₄₅₋₁₀₉₇.

Finally, Section 5.5 summarises and concludes the chapter with a summary of the results obtained and suggestions for future work.

5.2 Material and methods

5.2.1 Sample production and labelling for NMR solution studies

5.2.1.1 Large-scale protein expression

As described in Chapters 3 and 4, the *E. coli* Rosetta™ 2 (DE3) bacterial cells expressing either the His-EBA181₉₄₅₋₁₀₉₇ or the GST-4.1R_{10kDa} proteins and stored at -80°C were used here to initiate the bacterial cultures in minimum medium ('Ross' medium). The Ross minimum medium (*Ross et al., 2004*) was slightly modified and used to express single- and double-labelled (¹⁵N and ¹³C/¹⁵N) His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} proteins (Tables 5.1 and 5.2). After preliminary expression tests (not shown), the His-EBA181₉₄₅₋₁₀₉₇ and the GST-4.1R_{10kDa} proteins were expressed by inducing the bacterial cultures with 1 mM IPTG during 16 hours at 30°C.

5.2.1.2 Protein purification

As previously described in Chapter 3, the EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} proteins were firstly purified by affinity chromatography (IMAC) using respectively a nickel or a GST resin. Their tags were cleaved using thrombin protease (followed by a second affinity

| Modified Ross medium pH 6.8 | Concentration |
|--|----------------------|
| Glucose monohydrate (¹² C / ¹³ C-labelled) | 3.41 g/l |
| Magnesium sulphate | 1.2 g/l |
| Potassium dihydrogen phosphate | 8.7 g/l |
| 85% phosphoric acid | 7.38 g/l |
| Ammonium chloride (¹⁵ N labelled) | 1.94 g/l |
| 6 M sodium hydroxide | 6.67 ml/l |
| Thiamine hydrochloride | 9 mg/l |
| Trace elements* | 1 ml/l |

Table 5.1: Modified Ross medium preparation used to label the proteins (initially from Ross et al., 2004).

| *trace elements (x1000) | Concentration |
|---|----------------------|
| Fe ₂ (SO ₄) ₃ .H ₂ O | 82.3 g/l |
| CoSO ₄ .7H ₂ O | 4.7 g/l |
| MnSO ₄ .H ₂ O | 20.1 g/l |
| CuSO ₄ .5H ₂ O | 3.4 g/l |
| H ₃ BO ₃ | 4.7 g/l |
| Na ₂ MoO ₄ .2H ₂ O | 4.7 g/l |
| ZnSO ₄ .7H ₂ O | 53.0 g/l |
| EDTA | 14.1 g/l |
| Concentrated sulphuric acid (18M) | 4 ml/l |

Table 5.2: Trace element preparation required for the bacterial expression medium.

purification step for the EBA181₉₄₅₋₁₀₉₇ protein in order to remove the cleaved his tag). Finally, the proteins were purified on a size exclusion chromatography column (SuperdexTM 75). The two proteins were analysed in the following buffer conditions: 50 mM NaH₂PO₄ / Na₂HPO₄ at pH 6.0 + 50 mM NaCl.

5.2.2 Solution-state NMR experiments

5.2.2.1 Spectrometers

NMR experiments were performed on Varian/Agilent VNMRS 600 MHz and 800 MHz spectrometers using room temperature as well as cryogenically cooled triple-resonance HCN probes. Both of these spectrometers are located at the *Institut de Biologie Structurale* (IBS) located on the EPN campus. These NMR systems were used to carry out HSQC experiments on the samples made for this study (see more details in Section 2.2.4).

5.2.2.2 NMR spectroscopy

The NMR spectral assignment was obtained using a series of BEST-type triple resonance experiments: HNC0, intra-residue HN(CA)CO, HN(CO)CA, intra-residue HNCA, HNCOCACB and intra-residue HNCACB. The spectra were processed in NMRPipe (*Delaglio et al., 1995*), analysed using the program SPARKY (*T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco*) and automatic assignment of spin systems was achieved using MARS (*Jung et al., 2004*), followed by manual verification. Secondary chemical shifts were calculated using the random coil values from RefDB (*Zhang et al., 2003*). Site specific ^{15}N R_1 , R_2 and $\{^1\text{H}\}$ - ^{15}N heteronuclear nOes were measured using pulse sequences. The heteronuclear nuclear Overhauser effect (nOes) were obtained as the ratio between the two spectra acquired with and without initial proton saturation. The experiment was repeated twice and the values of the nOes were taken as the average between these two measurements. Residual dipolar couplings (RDCs) were measured in a liquid crystal composed of poly-ethylene glycol and 1-hexanol.

5.2.2.3 Structural ensemble selections on the basis of experimental NMR data

Ensemble selections of EBA181₉₄₅₋₁₀₉₇ were carried out by generating a large pool of statistical coil conformers (10000 structures) using Flexible-Meccano (*Ozenne et al., 2012*). From this pool, five representative ensembles of 200 conformers each were selected using ASTEROIDS (*Salmon et al., 2010*) on the basis of the experimental chemical shifts. The backbone dihedral angles were extracted from the five representative ensembles giving 1000 ϕ/ψ pairs for each residue. These dihedral angle distributions were used to generate a new pool of conformers of EBA181₉₄₅₋₁₀₉₇ from which a new round of selections of five ensembles was carried out. This procedure was repeated five times (five iterations) until convergence with respect to the experimental data was achieved.

The five ensembles obtained at the end of the fifth iteration were analysed in terms of site-specific populations in four distinct regions of Ramachandran space defined as: αL $\{\phi > 0^\circ\}$; αR $\{\phi < 0, -120^\circ < \psi < 50^\circ\}$; βP $\{-100^\circ < \phi < 0^\circ, \psi > 50^\circ \text{ or } \psi < -120^\circ\}$; βS $\{-180^\circ < \phi < -100^\circ, \psi > 50^\circ \text{ or } \psi < -120^\circ\}$. The populations of these quadrants are respectively denoted $p(\alpha\text{L})$, $p(\alpha\text{R})$, $p(\beta\text{P})$ and $p(\beta\text{S})$.

5.3 NMR solution spectroscopy measurements from EBA181₉₄₅₋₁₀₉₇

5.3.1 Structural and dynamic characterisation of EBA181₉₄₅₋₁₀₉₇

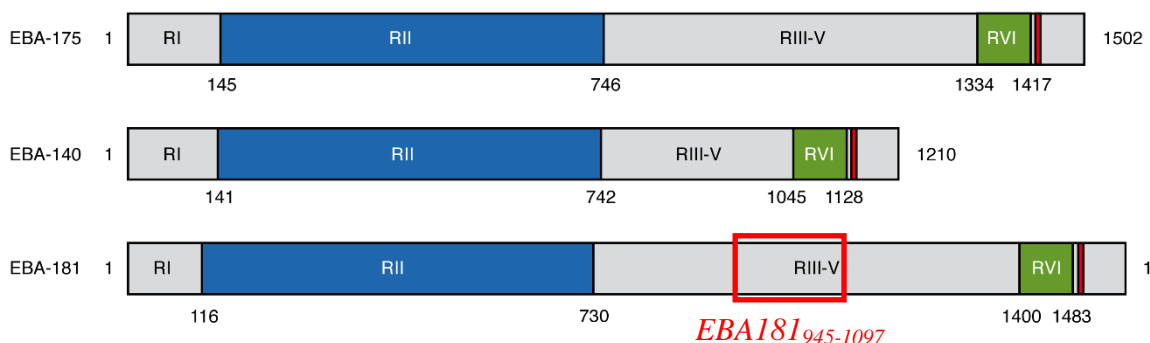
An analysis of the amino acid composition of RIII-V domains of the EBL proteins (Figure 5.1, A) reveals an abundance of charged amino acids (lysines, glutamic and aspartic acids), as well as hydrophilic residues (threonine and serine). Moreover, the EBL proteins are depleted in hydrophobic amino acids necessary for forming stable hydrophobic cores in folded proteins. A comparison with the amino acid composition of the DisProt database (<http://www.disprot.org/>) reveals that the disordered regions of the EBL proteins are richer in asparagines, serines and glutamic acids, while lacking alanines that normally represent one of the most common amino acid types in intrinsically disordered proteins (Figure 5.1, B). The functional role of RIII-V remains to be elucidated; however, they may function as interaction platforms for different cellular partners (see Chapter 3, prediction of interaction sites carried out using the server ANCHOR).

5.3.1.1 ¹H-¹⁵N HSQC and ¹³C-¹⁵N HSQC NMR spectra of labelled EBA181₉₄₅₋₁₀₉₇

Functionally disordered proteins often possess transiently populated secondary structures used by other proteins as molecular recognition elements. Once bound to their protein partners, these molecular recognition elements usually undergo folding into specific conformations. This pre-formation of secondary structure may facilitate the binding process by lowering the entropic cost of folding from a completely unfolded chain. In order to obtain information about transiently populated secondary structures,

EBA181₉₄₅₋₁₀₉₇ was labelled (¹⁵N and ¹⁵N-¹³C) and characterised at near atomic resolution using NMR spectroscopy.

A)



B)

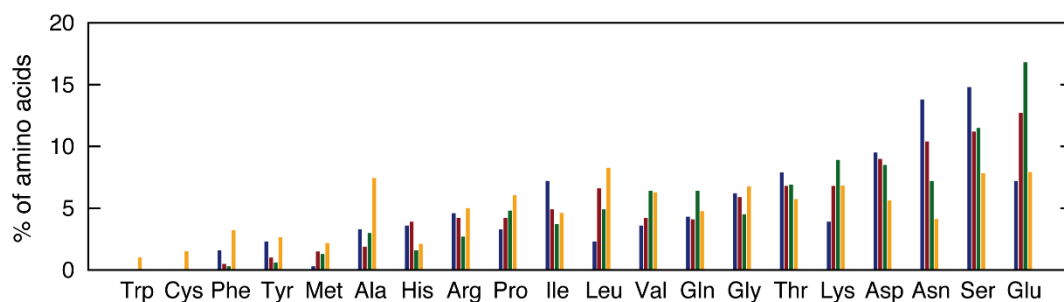
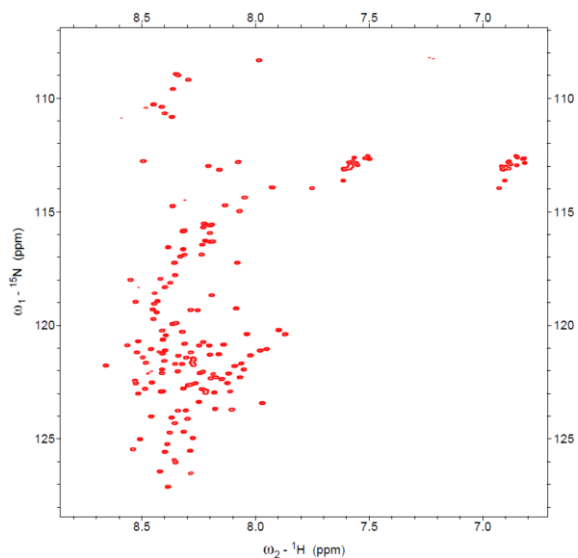


Figure 5.1: **(A)** Domain organisation of the three erythrocyte binding-like proteins EBA-175, EBA-140 and EBA-181. The blue and green colour indicates the folded regions (homology compared to the crystal structures of regions II and RVI available for EBA-175) and the positions of the transmembrane regions are indicated in red. The red box indicates the location of the EBA181₉₄₅₋₁₀₉₇ domain. **(B)** Amino acid composition within the RIII-V regions of the EBL proteins EBA-140 (blue), EBA-175 (red) and EBA-181 (green) compared to the composition of the DisProt database release 6.02 (orange).

Determination of the oligomeric state is a prerequisite required for the analysis of the NMR experiments. It was shown using size exclusion chromatography combined with detection by multi-angle laser light scattering and refractometry that EBA181₉₄₅₋₁₀₉₇ is monomeric in solution (Chapter 3). The far-UV circular dichroism spectrum of the protein displayed a typical signature of random coil conformations with a minimum around 202 nm, and the initial 1D NMR spectrum carried out also showed the unfolded state of EBA181₉₄₅₋₁₀₉₇ (Chapter 3). In order to increase the resolution of the initial 1D-NMR spectrum presented in Chapter 3, and to ensure the feasibility of the next NMR solution experiments involving labelling using ¹³C and ¹⁵N isotopes, the first NMR experiment was carried out on single labelled ¹H-¹⁵N EBA181₉₄₅₋₁₀₉₇. Thus, the disordered state of the EBA181₉₄₅₋₁₀₉₇ protein was demonstrated with certitude as shown by the HSQC spectrum and its limited signal dispersion in the ¹H dimension, characteristic of an intrinsically disordered protein (Figure 5.2, A).

In order to assign each amino acid to the different peak of the ¹H-¹⁵N HSQC spectrum, and to be able to carry out the full set of solution state NMR experiments, the EBA181₉₄₅₋₁₀₉₇ protein was produced as double labelled (¹³C and ¹⁵N). Figure 5.2 A displays the ¹H-¹⁵N HSQC spectrum on single ¹⁵N labelled EBA181₉₄₅₋₁₀₉₇ protein. The ¹³C-¹⁵N HSQC spectrum on double ¹⁵N-¹³C labelled EBA181₉₄₅₋₁₀₉₇ protein was also acquired but is not shown here. The complete spectral assignment of the ¹H^N, ¹⁵N, ¹³C α , ¹³C γ and ¹³C β nuclei of the protein was carried out using a set of standard triple resonance experiments (Appendix C, Table C.1) and the final HSQC spectrum with the complete attribution is presented in Figure 5.2 B.

A)



B)

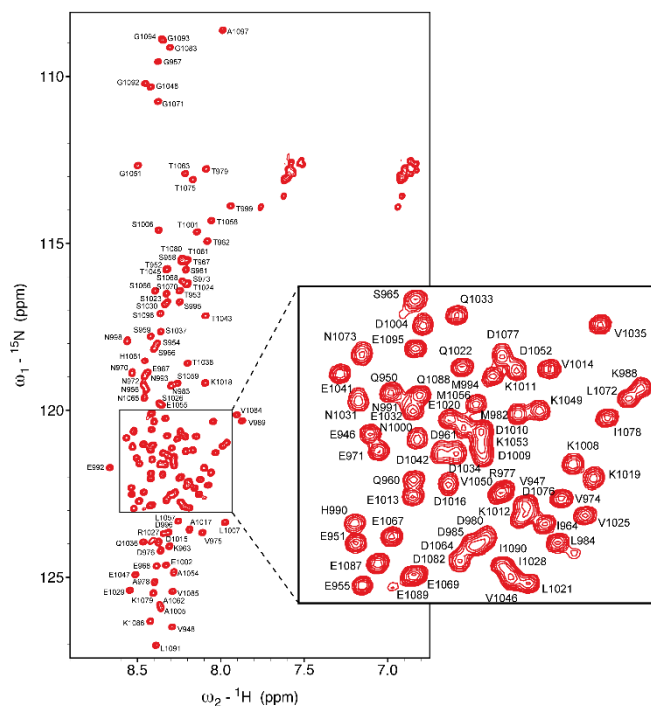


Figure 5.2: **A:** ^1H - ^{15}N HSQC on single labelled EBA181₉₄₅₋₁₀₉₇ (298 K) and **B:** ^1H - ^{15}N HSQC spectrum of EBA181₉₄₅₋₁₀₉₇ (298K) showing the spectral assignment of the NMR resonances.

5.3.1.2 Characterisation of the structure and dynamics of EBA181₉₄₅₋₁₀₉₇

In order to characterise further the transient structure and dynamics of the EBA181₉₄₅₋₁₀₉₇ protein, chemical shifts were measured. As shown in Figure 5.3 A, the secondary $\text{C}\alpha$ chemical shifts, which are the difference between the experimental $\text{C}\alpha$ chemical shifts and the chemical shifts expected for a random coil chain, showed the protein to be intrinsically disordered without significantly populated secondary structure elements. These results are

in agreement with the results from Chapters 3 and 4. The backbone dynamics of the protein were characterised (pico- to nano-second time scale) through the measurements of ^{15}N spin relaxation. With values close to zero at a ^1H frequency of 600 MHz, the $\{^1\text{H}\}\text{-}^{15}\text{N}$ heteronuclear nOes (Figure 5.3 B) displayed a profile characteristic of a disordered protein, in agreement with the previous results. The strong decrease of the nOes values from residues 1088 showed increased dynamics towards the C-terminus of the protein. On the contrary, the region encompassing residues 980-1025 shows on average larger heteronuclear nOes, with two short stretches having positive values (residues 986-988 and 1016-1020) corresponding to a rigidification of these parts of the protein compared to the remainder of the chain. The ^{15}N transverse and longitudinal spin relaxation rates (Figure 5.3 C and D) followed a similar pattern to ^{15}N R_2 rates, enhanced in the regions also showing positive heteronuclear nOes.

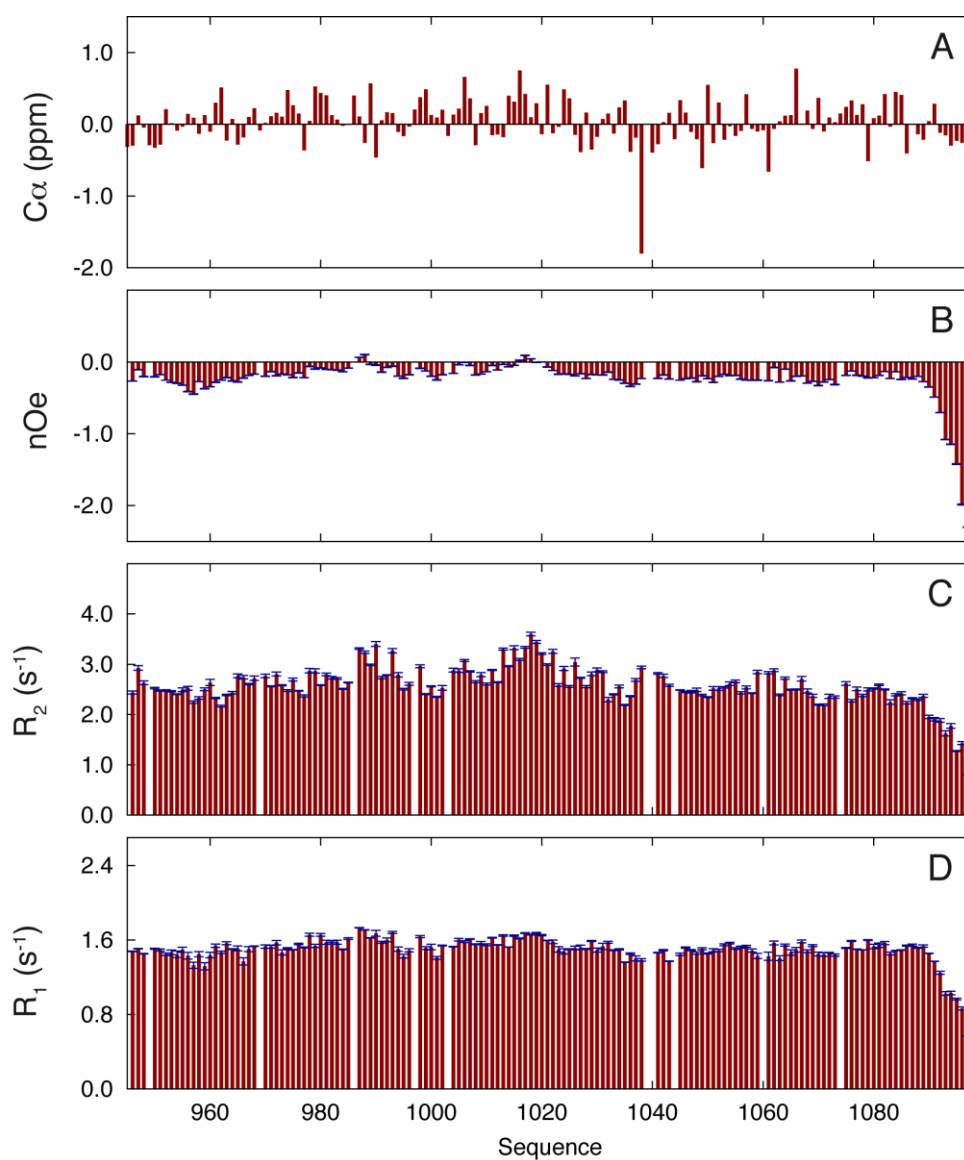


Figure 5.3: Characterisation of the structure and dynamics of EBA181₉₄₅₋₁₀₉₇. **(A)** Secondary C α chemical shifts of EBA181₉₄₅₋₁₀₉₇. **(B)** $\{^1\text{H}\}$ - ^{15}N heteronuclear nOes. **(C)** ^{15}N R_2 spin relaxation rates. **(D)** ^{15}N R_1 spin relaxation rates. All relaxation data were obtained at a ^1H frequency of 600 MHz and at 298 K.

5.3.2 Structural ensemble description of EBA181₉₄₅₋₁₀₉₇ from experimental chemical shifts

In order to obtain more details about the local conformational sampling of EBA181₉₄₅₋₁₀₉₇, ensemble selections of the protein were carried out based on the experimental chemical shifts. Thus, 10000 structures were generated using the statistical coil generator Flexible-Meccano to create a large pool of random coil conformers of the EBA181₉₄₅₋₁₀₉₇ protein. The genetic algorithm ASTEROIDS was finally used to select representative sub-ensembles of EBA181₉₄₅₋₁₀₉₇ based on the measured chemical shifts (200 conformers). Excellent agreement was obtained between the experimental chemical shifts and those back-calculated from the ASTEROIDS ensembles (Figure 5.4).

Following this, a detailed analysis of the site-specific conformational sampling in these selected ASTEROIDS ensembles revealed around 15 % depletion of the β -strand region of Ramachandran space compared to statistical coil sampling, as well as an increase in the poly-proline II (β P) populations in several regions of the protein. The increase reached 20% additional sampling in the β P region compared to statistical coil populations for some of the proline regions in the sequence, for example around P945, P949, P1003, P1039, P1040 and P1044 (Figure 5.5). In other regions, poly-proline II sampling was also enhanced but independently of the presence of proline residues (for example for the C-terminal residues 1085-1092).

Further analysis of the conformational sampling of EBA181₉₄₅₋₁₀₉₇ showed an increase of the alpha-helical population in four different regions of the protein: residues 979-982,

986-988, 1006-1009 and 1016-1019, corresponding to the region displaying the largest heteronuclear nOes and transverse spin relaxation rates as previously described. These regions could adopt turn motifs or short α -helices with populations of around 15-20 %, leading to a rigidification of the protein. This behaviour is also displayed by the calculation of the secondary structure propensity (SSP, *Kragelj et al., 2013*), a combined analysis of $C\alpha$ and $C\beta$ chemical shifts. The SSP is equal to 1 for 100 % α -helices and to -1 for 100 % β -sheets. As shown in Figure 5.6, the propensity to form secondary structure is very low, in the order of 10 to 20 %. Thus, the EBA181₉₄₅₋₁₀₉₇ protein is essentially intrinsically disordered, with a weak helical propensity (up to 20%) distributed all over the sequence.

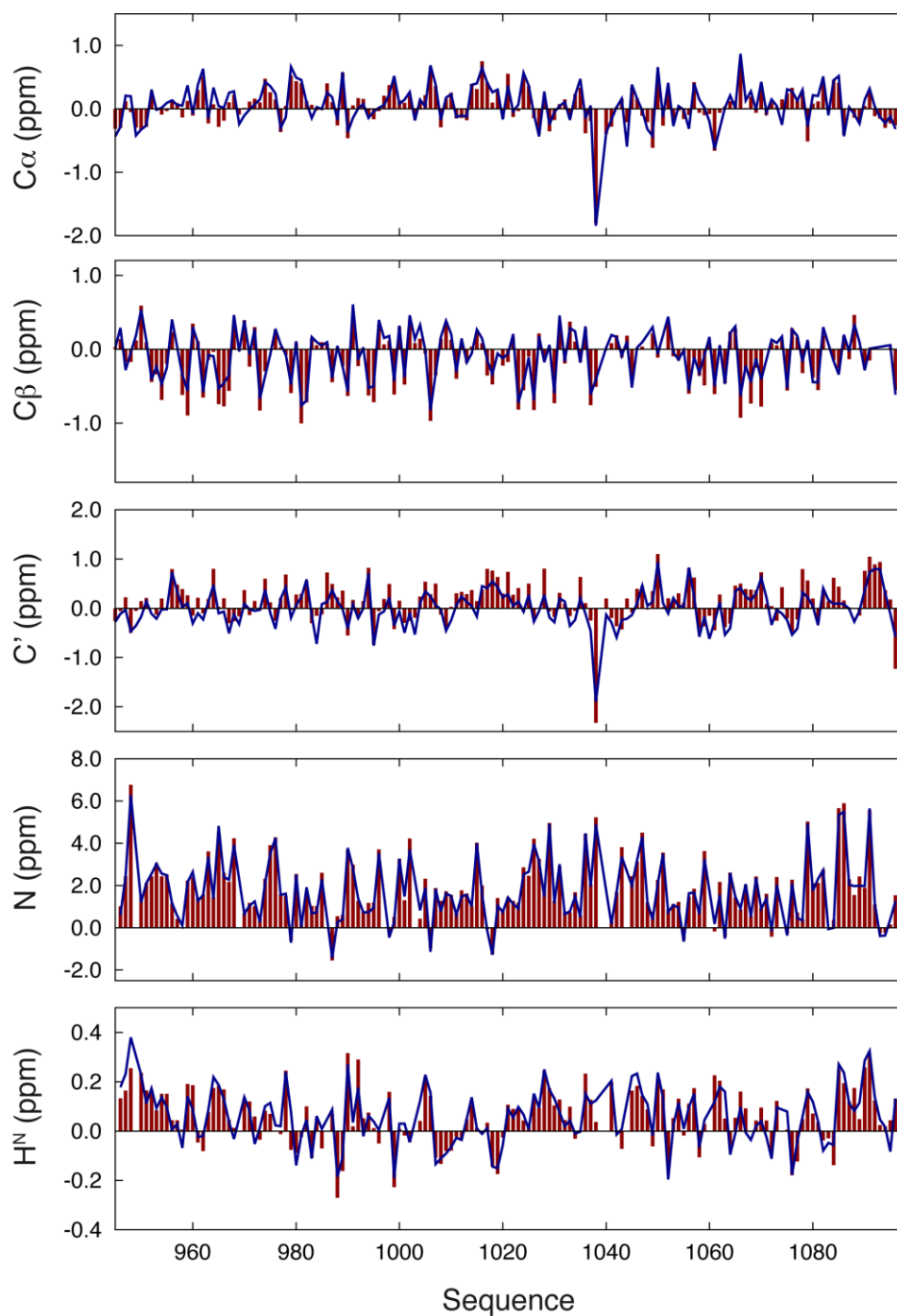


Figure 5.4: Generation of representative structural ensembles of EBA181₉₄₅₋₁₀₉₇ on the basis of experimental chemical shift values. Comparison of experimental secondary chemical shifts (red) and back-calculated values from a selected ASTEROIDS ensemble comprising 200 conformers (blue). From top to bottom: $C\alpha$, $C\beta$, C' , N and H^N secondary chemical shifts.

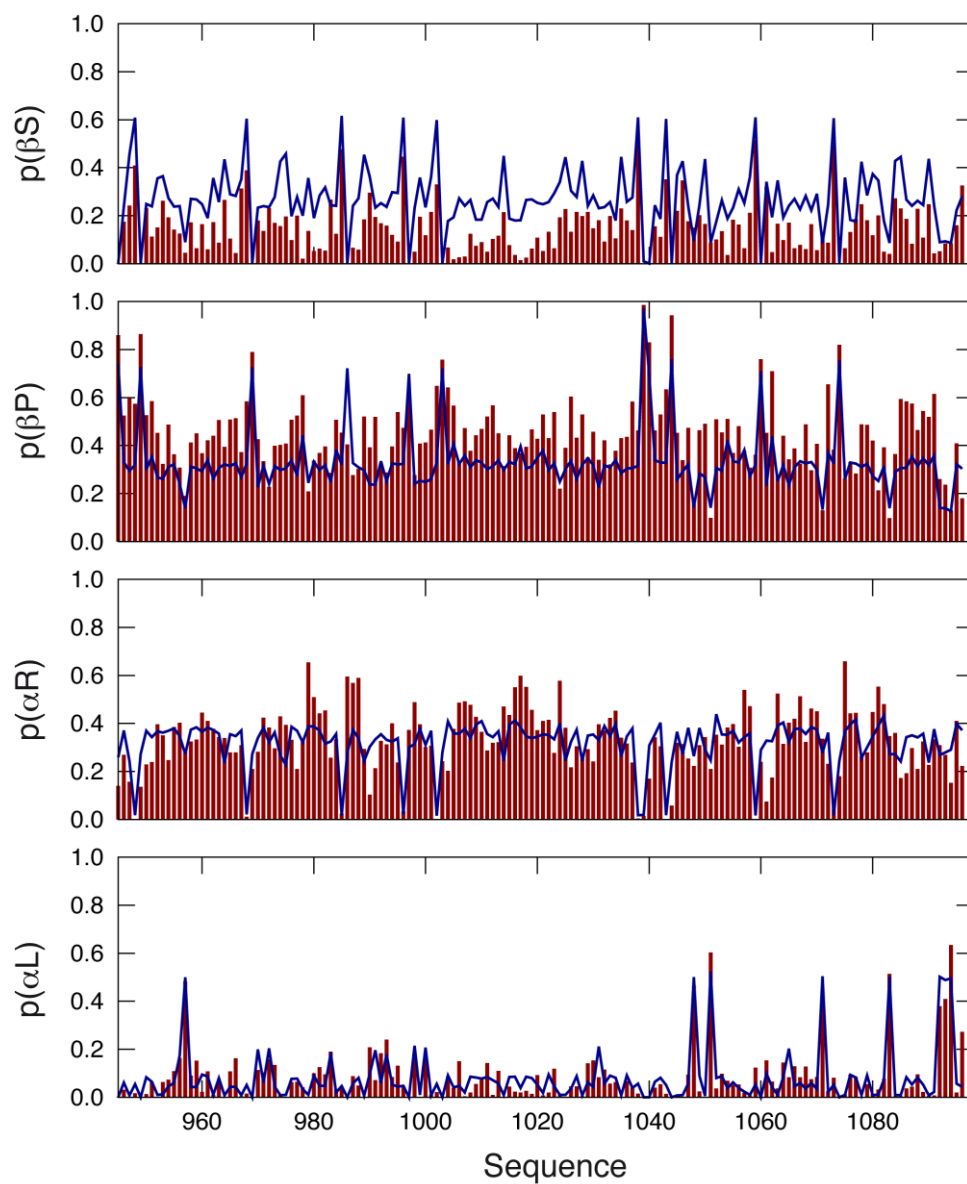


Figure 5.5: Site-specific populations of EBA181₉₄₅₋₁₀₉₇ in four distinct regions of Ramachandran space derived from ASTEROIDS ensembles selected on the basis of experimental chemical shift. Red bars indicate populations derived from the experimental chemical shifts, while blue lines correspond to statistical coil populations.

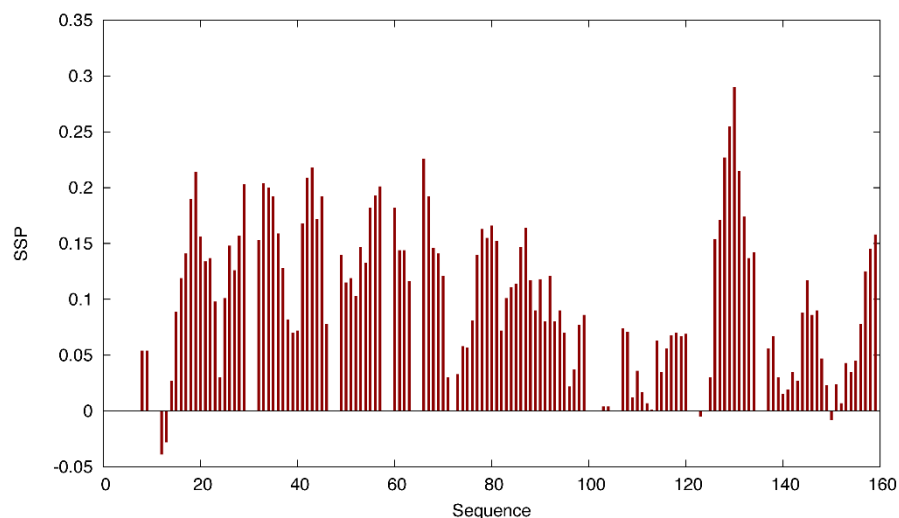


Figure 5.6: EBA181₉₄₅₋₁₀₉₇ secondary structure propensity (SSP), combined analysis of C α and C β chemical shifts. SSP: +1 for 100 % α -helices and -1 for 100 % β -sheets.

5.3.3 Validation of structural ensembles by experimental residual dipolar couplings

To confirm the previously selected structural ensembles, $^1D_{NH}$ residual dipolar couplings were measured on EBA181₉₄₅₋₁₀₉₇ in a liquid crystal composed of poly-ethylene glycol and 1-hexanol. Helical elements or turn motifs gives rise to a sign inversion of the RDCs compared to random coil conformations; residual dipolar couplings are thus a powerful tool to sample the backbone conformation of disordered proteins. As expected for a disordered protein which mainly samples random coil conformations, the experimental $^1D_{NH}$ RDCs were negative, except for four short stretches encompassing the residues 987-988, 998, 1006-1007 and 1016-1019, for which the RDCs were positive (Figure 5.7). The experimental RDCs matched reasonably well with those predicted from a statistical

coil ensemble generated using Flexible-Meccano (Figure 5.7 A). The reproduction of the experimental RDCs using the conformational sampling of EBA181₉₄₅₋₁₀₉₇ from the experimental chemical shifts gave a slight improvement of the results (Figure 5.7 B). The biggest improvement in the reproduction of the RDCs was observed within the turn or short helical motifs identified by the experimental chemical shifts and leading to positive back-calculated RDCs. Moreover, other improvements were observed in the poly-proline II regions of the protein (towards the N- and C-termini of the protein as well as around P1039-P1044). The improvement in the reproduction of the RDCs compared to a standard statistical coil ensemble provides an independent verification of the validity of the ASTEROIDS ensembles and testifies to the predictive nature of these ensembles.

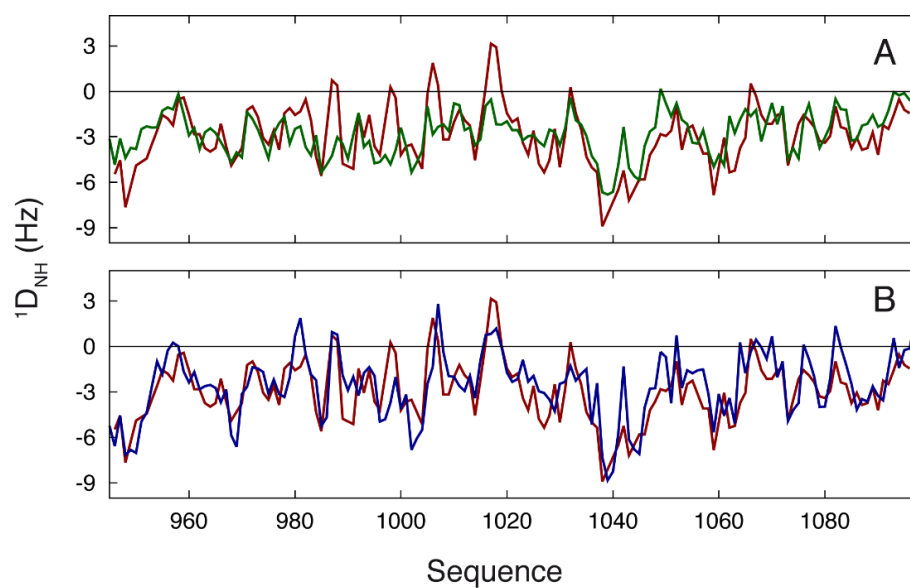


Figure 5.7: Validation of selected ASTEROIDS ensembles of EBA181₉₄₅₋₁₀₉₇ using $^1D_{NH}$ residual dipolar couplings. (A) Comparison of experimental (red) and predicted (green) $^1D_{NH}$ RDCs from a statistical coil ensemble. (B) Comparison of experimental (red) and predicted (blue) $^1D_{NH}$ RDCs from the representative ensembles of EBA181₉₄₅₋₁₀₉₇ derived on the basis of experimental chemical shifts only.

5.4 NMR solution spectroscopy results of the complex

Titration NMR spectroscopy was used to study the structure and binding properties of the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} complex. The titration experiments were performed in both directions, firstly with 4.1R_{10kDa} as the titrant being added to ¹⁵N/¹³C EBA181₉₄₅₋₁₀₉₇ and secondly, with EBA181₉₄₅₋₁₀₉₇ as the titrant being added to ¹⁵N/¹³C GST-4.1R_{10kDa}. Tagged 4.1R_{10kDa} domain (GST-4.1R_{10kDa}) was used in order to maximise the final quantity of the protein because of the already low expression yield coupled to the expensive materials used for labelling.

In the first titration experiment, the HSQC spectrum of labelled ¹⁵N/¹³C EBA181₉₄₅₋₁₀₉₇ was acquired (Figure 5.8, signals in red). Then, a 7.5 fold excess of non-labelled 4.1R_{10kDa} was added. A second spectrum was acquired (Figure 5.8, blue signals) aiming to follow the spatial displacement of the signals of labelled EBA181₉₄₅₋₁₀₉₇ on the ¹H-¹⁵N HSQC spectrum. No chemical shift changes nor modulation of the intensities were observed (Figure 5.8). Two explanations can be put forwards: either 1) the complex between EBA181 and 4.1R was not formed, or 2) there was no change in the chemical environment of EBA181₉₄₅₋₁₀₉₇ when bound to 4.1R_{10kDa} which is however unlikely to happen. This experiment was performed twice, with the same result.

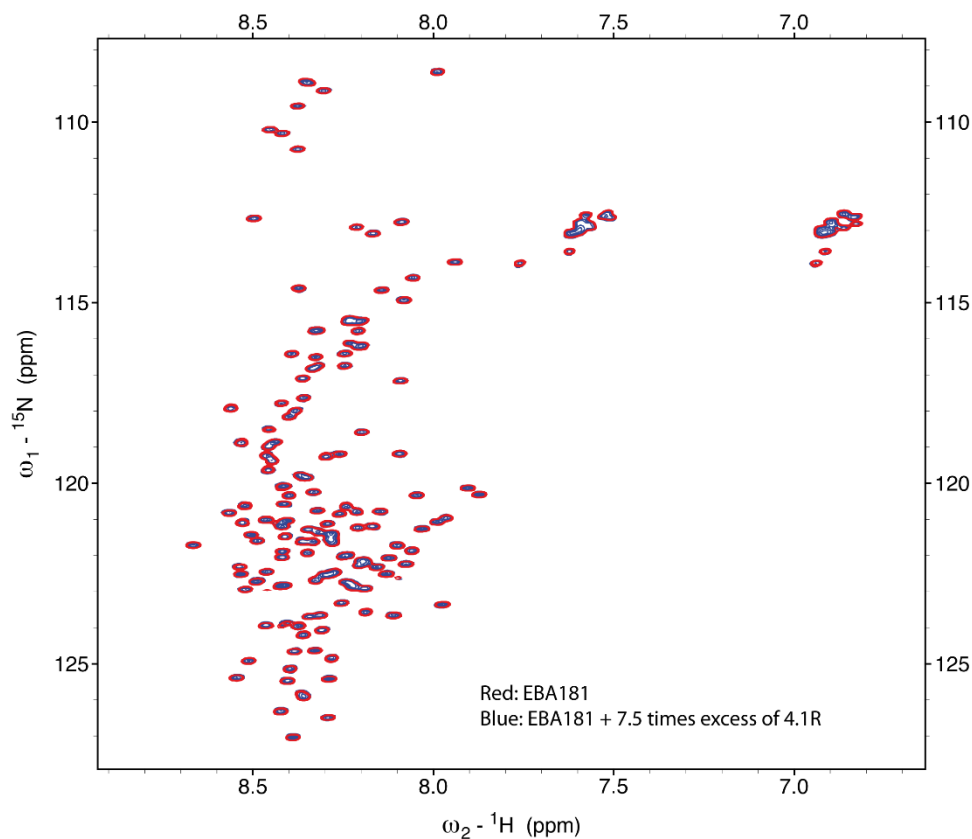


Figure 5.8: ^{13}C - ^{15}N HSQC spectrum. Titration of double labelled EBA181₉₄₅₋₁₀₉₇ with unlabelled 4.1R_{10kDa}. Superposition of the spectra of EBA181₉₄₅₋₁₀₉₇ without 4.1R_{10kDa} (red) and with 7.5 fold excess of 4.1R_{10kDa} (blue).

In the second titration experiment, the HSQC spectrum of double labelled $^{15}\text{N}/^{13}\text{C}$ GST-4.1R_{10kDa} was first acquired (Figure 5.9, signals in red). Following this, EBA181₉₄₅₋₁₀₉₇ was added up to a 1 to 1 molar ratio (Figure 5.9, blue signals) and then in a 4 to 1 ratio to reach an excess of EBA181₉₄₅₋₁₀₉₇. In this experiment changes of chemical shifts and modulation of intensities were noticed (Figure 5.10), indicating that this time the interaction between EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} is observed. Unfortunately these variations were too low in magnitude to continue further the characterisation.

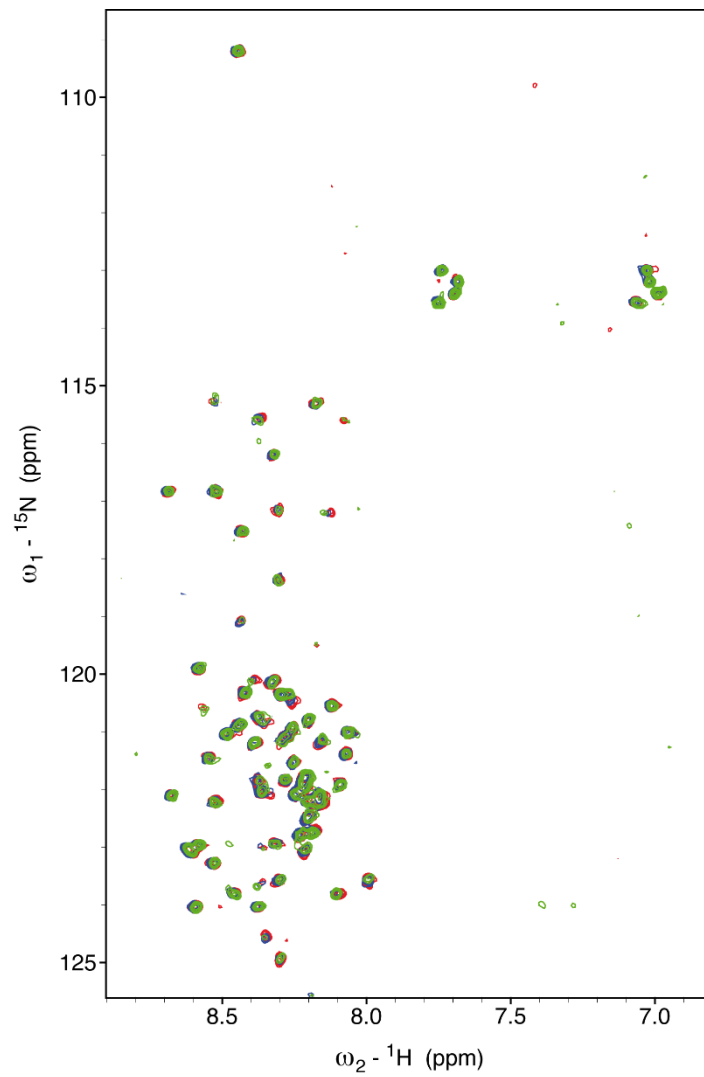


Figure 5.9: HSQC spectrum. Titration of labelled GST-4.1R_{10kDa} with unlabelled EBA181₉₄₅₋₁₀₉₇. **Red:** 4.1R_{10kDa} without EBA181₉₄₅₋₁₀₉₇, **blue:** 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ in 1:1 ratio, **green:** 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ in 1:4 ratio.

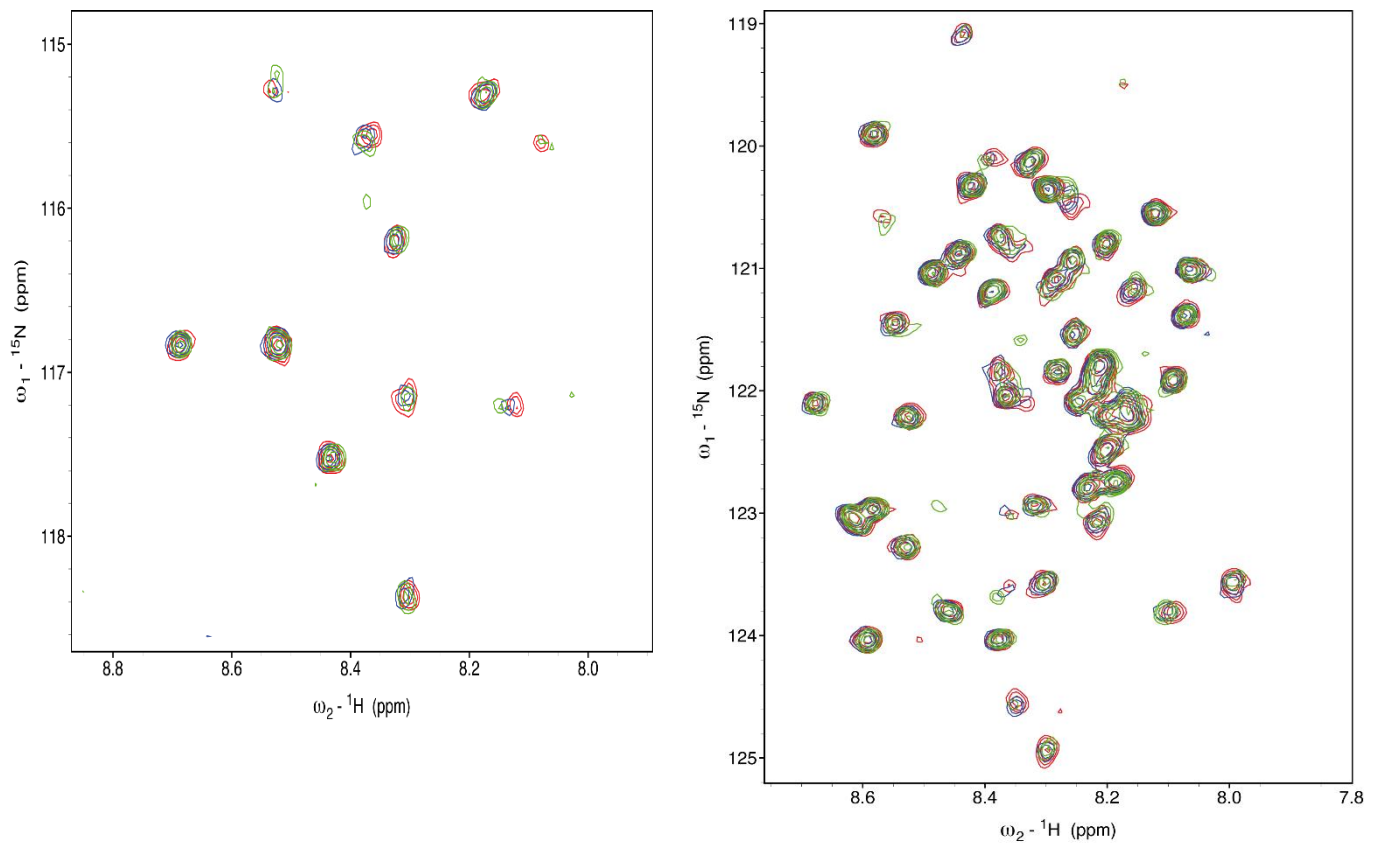


Figure 5.10: HSQC spectra (zoom in two regions). Titration of labelled GST-4.1R_{10kDa} with unlabelled EBA181₉₄₅₋₁₀₉₇. **Red:** 4.1R_{10kDa} without EBA181₉₄₅₋₁₀₉₇, **blue:** 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ in 1:1 ratio, **green:** 4.1R_{10kDa} and EBA181₉₄₅₋₁₀₉₇ in 1:4 ratio.

5.5 Discussion and conclusions

In the relatively recent past, intrinsically disordered proteins (IDPs) have appeared as a new class of proteins which challenge the classical structure-function paradigm. They do not adopt well-defined folded structures, and sample different conformations in solution. This is tremendously important for their structural and dynamic characterisation. One of the most significant characteristics of IDPs is their ability to interact with multiple partner proteins.

In addition to chapter 3, a specific bioinformatics analysis of the EBL proteins using a database on disordered proteins was carried out and revealed the intrinsically disordered nature of the RIII-V regions, encompassing multiple predicted interaction sites. Only a few interaction partners to the EBL proteins have been identified so far, leaving the role that these regions play in the invasion process largely unknown. Considering the variable length of RIII-V within the three EBL proteins, it can be hypothesised that these disordered regions direct the EBL proteins towards different cellular functions by discriminating between different partner proteins. This theory is supported by the fact that the three proteins share very little sequence conservation within RIII-V, and this also provides a means of diversifying the invasion process across the EBL family. Thus, it can be imagined that RIII-V binds to proteins that are subsequently sequestered and inactivated, thereby enabling a simpler invasion. RIII-V regions could also cooperate with proteins that play a more direct role in the invasion process. For example, interactions with host erythrocyte proteins, such as protein 4.1R, could destabilise the membrane skeleton and may also weaken host membrane repair pathways.

In order to obtain more insight into the conformational state within the RIII-V of the EBA-181 protein, structural and dynamic characterisation of EBA181₉₄₅₋₁₀₉₇ was performed using NMR spectroscopy. The data demonstrated that EBA181₉₄₅₋₁₀₉₇ is essentially a random coil with a number of turns or short alpha-helical motifs. Three of the four motifs are located in the 30 amino acid binding zone with 4.1R (as identified by Lanzilotti *et al.* in 2005) and could function as molecular recognition elements and in that way play a role in the biological function of the protein. Nevertheless, this remains to be elucidated. The spectral assignment of the protein has provides a new basis for mapping and identifying interactions with different partner proteins in the future, and that at amino acid resolution. This could therefore provide valuable information on host-pathogen interactions during the invasion of erythrocytes.

6. DISCUSSION, CONCLUSIONS and FUTURE WORK

The main results from this thesis work relate to the malarial parasite interaction with red blood cells either via the sialic acid dependent or independent pathways. Here, key interactions involve the EBA-175, EBA-140 and EBA-181 membrane receptors. Of these, both EBA-175 and EBA-140 bind to sialic acid and then to glycoporphins A and C respectively, whilst EBA-181 binds to sialic acid and possibly to an unknown second receptor on the erythrocyte surface. Collectively, these interactions are likely to be important for the ability of *P. falciparum* to evade the immune system upon its entry to the erythrocyte.

In this thesis work, the key objective was to bring a wide range of biochemical and biophysical techniques together to achieve a consensus view on the structure of the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins as well as the complex formed between them. The specific focus of the results summarised in Chapters 3, 4 and 5 is the interaction between the EBA-181 protein (located at the surface of the parasite *in vivo*) and the 4.1R protein (located within the red blood cell *in vivo*). This interaction, which was first identified using phage display, was confirmed during all of the major biochemical and biophysical experiments carried out during the project. One of the major challenges was, preliminary to the biochemical and biophysical characterisation, to optimise the quantities of proteins produced from the initial microgram scale to the milligram level required for structural work. Another point of importance was the production of labelled proteins (isotopes, including deuterium and ¹⁵N/¹³C labelling). Finally, the optimisation of protein conditions

(in particular buffers and aggregation states) as well as storage conditions was critical to the viability of the biophysical/biochemical characterisation.

In Chapter 3, the biochemical studies and pull-down analyses confirmed the interaction between the two proteins, and enabled further characterisation of the complex under a wide range of conditions. The affinity between EBA-181 and 4.1R is in the micro-molar range and this interaction was proved to resist high salt concentration (tested up to 2 M NaCl), implying that it is likely to resist high local salt concentrations in the cell. The interaction is also resistant to pH variation (tested from pH 5.0 to 9.0), showing the capacity of the parasite to resist different stresses. In addition, none of the short EBA-based peptides used as competitors to the EBA-181 / 4.1R complex formation disrupt the interaction between EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}, even with a 100 fold excess of peptide. This is an important result and suggests that even if the shorter interacting sequence is sufficient to bind to the 4.1R_{10kDa} protein, the whole EBA181₉₄₅₋₁₀₉₇ sequence is important for the binding affinity. It would be interesting for future work to clone and express longer constructs of EBA-181 in order to control the effects on affinity and binding to 4.1R_{10kDa}, especially by comparing it to the EBA181₉₄₅₋₁₀₉₇ construct.

The biophysical characterisation highlighted the intrinsically disordered state of EBA181₉₄₅₋₁₀₉₇, as well as the low level of folding in 4.1R_{10kDa}, which contains some alpha-helical structure. Furthermore, the small-angle scattering studies described in Chapter 4, using both neutron and X-ray approaches, demonstrated again the interaction and characterised it in more detail, providing the respective sizes and molecular shapes of the individual proteins and the complex in solution (Table 6.1) and showing this complex

to be a 1 to 1 complex. Moreover, the results were consistent between the SAXS and SANS analyses and have provided good agreement and complementarity in terms of the solution state modelling of the individual proteins as well as the complex formed between them. It is also interesting to compare the resulting sizes of the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins both from the SAXS analysis and the structural predictions performed by informatics. The structural prediction of EBA181₉₄₅₋₁₀₉₇ results in an apparent R_g of 5.3 nm, for an R_g of only 2.35 nm from the SAXS experiment. This difference is due to the unstructured character of the protein which appears as a long, coiled and fully disordered sequence by bioinformatics, resulting in a bigger R_g . The prediction may have not taken into account the dynamics and the local secondary structures that could rigidify or impose some local conformations on the protein. Bioinformatics structural predictions have to be considered with caution since it is very difficult for these software to predict the behaviour and the molecular shape of intrinsically disordered proteins. It is different for 4.1R_{10kDa}. Its structural prediction results in an R_g of 1.6 nm which is close to the value of 2.0 nm obtained from the SAXS data. The prediction of 4.1R_{10kDa} is probably more relevant than that for EBA181₉₄₅₋₁₀₉₇ since the protein contains some secondary structure and it might thus be easier to predict its 3D structure.

The nuclear magnetic resonance spectroscopy work performed on EBA181₉₄₅₋₁₀₉₇ and described in Chapter 5 enabled a detailed structural and dynamic characterisation of the protein at near atomic resolution and allowed preliminary work on the complex with 4.1R to be carried out. In this study, EBA181₉₄₅₋₁₀₉₇ was shown to be essentially a random coil, with the presence of short alpha-helical motifs. These structural motifs could function as molecular recognition elements and thereby play a role in the biological function of the

protein during erythrocyte invasion. Indeed, the majority of these short motifs appear in the EBA binding region for 4.1R identified by Lanzilotti *et al.* in 2005. It is interesting to note that the protein does not possess transiently populated secondary structures as commonly found for many intrinsically disordered proteins that fold via specific, pre-formed molecular recognition elements. Future work could include similar experiments on 4.1R_{10kDa}, complementing the CD analysis described in Chapter 3 and providing insight to the presence and localisation of structural elements or at least transiently populated secondary structures. The NMR results obtained for the complex were unfortunately insufficient to further characterise it, despite the fact that several experiments were performed, involving two complementary isotope labelling schemes with titrations of both proteins. Future work in this area is therefore highly desirable to resolve this situation.

In terms of linking these observations to the wider context of the interaction between the two proteins *in vivo*, it is obviously hard to bridge the molecular to the microscopic and cellular lengthscales, and considerable caution needs to be exercised to avoid over-interpretation.

The results show both proteins to be disordered. It is now clear that IDPs are of fundamental importance in nature. IDPs have been found in bacteria, archaea and eukaryotes, and their presence in the literature is rising steadily (www.disprot.org, Sickmeier *et al.*, 2007). To our knowledge, this is the first example of small-angle solution scattering experiments involving two intrinsically disordered proteins remaining disordered after the formation of their complex. Binding of an IDP to its protein partner is very often


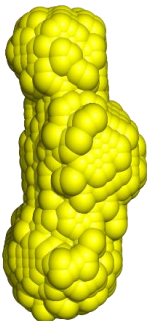
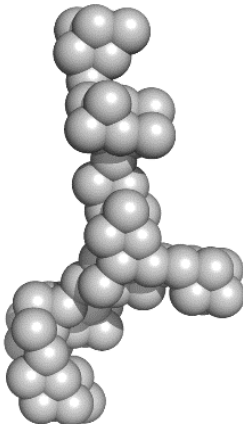
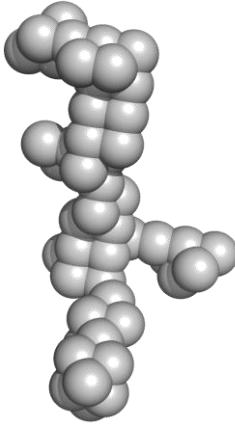
| Structural parameters | 4.1R _{10kDa} | EBA181 ₉₄₅₋₁₀₉₇ | Untagged complex | |
|-----------------------------------|---|---|--|---|
| | SAXS | SAXS | SAXS | SANS (0 % D ₂ O) |
| Buffer | NaP pH 6.0 | NaP pH 6.0 | NaP pH 6.0 | NaP pH 6.0 |
| Theoretical MM | 8 kDa | 17 kDa | 25 kDa | 25 kDa |
| MM - $I(0)$ from Guinier analysis | 10 kDa | 21 kDa | 26 kDa | 26 kDa |
| R_g - from Guinier analysis | 2.02 nm | 2.35 nm | 3.54 nm | 3.95 nm |
| D_{max} | 6.0 nm | 7.9 nm | 10.4 nm | 10.8 nm |
| R_g - from $P(r)$ analysis | 1.90 nm | 2.30 nm | 3.53 nm | 3.90 nm |
| <i>Ab-initio modelling</i> |  |  |  |  |

Table 6.1: Experimentally determined structural parameters of the best quality SAXS and SANS scattering curves used for *ab-initio* modelling of untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins as well as the complex.

followed by a conformational change in which the disordered protein becomes folded as part of the complex. This does not seem to be the case for the 4.1R_{10kDa} / EBA181₉₄₅₋₁₀₉₇ complex. This type of structural disorder as the functional native state of proteins, and as a property of an IDP complex, is defined as fuzziness. This concept of fuzziness has only been recently established (*Tompa et al., 2008*) and it has become more and more clear that fuzzy complexes play a distinct role in a wide range of biological processes (*Fuxreiter and Tompa, 2011*). In these cases, even if high-resolution structural studies are not feasible, low-resolution methods coupled to biochemical approaches can provide valuable information, as emphasised during this work.

One of the reasons that the two domains involved in the interaction are structurally disordered may be that the individual proteins have multiple binding partners under different functional contexts. This is plausible for the 10 kDa domain of the erythrocyte 4.1R protein which interacts with a wide variety of cytoskeleton proteins. For the parasite EBA181₉₄₅₋₁₀₉₇, it is conceivable that its intrinsically disordered structure suppresses the formation of recognition elements that could then be recognised by the human immune system. Indeed, the disordered state of EBA-181 and other parasite proteins could be of fundamental biological importance for *Plasmodium falciparum* in escaping the immune system. For example, Uversky and colleagues showed in 2005 that the intrinsic immunogenicity of disordered protein sequences are lower than that of more conserved and ordered regions (*Uversky et al., 2005*). Moreover, this lack of secondary structure could also be an advantage in terms of the interaction with the 4.1R protein, which is set in a complex and dense cytoskeleton network. The intrinsically disordered state of EBA-181 increases its flexibility and may well change its dynamics. IDPs have high specificity in

target selection, and a fast association/dissociation rate (*Tompa, 2005*), which is likely to be important for EBA-181 in its interaction with 4.1R and/or other human proteins.

Given that the 4.1R protein is inside the erythrocyte, it is perhaps surprising that the parasite membrane protein EBA-181 interacts with human 4.1R. From the results described in this thesis, the following mechanism of action is speculatively proposed (Figure 6.1).

Since the 4.1R protein is inside the red blood cell, the interaction between the two proteins cannot be immediately direct. After initial binding to the erythrocyte surface and reorientation, the parasite brings its apical end in contact with the red blood cell membrane (*Gaur et al., 2004*), signalling the secretion of the microneme and rhoptry proteins onto the parasite surface. From them, the DBL proteins (EBA-140, EBA-175 and EBA-181) are translocated onto the merozoite membrane through their C-terminal hydrophobic and transmembrane domains, with their DBL domains being free to interact with the erythrocyte receptors (*Gilberger et al., 2003*). A tight junction is thus established and ultimately connected to the actinomyosin motor which will drive the inwards motion inside the erythrocyte. Following this, the parasite will lose all its membrane receptors through the activity of proteases and becomes invisible to the immune system (*Ejigiri et al., 2012*). The disordered state of EBA-181 could be a helpful characteristic to favour its own release through the action the proteases during this process of removal of the parasite membrane receptors. Finally, the membranes seal once the tight junction reaches the posterior end of the parasite, concluding the invasion process. During or directly after this process, the EBA-181 protein could be secreted in the red blood cell in order to interact with 4.1R.

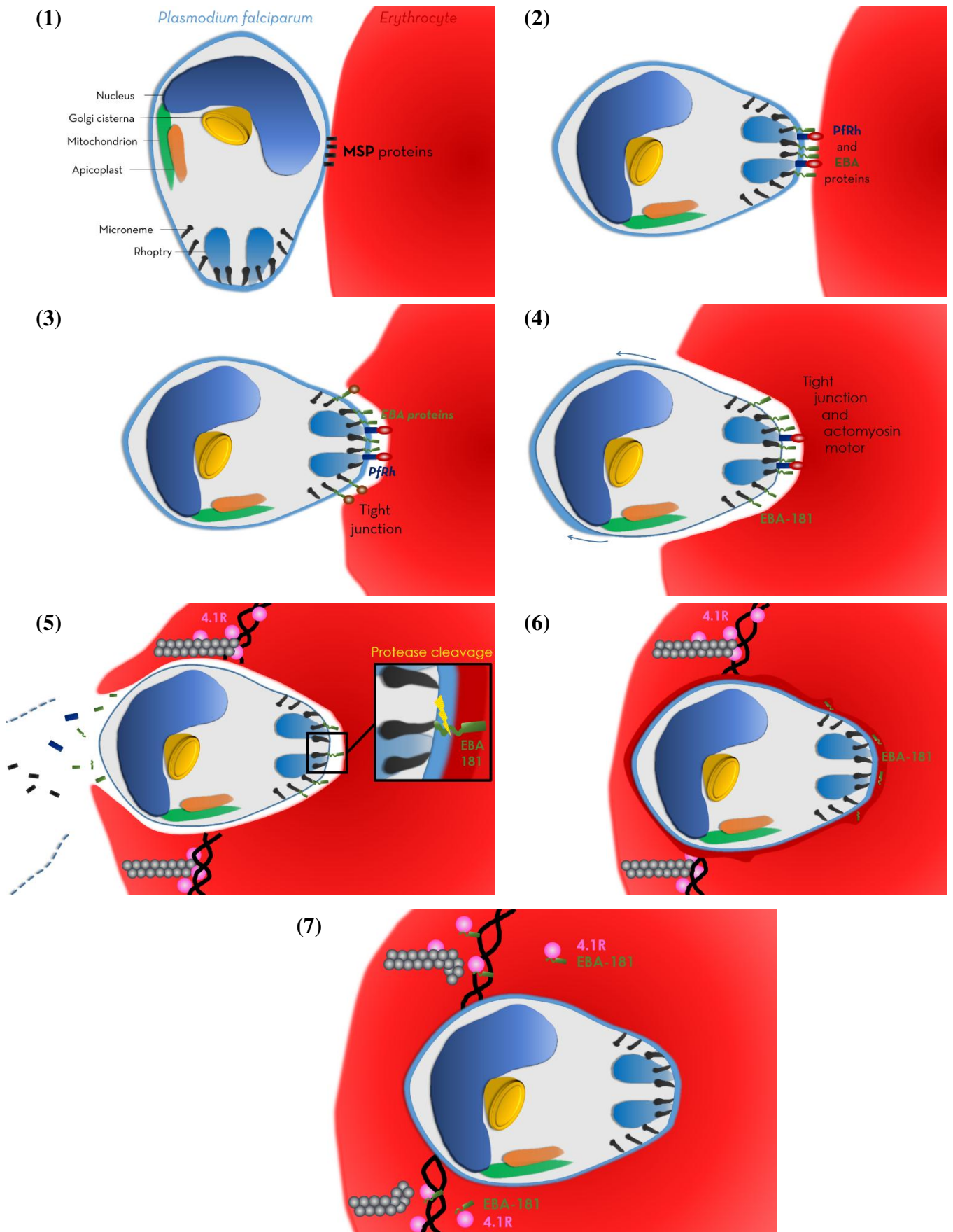


Figure 6.1: Putative mechanism for erythrocyte invasion by *P. falciparum*. **(1)** Adhesion. **(2)** Reorientation. **(3)** Tight junction formation. **(4)** Invasion. **(5)** Shedding of the surface

coat and possible role of the disordered region of EBA-181 with protease cleavage. **(6)** Resealing of the erythrocyte and parasitophorous vacuole membranes. **(7)** Interaction between EBA-181 and 4.1R proteins. EBA-181 is released in the red blood cell and is thus free to interact with the 4.1R protein to disrupt/weaken the cytoskeleton and prevent the activation of the repair machinery.

As previously described (*Fowler et al., 1985; Pasternack et al., 1989 and Conboy et al., 1993*), the 10 kDa domain of 4.1R maintains the plasticity of the red blood cell by facilitating the interaction between spectrin and actin, and it is also known to bind and regulate myosin activity. This whole network is crucial to regulate the shape of the erythrocytes by an ATP-dependent process (*Fowler et al., 1985*). Since 4.1R is vital in stabilising the red blood cell structure and controlling the erythrocyte membrane integrity, disrupting the interaction between the 4.1R protein and the structural spectrin, actin and glycophorin C proteins will perturb the cytoskeleton of the cell, favouring the parasite exploitation of the human cell to its own advantage. This could be an important role of the EBA-181 protein, disturbing the erythrocyte conformation by interacting with 4.1R. The disordered nature of EBA-181 may also increase its flexibility and dynamics; this could be important for its mobility in the purpose of interacting with 4.1R, given that it is set in this dense cell network. Thus, measurement of the dynamics of EBA181₉₄₅₋₁₀₉₇ using neutrons for example, could lead to a better understanding of the behaviour of the protein.

Additionally, EBA-181 may inhibit potential host membrane repair pathways subsequent to mechanical or chemical stress which damage the erythrocyte membrane. After

disruption of the skeletal network the cytosolic myosin is shifted to the erythrocyte bilayer, to form an actomyosin complex with the actin protofilaments and initiate the repair of the damaged zone (*Cibert et al., 1999*). In this sense, the damage caused by the invading parasites might initiate this actomyosin repair process and provide, with the restoration of the membrane, a barrier for the entry of other parasites. It is thus possible that the interaction between EBA-181 and 4.1R prior to/during the parasite entry is an evolutionary adaptation of *P. falciparum* to ensure its own survival, whereby EBA-181 would destabilise the linkages between 4.1R, spectrin, actin and myosin to prevent the activation of the actomyosin repair machinery and the restoration of the erythrocyte membrane.

A recent study by Persson *et al.* has shown that the invasion-inhibitory activity of acquired antibodies from children exposed to malaria was significantly reduced when using *P. falciparum* parasites that did not express the EBA-181 protein, compared to unmodified parasites (*Persson et al., 2013*). The same phenomenon was as well noticed for the two other EBA-140 and EBA-175 homologous proteins. In a near future the EBA proteins could thus be valuable targets for the development of vaccines to malaria disease. Nevertheless, the diversity in term of expression and use of EBA proteins (as well as PfRh proteins) by the parasite clearly implies that any vaccine that targets these invasion proteins would require, to be effective, the inclusion of multiple antigens, and thus combine both the EBA and PfRh proteins. This might help to overcome the capacity of the parasite to evade the human immune system.

In the context of this thesis and longer term perspectives, future work is needed to obtain further biophysical and structural information that would aim to permit the development of

novel therapeutic strategies for malaria prevention. A long term goal could be to express the full-length EBA-181 protein by overcoming the issues of expression and solubility of this membrane protein (for example using emerging cell free expression systems) in order to perform combined structural characterisation and understand the impact of the whole protein on binding to the 4.1R cellular skeletal protein. A further direction linking to the cellular / organism would be to take advantage of newly developed genome editing technology that have recently been demonstrated to be usable in *Plasmodium* genome (Ghorbal *et al.*, 2014). Here, one could imagine making point mutations in the EBA gene and studying the outcome in function. Another angle would be to perform *in vivo* studies to follow the process of invasion and to determine if the putative pathway shown in Figure 6.1 is correct. This could for example be achieved by GFP-labelling of the full-length protein and by performing invasion studies to follow its localisation in the cell, using high-resolution fluorescence microscopy. Further to structural characterisation, high-throughput screening methods using mass spectrometry and fragment libraries can be used to explore molecules able to bind and disrupt the complex formation, generating leads for potential therapeutics.

APPENDIX A: ADDITIONAL MATERIAL

A.1 Bacterial media

Below are presented the composition of the bacterial growth media used for the expression of hydrogenated protein.

LB broth medium

The LB broth medium was developed by Giuseppe Bertani in 1951 (*Bertani, 1951*). The LB medium was used for both small and high scale cultures of proteins.

For a final volume culture of 1 litre, 10 g of bacto-tryptone, 5 g of yeast extract and 10 g of NaCl are mixed together. Each flask are sterilised by autoclaving and stored at room temperature.

LB-agar

The LB-agar medium was used to make Petri dishes containing bacterial clones expressing the GST-4.1R_{10kDa} and His-EBA181₉₄₅₋₁₀₉₇ proteins.

15 g of agar powder was added to the previous LB medium. Each bottle of LB-agar were then sterilised by autoclaving (this process will also melt the agar), then left cool down to around 30 degrees prior adding the right antibiotics (30 µg/ml for kanamycin or 100 µg/ml

for ampicillin, and 34 µg/ml for chloramphenicol). Finally 20 ml of LB agar was poured into each Petri dish and stored at 4°C prior to be used.

Super optimal broth with catabolite repression

The super optimal broth with catabolite repression (SOC medium) was developed by Douglas Hanahan in 1983 (*Hanahan, 1983*). This nutrient-rich bacterial growth medium is an adjusted version of the commonly used LB media.

For a final culture of 1 litre, the SOC medium contains 2% w/v tryptone (20 g), 0.5% w/v Yeast extract (5 g), 10mM NaCl (0.584 g), 2.5mM KCl (0.186 g), 10mM MgCl₂ (anhydrous 0.952 g) or 10mM MgSO₄ (heptahydrate 2.408 g) and 20mM glucose (3.603 g).

pH adjustment

For maximum effectiveness, SOB/SOC media should have its pH adjusted to 7.0 by adding concentrated sodium hydroxide. The original literature states that the pH of the final medium should be between 6.8 and 7.0.

Sterilisation

Every media are either autoclaved at 121 °C or filtered through a 0.22 µm filter to ensure sterility.

A.2 Composition of buffers

Buffers used for protein purification and storage

The composition of every buffers used to purify and characterise the different proteins is presented in Tables A.1 and A.2.

| | His-EBA181₉₄₅₋₁₀₉₇ | EBA181₉₄₅₋₁₀₉₇ | GST-4.1R_{10kDa} | 4.1R_{10kDa} |
|---|--|--|--|---|
| Lysis and loading | 20 mM Tris, pH 8.0 +250 mM NaCl +5 mM βME +10 mM Imidazole +Complete antiprotease EDTA free (ROCHE®) | 20 mM Tris, pH 8.0 +250 mM NaCl +5 mM βME +10 mM Imidazole +Complete antiprotease EDTA free (ROCHE®) | Phosphate buffered saline, pH 7.4 +5 mM βME +Complete antiprotease (ROCHE®) | Phosphate buffered saline, pH 7.4 +5 mM βME +Complete antiprotease (ROCHE®) |
| Washing | 20 mM Tris, pH 8.0 +1M NaCl +5 mM βME +15k mM Imidazole | 20 mM Tris, pH 8.0 +1M NaCl +5 mM βME +15k mM Imidazole | Phosphate buffered saline, pH 7.4 +1M NaCl +5 mM βME | Phosphate buffered saline, pH 7.4 +1M NaCl +5 mM βME |
| Elution | 20 mM Tris, pH 8.0 +100 mM NaCl +5 mM βME +150 mM Imidazole | 20 mM Tris, pH 8.0 +100 mM NaCl +5 mM βME +150 mM Imidazole | 20 mM Tris, pH 8.0 +250 mM NaCl +5 mM βME +10 mM GSH | 20 mM Tris, pH 8.0 +250 mM NaCl +5 mM βME +10 mM GSH |
| Dialysis and size exclusion chromatography | 50 mM NaP, pH 6.0 +50 mM NaCl +5 mM βME | 50 mM NaP, pH 6.0 +50 mM NaCl +5 mM βME | Phosphate buffered saline, pH 7.4 +5 mM βME | 50 mM NaP, pH 6.0 +50 mM NaCl +5 mM βME |

Table A.1: buffer composition for purification steps of tagged His-EBA181₉₄₅₋₁₀₉₇ and GST-4.1R_{10kDa} or untagged EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins.

| Salts | Concentration (mM) | Concentration (g/l) |
|---|-------------------------------|--------------------------------|
| NaCl | 137 | 8.01 |
| KCl | 2.7 | 0.20 |
| Na ₂ HPO ₄ (2 H ₂ O) | 10 | 1.78 |
| KH ₂ PO ₄ | 2.0 | 0.27 |
| <i>pH 7.4</i> | | |

Table A.2: Composition of phosphate buffered saline (PBS).

Native polyacrylamide gel electrophoresis

pH buffers:

4x acetic acid gel buffer (200 mM acetic acid, pH 3.7 to 5.6)

4x phosphate gel buffer (400 mM sodium phosphate, pH 5.8 to 8.0)

4x tris gel buffer (200 mM Tris-Cl, pH 7.1 to 8.9)

4x glycine gel buffer (200 mM glycine, pH 8.6 to 10.6)

pH range 3.7 to 5.6

| Stock solution | Final acrylamide concentration in gel |
|-------------------------------------|---------------------------------------|
| | 12.5 % |
| 30% acrylamide / 0.8% bisacrylamide | 16.8 ml |
| 300 mM sodium sulphite | 0.4 ml |
| 4x acetic acid gel buffer | 10 ml |
| H ₂ O | 12.48 ml |
| 10% (w/v) ammonium persulphate | 0.3 ml |
| TEMED | 0.02 ml |

pH range 5.8 to 8.0

| Stock solution | Final acrylamide concentration in gel |
|-------------------------------------|---------------------------------------|
| | 12.5 % |
| 30% acrylamide / 0.8% bisacrylamide | 16.8 ml |
| 4x phosphate gel buffer | 10 ml |
| H ₂ O | 12.98 ml |
| 10% (w/v) ammonium persulphate | 0.2 ml |
| TEMED | 0.02 ml |

pH range 7.1 to 8.9

| Stock solution | Final acrylamide concentration in gel |
|-------------------------------------|---------------------------------------|
| | 12.5 % |
| 30% acrylamide / 0.8% bisacrylamide | 16.8 ml |
| 4x Tris gel buffer | 10 ml |
| H ₂ O | 12.98 ml |
| 10% (w/v) ammonium persulphate | 0.2 ml |
| TEMED | 0.02 ml |

pH range 8.6 to 10.6

| Stock solution | Final acrylamide concentration in gel |
|-------------------------------------|---------------------------------------|
| | 12.5 % |
| 30% acrylamide / 0.8% bisacrylamide | 16.8 ml |
| 4x glycine gel buffer | 10 ml |
| H ₂ O | 12.98 ml |
| 10% (w/v) ammonium persulphate | 0.2 ml |
| TEMED | 0.02 ml |

Table A.3: Buffer conditions for the preparation of native polyacrylamide gels.

Samples:

The protein samples to be analysed are solubilised using 5% (w/v) sucrose (using 2x stock solution at 10% (w/v) sucrose in the chosen gel buffer).

Separation:

The current is set to 15 mA and the run is performed at 8°C.

If the protein is negatively charged under the separation conditions, then the standard SDS-PAGE electrode polarity should be used (proteins will migrate to the anode as positive electrode). If the protein is positively charged, then the electrodes should be reversed at the power supply so the positively charged protein migrates to the negative cathode.

APPENDIX B: EBA-181 AND 4.1R PROTEINS

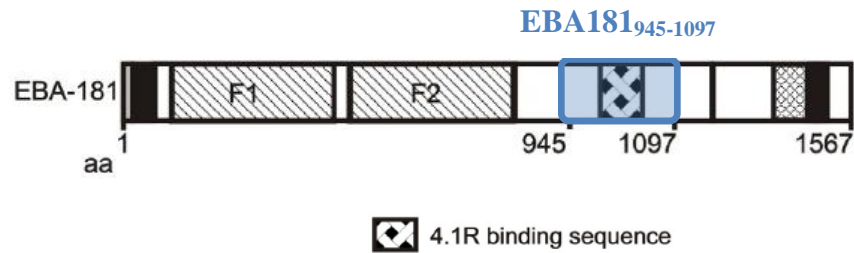
B.1 The cloning strategy

Protein EBA181₉₄₅₋₁₀₉₇

The chosen DNA sequence for EBA-181 contains the coding sequence for the amino acids from 945 to 1097 ('EBA181₉₄₅₋₁₀₉₇', Figure B.1 A, blue rectangle) which contains the 4.1R binding sequence as demonstrated by our collaborators (*Lanzilotti et al., 2005*). The gene corresponding to this sequence has been cloned into a pET15b vector which adds an N-terminal polyhistidine (6x histidines) tag in order to facilitate the purification of the protein.

The pET15b vector (Figure B.1 B) is a system for cloning and expressing recombinant proteins in *E. coli*. Protein expression is driven by a T7 promoter (T7 prom), which contains a downstream *lac* operator (*lac o*) sequence. The plasmid houses a multiple cloning site (MCS) where the gene for EBA181₉₄₅₋₁₀₉₇ was cloned with an N-terminal six-histidine tag which can be cleaved by the thrombin protease. The pET15b contains a T7 terminator sequence (T7 term), a *lacI* repressor sequence and an ampicillin resistance gene (Amp). The origin of replication (*ori*) is also indicated in the vector map.

A)



B)

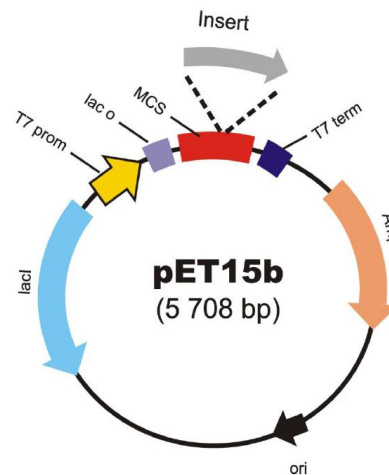
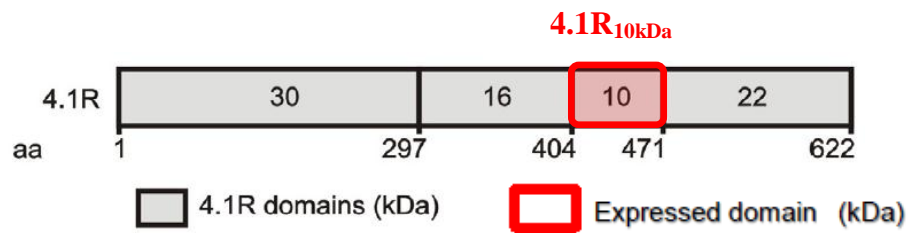


Figure B.1: Schematic of **A)** the EBA-181 protein with its EBA181₉₄₅₋₁₀₉₇ domain and **B)** the map of the pET15b cloning vector adding the N-terminus histidine tag.

Protein 4.1R_{10kDa}

The DNA sequence coding for the 10 kDa domain of 4.1R (amino acids from 405 to 471, Figure B.2 A, red rectangle) was cloned into a pGEX vector which add the glutathion-S-transferase (GST) protein in fusion with the cloned protein. The GST has two interests: firstly to enhance the solubility of the fused protein (here the 10 kDa domain of 4.1R) and secondly to facilitate the purification.

A)



B)

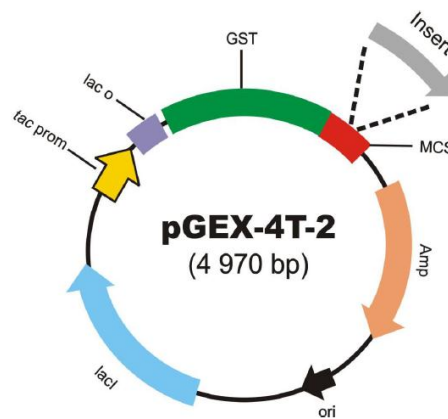


Figure B.2: Schematic of **A)** the 4.1R protein with its 4.1R_{10kDa} domain and **B)** the map of the pGEX-4T-2 cloning vector adding the N-terminus GST tag.

The pGEX-4T-2 vector (Figure B.2 B) offers a *tac* promoter (*tac* prom) for chemically inducible, high level expression of recombinant proteins. A *lac* operator sequence (*lac* o) is situated downstream of the promoter followed by a region encoding the GST, before the multiple cloning site (MCS) where the gene for the 10 kDa domain of 4.1R was cloned. A thrombin site is located immediately after the GST region allowing the cleavage of the tag. The β -lactamase gene region (Amp), the *lacI* repressor region and the origin of plasmid replication (*ori*) are also indicated in Figure B.2 B.

As part of this thesis work, the antibiotic resistance of the expression vectors for both proteins was modified from ampicillin to kanamycin by the *in-vitro* transposition method. This method is based on a transposon containing the gene for kanamycin resistance and which is inserted randomly into the vector sequence. The clones into which the transposon is inserted in the ampicillin resistance sequence are selected and screened using different ampicillin or kanamycin agar plates. The commercial *E. coli* BL21 Rosetta™ 2 (DE3) bacteria from Novagen® were finally used in order to express the two proteins (Figure B.3 and Table B.1).

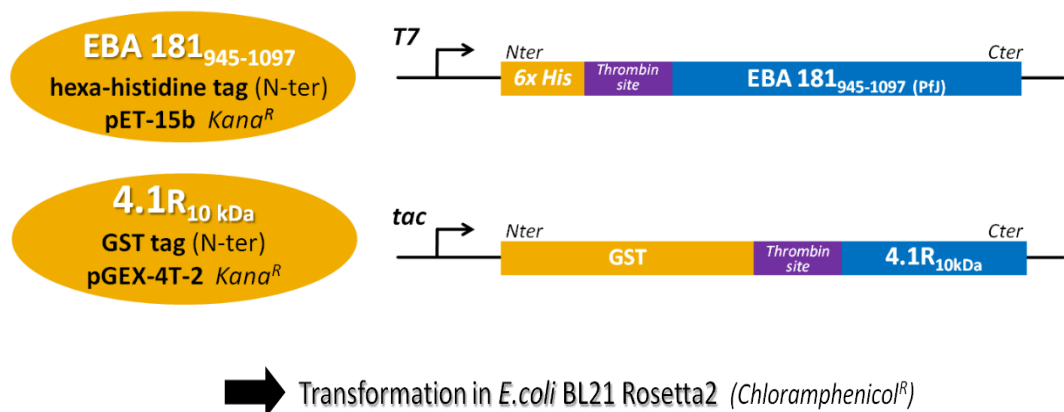
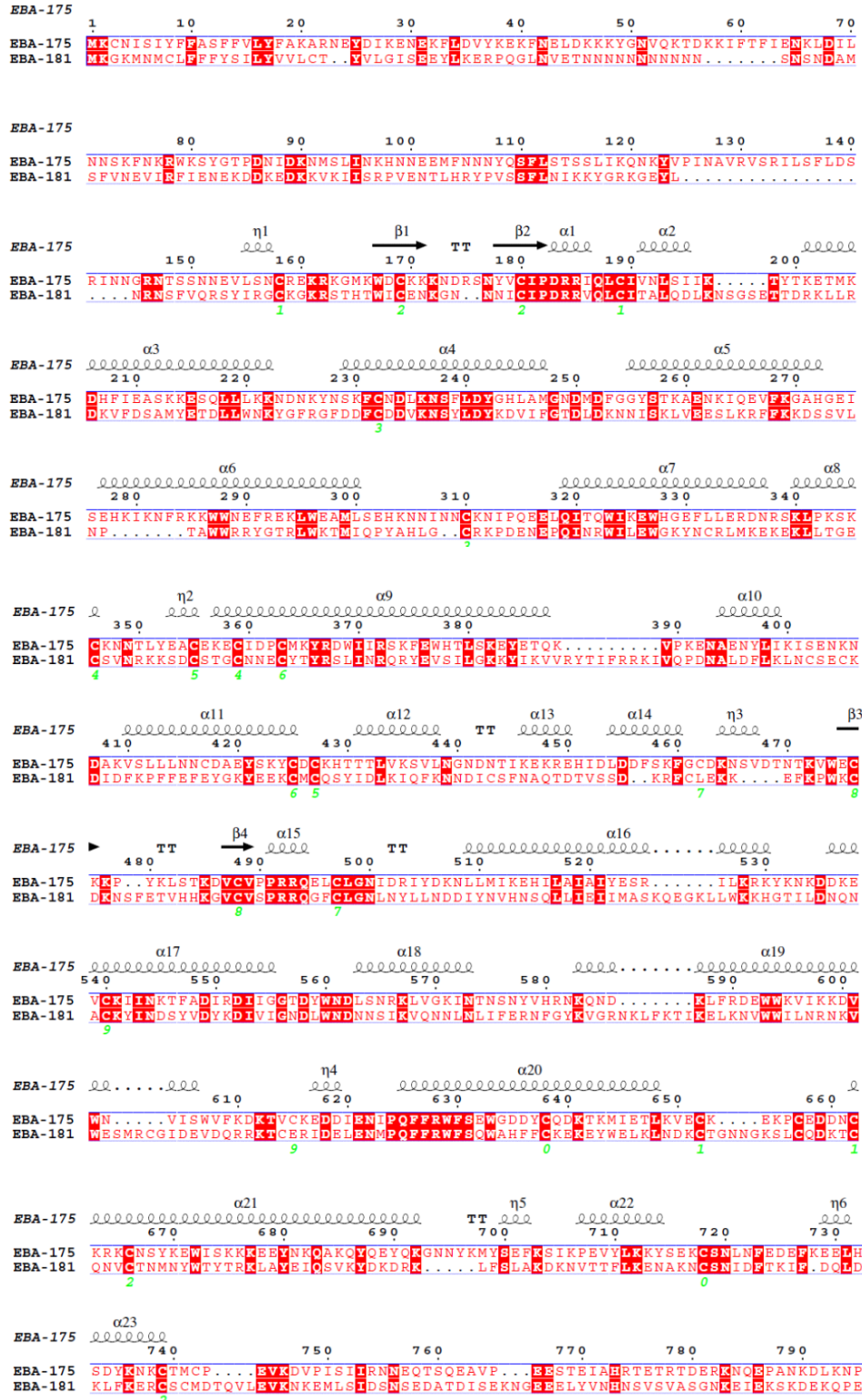


Figure B.3: Summary of the cloning strategies for EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa}.

| | + tag | | - tag | |
|--|--------------|-----------|--------------|-----------|
| | <i>MW</i> | <i>pI</i> | <i>MW</i> | <i>pI</i> |
| EBA181 ₉₄₅₋₁₀₉₇ | 19 kDa | 4.2 | 17 kDa | 4.2 |
| 4.1R _{(10kDa) 405-471} | 38 kDa | 6.6 | 8 kDa | 8.3 |

Table B.1: Summary of the molecular weights and isoelectric points of the EBA181₉₄₅₋₁₀₉₇ and 4.1R_{10kDa} proteins with and without their respective tags.

B.2 EBA-181 / EBA-175 sequence alignment



25% identical (red boxes) - 50% similar

The EBA181₉₄₅₋₁₀₉₇ domain is highlighted in orange

▲ : Interaction sequence with 4.1R_{10kDa}

Input parameters:

Alignment Method: slow

Gap open: 10

KTUP: 1

Gap extend: 0.1

Score: percent

DNA Weight Matrix: iub

Gap extend: 0.2

Sequence type: protein

DNA Weight Matrix: iub

Gap distance: 5

Iteration type: none

Pairgap: 3

Protein Weight Matrix: gonnet

End gap: false

Top Diagonals: 5

Number Iterations: 1

Output order: aligned

Clustering: NJ

Protein Weight Matrix: gonnet

Gap open: 10

Window: 5

Program: clustalw2

Alignment format: aln1

B.3 Protein EBA181₉₄₅₋₁₀₉₇ sequence and parameters

10 20 30 40 50 60
GSCMHMPEVV PQETTSENGS SQDTKISSTE PNENSVVDRA TDSMNLDPK VHNENMSDPN
70 80 90 100 110 120
TNTEPDASLK DDKKEVDDAK KELQSTVSRI ESNEQDVQST PPEDTPTVEG KVGDKAEMLT
130 140 150
SPHATDNSES ESGLNPTDDI KTTDGVVKEQ EILGGGESA

Number of amino acids: 159

Molecular weight: 17075.1 Da

Theoretical pI: 4.15

Amino acid composition:

| | | |
|---------|----|-------|
| Ala (A) | 6 | 3.8% |
| Arg (R) | 2 | 1.3% |
| Asn (N) | 11 | 6.9% |
| Asp (D) | 17 | 10.7% |
| Cys (C) | 1 | 0.6% |
| Gln (Q) | 6 | 3.8% |
| Glu (E) | 20 | 12.6% |
| Gly (G) | 9 | 5.7% |
| His (H) | 3 | 1.9% |
| Ile (I) | 4 | 2.5% |
| Leu (L) | 6 | 3.8% |
| Lys (K) | 11 | 6.9% |
| Met (M) | 5 | 3.1% |
| Phe (F) | 0 | 0.0% |
| Pro (P) | 11 | 6.9% |
| Ser (S) | 19 | 11.9% |
| Thr (T) | 16 | 10.1% |
| Trp (W) | 0 | 0.0% |
| Tyr (Y) | 0 | 0.0% |
| Val (V) | 12 | 7.5% |
| Pyl (O) | 0 | 0.0% |
| Sec (U) | 0 | 0.0% |

Total number of negatively charged residues (Asp + Glu): 37

Total number of positively charged residues (Arg + Lys): 13

Atomic composition:

| | | |
|----------|---|------|
| Carbon | C | 698 |
| Hydrogen | H | 1127 |
| Nitrogen | N | 199 |
| Oxygen | O | 286 |
| Sulfur | S | 6 |

Formula: $C_{698}H_{1127}N_{199}O_{286}S_6$
Total number of atoms: 2316

Extinction coefficients:

This protein does not contain any Trp residues.

B.4 Protein 4.1R_{10kDa} sequence and parameters

 10 20 30 40 50 60
WKKKRERLDG ENIYIRHSNL MLEDLKDSQE EIKKHHSIS ELKKNFMESV PEPRPSEWDK

RLSTHSP

Number of amino acids: 67

Molecular weight: 8084.1 Da

Theoretical pI: 8.25

Amino acid composition:

| | | |
|---------|---|-------|
| Ala (A) | 1 | 1.5% |
| Arg (R) | 5 | 7.5% |
| Asn (N) | 3 | 4.5% |
| Asp (D) | 4 | 6.0% |
| Cys (C) | 0 | 0.0% |
| Gln (Q) | 1 | 1.5% |
| Glu (E) | 9 | 13.4% |
| Gly (G) | 1 | 1.5% |
| His (H) | 4 | 6.0% |
| Ile (I) | 4 | 6.0% |
| Leu (L) | 6 | 9.0% |
| Lys (K) | 9 | 13.4% |

| | | |
|---------|---|-------|
| Met (M) | 2 | 3.0% |
| Phe (F) | 1 | 1.5% |
| Pro (P) | 4 | 6.0% |
| Ser (S) | 8 | 11.9% |
| Thr (T) | 1 | 1.5% |
| Trp (W) | 2 | 3.0% |
| Tyr (Y) | 1 | 1.5% |
| Val (V) | 1 | 1.5% |
| Pyl (O) | 0 | 0.0% |
| Sec (U) | 0 | 0.0% |

Total number of negatively charged residues (Asp + Glu): 13

Total number of positively charged residues (Arg + Lys): 14

Atomic composition:

| | | |
|----------|---|-----|
| Carbon | C | 354 |
| Hydrogen | H | 565 |
| Nitrogen | N | 105 |
| Oxygen | O | 108 |
| Sulfur | S | 2 |

Formula: $C_{354}H_{565}N_{105}O_{108}S_2$

Total number of atoms: 1134

Extinction coefficients:

Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water.

Ext. coefficient 12490

Abs 0.1% (=1 g/l) 1.545

APPENDIX C: NMR supplemental information

C.1 Assigned chemical shifts of EBA181₉₄₅₋₁₀₉₇

| Residue | | HN | N | C α | C β | C' |
|---------|---|-------|---------|------------|-----------|---------|
| 945 | P | - | - | 63.154 | 31.964 | 176.632 |
| 946 | E | 8.503 | 121.446 | 56.572 | 30.335 | 176.374 |
| 947 | V | 8.204 | 122.210 | 62.182 | 32.539 | 175.882 |
| 948 | V | 8.295 | 126.544 | 59.752 | 32.428 | 174.403 |
| 949 | P | - | - | 63.177 | 32.059 | 176.838 |
| 950 | Q | 8.466 | 121.035 | 55.791 | 29.730 | 176.042 |
| 951 | E | 8.535 | 122.547 | 56.585 | 30.296 | 176.638 |
| 952 | T | 8.327 | 115.803 | 61.851 | 69.678 | 174.705 |
| 953 | T | 8.244 | 116.419 | 61.655 | 69.782 | 174.575 |
| 954 | S | 8.381 | 117.990 | 58.292 | 63.346 | 174.689 |
| 955 | E | 8.521 | 122.952 | 56.837 | 30.003 | 176.426 |
| 956 | N | 8.444 | 119.387 | 53.375 | 38.783 | 175.874 |
| 957 | G | 8.374 | 109.553 | 45.599 | - | 174.367 |
| 958 | S | 8.236 | 115.534 | 58.246 | 63.411 | 174.878 |
| 959 | S | 8.421 | 117.788 | 58.507 | 63.135 | 174.751 |
| 960 | Q | 8.416 | 121.878 | 56.017 | 29.487 | 175.752 |
| 961 | D | 8.314 | 121.381 | 54.480 | 40.960 | 176.521 |
| 962 | T | 8.079 | 114.942 | 62.152 | 69.468 | 174.600 |
| 963 | K | 8.307 | 124.066 | 56.362 | 32.645 | 176.527 |
| 964 | I | 8.155 | 122.327 | 61.101 | 38.615 | 176.367 |
| 965 | S | 8.414 | 120.087 | 58.097 | 63.289 | 174.520 |
| 966 | S | 8.399 | 118.159 | 58.196 | 63.258 | 174.695 |
| 967 | T | 8.206 | 115.531 | 61.743 | 69.557 | 174.397 |
| 968 | E | 8.383 | 124.669 | 54.423 | 29.600 | 174.637 |
| 969 | P | - | - | 63.383 | 31.946 | 176.730 |
| 970 | N | 8.534 | 118.916 | 53.248 | 38.942 | 175.450 |
| 971 | E | 8.490 | 121.613 | 56.983 | 29.964 | 176.294 |
| 972 | N | 8.459 | 119.251 | 53.389 | 38.849 | 175.226 |

| | | | | | | |
|------|---|-------|---------|--------|--------|---------|
| 973 | S | 8.195 | 116.187 | 58.484 | 63.200 | 174.457 |
| 974 | V | 8.122 | 122.084 | 62.535 | 32.413 | 176.258 |
| 975 | V | 8.110 | 123.684 | 62.323 | 32.717 | 175.777 |
| 976 | D | 8.359 | 124.237 | 54.327 | 41.106 | 176.059 |
| 977 | R | 8.238 | 121.999 | 56.057 | 30.685 | 176.226 |
| 978 | A | 8.395 | 125.123 | 52.888 | 18.982 | 178.353 |
| 979 | T | 8.084 | 112.761 | 62.166 | 69.527 | 174.666 |
| 980 | D | 8.272 | 122.493 | 54.616 | 40.931 | 176.589 |
| 981 | S | 8.206 | 115.780 | 58.785 | 63.028 | 174.785 |
| 982 | M | 8.280 | 121.396 | 55.795 | 32.641 | 175.898 |
| 983 | N | 8.292 | 119.260 | 53.295 | 38.694 | 174.777 |
| 984 | L | 8.130 | 122.503 | 54.899 | 42.436 | 176.740 |
| 985 | D | 8.290 | 122.554 | 52.188 | 40.996 | 174.880 |
| 986 | P | - | - | 63.872 | 32.017 | 177.613 |
| 987 | E | 8.438 | 118.887 | 56.978 | 29.756 | 176.921 |
| 988 | K | 7.960 | 120.999 | 56.332 | 32.790 | 176.564 |
| 989 | V | 7.878 | 120.385 | 62.631 | 32.482 | 176.020 |
| 990 | H | 8.526 | 122.407 | 55.397 | 29.339 | 174.277 |
| 991 | N | 8.419 | 121.211 | 53.284 | 38.966 | 175.246 |
| 992 | E | 8.660 | 121.720 | 57.037 | 29.973 | 176.235 |
| 993 | N | 8.452 | 118.974 | 53.385 | 38.532 | 175.223 |
| 994 | M | 8.255 | 120.840 | 55.563 | 32.736 | 176.167 |
| 995 | S | 8.242 | 116.736 | 58.216 | 63.313 | 173.758 |
| 996 | D | 8.310 | 123.666 | 52.167 | 41.225 | 174.941 |
| 997 | P | - | - | 63.676 | 32.007 | 177.076 |
| 998 | N | 8.559 | 117.924 | 53.605 | 38.662 | 175.571 |
| 999 | T | 7.933 | 113.878 | 62.126 | 69.509 | 174.275 |
| 1000 | N | 8.409 | 121.492 | 53.359 | 38.852 | 175.232 |
| 1001 | T | 8.142 | 114.668 | 61.734 | 69.641 | 174.403 |
| 1002 | E | 8.326 | 124.653 | 54.401 | 29.587 | 174.640 |
| 1003 | P | - | - | 63.309 | 32.023 | 176.698 |
| 1004 | D | 8.401 | 120.380 | 54.314 | 40.993 | 176.548 |
| 1005 | A | 8.360 | 125.939 | 53.056 | 19.003 | 178.206 |
| 1006 | S | 8.372 | 114.592 | 59.040 | 63.062 | 174.771 |
| 1007 | L | 7.972 | 123.360 | 55.281 | 42.027 | 177.389 |

| | | | | | | |
|------|---|-------|---------|--------|----------|---------|
| 1008 | K | 8.097 | 121.727 | 56.298 | 32.932 | 176.212 |
| 1009 | D | 8.279 | 121.673 | 54.335 | 41.183 | 175.916 |
| 1010 | D | 8.281 | 121.511 | 54.435 | 40.978 | 176.337 |
| 1011 | K | 8.207 | 121.222 | 56.439 | 32.390 | 176.644 |
| 1012 | K | 8.198 | 122.222 | 56.446 | 32.889 | 176.674 |
| 1013 | E | 8.418 | 122.051 | 56.691 | 30.019 | 176.716 |
| 1014 | V | 8.147 | 120.795 | 62.456 | 32.745 | 176.030 |
| 1015 | D | 8.373 | 123.975 | 54.493 | 41.084 | 176.453 |
| 1016 | D | 8.348 | 121.952 | 54.932 | 40.931 1 | 76.686 |
| 1017 | A | 8.184 | 123.553 | 53.262 | 18.706 | 178.471 |
| 1018 | K | 8.089 | 119.189 | 56.689 | 32.314 | 177.102 |
| 1019 | K | 8.056 | 121.867 | 56.884 | 32.785 | 176.972 |
| 1020 | E | 8.344 | 121.284 | 56.730 | 29.977 | 176.728 |
| 1021 | L | 8.187 | 122.911 | 55.473 | 42.208 | 177.630 |
| 1022 | Q | 8.320 | 120.757 | 55.993 | 29.297 | 176.172 |
| 1023 | S | 8.318 | 116.743 | 58.345 | 63.214 | 174.898 |
| 1024 | T | 8.203 | 116.226 | 62.128 | 69.566 | 174.629 |
| 1025 | V | 8.073 | 122.233 | 62.422 | 32.612 | 176.159 |
| 1026 | S | 8.367 | 119.769 | 58.234 | 63.209 | 174.272 |
| 1027 | R | 8.343 | 123.691 | 56.031 | 30.872 | 176.019 |
| 1028 | I | 8.224 | 122.819 | 61.194 | 38.528 | 176.374 |
| 1029 | E | 8.544 | 125.391 | 56.516 | 30.196 | 176.370 |
| 1030 | S | 8.334 | 116.824 | 58.205 | 63.302 | 174.311 |
| 1031 | N | 8.528 | 121.119 | 53.304 | 38.897 | 175.395 |
| 1032 | E | 8.415 | 121.211 | 57.018 | 30.008 | 176.523 |
| 1033 | Q | 8.329 | 120.220 | 55.989 | 29.514 | 175.744 |
| 1034 | D | 8.329 | 121.634 | 54.416 | 40.953 | 176.259 |
| 1035 | V | 8.044 | 120.348 | 62.392 | 32.541 | 176.294 |
| 1036 | Q | 8.463 | 123.928 | 55.736 | 29.408 | 176.001 |
| 1037 | S | 8.357 | 117.619 | 58.192 | 63.275 | 174.241 |
| 1038 | T | 8.197 | 118.597 | 59.840 | 69.611 | 172.375 |
| 1039 | P | - | - | - | - | - |
| 1040 | P | - | - | 63.074 | 31.927 | 177.087 |
| 1041 | E | 8.568 | 120.832 | 56.592 | 30.283 | 176.235 |
| 1042 | D | 8.360 | 121.634 | 54.208 | 41.040 | 175.946 |

| | | | | | | |
|------|---|-------|---------|--------|--------|---------|
| 1043 | T | 8.088 | 117.181 | 59.960 | 69.690 | 172.774 |
| 1044 | P | - | - | 63.261 | 32.122 | 177.089 |
| 1045 | T | 8.324 | 115.802 | 61.975 | 69.678 | 174.619 |
| 1046 | V | 8.223 | 122.914 | 62.224 | 32.722 | 176.053 |
| 1047 | E | 8.512 | 124.936 | 56.764 | 30.237 | 176.881 |
| 1048 | G | 8.418 | 110.321 | 45.302 | - | 173.866 |
| 1049 | K | 8.168 | 121.203 | 55.977 | 33.004 | 176.690 |
| 1050 | V | 8.249 | 122.038 | 62.607 | 32.597 | 176.760 |
| 1051 | G | 8.499 | 112.693 | 45.247 | - | 173.912 |
| 1052 | D | 8.209 | 120.809 | 54.483 | 41.227 | 176.512 |
| 1053 | K | 8.283 | 121.577 | 56.375 | 32.696 | 176.500 |
| 1054 | A | 8.281 | 124.838 | 52.810 | 18.909 | 177.975 |
| 1055 | E | 8.352 | 119.848 | 56.710 | 30.146 | 176.542 |
| 1056 | M | 8.291 | 121.130 | 55.575 | 32.758 | 176.160 |
| 1057 | L | 8.254 | 123.321 | 55.340 | 42.262 | 177.511 |
| 1058 | T | 8.054 | 114.315 | 61.577 | 69.760 | 174.285 |
| 1059 | S | 8.259 | 119.181 | 56.301 | 62.811 | 172.737 |
| 1060 | P | - | - | 63.388 | 32.017 | 176.738 |
| 1061 | H | 8.436 | 118.559 | 55.199 | 29.365 | 174.382 |
| 1062 | A | 8.354 | 125.786 | 52.777 | 19.003 | 177.946 |
| 1063 | T | 8.211 | 112.927 | 61.679 | 69.661 | 174.316 |
| 1064 | D | 8.301 | 122.560 | 54.301 | 41.084 | 175.996 |
| 1065 | N | 8.454 | 119.661 | 53.356 | 38.827 | 175.539 |
| 1066 | S | 8.390 | 116.418 | 59.154 | 63.105 | 174.989 |
| 1067 | E | 8.462 | 122.457 | 56.881 | 29.926 | 176.819 |
| 1068 | S | 8.227 | 116.115 | 58.571 | 63.296 | 174.870 |
| 1069 | E | 8.413 | 122.856 | 56.807 | 29.953 | 176.794 |
| 1070 | S | 8.325 | 116.503 | 58.747 | 63.253 | 175.221 |
| 1071 | G | 8.373 | 110.745 | 45.410 | - | 173.971 |
| 1072 | L | 7.986 | 121.055 | 55.015 | 42.467 | 176.931 |
| 1073 | N | 8.522 | 120.629 | 51.331 | 38.752 | 173.346 |
| 1074 | P | - | - | 63.620 | 32.058 | 177.317 |
| 1075 | T | 8.166 | 113.079 | 61.881 | 69.560 | 174.500 |
| 1076 | D | 8.181 | 122.226 | 54.512 | 41.132 | 175.839 |
| 1077 | D | 8.237 | 120.671 | 54.309 | 41.013 | 176.081 |

| | | | | | | |
|------|---|-------|---------|--------|--------|---------|
| 1078 | I | 8.029 | 121.269 | 61.307 | 38.328 | 176.365 |
| 1079 | K | 8.403 | 125.480 | 56.076 | 32.887 | 176.905 |
| 1080 | T | 8.231 | 115.460 | 61.724 | 69.738 | 174.897 |
| 1081 | T | 8.196 | 115.488 | 61.758 | 69.568 | 174.420 |
| 1082 | D | 8.323 | 122.718 | 54.603 | 41.040 | 176.674 |
| 1083 | G | 8.301 | 109.132 | 45.477 | - | 174.030 |
| 1084 | V | 7.902 | 120.149 | 62.512 | 32.561 | 176.278 |
| 1085 | V | 8.289 | 125.438 | 62.469 | 32.451 | 176.097 |
| 1086 | K | 8.424 | 126.342 | 56.183 | 32.960 | 176.499 |
| 1087 | E | 8.490 | 122.736 | 56.861 | 30.064 | 176.448 |
| 1088 | Q | 8.405 | 121.040 | 55.980 | 29.603 | 175.893 |
| 1089 | E | 8.419 | 122.861 | 56.653 | 30.216 | 176.285 |
| 1090 | I | 8.238 | 122.760 | 61.073 | 38.411 | 176.329 |
| 1091 | L | 8.389 | 127.040 | 55.207 | 42.229 | 177.936 |
| 1092 | G | 8.455 | 110.223 | 45.393 | - | 174.777 |
| 1093 | G | 8.354 | 108.876 | 45.356 | - | 174.831 |
| 1094 | G | 8.348 | 108.894 | 45.210 | - | 174.259 |
| 1095 | E | 8.414 | 120.578 | 56.640 | 30.223 | 176.607 |
| 1096 | S | 8.361 | 117.093 | 58.120 | 63.477 | 173.263 |
| 1097 | A | 7.986 | 131.633 | 53.864 | 19.911 | 176.082 |

Table C.1: Assigned chemical shifts (ppm) of EBA181₉₄₅₋₁₀₉₇ obtained at 298K.

REFERENCES

- Baruchel J., Hodeau J.L., Lehmann M.S., Regnard J.R., Schlenker C. – Volume 1 Theory, instruments and methods. Neutron and synchrotron radiation for condensed matter studies. EDP Sciences SPRINGER VERLAG, 1993.
- Baum J., Thomas A.W. and Conway D.J. – Evidence for diversifying selection on erythrocyte-binding antigens of *Plasmodium falciparum* and *P. vivax*. *Genetics* 163:1327–1336, 2003.
- Baum J., Tonkin C.J., Paul A.S., Rug M., Smith B.J., Gould S.B., Richard D., Pollard T.D., Cowman A.F. – A malaria parasite formin regulates actin polymerization and localizes to the parasite–erythrocyte moving junction during invasion. *Cell. Host Microbe*. 3, 188–198, 2008.
- Bei A.K., Membi C.D., Rayner J.C., Mubi M., Ngasala B., Sultan A.A., Premji Z., Duraisingh M.T. – Variant merozoite protein expression is associated with erythrocyte invasion phenotypes in *Plasmodium falciparum* isolates from Tanzania. *Mol. Biochem. Parasitol.* 153, 66–71, 2007.
- Bernado et al. – Mass retrieval from Guinier analysis. *Biophys J*, 97 (10), 2839-2845, 2009.
- Bertani G. – Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* 62:293-300, 1951.
- Blackman M.J. – Proteases involved in erythrocyte invasion by the malaria parasite: function and potential as chemotherapeutic targets. *Curr Drug Targets*, 1(1):59-83, 2000.
- Blanc M., Coetzer T.L., Blackledge M., Haertlein M., Mitchell E.P., Forsyth V.T. and Jensen M.R. – Intrinsic disorder within the erythrocyte binding-like proteins from *Plasmodium falciparum*. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, Volume 1844, Issue 12, 2059-2366, 2014.

- Blanc M., Coetzer T.L., Haertlein M., Mitchell E.P. and Forsyth V.T. – *Plasmodium falciparum* and erythrocyte infection: a SAXS and SANS study of the interaction of *P. falciparum* EBA-181 with human 4.1R.. *Malaria Journal*, 2015.
- Bodenhausen G. and Ruben D.J. – Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* 69 (1): 185-189, 1980.
- Bradford M.M. – A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248–54, 1976.
- Brahms S., and Brahms J. – Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.* 138:149-178, 1980.
- Braun Breton C, Pereira da Silva L.H. – Malaria proteases and red blood cell invasion. *Parasitol Today*, 9(3):92-96, 1993.
- Brown R. – A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine* 4, 161-173, 1828.
- Bruce-Chwatt, L.J. – *Essential Malariology*, Oxford Univ. Press, New York, 1993.
- Bustamante L. Y. et al. – A full-length recombinant *Plasmodium falciparum* PfRH5 protein induces inhibitory antibodies that are effective across common PfRH5 genetic variants. *Vaccine* 31, 373–379, 2013.
- Chayen N.E. – *Comparative Studies of Protein Crystallization by Vapour-Diffusion and Microbatch Techniques*. *Acta Cryst. D* 54, 8-15, 1998.
- Chirenda, J., Siziya, S., and Tshimanga, M. – Association of HIV infection with the development of severe and complicated malaria cases at a rural hospital in Zimbabwe. *Cent. Afr. J. Med.* 46:5–9, 2000.
- Chitnis C.E., Blackman M.J. – Host cell invasion by malaria parasites, *Parasitol. Today* 16, 411-415, 2000.
- Chitnis C.E. – Molecular insights into receptors used by malaria parasites for erythrocyte invasion, *Curr. Opin. Hematol.* 8, 85–91, 2001.

Cibert C, Pruliere G, Lacombe C, Deprette C, Cassoly R – Calculation of a Gap restoration in the membrane skeleton of the red blood cell: possible role for myosin II in local repair. *Biophys J.*, 76(3):1153-1165, 1999.

Combet C., Blanchet C., Geourjon C. and Deléage G. – Sequence comparison: NPS@: Network Protein Sequence Analysis. *TIBS March* Vol. 25, No 3 [291]:147-150, 2000.

Conboy J.G.: Structure, function, and molecular genetics of erythroid membrane skeletal protein 4.1 in normal and abnormal red blood cells. *Semin Hematol*, 30(1):58-73, 1993.

Conboy J., Kan Y.W., Shohet S.B., and Mohandas N. – Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Proc. Nati. Acad. Sci. USA, Cell Biology*, Vol. 83,. 9512-9516, 1986.

Cooke BM, Mohandas N, Coppel RL – Malaria and the red blood cell membrane. *Semin Hematol* 41(2), 173-188, 2004.

Cowman A.F., Crabb B.S. – Invasion of red blood cells by malaria parasites. *Cell* 124, 755–766, 2006.

Crosnier C. et al. – Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature* 480, 534–537, 2011.

Delaglio F., Grzesiek S., Vuister G.W., Zhu G., Pfeifer J. and Bax A. – NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6, 277-293, 1995.

DeSimone T.M., Bei A.K., Jennings C.V., and Duraisingh M.T. – Genetic analysis of the cytoplasmic domain of the PfRh2b merozoite invasion protein of *Plasmodium falciparum*. *Int. J. Parasitol.* 39:399–405, 2009.

DeSimone T.M – Cooperativity between *Plasmodium falciparum* adhesive proteins for invasion into erythrocytes. *Mol. Microbiol.* 72:578–589, 2009.

Diakowsk W., Grzybek M. and Sikorski A.F. – Protein 4.1, a component of the erythrocyte membrane skeleton and its related homologue proteins forming the protein 4.1/FERM superfamily. *FOLIA HISTOCHEMICA ET CYTOBIOLOGICA.* Vol. 44, No. 4, 231-248, 2006.

DISOPRED:

<http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1>

Dolan S.A., Miller L.H. and Wellem's T.E. – Evidence for a switching mechanism in the invasion of erythrocytes by *Plasmodium falciparum*. *J. Clin. Invest.* 86, 618–624, 1990.

Dosztányi Z., Csizsók V., Tompa P. and Simon I. – IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21, 3433–3434, 2005.

Douglas A. D. et al. – The blood-stage malaria antigen PfrH5 is susceptible to vaccine-inducible cross-strain neutralizing antibody. *Nature Commun.* 2, 601, 2011 .

Douglas A. D. et al. – Neutralization of *Plasmodium falciparum* merozoites by antibodies against PfrH5. *J. Immunol.* 192, 245–258, 2014.

Dowse T. and Soldati D. – Host cell invasion by the apicomplexans: the significance of microneme protein proteolysis. *Curr Opin Microbiol.* 7(4):388–396, 2004.

Dunker, Lawson, Brown, Williams, Romero, Oh, Oldfield, Campen, Ratliff, Hipps, Ausio, Nissen, Reeves, Kang, Kissinger, Bailey, Griswold, Chiu, Garner, Obradovic – Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, Volume 19, Issue 1, 2001.

Dunker, Oldfield, Meng, Romero, Yang, Chen, Vacic, Obradovic and Uversky – The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*, 2008.

Duraisingh M.T. – Phenotypic variation of *Plasmodium falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes. *EMBO J.* 22:1047–1057, 2003.

Duraisingh M.T., Maier A.G., Triglia T. and Cowman A.F. – Erythrocyte-binding antigen 175 mediates invasion in *Plasmodium falciparum* utilizing sialic acid-dependent and -independent pathways. *Proc. Natl. Acad. Sci. USA* 100, 4796–4801, 2003.

Edman P., Högfeldt E., Gunnar S.L. and Per-Olof K. – Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand.* 4: 283–293. doi:10.3891/acta.chem.scand.04-0283, 1950.

- Ejigiri I. et al. – Shedding of TRAP by a rhomboid protease from the malaria sporozoite surface is essential for gliding motility and sporozoite infectivity. *PLoS Pathog* 8(7):e1002725, 2012.
- Enfors S.-O. and Castan A. – Characteristics of DO-controlled fed-batch processes. *Bioprocess Eng.* 22, 509-515, 2000.
- Erickson H.P. – Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol Proced Online*, 15;11:32-51. 2009.
- Fanchon E., Geissler E., Hodeau J.L., Regnard J.R. and Timmins P.A. – Structure and dynamics of biomolecules. *Neutron and synchrotron radiation for condensed matter studies*, Oxford University Press, 2002.
- Franke, D. and Svergun, D.I. –DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* 42, 342-346, 2009.
- Fowler V.M., Davis J.Q. and Bennett V. – Human erythrocyte myosin: identification and purification. *J Cell Biol*, 100(1):47-55, 1985.
- Fuxreiter M. and Tompa P. – Book “Fuzziness: Structural Disorder in Protein Complexes”, chapter “Fuzzy Complexes: A More Stochastic View of Protein Function”, 2011.
- Gao X. et al. – Antibodies targeting the PfrH1 binding domain inhibit invasion of *Plasmodium falciparum* merozoites. *PLoS Pathog.* 4:e1000104, 2008.
- Garman E.F. and Mitchell E.P. – Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions. *J. Appl. Cryst.* 29, 584-587, 1996.
- Garman E. and Nave C. - Radiation damage to crystalline biological molecules: current view. *J. Synchrotron Rad.* 9, 327-328, 2002.
- Gaur D., Mayer D.C. and Miller L.H. – Parasite ligand-host receptor interactions during invasion of erythrocytes by *Plasmodium merozoites*. *Int J Parasitol*, 34(13-14):1413-1429, 2004.
- Gaur D. – Upregulation of expression of the reticulocyte homology gene 4 in the *Plasmodium falciparum* clone Dd2 is associated with a switch in the erythrocyte invasion pathway. *Mol. Biochem. Parasitol.* 145: 205–215, 2006.

- Gaur D., Singh S., Singh S., Jiang L., Diouf A. and Miller L.H., – Recombinant *Plasmodium falciparum* reticulocyte homology protein 4 binds to erythrocytes and blocks invasion. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17789–17794, 2007.
- Ghorbal M., Gorman, Macpherson, Martins, Scherf and Lopez-Rubio – Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nature Biotechnology*, 2014.
- Gilberger T.W. – A novel erythrocyte binding antigen-175 paralogue from *Plasmodium falciparum* defines a new trypsin-resistant receptor on human erythrocytes. *J. Biol. Chem.* 278:14480–14486, 2003.
- Gilberger T.W., Thompson J.K., Reed M.B., Good R.T. and Cowman A.F. – The cytoplasmic domain of the *Plasmodium falciparum* ligand EBA-175 is essential for invasion but not protein trafficking. *J Cell Biol*, 162(2):317-327, 2003.
- Gilson P.R., Nebl T., Vukcevic D., Moritz R.L., Sargeant T., Speed T.P., Schofield L. and Crabb, B.S. – Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum*. *Mol. Cell. Proteomics* 5, 1286–1299, 2006.
- Gilson P.R. and Crabb B.S. – Morphology and kinetics of the three distinct phases of red blood cell invasion by *Plasmodium falciparum* merozoites. *Int. J. Parasitol.* 39, 91–96, 2009.
- Goel V.K., Li X., Chen H., Liu S.C., Chishti A.H., and Oh S.S – Band 3 is a host receptor binding merozoite surface protein 1 during the *Plasmodium falciparum* invasion of erythrocytes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5164–5169, 2003.
- Goryshin I.Y. and Reznikoff W.S. – Tn5 in vitro transposition. *J Biol Chem*, 1998
- Grimwade K., et al. – Childhood malaria in a region of unstable transmission and high human immunodeficiency virus prevalence. *Pediatr. Infect. Dis. J.* 22:1057–1063, 2003.
- Hao J.S., Perry T.A.J., Akutsu T., Webb G.I. and Whisstock J.C. – Pike R.N. PROSPER: An Integrated Feature-Based Tool for Predicting Protease Substrate Cleavage Sites. *PLoS ONE* 7(11): e50300, 2012.

- Han B.G., Nunomura W., Takakuwa Y., Mohandas N. and Jap B.K. – Protein 4.1R core domain structure and insights into regulation of cytoskeletal organization. *Nat.Struct.Biol* 7: 871-875, 2000.
- Hanahan D. – Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 166 (4): 557–580, 1983.
- Iakoucheva L., Kimzey A.L., Masselon C.D., Smith R.D., Keith Dunker A., and Ackerman E.J. – Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Science* 10:1353–1362, 2001.
- Ibel K. and Stuhrmann H.B. – Comparison of neutron solutions. *J Mol Biol*, 1975.
- IUPRED: <http://iupred.enzim.hu>
- Jacques D.A. and Trewhella J. – Small-angle scattering for structural biology, Expanding the frontier while avoiding the pitfalls. *Protein Sci.* 19(4): 642–657, 2010.
- Jung Y.S. and Zweckstetter M. – robust automatic backbone assignment of proteins. *Biomol. NMR*, 30, 11-23, 2004.
- Kafsack B.F.C. et al. – A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*, 2014.
- Kappe S.H.I., Vaughan A.M., Boddey J.A. and Cowman A.F. – That Was Then But This Is Now: Malaria Research in the Time of an Eradication Agenda. *Science*: 862-866, 2010.
- Kokhanovsky A. – Light Scattering and Radiative Transfer. Series: Springer Praxis Books, Light Scattering Reviews, Vol. 9, 2014.
- Konarev P.V., Volkov V.V., Sokolova A.V., Koch M.H.J. and Svergun D.I. – PRIMUS a Windows-PC based system for small-angle scattering data analysis. *J Appl Cryst.* 36, 1277-1282, 2003.
- Korenromp E.L., Williams B.G., Gouws E., Dye C. and Snow R.W. – Measurement of trends in childhood malaria mortality in Africa: an assessment of progress toward targets based on verbal autopsy. *Lancet Infect. Dis.* 3:349–358, 2003.
- Kragelj J., Ozenne V., Blackledge M. and Jensen M.R. – Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. *Chemphyschem* 16;14(13):3034-45, 2013.

Kratky O., Porod G., Sekora A. and Paletta B. – Determination of the shape of gamma-globulin by means of the x-ray low-angle method. *Journal of Polymer Science, Volume 16, Issue 82, pages 163–175, 1955.*

Laemmli U.K. – Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227 (5259): 680–685, 1970.*

Lanzillotti R., Coetzer T.L. – The 10 kDa domain of human erythrocyte protein 4.1 binds the *Plasmodium falciparum* EBA-181 protein. *Malaria Journal. 5, 100, 2005.*

Lasonder E., Ishihama Y., Andersen J.S., Vermunt A.M., Pain A., Sauerwein R.W., Eling W.M., Hall N., Waters A.P., Stunnenberg H.G. and Mann M. – Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature 419(6906): 537-542, 2002.*

Lauterbach S.B., Lanzillotti R. and Coetzer T.L. – Construction and use of *Plasmodium falciparum* phage display libraries to identify host parasite interactions. *Malar J, 2(1):47, 2003.*

Leto T.L. and Marchesi V.T. – A Structural Model of Human Erythrocyte Protein 4.1. *THE JOURNAL OF BIOLOGICAL CHEMISTRY, 1984.*

Levine N.D. – Progress in taxonomy of the Apicomplexan protozoa. *J. Protozool. 35, 518–520, 1988.*

LOMETS-ITASSER:

<http://zhanglab.ccmb.med.umich.edu>

Maier A.G. et al. – *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nat. Med. 9:87–92, 2003.*

Maier A.G., Baum J., Smith B., Conway D.J. and Cowman A.F. – Polymorphisms in erythrocyte binding antigens 140 and 181 affect function and binding but not receptor specificity in *Plasmodium falciparum*. *Infect. Immun. 77:1689–1699, 2009.*

Marsh K. – Malaria disaster in Africa. *Lancet.352:924–925, 1998.*

Mayer D.C., Kaneko O., Hudson-Taylor D.E., Reid M.E. and Miller L.H. – Characterization of a *Plasmodium falciparum* erythrocyte-binding protein paralogous to EBA-175. *Proc. Natl. Acad. Sci. U. S. A.* 98:5222–5227, 2001.

Mayer D.C. – Polymorphism in the *Plasmodium falciparum* erythrocyte-binding ligand JESEBL/EBA-181 alters its receptor specificity. *Proc. Natl. Acad. Sci. U.S.A.* 101:2518–2523, 2004.

Mayer D.C. – Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1. *Proc. Natl. Acad. Sci. U.S.A.* 106:5348–5352, 2009.

McClintock B. – The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U.S.A.* 36 (6): 344–55, 1950.

METASERVER:

http://meta.bioinfo.pl/submit_wizard.pl

Michalakis Y. & Renaud F. – Malaria: Evolution in vector control. *Nature* 462, 298-300, 2009.

Miller L.H., Baruch D.I., Marsh K. and Doumbo O.K. – The pathogenic basis of malaria. *Nature* 415, 673–679, 2002.

Mitchell G.H., Thomas A.W., Margos G., Dluzewski A.R. and Bannister L.H. – Apical membrane antigen 1, a major malaria vaccine candidate, mediates the close attachment of invasive merozoites to host red blood cells. *Infect. Immun.* 72, 154–158, 2004.

Nery S., Deans A.M., Mosobo M., Marsh K., Rowe J.A. and Conway D.J. – Expression of *Plasmodium falciparum* genes involved in erythrocyte invasion varies among isolates cultured directly from patients. *Mol. Biochem. Parasitol.* 149, 208–215, 2006.

Nosten F. et al. – Effects of artesunate-mefloquine combination on incidence of *Plasmodium falciparum* malaria and mefloquine resistance in western Thailand: a prospective study. *Lancet.* 356:297–302, 2000.

O'Donnell R.A., De Koning-Ward T.F., Burt R.A., Bockarie M., Reeder J.C., Cowman A.F. and Crabb, B.S. – Antibodies against merozoite surface protein (MSP)- 1(19) are a major component of the invasion-inhibitory response in individuals immune to malaria. *J. Exp. Med.* 193, 1403–1412, 2001.

Orlandi, P.A., Sim K.L., Chulay J.D. and Haynes J.D. – Characterization of the 175-kilodalton erythrocyte binding antigen of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 40:285–294, 1990.

Orlandi P.A., Klotz F.W. and Haynes J.D. – A malaria invasion receptor, the 175- kilodalton erythrocyte binding antigen of *Plasmodium falciparum* recognizes the terminal Neu5Ac(alpha 2-3)Gal-sequences of glycophorin *Am. J. Biol. Chem.* 116, 901–909, 1992.

Ozenne V., Bauer F., Salmon L., Huang J.R., Jensen M.R., Segard S., Bernadó P., Charavay C. and Blackledge M. – Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* 1;28(11):1463-70, 2012.

Pasternack G.R., Racusen R.H. – Erythrocyte protein 4.1 binds and regulates myosin. *Proc Natl Acad Sci USA*, 86(24):9712-9716, 1989.

Pernot P., Round, Barrett, De Maria Antolinos, Gobbo, Gordon, Huet, Kieffer, Lentini, Mattenet, Morawe, Mueller-Dieckmann, Ohlsson, Schmid, Surr, Theveneau, Zerrad and McSweeney –

Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Rad.* 20, 660-664, 2013.

Persson, Fowkes, McCallum, Gicheru, Reiling, Richards, Wilson, Lopaticki, Cowman, Marsh and Beeson – Erythrocyte-Binding Antigens of *Plasmodium falciparum* Are Targets of Human Inhibitory Antibodies and Function To Evade Naturally Acquired Immunity. *The Journal of Immunology* vol. 191 no. 2 785-794, 2013.

PlasmoDB: <http://www.plasmodb.org>

Preiser P., Kaviratne M., Khan S., Bannister L. and Jarra W. – The apical organelles of malaria merozoites: host cell selection, invasion, host immunity and immune evasion. *Microbes Infect*, 2(12):1461-1477, 2000.

Rayner J.C., Vargas-Serrato E., Huber C.S., Galinski M.R and Barnwell J.W. – A *Plasmodium falciparum* homologue of *Plasmodium vivax* reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway. *J. Exp. Med.* 194:1571–1581, 2001.

- Reddy K. S. et al. – Bacterially expressed full-length recombinant *Plasmodium falciparum* RH5 protein binds erythrocytes and elicits potent strain-transcending parasite-neutralizing antibodies. *Infect. Immun.* 82, 152–164 (2014).
- Reed M.B., Caruana S.R., Batchelor A.H., Thompson J.K., Crabb B.S., Cowman A.F. – Targeted disruption of an erythrocyte binding antigen in *Plasmodium falciparum* is associated with a switch toward a sialic acid-independent pathway of invasion. *Proc. Natl. Acad. Sci. USA.* 97, 7509–7514, 2000.
- Ross A., Kessler, Krumme, Menge, Wissing, van den Heuvel, Flohé. – Optimised fermentation strategy for 13 C / 15 N recombinant protein labelling in *Escherichia coli* for NMR-structure analysis. *J. Biotech.*, 2004.
- Roy A., Kucukural A and Zhang Y. – I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, vol 5, 725-738, 2010.
- Roy A., Yang J. and Zhang Y. – COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research*, vol 40, W471-W477, 2012.
- Salmon L., Nodet G., Ozenne V., Yin G., Jensen M. R., Zweckstetter M. and Blackledge M. – NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc.* 23;132(24):8407-18, 2010.
- Sambrook and Russell – *Molecular Cloning: A Laboratory Manual* (3rd ed.). Cold Spring Harbor Laboratory Press, ISBN 978-087969577-4, 2001.
- Schaegger H. and Von Jagow G. – Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem Vol.* 166, N°2, P. 368–379, 1987.
- Schischmanoff P.O., Winardi R., Discher D.E., Parra M.K., Bicknese S.E., Witkowska H.E., Conboy J.G. and Mohandas N. – Defining of the Minimal Domain of Protein 4.1 Involved in Spectrin-Actin Binding. *The Journal of Biological Chemistry*, 270, 21243-21250, 1995
- Serdyuk I.N., Zaccai N.R. and Zaccai J. – *Methods in Molecular Biophysics: Structure, Dynamics, Function.* Cambridge University Press, 2007.

- Sickmeier M., Hamilton J.A., LeGall T., Vacic V., Cortese M.S., Tantos A., Szabo B., Tompa P., Chen J., Uversky V.N., Obradovic Z. and Dunker A.K. – *DisProt: the Database of Disordered Proteins*. *Nucleic Acids Res.* 35, D786-93, 2007.
- Silverstein R.M., Bassler G.C. and Morrill T.C. – *Spectrometric Identification of Organic Compounds*, 5th Ed., Wiley, 1991.
- Sim B.K.L., et al. – *Primary structure of the 175K Plasmodium falciparum erythrocyte binding antigen and identification of a peptide which elicits antibodies that inhibit malaria merozoite invasion*. *J. Cell Biol.* 111:1877–1884, 1990.
- Sim B.K.L., Chitnis C.E., Wasniowska K., Hadley T.J. and Miller L.H. – *Receptor and ligand domains for invasion of erythrocytes by Plasmodium falciparum*. *Science* 264:1941–1944, 1994.
- Sim B.K.L. – *Plasmodium falciparum: further characterization of a functionally active region of the merozoite invasion ligand EBA-175*. *Exp. Parasitol.* 78, 259–268, 1994.
- Singh A.P. et al. – *Targeted deletion of Plasmodium knowlesi Duffy binding protein confirms its role in junction formation during invasion*. *Mol. Microbiol.* 55:1925–1934, 2005.
- Singh B. – *A large focus of naturally acquired Plasmodium knowlesi infections in human beings*. *Lancet* 363, 1017–1024, 2004.
- Singh N., Preiser P., Renia L., Balu B, Barnwell J., Blair P., Jarra W., Voza T., Landau I. and Adams J.H. – *Conservation and developmental control of alternative splicing in maebl among malaria parasites*. *J Mol Biol* 343(3):589-599, 2004.
- Smith P.K. et al. – *Measurement of protein using bicinchoninic acid*. *Anal. Biochem.* 150 (1): 76–85, 1985.
- Snow R.W., Guerra C.A., Noor A.M., Myint H.Y. and Hay S.I. – *The global distribution of clinical episodes of Plasmodium falciparum malaria*. *Nature* 434, 214–217, 2005.
- Stubbs J. et al. – *Molecular mechanism for switching of P. Falciparum invasion pathways into human erythrocytes*. *Science* 309:1384–1387, 2005.

- Su X., Hayton K. & Wellem T.E. – Genetic linkage and association analyses for trait mapping in *Plasmodium falciparum*. *Nature Reviews Genetics* 8, 497-506, 2007.
- Sugase K., Dyson H.J. and Wright P.E. – Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447, 1021-1025, 2007.
- Svergun D.I. – Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25, 495-503, 1992.
- Svergun D.I., Barberato C. and Koch M.H.J. – CRYSOLO - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates *J. Appl. Cryst.* , 28, 768-773, 1995.
- Svergun D.I. – Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* 2879-2886, 1999.
- Tham W.H. et al. – Antibodies to reticulocyte binding protein-like homologue 4 inhibit invasion of *Plasmodium falciparum* into human erythrocytes. *Infect. Immun.* 77:2427–2435, 2009.
- Tolia N.H., Enemark E.J., Sim B.K., Joshua-Tor L. – Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. *Cell* 122, 183–193, 2005.
- Tompa P. – The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* 13;579(15):3346-54, 2005.
- Tompa P. and Fuxreiter M. – Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci.* 33(1):2-8., 2008.
- Tompa P. and Han K. – *Intrinsically disordered proteins*. American Institute of Physics, 2012.
- Topolska A.E., Wang L., Black C.G. and Coppel R.L. – Merozoite cell biology. In *Malaria parasites: genomes and molecular biology*. Edited by: Waters A.P., Janse C.J., Wymondham, UK, Caister Academic Press, 2004.

- Triglia T., Duraisingh M.T., Good R.T., and Cowman A.F. – Reticulocyte- binding protein homologue 1 is required for sialic acid-dependent invasion into human erythrocytes by *Plasmodium falciparum*. *Mol. Microbiol.* 55:162–174, 2005
- Triglia T., Tham W.H., Hodder A. and Cowman A.F. – Reticulocyte binding protein homologues are key adhesins during erythrocyte invasion by *Plasmodium falciparum*. *Cell. Microbiol.* 11:1671–1687, 2009.
- Triglia T. – Identification of proteins from *Plasmodium falciparum* that are homologous to reticulocyte binding proteins in *Plasmodium vivax*. *Infect. Immun.* 69:1084–1092, 2001.
- Triglia T., Thompson J.K. and Cowman A.F. – An EBA175 homologue which is transcribed but not translated in erythrocytic stages of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 116:55–63, 2001.
- Uversky V.N., Oldfield C.J. and Dunker A.K. – Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 18: 343–384, 2005.
- Ventura L. and Blier B. – *Les Tontons flingueurs*. Georges Lautner, 1963.
- Volkov V.V. and Svergun D.I. – Uniqueness of ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* 36, 860-864, 2003.
- Ward J.J., Sodhi J.S., McGuffin L.J., Buxton B.F. and Jones D.T. – Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* 337, 635-645, 2004.
- Williams A. R. et al. – Enhancing blockade of *Plasmodium falciparum* erythrocyte invasion: assessing combinations of antibodies against PfRH5 and other merozoite antigens. *PLoS Pathog.* 8, e1002991, 2012.
- World Health Organization – World Malaria Report, from 2005 to 2014. http://www.who.int/malaria/publications/world_malaria_report/en/
- World Health Organization – The World Health Report, 2005.
- Zhang H., Neal S. and Wishart D. – RefDB: A database of uniformly referenced protein chemical shifts. *Journal of Biomolecular NMR*, 25: 173-195, 2003.

Zhang Y. – I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9:40, 2008.