

**Linking Structure and Function of the
Asialoglycoprotein Receptor H1-CRD
using
Site-Directed Mutagenesis and Isotope Labeling**

Inauguraldissertation
zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Anna Karin Johansson
aus Stockholm, Schweden

Budapest, 2007

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Beat Ernst, Institute of Molecular Pharmacy, University of Basel

Prof. Dr. Martin Spiess, Growth and Development, Biocenter, University of Basel

Prof. Dr. Hans-Peter Hauri

Dekan

Basel, den 26. Juni 2007

“Imagination is more important than knowledge.

For knowledge is limited to all we know and understand, while imagination embraces the entire world, and all there ever will be to know and understand.”

(Albert Einstein, 1879-1955)

Acknowledgements

It feels slightly overwhelming to realize that my PhD is finished, at last. The lab work has come to an end, the results have been summarized and evaluated and all the work I have done for the past three years and nine months have been condensed here – in the thesis! It also makes me feel very nostalgic (as I am sure comes as no surprise to those who know me) since I truly enjoyed my time as a PhD student, not only because of the scientific work and my experimental endeavors, but also for having the privilege to meet and work with so many fantastic and great persons along the way. Without their help, support and friendship, my PhD would just not have been the same!

First of all I would like to thank Prof. Beat Ernst for giving me the opportunity to make my PhD in his group at the Institute of Molecular Pharmacy. He has created a great institute, which not only offers a multidisciplinary environment, but also exudes a friendly atmosphere. I am very grateful for his supervision and encouragement during these past four years.

I sincerely thank Prof. Martin Spiess for kindly accepting to be the co-referee of my thesis.

I would also like to gratefully acknowledge Dr. Hannelore Peters, who welcomed me in her lab at the Institute of Chemistry at the University of Lübeck, and shared her knowledge and expertise on isotope labeling with me.

I would like to thank all members of the Institute of Molecular Pharmacy, both past and present, for making my time at the university as well as in Basel so special. As I previously said, the Institute of Molecular Pharmacy is a truly extraordinary place, much thanks to the people, who are always willing to help, share their experience, discuss problems or just listen. Special thanks to my colleagues in the Asialoglycoprotein receptor project team; Rita Born, Daniel Ricklin, Daniela Stokmaier, Claudia Riva and Oleg Khorev, for constant support and encouragement. I would also like to thank Dr. Brian Cutting, who always made time for me, explaining NMR experiments and results with much patience, or to just discuss science in general. Tina Weber, thank you for always being there! Also, thanks to Zorica Dragic and Tamara

Visekruna, who helped me get started when I first arrived at the institute, showing genuine interest and offering helpful advice. I thank Morena Spreafico for taking time and kindly providing me with beautiful pictures of H1-CRD.

Fellow researchers of 4020 – it has been a pleasure to share the lab with you!

My special thanks goes to Cornelia Blaser, Marlen Schneider, Franchina Purtschert and Daniela Abgottspon, my four diploma students, who have been of tremendous support. These girls were not only excellent students and great scientists, but also proved to be very good friends.

On a more personal level, I would like to thank the people who have always stood by me, supported and encouraged me, even when that meant me leaving them, my family. Mamma, Pappa och Gunilla, tack för att ni alltid finns där för mig. Och för att ni alltid säger de rätta sakerna! I would also like to thank my second family, klanen Jacobson. Jag är väldigt glad över att vara en del av er!

Finally, I would like to thank the person who means the most to me, my boyfriend Jonas, who always encourages and supports me, shares my ups and downs and makes me want to be the best person I can be. Jonas, du är min hjälte!

Abstract

The asialoglycoprotein receptor (ASGP-R) is a C-type lectin, abundantly expressed on hepatocytes. It mediates the clearance of desialylated glycoproteins carrying terminal galactose (Gal) or *N*-acetyl-galactosamine (GalNAc) residues through endocytosis. The receptor consists of two subunits, H1 and H2, both containing a carbohydrate recognition domain (CRD), responsible for ligand binding. Organ specific expression and high ligand specificity makes the ASGP-R a potential candidate for targeted drug delivery to the liver.

The aim of this work has been to improve the understanding of H1-CRD by linking function and structure of the subunit. Site-directed mutagenesis was used to deduce the role of three *N*-terminal cysteines in dimer formation seen upon expression of H1-CRD *in vitro*. Step-wise substitution of the cysteines by serines proved to reduce and even completely abolish the dimerization. However, GalNAc affinity of the mutant proteins was impaired as a result of the modification(s).

Site-directed mutagenesis was applied in a second study to investigate the functional importance of selected amino acid residues in the binding site of H1-CRD. Five single mutant proteins were created, identifying one residue of major importance for GalNAc binding. Two other residues displayed only minor influence on ligand binding, while one mutation was seen to result in an improvement of the affinity for GalNAc. In addition, one mutant was created to investigate the role of a histidine vs. a glutamate in pH-dependent ligand binding exhibited by H1-CRD.

Finally, a method for high efficiency isotope labeling of H1-CRD was established. The method was shown to yield protein with high incorporation levels of both ^{13}C and ^{15}N , providing a good basis for a future structure determination of H1-CD by NMR.

Abbreviations

Ab	Antibody
ABTS	2,2'-azino-di-[3-ethylbenzthiazoline-6-sulfonic acid]
Ala	Alanine
APS	Ammonium persulfate
Arg	Arginine
ASF	Asialofetuin
ASGP-R	Asialoglycoprotein receptor
Asn	Asparagine
ASOR	Asialoorosmucoid
Asp	Aspartate
AU	Absorbance units
bp	Base pair
BSA	Bovine serum albumine
CRD	Carbohydrate recognition domain
Cys	Cysteine
DEAE	Diethylaminoethyl
DNA	Deoxyribonucleic acid
DTNB	5,5'-dithiobis-(2-nitrobenzoate)
DTT	Dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
FPLC	Fast protein liquid chromatography
Gal	D-Galactose
GalNAc	<i>N</i> -Acetyl-D-galactosamine

Glu	Glutamate
Gly	Glycine
H1	Asialoglycoprotein receptor subunit 1
H2	Asialoglycoprotein receptor subunit 2
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HMQC	Heteronuclear multiple quantum coherence
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum correlation
IC ₅₀	50% inhibition constant
IEC	Ion exchange chromatography
IMP	Institute of Molecular Pharmacy
IPTG	Isopropyl- β -D-thiogalactopyranoside
K _D	Equilibrium dissociation constant
k _{off}	Dissociation rate constant
k _{on}	Association rate constant
LacS	Lactose transport protein
LB medium	Luriani bertani medium
MBP	Mannose binding protein
Met	Methionine
MGR	Macrophage galactose receptor
MHL-1	Mouse asialoglycoprotein receptor subunit 1
MHL-2	Mouse asialoglycoprotein receptor subunit 2
MS	Mass spectrometry
NEM	<i>N</i> -Ethylmaleimide
NMR	Nuclear magnetic resonance

OD	Optical density
O/D	Overday
OE-PCR	Overlap extension polymerase chain reaction
O/N	Overnight
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pro	Proline
RBS	Ribosome binding site
RHL-1	Rat asialoglycoprotein receptor subunit 1
R _{max}	Maximum response
RP	Reversed phase
rpm	Rounds per minute
RT	Room temperature
RU	Resonance unit
SDM	Site-directed mutagenesis
SDS	Sodium dodecylsulfate
Ser	Serine
SPR	Surface plasmon resonance
STD	Saturation transfer difference
TB medium	Terrific broth medium
TCA	Trichloroacetic acid
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFA	Trifluoroacetic acid
T _m	Melting temperature

Tris	Tris(hydroxymethyl)aminomethane
w/o	Without
WT	Wildtype

1	Introduction.....	1
1.1	Background.....	1
1.2	The Asialoglycoprotein Receptor (ASGP-R).....	3
1.2.1	Structural organization of ASGP-R.....	3
1.2.2	Ligand binding properties of ASGP-R.....	4
1.2.3	Physiological functions of ASGP-R.....	5
1.2.4	ASGP-R – a model for endocytosis.....	6
1.2.5	Extrahepatic occurrence of ASGP-R.....	7
1.2.6	H1-CRD – the crystal structure.....	8
1.2.7	The sugar binding site of H1-CRD.....	9
1.2.8	Sugar binding to H1-CRD.....	11
1.2.9	ASGP-R – a candidate for targeted drug delivery.....	13
1.3	Recombinant protein production.....	13
1.3.1	Starting at the DNA-level.....	14
1.3.2	The pET expression system.....	16
1.3.3	Protein expression.....	17
1.3.4	Protein purification.....	18
1.3.5	General concluding remarks.....	19
1.4	Site-directed mutagenesis.....	19
1.4.1	PCR-based mutagenesis strategies.....	20
1.4.2	Planning a mutation.....	23
1.4.3	Alanine-scanning mutagenesis.....	23
1.5	Receptor-ligand binding assays.....	24
1.5.1	A solid-phase competition assay.....	24
1.5.2	Surface plasmon resonance (SPR).....	25
1.6	Isotope labeling of recombinant proteins.....	27
1.6.1	Uniform labeling.....	28
1.6.2	NMR as a tool to detect and investigate protein-ligand interactions.....	29
1.7	Scope of thesis.....	30
1.7.1	The role of the three cysteines at the N-terminus.....	30
1.7.2	Investigation of the binding site.....	31
1.7.3	Isotope labeling of H1-CRD.....	33
1.7.4	Aim.....	34
2	Material and Methods.....	35
2.1	Material and reagents.....	35
2.2	Equipment.....	36
2.3	General methods for working with <i>E.coli</i> as expression system.....	37
2.3.1	<i>E.coli</i> strains used for cloning and protein expression.....	37
2.3.2	Bacterial growth and expression medium.....	38
2.3.3	Competent cells.....	39
2.3.4	Electrocompetent cells.....	39
2.3.5	Calciumcompetent cells.....	40
2.4	General methods for working with recombinant DNA.....	40
2.4.1	Plasmid.....	40
2.4.2	Agarose gel electrophoresis analysis.....	41

2.4.3	Restriction digestion	41
2.5	Expression, purification and analysis of ASGP-R H1-CRD	42
2.5.1	Confirming the sequence of H1-CRD DNA.....	42
2.5.2	Expression of WT H1-CRD	42
2.5.3	Material for the purification of WT H1-CRD	43
2.5.4	Solubilization and renaturation of WT H1-CRD	44
2.5.5	Purification by FPLC affinity chromatography	44
2.5.6	Protein analysis by SDS-PAGE	45
2.5.7	Western Blot analysis	46
2.5.8	Separation of monomer and dimers by HPLC ion exchange chromatography	48
2.5.9	Separation of monomers and dimers by HPLC reversed phase chromatography	49
2.5.10	Purification by HPLC affinity chromatography	49
2.5.11	Concentration determination by the Bradford assay	49
2.6	Cloning, expression and characterization of mutants of H1-CRD.....	50
2.6.1	Isolation of pET-3bH1.....	50
2.6.2	Preparation of template H1-CRD DNA	50
2.6.3	Quantification of DNA	52
2.6.4	Preparing plasmid pET-3b.....	52
2.7	Site-directed mutagenesis.....	53
2.7.1	Primers.....	53
2.7.2	ExSite PCR-based site-directed mutagenesis system.....	54
2.7.3	Conventional PCR with mutagenic primers.....	57
2.7.4	Overlap extension PCR.....	59
2.7.5	Ligation	61
2.7.6	Transformation by electroporation	61
2.7.7	Big dye 1.1 terminator sequencing.....	62
2.7.8	Glycerol stocks.....	63
2.7.9	Transformation into calcium competent cells.....	63
2.7.10	Small-scale expression analysis	64
2.7.11	Determination of target protein solubility.....	64
2.8	Large-scale expression of mutant H1-CRD	65
2.8.1	Solubilization and renaturation of mutant H1-CRD.....	65
2.8.2	Purification by FPLC affinity chromatography	66
2.8.3	Separation of monomers and dimers by HPLC IEC.....	66
2.8.4	Purification by HPLC affinity chromatography	66
2.9	Evaluation of WT and mutant H1-CRD	67
2.9.1	The solid phase competition assay.....	67
2.9.2	The Biacore assay.....	68
2.10	NMR studies of H1-CRD	69
2.10.1	T1rho measurements	69
2.10.2	Stability measured by T1rho	70
2.10.3	Saturation transfer difference measurements	70
2.11	Expression of isotope labeled H1-CRD	71
2.11.1	Minimal medium for bacterial growth and protein expression.....	71
2.11.2	Expression of isotope labeled protein using a two-stage protocol	72

2.11.3	Expression of isotope labeled protein by exclusive use of minimal medium.....	75
2.11.4	Sample preparation for SDS-PAGE electrophoresis.....	76
2.11.5	Solubilization and renaturation of isotope labeled H1-CRD	76
2.11.6	Purification by FPLC affinity chromatography	77
2.11.7	Separation of monomers and dimers by HPLC IEC.....	77
2.11.8	Separation of monomers and dimers by HPLC RP	77
2.11.9	Mass spectrometry analysis.....	77
2.11.10	NMR studies of isotope labeled H1-CRD	77
3	Results.....	80
3.1	Expression, purification and analysis of H1-CRD.....	80
3.1.1	Expression and purification of H1-CRD	80
3.1.2	Reference batch of H1-CRD	85
3.1.3	Separation of monomers and dimers using HPLC reversed phase chromatography	86
3.1.4	Comparison of the binding affinity of monomers purified either by HPLC IEC or RP.....	87
3.2	Site-directed mutagenesis of H1-CRD	92
3.2.1	Cloning of H1-CRD	92
3.2.2	Site-directed mutagenesis	94
3.2.3	Transformation by electroporation	101
3.2.4	Sequencing results.....	102
3.2.5	Transformation into calcium competent cells.....	103
3.2.6	Determination of target protein solubility.....	104
3.3	Expression and characterization of cysteine mutants of the ASGP-R H1-CRD	104
3.3.1	Expression and purification of cysteine mutants of H1-CRD.....	105
3.3.2	Analysis of dimer content	105
3.3.3	Separation of monomers and dimers.....	106
3.3.4	Evaluation of the binding affinity of the cysteine mutants.....	108
3.4	Investigation of the binding site of the ASGP-R H1-CRD.....	113
3.4.1	Expression and purification of binding site mutants of H1-CRD	113
3.4.2	Evaluation of the binding site mutants of H1-CRD using the solid-phase competition assay.....	114
3.4.3	Evaluation of the binding site mutants of H1-CRD using the Biacore assay	117
3.4.4	Probing pH-dependent ligand binding of mutant H256E.....	118
3.5	Expression and purification of isotope labeled H1-CRD	121
3.5.1	Expression of isotope labeled H1-CRD using a two-stage protocol....	121
3.5.2	Analysis of the isotopic enrichment using NMR	128
3.5.3	Mass spectrometry analysis of isotope labeled H1-CRD produced by the two-stage protocol	130
3.5.4	HSQC of ¹⁵ N labeled H1-CRD expressed according to the two-stage protocol	132

3.5.5	HMQC of $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD expressed according to the two-stage protocol.....	134
3.5.6	Alternative expression set-up of the two-stage protocol.....	134
3.5.7	Expression of isotope labeled H1-CRD by exclusive use of minimal medium.....	135
3.5.8	HSQC of ^{15}N labeled H1-CRD produced by exclusive use of minimal medium.....	138
4	Discussion.....	142
4.1	Expression and purification of H1-CRD – protein production interrupted	142
4.2	Expression and characterization of cysteine mutants of ASGP-R H1-CRD	145
4.2.1	Dimer formation can be stopped by site-directed mutagenesis... ..	146
4.2.2	...but at the expense of protein affinity and stability.....	147
4.3	Investigation of the binding site of ASGP-R H1-CRD.....	149
4.3.1	Identification of residues important for the functionality of ASGP-R H1-CRD	149
4.3.2	Histidine 256 – responsible for pH dependent ligand binding.....	153
4.4	Isotope labeling of ASGP-R H1-CRD.....	154
4.4.1	H1-CRD expressed by the two-stage protocol – a vanishing act of ^{13}C and ^{15}N	155
4.4.2	Expression of isotope labeled H1-CRD	157
4.5	Summary and outlook	157
4.5.1	Role of the three N-terminal cysteines	157
4.5.2	Investigation of the binding site of H1-CRD	158
4.5.3	Isotope labeling of H1-CRD.....	159
5	References	161

1 Introduction

1.1 Background

It has been roughly 100 years since Paul Ehrlich suggested the concept of drug targeting by coining the expression “the magic bullet”. Ehrlich, a German bacteriologist, saw the magic bullet as an entity consisting of two components with separate actions and objectives. One part of the bullet/drug should recognize and selectively bind the predetermined target, while the purpose of the second part was to provide a therapeutic action. Such a drug would theoretically be devoid of any side effects, as it would only attack invading pathogens or bind to the anticipated receptor, effectively fight the disease but leave the patient unharmed [1,2]. The magic bullet, as Ehrlich envisaged it, is in essence targeted drug delivery and remains a trait modern drug discovery and development strive for.

Design of drugs for targeted delivery makes use of molecules capable of specific recognition and binding to a predestined site. Such molecules, also referred to as target moieties, can for instance be antibodies, lectins, lipoproteins, saccharides *etc.* The actual target in turn can be any receptor or compound/antigen that is specific for the organ or tissue of interest [1].

An attractive target for drug delivery is the liver, or more specifically the hepatocytes. These cells are of great importance as they play a major role in many aspects of lipid and carbohydrate metabolism, and are the main site for the synthesis of numerous serum proteins. In addition, hepatocytes can produce inflammatory mediators upon damage, leading to pathological cascades. Drug targeting of the liver presents interesting opportunities, e.g. to influence hepatocyte related metabolism, correct genetic defects and attenuate liver injury. A potential target candidate suggested for drug delivery directed to the liver is the asialoglycoprotein receptor (ASGP-R) [3]. The receptor is abundantly expressed on mammalian hepatocytes and recognizes ligands bearing terminal galactose (Gal) or *N*-acetylgalactosamine (GalNAc) residues with high specificity. Following binding, the receptor-ligand complex is internalized through endocytosis, whereupon the ligand is released into the endosome and the receptor is recycled back to the plasma membrane [4].

ASGP-R is composed of two subunits, the major H1 and the minor H2, both containing a carbohydrate recognition domain (CRD) responsible for the ligand binding. The crystal structure of H1-CRD was determined in 2000, providing valuable information on the structure of the binding site and folding of the protein [5].

The aim of this work has been to further investigate functional and structural elements of H1-CRD in an attempt to gain deeper insight into the functionality of the subunit. Three main projects, addressing different aspects of the question at hand, have been undertaken.

- 1) Elucidation of the importance of the three cysteines (Cys, C) 152, 153 and 164 at the *N*-terminus for the functionality of the subunit. The cysteines in question represent a probable cause for the subunit dimer formation seen upon expression *in vitro*. A strong incentive for this study was to investigate the possibility of substituting the cysteines or even shortening the protein, and thereby preventing the dimerization.
- 2) Investigation of specific amino acid residues in the actual binding site of H1-CRD to determine their individual contributions to ligand binding. The binding site of the rat ASGP-R H1 (RHL-1) was the focus of numerous mutagenesis studies during the mid nineties, revealing the basis for the specificity of ASGP-R CRD for Gal/GalNAc over other sugars [6-8]. However, the studies stopped before a complete overview of the binding site could be established. This project aimed to provide a finalized summary of the binding site, assigning function and relevance to each residue.
- 3) Development of a method for isotope labeling of H1-CRD to enable studies by three-dimensional heteronuclear nuclear magnetic resonance (NMR). Multidimensional NMR is a most valuable method for studying proteins in solution under native-like conditions, e.g. to determine the structure or study protein-ligand interactions [9]. However, the method necessitates proteins enriched with stable isotopes such as ¹³C and/or ¹⁵N, rationalizing the need for isotope labeling of H1-CRD.

In summary, all three projects aim to characterize the ASGP-R H1-CRD by establishing a structure-function relationship. A secondary, long-term goal and expectation is that the information obtained during this work ultimately will aid the

development of targeted drug delivery to the liver through the asialoglycoprotein receptor.

1.2 The Asialoglycoprotein Receptor (ASGP-R)

The ASGP-R was the first mammalian lectin to be described, identified by Ashwell and Morell in the mid sixties [10,11]. The receptor is a member of the C-type lectin family, which is defined by their characteristics to require Ca^{2+} for ligand binding and to contain disulfide bridges in the carbohydrate recognition domain [12-14]. ASGP-R is an integral membrane protein, predominately expressed on the sinusoidal surface of mammalian hepatocytes. Although the physiological function of the receptor has yet to be established, it is known to mediate the clearance of desialylated glycoproteins carrying terminal galactose or *N*-acetylgalactosamine residues from the circulation through endocytosis. The receptor is an excellent model and one of the best-characterized systems for receptor-mediated endocytosis via the clathrin-coated pit pathway [4,15].

1.2.1 Structural organization of ASGP-R

The functional ASGP-R *in vivo* is a hetero-oligomer composed of two homologous subunits, the major H1 and the minor H2, sharing 55% sequence identity. Each subunit consists of four domains, a *N*-terminal cytoplasmic domain, a single-pass transmembrane domain, an extracellular stalk segment and a carbohydrate recognition domain (CRD), the latter being responsible for the ligand binding (*Figure 1A*) [4,16]. A notable difference between H1 and H2 is an 18-amino acid insert present in the cytoplasmic domain of only H2 [4,17]. Furthermore, while H1 only occurs as one protein isoform, three splice variants of H2, designated H2a, H2b and H2c, have been isolated from human liver and HepG2 cells. However, H2a does not occur in native ASGP-R complexes. In contrast, both H2b and H2c associate with H1 in functional ASGP-Rs, but never together in the same receptor complex [18].

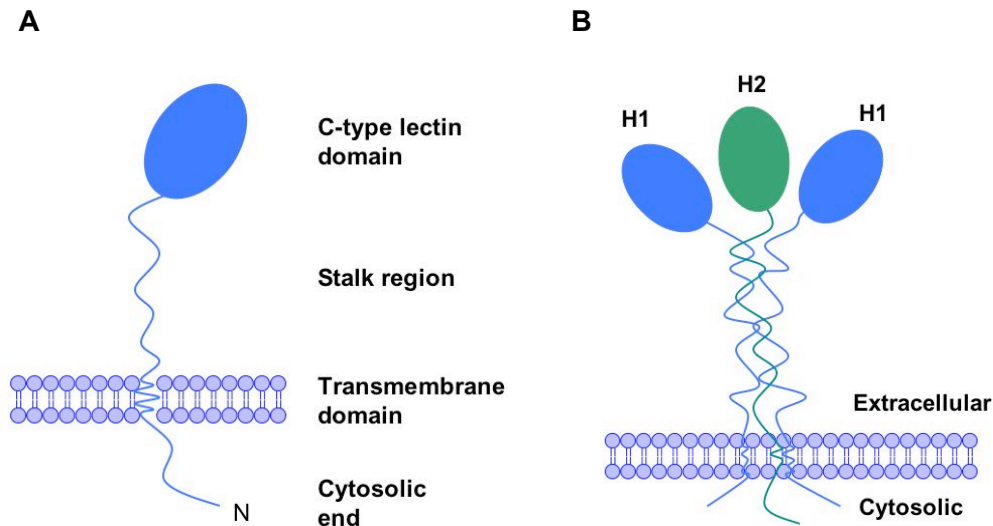


Figure 1. Schematic presentation of ASGP-R [19]. A) Each subunit, H1 and H2, consists of four domains; a N-terminal cytoplasmic domain, a single-pass transmembrane domain, an extracellular stalk segment and a carbohydrate recognition domain. B) A hetero-oligomeric complex of two H1 and one H2 subunits has been proposed as the minimum size of an active ASGP-R.

Presence of both subunits is a prerequisite for the ASGP-R to reach full functionality. Nevertheless, the subunit composition of the native receptor remains unspecified [4]. A stoichiometry of 2:1 of H1 and H2 has been proposed as the minimum size of an active ASGP-R complex, exhibiting high affinity ligand binding (*Figure 1B*) [20,21]. This subunit ratio is partially supported by another study in which the functional receptor is suggested to be a 2:2 heterotetramer. Extracellular stalk segments of H1 and H2, responsible for the subunit oligomerization, were expressed and shown to associate in homo- and heterooligomers *in vitro* [22]. Conditions mimicking the cellular expression levels of H1 and H2 at a molar ratio of 3:1 favored formation of the 2:2 heterotetramers.

1.2.2 Ligand binding properties of ASGP-R

The ASGP-R specifically binds terminal Gal and GalNAc residues (*Figure 2*), present on desialylated glycoproteins. Studies have shown that the binding affinity is highly dependent on the number, distance and three-dimensional arrangement of the sugar moieties. A single Gal residue exhibits only modest binding affinity with a dissociation constant (K_D) in the order of 1 mM. Additional sugars, going from a mono- to a triantennary structure, result in a significant increase in affinity with dissociation constants in the nanomolar range. Hence, high affinity binding is achieved through multiple interactions between the CRDs in a receptor complex and multiple sugar residues [23,24]. Furthermore, the sugar spacing has been shown to be important for

optimal receptor recognition. A spacer length of $>20\text{\AA}$, separating the sugar moieties from the branching point of the ligand, results in high affinity binding to the ASGP-R of Gal containing ligand constructs [24,25]. In addition, the ASGP-R clearly favors binding of GalNAc over Gal, showing an approximate 50-fold higher affinity for the former [25,26].

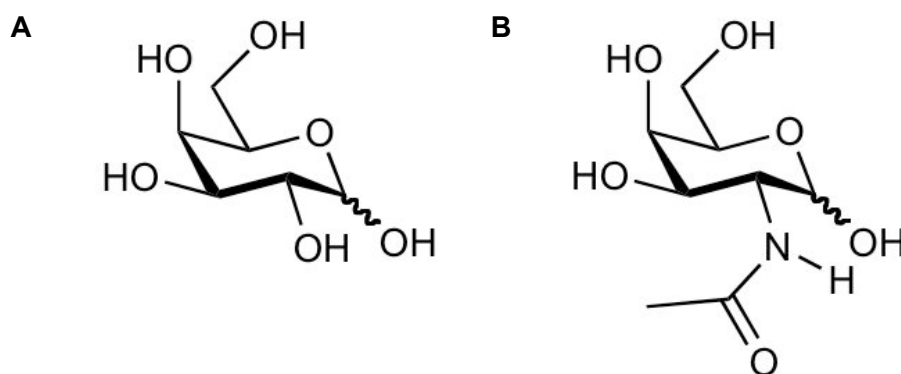


Figure 2. Monosaccharide ligands of H1-CRD A) D-Galactose B) *N*-Acetyl-D-galactosamine

Two of the best-characterized and most commonly used ligands for the ASGP-R are asialo-orosmuroid (ASOR) and asialofetuin (ASF), two glycoproteins with five and three *N*-linked glycans respectively [27]. ASOR displays a K_D of ~ 10 nM for the native receptor. It has also been observed to bind H1, overexpressed in transfected COS-7 cells, in the absence of H2 with a K_D of ~ 40 nM [23].

1.2.3 Physiological functions of ASGP-R

The primary function of the ASGP-R has been considered to be the removal and degradation of desialylated glycoproteins from the circulation. Normally, many oligosaccharide chains on glycoproteins carry terminal sialic acid residues. Upon removal of the sialic acid, caused by the action of neuraminidases, penultimate galactose residues are exposed and recognized by the ASGP-R [4,28].

As previously noted, both H1 and H2 are required to reach full functionality of the receptor. In addition, the expression of the two subunits has been shown to be interdependent. Studies of mice lacking the minor subunit (MHL-2), due to disruption of the corresponding gene, also showed clearly decreased levels of the major subunit (MHL-1) and were unable to clear asialo-orosmuroid from the plasma [29]. Further experiments with a reversed approach revealed that MHL-1-deficient mice were unable to express detectable levels of MHL-2 and also incapable of clearing ASOR from the

plasma. In summary, both subunits are required for the stable expression of the ASGP-R, but HL-1 appears to be more strictly required than HL-2. Interestingly, even though the clearance of ASOR by ASGP-R was severely impaired in both MHL-1- and MHL-2-deficient mice, neither accumulation of plasma glycoproteins nor any obvious phenotype could be observed [28]. These results suggest that ASGP-R is not ultimately responsible for the clearance of plasma glycoproteins, but is likely to possess other functions.

ASGP-R has been proposed to be involved in the metabolism of plasma lipoproteins [30] and cellular fibronectin [31]. However, studies with ASGP-R knock-out mice could not confirm this theory, as the plasma levels of fibronectin and lipoprotein appeared unaffected in the absence of the receptor [28]. ASGP-R has also been implicated in the clearance of apoptotic cells by the liver. Studies showed that the uptake of apoptotic bodies was blocked in the presence of an ASGP-R specific antibody or following the addition of receptor specific ligands such as ASF or GalNAc [32]. Immunoglobulin A has also been proposed as a ligand for the receptor [33].

Finally, ASGP-R is thought to act as an entry point for a few specified pathogenic viruses to the hepatocytes. Experimental data indicate that the Marburg virus [34], the hepatitis B virus [35,36] and the hepatitis C virus [37] are capable of binding to the receptor, followed by infection of the host cell.

1.2.4 ASGP-R – a model for endocytosis

Receptor-mediated endocytosis serves as a mechanism by which cells can internalize macromolecules like peptides and proteins. The ASGP-R has been the focus of several studies aiming to understand endocytosis via the clathrin-coated pit pathway, see figure 3. Upon ligand binding, the ASGP-Rs cluster into clathrin-coated domains of the membrane, which in turn invaginates. Clathrin-coated pits are then formed and subsequently turned into coated vesicles, which are internalized by the cell. Following the uptake, the coat detaches and the vesicle fuses with intracellular endosomes. The acidic environment in the endosomes will cause the receptor and ligand to dissociate and part ways. The ligand-containing endosomes proceed to fuse with lysosomes, where the ligands are degraded. The ASGP-Rs meet a less grim fate as they are recycled and transported back to the plasma membrane [38,39].

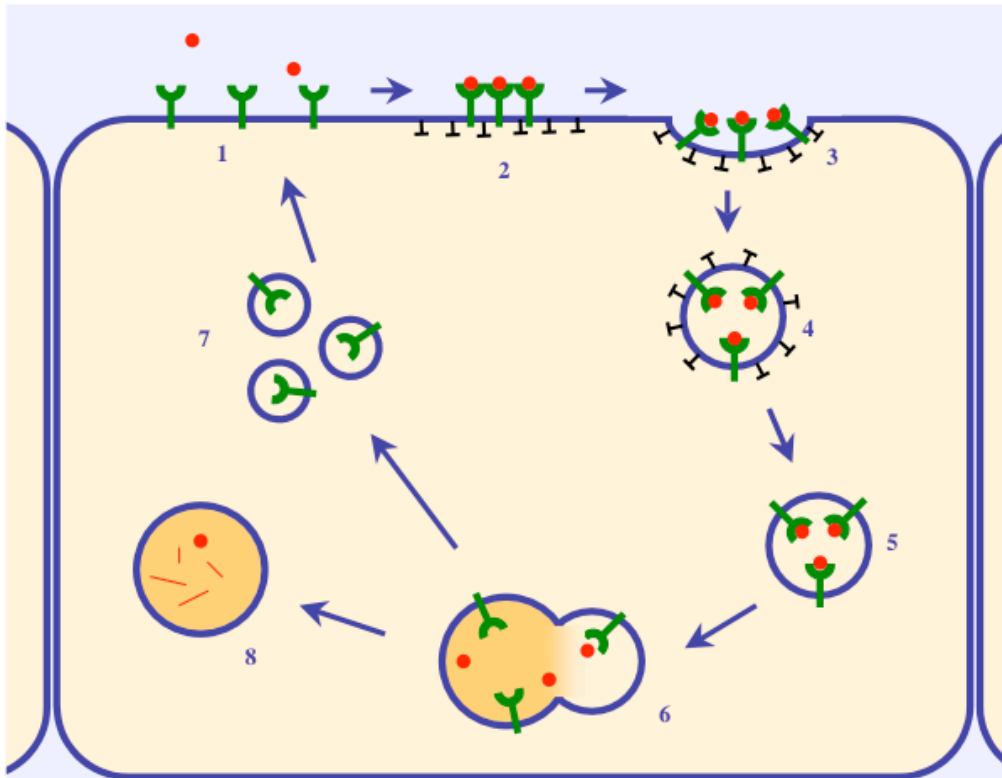


Figure 3. Schematic representation of receptor-mediated endocytosis [40]. Membrane-bound receptors (1) cluster upon ligand binding in clathrin-coated domains (2), which proceed to form a pit (3) and subsequently a vesicle that is internalized (4). The vesicle loses the clathrin coat (5) and fuses with an endosome where the ligand is released (6). The receptor is recycled back to the surface (7), while the ligand is degraded (8).

Recycling of the ASGP-R is a continuous process. It occurs several hundred times during the life span of an individual receptor, the time of which is estimated to approximately 30 hours [41]. Internalization of the ASGP-R takes place independently of ligand binding, but was shown to increase 2-fold upon binding of ASOR. Speculations attribute this enhanced internalization rate to a possible conformational change induced by ligand binding [27]. Furthermore, a tyrosine residue in H1 has been shown to be of critical importance for efficient endocytosis of the ASGP-R. The corresponding residue in H2, a phenylalanine, does not appear to contribute to the internalization of hetero-oligomeric receptor complexes [42,43].

1.2.5 Extrahepatic occurrence of ASGP-R

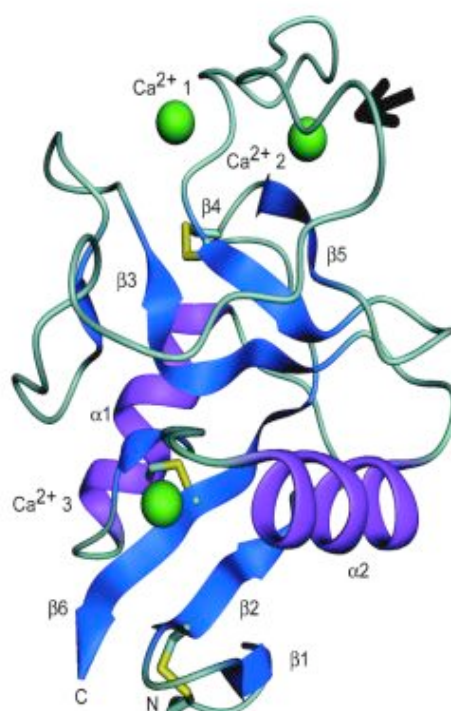
ASGP-R is predominantly expressed by hepatocytes, at an estimated density of 100'000-500'000 binding sites per cell [44,45]. However, the receptor has also been detected on extrahepatic cells. Studies using antibodies raised in mice against H1 were

able to provide evidence of expression of the receptor in the T-cell line Jurkat. Moreover, Tera-1 cells derived from human testis tissue gave rise to a weak signal, indicating presence of ASGP-R when analyzed by fluorescence activated flow cytometry [46]. Further studies support these findings, and also report on ASGP-R expression in cells originating from human bone intestine and kidney [47,48].

1.2.6 H1-CRD – the crystal structure

The structure of ASGP-R H1-CRD was solved in 2000 by X-ray crystallography [5], see figure 4. It is a globular protein, containing two α -helices and eight β -strands. The β -strands are arranged in a bent plane and form the core of the protein, while the α -helices are positioned on either side of the plane. Three calcium ions can be seen in the structure, forming an integral part of the protein as they pin together several loops. The two calcium ions at site 1 and 2 are seen in close proximity, both coordinated by glutamate residue 252. Site 2 is of particular interest as it is essential for sugar binding and is present in all C-type lectins. Calcium binding site 3 is found close to the *N*- and *C*-terminus of the protein.

A



B

146	MGSERTCPV	NWVEHERSCY	165
166	WFSRSGKAWA	DADNYCRLED	185
186	AHLVVVTSWE	EQKFVQHHIG	205
206	PVNTWMGLHD	QNGPWKWVDG	225
226	TDYETGFKNW	RPEQPDDWYG	245
246	HGLGGGEDCA	HFTDDGRWND	265
266	DVCQRPYRWV	CETELDKASQ	285
286	EPPLL ⊗⊗		

Figure 4. A) Ribbon diagram of the H1-CRD [5]. The three calcium ions can be seen in green. The second calcium binding site, which is part of the sugar binding site, is denoted by a black arrow. The disulfide bridges are marked in yellow. The structure is deposited in the RCSB Protein Data Bank, entry code 1DV8. B) Primary sequence of H1-CRD. The disulfide bridges are high-lighted pairwise. The sugar binding site can be seen in bold.

C-type CRDs can be classified into short-form CRDs, with two disulfide bridges, or long-form CRDs, containing a conserved extension with an additional disulfide bond at the *N*-terminus. H1-CRD falls into the latter category, containing seven cysteines out of which six are engaged in three disulfide bridges. One bridge constitutes part of the sugar binding site, seen in the crystal structure between Cys254 and Cys268. A second bond is formed between Cys181 and Cys276, bringing the *C*- and *N*-terminus close together, and contributes to the tertiary structure of the subunit. Finally, a third typical long-form CRD disulfide bridge is found at the *N*-terminus between Cys153 and Cys164. The *N*-terminal residues 147-152, including the seventh odd cysteine (Cys152), and the *C*-terminal residues 281-290 could not be positioned into the electron density and hence, cannot be seen in the crystal structure [5].

1.2.7 The sugar binding site of H1-CRD

The binding characteristics of ASGP-R, high affinity and specificity for Gal and GalNAc, are properties ultimately dependent on the primary structure of the protein. The region around the sugar binding site in H1-CRD is formed by one continuous stretch of the polypeptide chain from Arginine (Arg) 236 to Cys268. Numerous mutagenesis studies have been carried out with the closely related protein RHL-1, aiming to deduce the residues giving rise to its ligand binding specificity. RHL-1 has been used as a reference in attempts to mimic its Gal- and GalNAc-binding properties

in other lectins, e.g. the mannose binding protein (MBP) [6] and the macrophage galactose receptor (MGR) [7] (*Figure 5*). Such studies form the basis for the explanation of H1-CRDs sugar preference as well as its distinction between high- and low-affinity ligands.

H1-CRD	200	VQHHIGPVNT	WMGLHDQNGP	WKWVDGTDYE	229
RHL-1	200	VQQHMGPL NT	WIGLTDQNGP	WKWVDGTDYE	229
MGR	222	LQTHMGSVVT	WIGLTDQNGP	WRWVDGTDYE	251
MBP	147	IQ-EVAKTSA	FLGITDEVTE	GQFMYVTGGR	175
H1-CRD	230	TGFKNWRPEQ	PDDWYGHGLG	GGEDCAHFTD	259
RHL-1	230	TGFKNWR PGQ	PDDWYGHGLG	GGEDCA HFTT	259
MGR	252	KGFTHWAPKQ	PDNWYGHGLG	GGEDCAHFTS	281
MBP	176	LTYSNWKKDE	PND---HG--	SGEDCVTIVD	200
H1-CRD	260	DGRWNDDVCQ	RPYRWVCETE	LDKASQEPPL	L
RHL-1	260	DGHWNDDVCR	RPYRWVCETE	LGKAN-----	-
MGR	282	DGRWNDDVCQ	RPYRWCEMK	LAKDS-----	-
MBP	201	NGLWNDISCQ	ASHTAVCEFP	-----	A

Figure 5. Sequence comparison of the binding site in CRDs of human H1, RHL-1, MGR and MBP. The numbers correspond to the first and last residue in each sequence. The highlighted residues in the RHL-1 sequence have been shown to contribute to the Gal- and GalNAc-specificity of the subunit.

Introduction of two point mutations in MBP, E185Q and N187D, corresponding to Q239 and D241 in RHL-1, was sufficient to achieve galactose binding in the protein, which is normally binding mannose [6]. However, as the Gal affinity of the MBP mutant, designated QPD, was rather low it was concluded that other residues in H1-CRD also must contribute to the ligand binding. An additional mutation, H189W, showed to increase the affinity significantly, making the galactose binding of mutant QPDW comparable to that of RHL-1. Selectivity for galactose over mannose was achieved by incorporating a glycine-rich stretch following the QPDW sequence, referred to as the QPDWG mutant. Stepwise mutations within the glycine rich loop, corresponding to residues Y244, G245 and H246 in RHL-1, showed that they all contribute to the selectivity for Gal [6]. However, the influence of the amino acids on the selectivity could equally well be attributed to stabilization and support of the protein structure as to an actual effect on the ligand binding.

RHL-1, as well as H1-CRD, shows preferential binding to GalNAc over Gal. The selectivity for GalNAc has been probed by mutations studies of MGR, which binds Gal and GalNAc with roughly equal affinity. The CRDs of RHL-1 and MGR are highly homologues (77%) and the difference in selectivity is likely to stem from divergences in the sequences. It was concluded that substitution of four amino acids in MGR, V230N, A258R, K260G and S281T, is sufficient to induce GalNAc binding comparable with that of RHL-1. N230 increased the selectivity 20-fold, while the contributions by R258 and G260 were less prominent. The latter two residues only showed marginal or no effect at all when substituted individually, but a 2-fold increase when present together. It is possible that glycine contributes to the affinity by positioning arginine, which has the potential for forming hydrogen bonds with the ligand. T281 significantly increased the selectivity for GalNAc when inserted into MGR simultaneously with R258 and G260, but is more likely to play an indirect role for ligand binding as it is predicted to be positioned at a considerable distance (10 Å). The residue appearing most important for achieving GalNAc binding is a histidine (His, H), found in both RHL-1 and MGR at position 256 and 278 respectively. Initial studies exchanging H278 in MGR for an alanine (Ala, A) resulted in an almost complete loss of GalNAc selectivity, but without any apparent effect on the galactose binding. The importance of the histidine was further investigated by repeating the substitution in RHL-1, causing a 25-fold loss in affinity for GalNAc, but without affecting that for Gal [7]. Corresponding residue in MBP, T202, was also substituted by histidine and resulted in a 14-fold increase in the relative affinity for GalNAc of the protein (referred to as the QPDWGH mutant) [49].

1.2.8 Sugar binding to H1-CRD

A model for sugar binding to H1-CRD has been proposed based on the mutagenesis experiments with MBP [8,49]. Crystallographic data of the QPDWG mutant could show that Gal and GalNAc bind directly to the Ca²⁺ in the second calcium binding site. The 3-OH and 4-OH groups of the sugar replaces two water molecules, normally coordinated by the calcium (*Figure 6A*). In addition, the same OH groups also forms hydrogen bonds with amino acid chains that are Ca²⁺ site 2 ligands (Q239, D241, E252, N264 in H1-CRD). Further interactions between the ligand and the protein are formed by stacking of the apolar face of the sugar against the side-chain of a tryptophan (W243 in H1-CRD) [8].

The GalNAc-specific MBP mutant QPDWGH was also analyzed by crystallography (Figure 6B). The results revealed direct interactions between the inserted histidine and the methyl group of the *N*-acetyl substituent of the sugar. The formed van der Waals contact serves as an explanation for the increase in GalNAc specificity exhibited by the mutant protein [49].

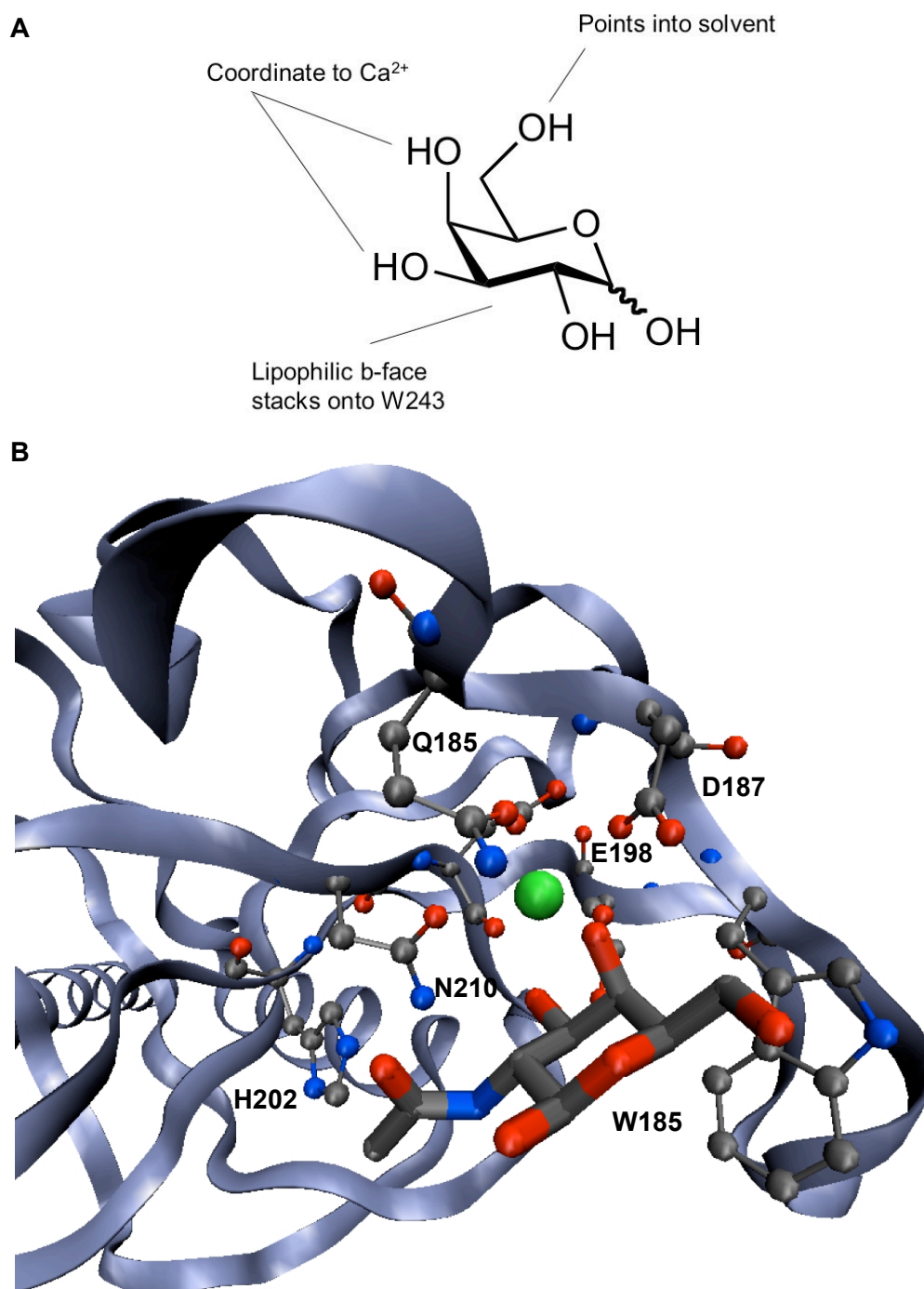


Figure 6. A) Schematic representation of interactions formed between Gal and H1-CRD. B) Ribbon diagram of the sugar binding site of MBP mutant QPDWGH (PDB entry code 1BCH) complexed with GalNAc [49]. Residues binding to the calcium ion (green) and the ligand are shown as ball-and-sticks [50].

1.2.9 ASGP-R – a candidate for targeted drug delivery

The ASGP-R presents a most appealing opportunity for targeted drug delivery to the liver. A recent review covers a number of different approaches for targeting liposomes and other carrier systems to the ASGP-R [3]. Labeling conventional liposomes with asialofetuin was seen to result in a significant increase of liver uptake, compared to unlabeled liposomes, after intravenous injection into mice. The enhanced uptake was most likely mediated by the ASGP-R [51]. Another study used glycolipids containing a cluster galactoside moiety for targeting to ASGP-R. The glycolipids were incorporated into liposomes at a ratio of 5% (w/w) and injected intravenously into mice. The liver uptake of the glycolipid-liposomes was estimated to exceed 80% compared to less than 10% for conventional liposomes 30 minutes after injection. Furthermore, pre-injection of ASF was seen to almost completely abolish the liver uptake of the glycolipid-enriched liposomes, implicating the ASGP-R as an active participant in the process [52].

To further aid the development of drug targeting to the liver aimed at ASGP-R it is important to understand the intricate relationship between structure and function of the receptor. Characterization of H1-CRD provides a good basis for expanding the knowledge and gaining a deeper understanding. Numerous methods can be employed in the pursuit of elucidating the relevance of structural elements or study protein-ligand interactions. The starting point for most methods is access to the protein and hence, recombinant protein production. A method for expression and purification of H1-CRD from *E.coli* was developed and reported in 2000 by Meier *et al.*, enabling structure determination of the subunit by X-ray crystallography [5].

1.3 Recombinant protein production

Pure, soluble and functional proteins are of high demand in modern biotechnology, e.g. for biochemical or three-dimensional analysis studies *etc.* Natural protein sources are rarely able to provide sufficient amounts, and even less in a cost-efficient manner, making recombinant protein production an appealing alternative [53]. Recombinant protein production involves several steps (*Figure 7*), starting with cloning of the gene into an expression vector, choosing an appropriate expression system and finally optimizing the protein production both in terms of expression yields and subsequent purification of the gene product [54].

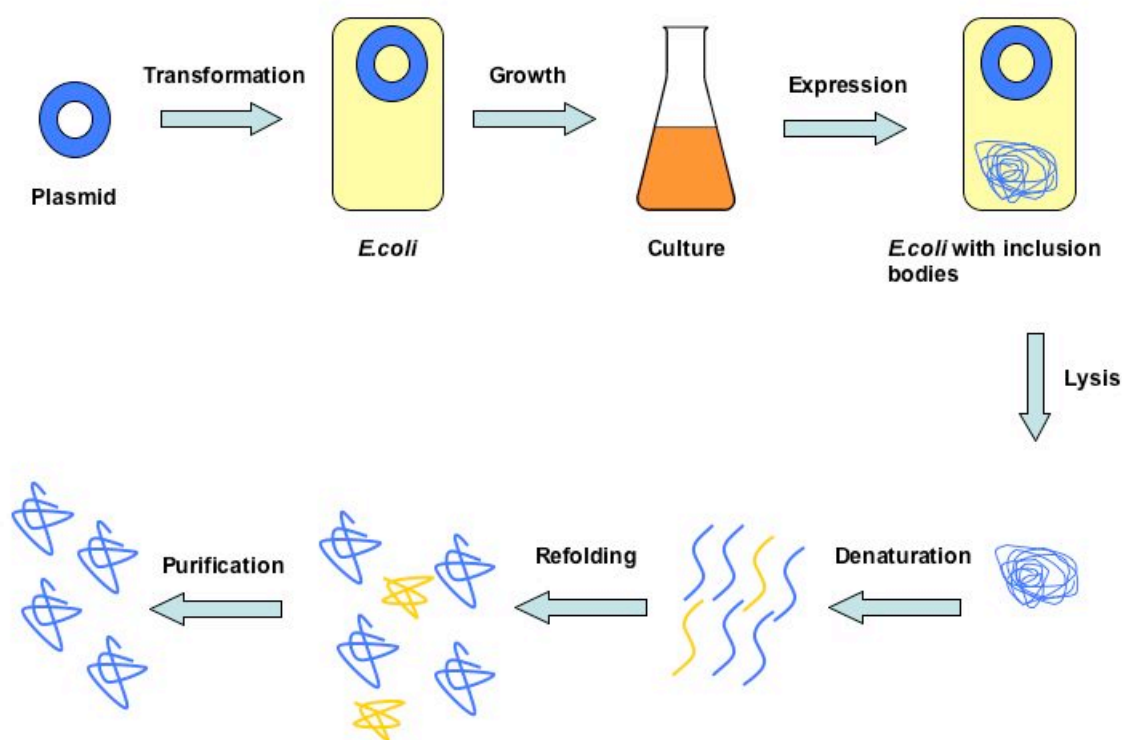


Figure 7. Schematic overview of recombinant protein production in *E. coli* [19]. The plasmid containing the target sequence is transformed into *E. coli* for expression. The expressed protein accumulates (most often) in the cells as inclusion bodies, which are released by lysis. In order to recover the protein and restore its native structure, the inclusion bodies are denatured and the protein is refolded. Finally, the protein is purified to remove contaminant bacterial proteins.

E. coli was the first organism used for production of recombinant proteins and still remains one of the most common expression systems [55]. Part of the bacterium's appeal surely lies in its ability to grow rapidly and at high cell densities on inexpensive substrates, its well-characterized genetics and the availability of an increasingly large number of cloning vectors and mutant strains [56]. Several informative and most useful reviews covering different aspects of the use of *E. coli* as expression system and offering advice on how to achieve high-level production have been published during the years [53,54,57,58]. This section will also focus on *E. coli* as it is the expression host used within this work.

1.3.1 Starting at the DNA-level

The gene of interest is initially cloned into an expression plasmid. For efficient expression of the target gene a number of structural elements has to be considered. The

most fundamental are a transcriptional promoter, a ribosome binding site (RBS) and a translational as well as a transcriptional terminator [53,58]. A promoter should be strong, resulting in high yields of the desired protein, but simultaneously showing low basal expression levels in the absence of an inducer. Transcription can be regulated through the presence of a repressor, encoded by a gene either integrated in the host chromosome or present on the vector itself. In addition, simple and cost-effective induction is desirable [57,58]. The RBS includes the Shine-Dalgarno site, which interacts with the 3' end of the 16S ribosomal RNA during translational initiation of the transcribed messenger RNA (mRNA). A translational terminator is found at the end of the gene encoding sequence, termination efficiency increasing with consecutive stop codons. Finally, a transcription terminator located downstream of the target sequence not only puts an end to the transcription, but also stabilizes the mRNA as an added benefit [58,53].

Additional elements not directly involved in the gene expression, but important for propagation and maintenance of the plasmid within the host cell are origin of replication and antibiotic resistance markers. The origin of replication controls the plasmid copy number, which in turn can influence the expression levels [58]. High-copy number vectors (e.g. pBR322, 15-60 copies) are often chosen to achieve maximum gene expression and might appear superior, but there are also drawbacks. Large amounts of foreign protein can prove to be toxic for the host cell, favoring use of medium- to low-copy number plasmids. Furthermore, protein aggregation can occur as a result of high expression levels, but can be prevented if the amount of protein is sufficiently reduced [54]. Antibiotic resistance markers are a valuable feature of plasmids. By conferring resistance to the host cell, the markers aid selection and propagation of the plasmid. Newly transformed bacteria are plated on selective agar-plates, allowing for growth only of the cells that were successful taking up the plasmid. Commonly used antibiotic markers are ampicillin, kanamycin, chloramphenicol and tetracycline. Ampicillin present in cultivation medium is especially susceptible to degradation. Plasmid evoked resistance towards ampicillin is accomplished by expression of β -lactamase from the *bla* gene. The enzyme is secreted into the periplasm, where it catalyzes hydrolysis of the antibiotic. Ampicillin can also be degraded by acidic conditions in high-density cultures, a risk that can be circumvented by using the less degradation prone ampicillin analog carbenicillin [53].

1.3.2 The pET expression system

There are several expression systems available for production of recombinant proteins, differing in their use of promoters and applications *etc.* One of the most widely used systems is the T7 based pET expression system, commercialized by Novagen [59]. The system makes use of T7 RNA polymerase, which is highly selective for its own promoters and shows a high rate of transcription. The T7 promoters, present on pET plasmids, do not normally occur in *E.coli*, nor are they recognized by *E.coli* RNA polymerase. Expression is achieved exclusively by T7 RNA polymerase, which can be provided by induction of the T7 *gene 1*. Host cells that are lysogenes of the bacteriophage DE3 carry a DNA fragment containing the T7 *gene 1* under the control of the *lacUV5* promoter, and also the *lacI* gene. The *lacI* gene gives rise to LacI monomers, forming tetramer complexes. The tetramers inhibit transcription of the T7 *gene 1* and hence, protein expression. When administered, isopropyl- β -D-thiogalactopyranoside (IPTG) binds the tetramer complexes, thus the T7 *gene 1* is transcribed and the T7 RNA polymerase can initiate transcription of the target gene, see figure 8. T7 RNA polymerase is so selective and active that under optimized conditions essentially all of the resources of the cell can be converted to target gene expression, resulting in yields as high as up to 50% or more of the total cell protein [60,53].

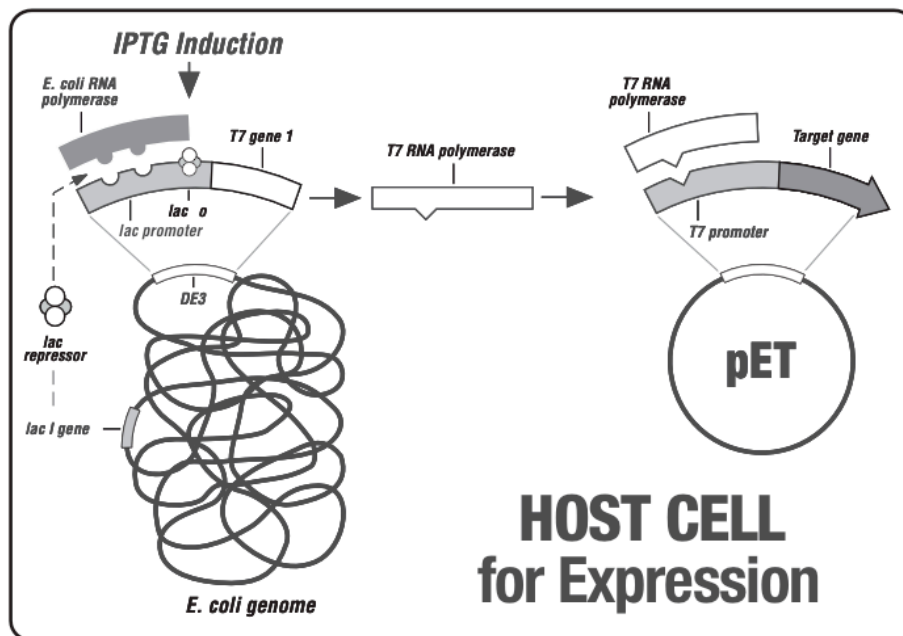


Figure 8. Protein expression initiated from the T7 promoter in a pET plasmid. The host cell, a DE3 lysogen, carries a copy of the T7 *gene 1* under the control of the *lacUV5* promoter. Addition of IPTG to a growing culture will induce T7 RNA polymerase production, which initiates the protein expression [59].

Adding to the advantages of the pET system are the many different vectors available, designed and developed for a variety of expression applications. The pET plasmids offer the option of different cloning sites and allows for incorporation of different fusion partners and tags [53,59].

1.3.3 Protein expression

The choice of expression strain can influence the protein production, both in terms of protein quality and yield. In general, expression strains should be deficient in the most harmful natural proteases and maintain the expression plasmid at stable levels. Depending on further requirements, various *E. coli* strains have been genetically engineered for different purposes such as promotion of cytoplasmic disulfide bond formation (e.g. Novagen Origami and AD494), enhancement of expression of proteins containing rare codons (e.g. Novagen Rosetta) or soluble expression of membrane or aggregation prone proteins (e.g. Avidis C41(DE3) and C43(DE3)) etc [53]. The most commonly used expression host is *E.coli* BL21 from Novagen, derived from *E.coli* B and as such deficient in *lon* and *ompT* proteases that can degrade proteins during purification [59].

Cultivation conditions should be carefully considered and optimized to achieve high expression levels of the target protein. The culture medium composition constitutes one important factor subject to optimization as it can influence both cell growth and protein production. The expression temperature is a second parameter worth investigating. Naturally, the final expression levels will also be dependent on the expression time [58,61].

Recombinant protein expressed intracellularly by *E.coli* can either accumulate as soluble protein in the cytoplasm or precipitate and form inclusion bodies, *i.e.* large insoluble aggregates of misfolded and inactive protein. Aggregation can occur when the expression rate is high, resulting in an accumulation of folding intermediates, or due to insufficient folding abilities, e.g. the inability to form native disulfide bonds. In order to avoid inclusion body formation, different strategies can be attempted, including optimization of expression temperature and rate, use of different *E.coli* strains, co-expression of molecular chaperones or use of affinity tags or fusion partners to increase the protein solubility etc [62]. However, there are also advantages

associated with the inclusion bodies. Normally toxic proteins cannot harm the host cell when present in aggregates, often resulting in high expression levels. Concurrently, the protein itself is protected from degradation by proteases. Nevertheless, the necessity to purify and refold the protein into its native structure remains [63].

1.3.4 Protein purification

Expression is followed by isolation and purification of the target protein. Soluble proteins accumulated in the cytoplasm of *E.coli* are initially released by cell lysis, through chemical treatment or mechanical-disruption techniques, followed by purification e.g. using different chromatography methods. In contrast, proteins accumulated as inclusion bodies require additional solubilization and renaturation steps [63].

There exists no universal protocol for optimal recovery of recombinant protein from inclusion bodies, but there are guidelines listing general measures [64,65,63]. A common procedure is to first isolate the inclusion bodies by cell lysis and centrifugation, which is easily done due to the high density of the protein aggregates. A washing step, utilizing detergents or low concentration of denaturants, is sometimes included to remove contaminating host proteins that may have absorbed onto the hydrophobic inclusion bodies. Next, the inclusion bodies are solubilized by use of high concentration of denaturants, such as guanidinium chloride and urea. If the recombinant protein contains cysteine residues, a reducing agent should also be included in the solubilization buffer and added in slight molar excess to ensure complete reduction of all cysteines and to prevent non-native disulfide bond formation. Typical reducing agents are beta-mercaptoethanol and dithiothreitol. The native structure and function of the protein is restored by renaturation, realized by the removal of denaturant and reducing agent by dialysis or dilution [65]. The renaturation conditions have to be carefully optimized to avoid incorrect folding or protein aggregation of folding intermediates. Correct folding is generally favored at low protein concentration, but might not always be practical. Different strategies have been attempted to facilitate correct folding, e.g. presence of denaturants at low non-denaturing concentrations or use of additives that prevent aggregation and/or enhance the stability and solubility of folding intermediates *etc.* In addition, if the target protein contains disulfide bonds, it is advisable to include a redox system in the renaturation buffer, allowing for formation and reshuffling of disulfide bonds [64].

1.3.5 General concluding remarks

E.coli is one of the most widely used expression systems for recombinant protein production. Even so, there exists no universal fail-proof protocol on how to design an expression set-up to obtain optimal yields, but it has to be determined empirically for each individual protein. Similarly, protein purification and refolding protocols also have to be developed on a case-by-case basis. If the final product cannot be recovered in sufficient amounts due to inefficient purification or refolding an alternative expression system, e.g. mammalian cells, ought to be considered. The same is true if post-translational modifications are required for the biological activity [63].

1.4 Site-directed mutagenesis

Site-directed mutagenesis (SDM) has proved to be of great value in the study of proteins. Its power lies in its ability to change a specific DNA target in a predetermined way, and consequently the gene product can be altered. As is known, the function of any given protein is highly dependent on its three-dimensional structure, which in turn is determined by the primary sequence. Therefore, SDM not only enables investigation of the protein structure and function, but also of the relation in-between [66,67].

SDM can be summarized as any of various techniques used to create specific alterations in the sequence of a gene. Such alterations do not merely refer to point mutation(s), which is the focus of this section, but also entail insertion or deletion of nucleotides. At the outset, SDM was performed with a single-stranded template (created by subcloning in M13), an oligonucleotide primer carrying the mutation and a thermolabile DNA polymerase. Although able to produce the desired mutation, this method was often time-consuming and not very efficient as also the wild-type sequence was produced. The introduction of the polymerase chain reaction (PCR) has therefore been of outstanding importance for the further development and improvement of SDM *in vitro*, making it a highly efficient process. Today, several techniques based on PCR are used for the creation of point mutations, a few of which are described below [67,68].

Regardless of the method used, four essential components are required in an *in vitro* DNA mutagenesis reaction. First, the template, *i.e.* the DNA sequence to be mutated, must be provided, and can be double- or single-stranded, linear or circular. Second,

primers of which at least one carries the mutation have to be included. The design of the primers is important to ensure successful mutagenesis. In general, it is recommended to place mismatches in the middle of the primer, but for some methods it is preferable to put the mutation at the 5'-terminus. Hence, the primer design should be carefully considered for each new method to be used. Third, building blocks in the form of the four dNTPs have to be supplied to enable DNA synthesis. Finally, a DNA polymerase, realizing the DNA synthesis reaction, is necessary. DNA polymerases used in PCR applications have to be thermostable to remain active under the high temperatures used during the reaction. As an example, denaturation of the template DNA is carried out around 95°C. Furthermore, when choosing a polymerase, the fidelity rate should be considered. Some polymerases, e.g. *Taq*, are more prone to cause unintended and non-desirable mutations compared to others, e.g. *Pfu* or *Vent*. The low fidelity of *Taq* is partly due to lack of 3'-5' exonuclease activity that "proofreads" and corrects sequence errors caused during DNA synthesis. As a final point, buffers and various other conditions also contribute to the success and fidelity of the DNA mutagenesis reaction. Following the mutagenesis reaction, the PCR product can be cloned into a vector for subsequent transformation and characterization [67].

1.4.1 PCR-based mutagenesis strategies

PCR was applied in SDM almost immediately after its establishment in 1985 and since then a number of techniques have been developed. Initial conditions, special requirements and the purpose of the mutagenesis reaction serve as guidelines when choosing a method. Three strategies used within this work are presented in the following sections.

1.4.1.1 Conventional PCR introducing mutation into the terminus of a sequence

If the intended mutation is located close to either end of the DNA sequence, it can be introduced by a conventional PCR set-up. Two primers are used, complimentary to the termini of the sequence, with the exception of the desired mutation(s) encoded by one of the primers (*Figure 9*).

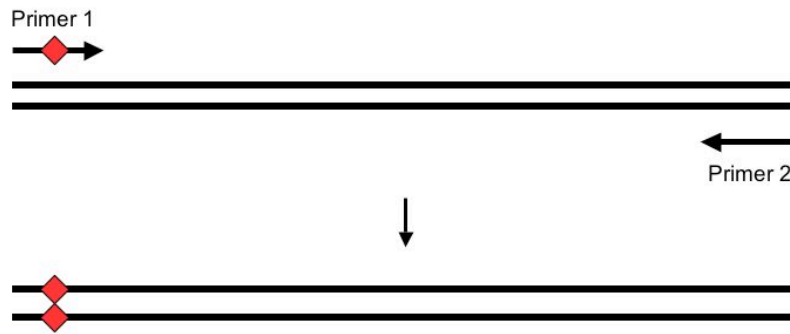


Figure 9. Introducing a point mutation into one of the termini. Primer 1 carries the mutation (marked as ♦), while primer 2 is completely complimentary to the second terminus of the target sequence.

The PCR reaction is run according to standard procedures, producing the mutated DNA. A pair of restriction sites is also usually built into the two primers, enabling subsequent cloning of the PCR product into a suitable vector [69].

1.4.1.2 Overlap extension PCR

To create mutations in the middle of a DNA sequence requires a different approach. One commonly used technique is overlap extension PCR (OE-PCR), consisting of two individual reactions. In a first step, two separate PCR reactions are performed, generating two fragments with overlapping sequences containing the mutation. The two fragments are produced by pairing a flanking forward primer with an internal reversed mutagenic primer and vice versa. In the second step, the two fragments are combined in a PCR reaction and the flanking primers are added. The overlapping sequences will allow annealing of the two fragments, giving rise to two heteroduplexes. Only the heteroduplex carrying two 3' termini at the joint can be extended by the polymerase to form full-length double-stranded DNA. The extended double-stranded DNA is further amplified by the flanking primers [70,67], see figure 10.

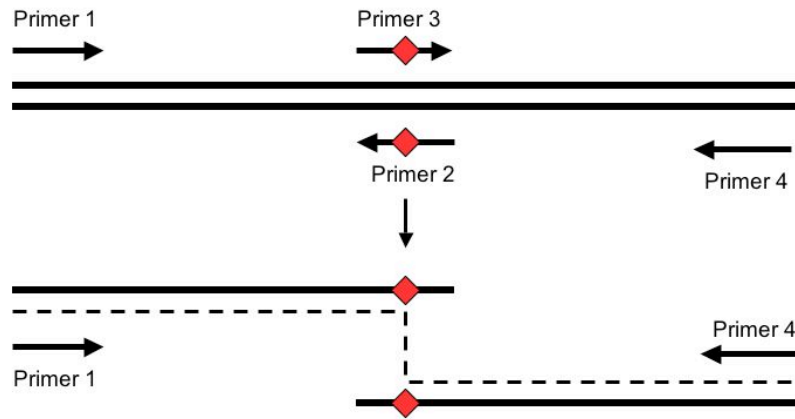


Figure 10. Introducing a point mutation into the middle of a sequence using overlap extension PCR. Primer 1 and 4 are completely complementary to the termini of the sequence, while primer 2 and 3 carry the mutation, marked as \blacklozenge . In a first step, two reactions are carried out simultaneously, generating two overlapping halves of the target sequence. In a second step, the combined fragments will anneal and extend to full-length double stranded DNA. Addition of the flanking primers will promote amplification of the DNA.

1.4.1.3 Inverse PCR

A third example of how to introduce mutation(s) is the inverse PCR method. The target sequence is located in a plasmid and a pair of tail-to-tail primers are used, the mutation located in one of the two. During the PCR reaction, the entire plasmid will be amplified by the primers, resulting in a linear double-stranded PCR product (Figure 11) [71].

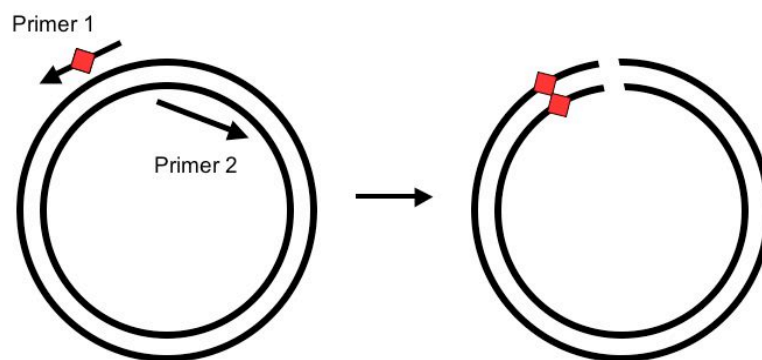


Figure 11. Introducing a point mutation into a sequence using inverse PCR. A pair of tail-to-tail primers are used, where primer 1 carries the mutation (marked as \blacklozenge). The primers will amplify the entire plasmid during the PCR reaction, while simultaneously introducing the mutation. The produced double-stranded linear DNA is ligated prior to transformation into *E.coli*.

The product is subsequently circularized through ligation and transformed into *E.coli*. Worth mentioning is that the ligation requires phosphorylated 5'-termini of the PCR products to succeed. This can be accomplished either by phosphorylating the primers in advance or the PCR products afterwards. Inverse PCR suffers from difficulties to separate the mutated product from the wild-type template, often leading to transformation of cells with a mixture and hence, resulting in lower mutagenesis efficiency [71,67].

1.4.2 Planning a mutation

The choice of a mutation is often based on previous knowledge about the protein under investigation, e.g. X-ray crystallographic data or genetic studies. Regions assumed to be important for the protein function, stability, pH tolerance or other properties are examined by changing strategically chosen residues [66]. Similarly, the mutant residue into which the wild-type amino acid is converted is also tactically selected to yield maximum information. As an example, cysteine residues probed by mutagenesis studies are often exchanged for alanine and/or serine residues [72,73]. While replacement with alanine conserves the hydrophobic character, a cysteine to serine conversion gives a relatively small change in size of the polar amino acid side chain. SDM is particularly useful for comparison of proteins, which might be related but with different functions. In such an event, the mutations are simply chosen on basis of the sequences differences to deduce the amino acids responsible for any dissimilar behavior.

1.4.3 Alanine-scanning mutagenesis

Alanine-substituted proteins can be used for effective identification of amino acid residues important for protein function, stability and structure. Conversion into alanine will result in the removal of all side chain atoms past the β -carbon. Systematic alanine replacement also referred to as alanine-scanning mutagenesis, will therefore make it possible to conclude the role of side chain functional groups at specific positions [74]. Alanine is the substitution residue of choice for several reasons. It does not change the backbone conformation, nor does it impose any extreme electrostatic or steric effects. Like alanine, glycine also invalidates the side chain, but the latter amino acid is not a good option as it could introduce conformational flexibility into the protein backbone [74,75]. Nonetheless, even though alanine-scanning mutagenesis is able to provide a detailed map of important functional residues, it is a very laborious method. Each

alanine-substituted protein must be separately created, expressed and purified before the loss of the side chain functionality can be evaluated.

1.5 Receptor-ligand binding assays

The interactions between a ligand and a receptor can be studied and evaluated by means of different assay formats. It is a central feature in drug discovery as well as in basic research for the determination of the behavior and biological function of a protein. Due to the wealth of assay technologies, it is possible to measure ligand binding under a multitude of conditions, e.g. in solution or in solid state, using isolated receptors or whole cells or organs, with or without labeling of the receptor or ligand, using fluorescence or enzyme-induced color development [76]. This section will introduce two assay formats used within this work.

1.5.1 A solid-phase competition assay

A solid-phase competition assay was developed within a diploma thesis by Daniela Stokmaier at the Institute of Molecular Pharmacy (IMP) (Pharmacenter, University of Basel, Switzerland), with the purpose to measure interactions between the ASGP-R H1-CRD and potential ligands [77]. The assay is based on a similar assay for E-selectin [78] and makes use of polyacrylamide-based glycoconjugates described by Bovin [79].

The assay consists of three basic steps (*Figure 12*). In the first, the protein is coated onto a 96-well plate. Second, a biotinylated polymer carrying sugar residues, in this case Gal or GalNAc, is added to the protein together with the substance under investigation. The substance is prepared as a dilution series while the polymer concentration is kept constant. The polymer is precomplexed with streptavidin-peroxidase, allowing for detection and evaluation of the binding, which is the third step. A substrate for horseradish peroxidase is added, here 2,2'-azino-di-[3-ethylbenzthiazoline-6-sulfonic acid] (ABTS), and a colored product is formed. Absorbance is measured at 415nm, quantifying the degree of binding [77].

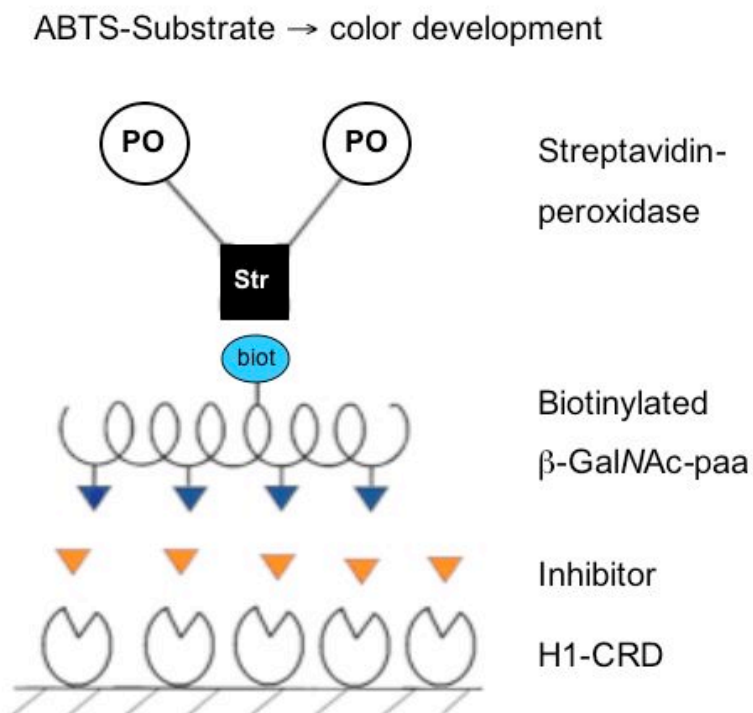


Figure 12. Schematic representation of the solid-phase competition assay. The protein is coated onto a 96-well plate. Inhibitor substance is added simultaneously with a biotinylated β -GalNAc-polymer precomplexed with streptavidin-peroxidase. ABTS substrate is added and an emerald green color appears.

The assay set-up using a constant concentration of protein and polymer while varying the inhibitor concentration permits the construction of an inhibition curve. An IC_{50} value, *i.e.* the substance concentration that inhibits 50% of the maximum polymer binding, can then be determined from the curve.

1.5.2 Surface plasmon resonance (SPR)

Surface plasmon resonance can be used to monitor molecular interactions in real-time and determine the equilibrium dissociation constant, K_D . The method relies on the measurement of changes in refractive index and resonance angle, which can occur upon ligand binding [80,81,76]. The principle components of Biacore, an SPR-based instrument, are a sensor chip, a flow cell, a light source and a detector. A schematic overview of the set-up can be seen in figure 13.

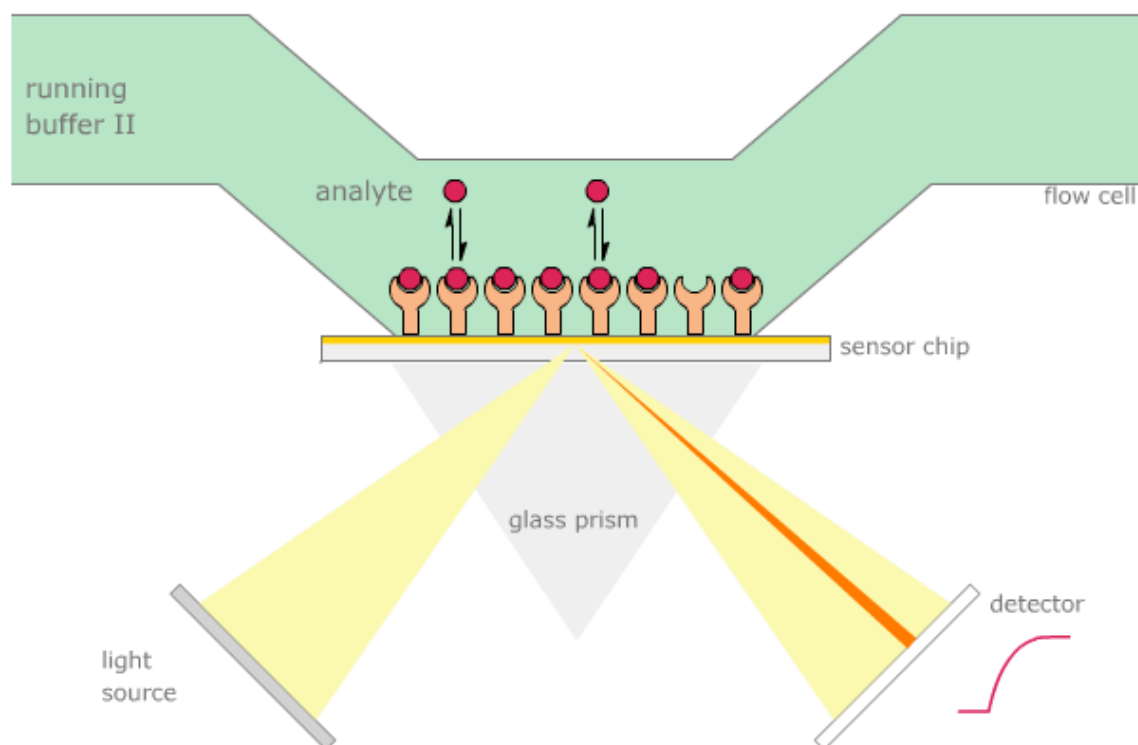


Figure 13. Experimental setup of a Biacore instrument. A target molecule is immobilized on the sensor chip. An analyte is injected into the flow cell and its binding is detected through changes in resonance angle [82].

In short, the protein (or the ligand) is immobilized on a modified gold surface, *i.e.* the sensor chip, which is illuminated with polarized light. Following immobilization, the ligand (or the protein) is injected into the flow cell, running over the sensor chip. Association between the receptor and the ligand will result in an increase in mass and hence, also a change in refractive index and a change in the angle of reflected light. Once the injection is stopped, washing continues with running buffer and the ligand will dissociate from the receptor. Anew the signal changes as the mass decreases and the angle of reflected light is restored to its original position. The shifts in resonance angle are recorded by the detector, and can be plotted in dependence of time. The signal-to-time plot, referred to as a sensogram, will visualize the binding behavior of the ligand, illustrated in figure 14 [81,76].

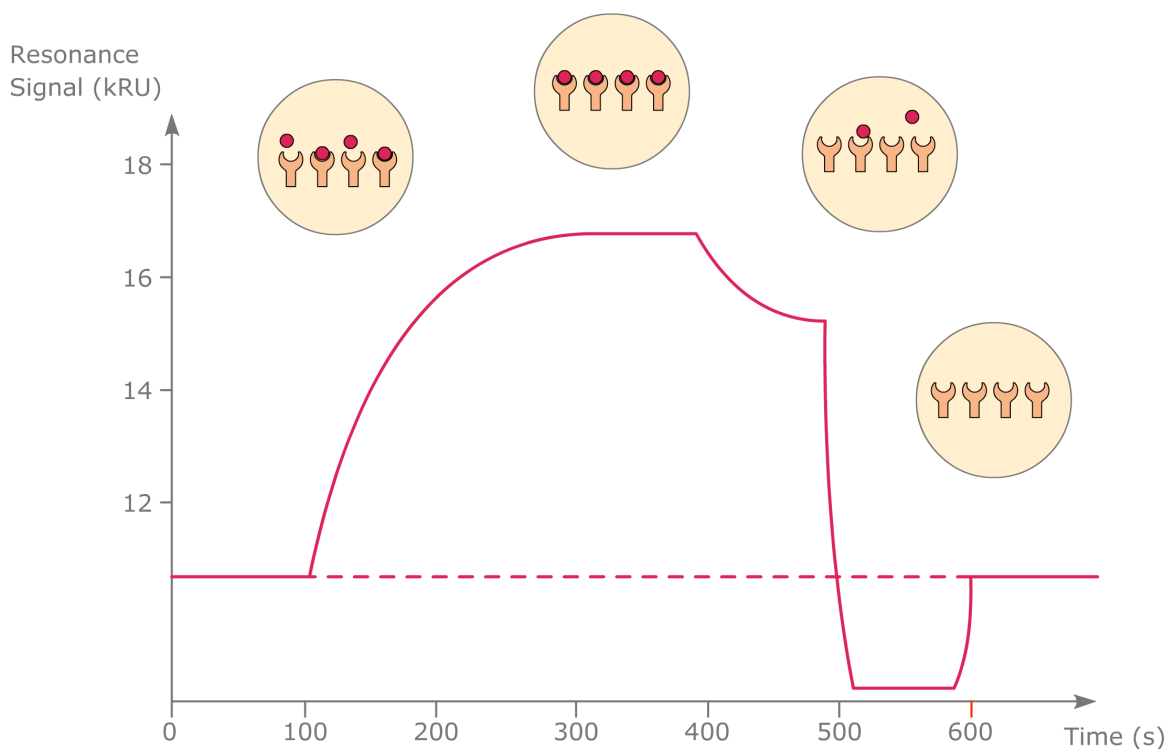


Figure 14. Schematic representation of a sensogram and the different stages of a binding event. Upon injection of the analyte, the signal increases as binding occurs until steady state is reached and the curve flattens out. Once the injection is stopped, running buffer will replace the bound analyte and the signal decreases. Some assays require an additional regeneration step to reach the baseline again [82].

Parameters such as the association and dissociation constants (k_{on} and k_{off} , respectively) can be derived from the binding curve. The equilibrium dissociation constant can be calculated by dividing k_{off} with k_{on} .

1.6 Isotope labeling of recombinant proteins

The need for proteins labeled with stable isotopes became apparent in parallel to the development of biomolecular NMR. There are many applications for NMR in protein studies, one of the most prominent being structure determination. Compared to X-ray crystallography, NMR holds the advantage that proteins can be studied in solution under native-like conditions and hence, no crystallization is needed. It also offers the possibility to monitor protein-ligand interactions [9], which is discussed further below. Still, NMR suffers from drawbacks, the two main problems being signal overlap and low signal-to-noise ratio. Both problems become increasingly acute as the size of the study object increases. These hurdles are partly surmounted by labeling of the protein with stable isotopes such as ^{13}C and ^{15}N [83].

The most abundant isotopes of carbon and nitrogen, *i.e.* ^{12}C and ^{14}N , cannot be used for NMR measurements as they either lack a net nuclear spin or, in the case of ^{14}N , are unable to produce strong, useful signals. In contrast, the less common isotopes ^{13}C and ^{15}N exhibit nuclei specific signals, making them suitable for NMR measurements and enabling multidimensional heteronuclear NMR experiments. In short, by labeling a protein with ^{13}C and/or ^{15}N , one or more additional dimensions are added to the spectra, further separating the NMR signals. Consequently, the signal overlap decreases, allowing for proteins on the order of 20 kDa to be studied by NMR when double-labeled with ^{13}C and ^{15}N [84,85].

1.6.1 Uniform labeling

For consistency and of relevance to this work, isotope labeling using *E.coli* as expression host is addressed in this section. Uniform labeling of recombinant proteins, *i.e.* labeling of all nitrogens and/or carbons with stable isotopes, can be achieved by growing *E.coli* cells and inducing protein expression in medium containing ^{13}C and/or ^{15}N . Rich medium containing isotope labeled nutrients is commercially available and represents one possibility, but can be rather expensive when used repeatedly or for large-scale production. Therefore, it is common procedure to grow *E.coli* in so-called “chemically defined” or “minimal” medium, where only one or a very limited number of carbon and nitrogen sources, isotopically labeled, are provided. The bacteria are forced to produce all metabolites themselves; including amino acids, and as a result the isotopes are incorporated into the recombinant protein during expression [86,87].

The composition of minimal medium can vary depending on the requirements of the expression strain. Nevertheless, commonly used standard or modified versions of M9 minimal medium [88] are based on a number of general and essential components. The M9 salts, supplying phosphorus atoms, consist of potassium dihydrogenphosphate, disodium hydrogenphosphate and sodium chloride. Glucose and/or glycerol often constitute the carbon source, while nitrogen is supplied in the form of ammonium chloride or ammonium sulfate [86]. Finally, magnesium sulfate, calcium chloride and a B-vitamin mixture are added to provide the bacteria with essential salt and nutrients. Some media also include a trace element solution, consisting of metal ions such as Co^{2+} , Cu^{2+} , Fe^{2+} , Zn^{2+} *etc.*, for the benefit of the bacteria [89].

Even though the production of isotope labeled proteins can appear rather straightforward, it can be both costly and time-consuming. Efforts are constantly being made to improve the yields and reduce the costs. One alternative approach to achieve efficient isotope labeling of recombinant proteins was presented by Marley *et al.*, in 2001 [90]. It is a two-stage protocol that generates cell mass using unlabeled rich medium followed by exchange into an equal or smaller volume labeled minimal medium. A short period of growth recovery and clearance of unlabeled metabolite is included before the cells are induced. This expression set-up is deemed both cost and time efficient. The isotope consumption is reduced as the majority of the cell mass is generated using unlabeled medium, while labeled minimal medium is introduced only at the stage of or shortly before induction. Furthermore, the growth rate of *E.coli* is generally faster in rich medium. Finally, by exchanging the cells into fresh medium prior to induction, byproducts, which can hamper both cell growth and protein expression, are removed. The volume of minimal medium may equal that of the starter culture or can be reduced to obtain higher cell density. The optimal cell density for induction has to be determined empirically for each protein [90]. A further suggestion for improvement of expression yields is addition of labeled rich medium to the minimal medium [91].

1.6.2 NMR as a tool to detect and investigate protein-ligand interactions

NMR is a valuable method for detection of protein-ligand binding, also without isotope labeled proteins. NMR experiments based solely on protons can provide qualitative measures whether or not binding takes place. However, multidimensional NMR and the use of isotope labeled protein also permits quantitative measurements of the binding, and furthermore, can ascribe the interactions to specific residues or involved regions. By performing a ^1H - ^{15}N heteronuclear single quantum correlation (HSQC) experiment, a two-dimensional spectrum is obtained, ideally containing one cross-peak for each amide in the protein. Addition of the ligand can cause a change in the signal distribution, if binding takes place. The signals of those amides involved in binding and whose environments are changed will experience obvious chemical shift changes. If resonance assignment and a 3D structure are available, these changes can be mapped onto the protein structure, indicating the location of the binding site [92]. In addition, a similar set-up can be used to determine the equilibrium dissociation constant by titration of ligand to the protein [93].

1.7 Scope of thesis

The overall aim of this work has been to investigate functional and structural elements of ASGP-R H1-CRD in order to gain a deeper insight into the functionality of the subunit. The work encompasses three main projects with the different focuses and backgrounds presented below.

1.7.1 The role of the three cysteines at the *N*-terminus

H1-CRD contains seven cysteines, six of which are engaged in disulfide bridges. Based on the crystal structure [5], two of the bridges appear to be of great importance for the overall structure and functionality of the subunit. Disulfide bond Cys254-Cys268 contributes to the structure of the binding site, while the second S-S bond, Cys181-Cys268, brings the two termini together, shaping the overall fold of the subunit. In contrast, the third bridge, Cys153-Cys164, is more likely to be of limited importance and only influence the local structure at the *N*-terminus. Here it should also be mentioned that an alternative pairing of the third disulfide bond was indicated in an earlier doctoral thesis at IMP by Rita Born [94]. Results derived from MS analysis indicated a disulfide bond between Cys152 and Cys153. Nevertheless, the seventh odd cysteine (Cys152 according to the crystal structure) lacks a thiol-group to react with within the subunit. It can, however, react with a free cysteine in a second subunit and hence, could be the cause of H1-CRD dimer formation seen upon expression *in vitro* [94]. The dimerization presents a problem as it is neither possible to control, nor reproducible, giving rise to in heterogeneous samples of H1-CRD. A further disadvantage of the dimers is the uncertainty of how they behave and whether they exhibit different binding affinities compared to the monomers. For consistency, receptor-ligand assays such as Biacore and NMR, aiming to determine ligand binding affinity, should therefore always be performed with homogenous samples of monomeric H1-CRD.

A method to separate monomers and dimers of H1-CRD was developed within a doctoral thesis of Daniel Ricklin at IMP [95]. HPLC ion-exchange chromatography (IEC) was employed for this purpose, resulting in a clear and reproducible separation of the two species. However, analysis of the monomer samples was seen to show an ever-present fraction of dimers, most likely formed after the separation as the result of an

equilibrium forming between monomers and dimers. Therefore, merely separation of monomers and dimers is not sufficient to obtain pure samples of monomeric H1-CRD.

If the dimer formation is caused by the seventh odd cysteine, it can theoretically be inhibited by replacing the cysteine with e.g. a serine through SDM. All three cysteines closest to the *N*-terminus could be implicated in the intermolecular disulfide bonding, as proposed by the conflicting indications concerning the pairing of the third disulfide bond [5,94]. Hence, it seems reasonable to exchange all three cysteines by serines to ensure that dimer formation is completely prevented. A second concern that has to be addressed is the activity of a mutant subunit. Even if the dimerization is abolished by mutations, it is absolutely crucial that the protein maintains its native behavior and binding affinity to be of use in future studies. As an alternative approach, shortening the protein could also be considered.

1.7.2 Investigation of the binding site

The binding characteristics of H1-CRD, high specificity and affinity for Gal and GalNAc, have been attributed to a number of residues in the binding site. Out of the 33 amino acid long binding site, stretching from Arg236 to Cys268, the functional importance of the great majority of the residues have been explored and elucidated [4-7,49,96]. However, the studies stopped before a complete overview of the binding site could be established. Therefore, previously uninvestigated residues, lacking assigned function and relevance, were probed by SDM and subsequent expression and evaluation within this work. The residues under investigation were chosen with respect to location, functional group of the side chain and, occasionally, based on previous studies (*Figure 15*).

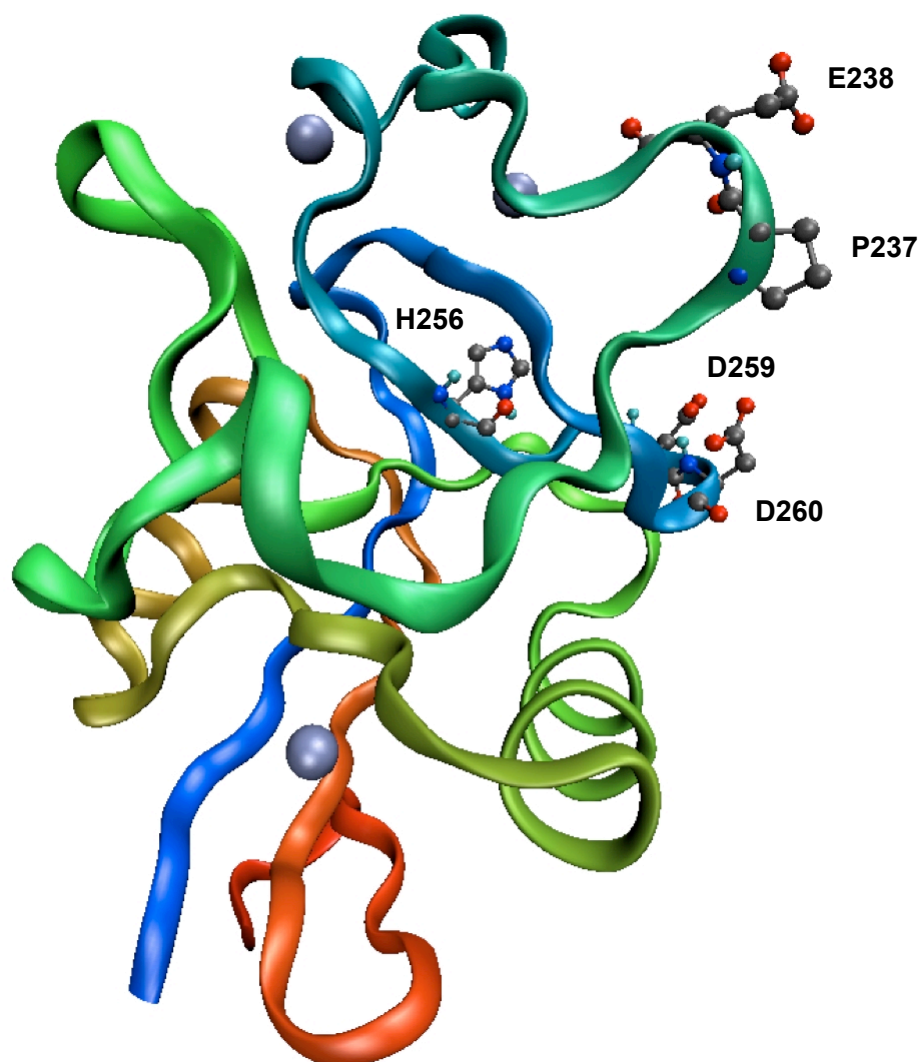


Figure 15. Ribbon diagram of H1-CRD. The residues that were mutated within the project are drawn as ball-and-sticks. The *N*-terminus of the backbone can be seen in red, while the *C*-terminus ends in blue [50].

Proline 237 and aspartate 260 are conserved residues in H1-CRD, whose functional contributions have not been investigated before. Both residues are conserved in H1-CRD of human, mouse and rat. In addition, Asp260 is also found in H2-CRD of the same species. In order to evaluate the significance of these residues, they were replaced by alanines.

Aspartate 259 was also substituted for alanine. The corresponding residue in RHL-1 is a threonine, which has been shown to contribute to the binding of GalNAc. As the threonine is predicted to be rather distant from the bound sugar, it is more likely to be

of structural than functional importance [7]. Investigations to elucidate the role of aspartate in this position have not been pursued up till now.

The influence of glutamate 238 in H1-CRD has not been previously determined. Mutagenesis studies have identified the corresponding residue in RHL-1, a glycine, as an important residue for the selectivity of GalNAc [7]. However, the distinction in activity between glutamate and glycine has not been established.

Histidine 256 has been revealed to be of crucial importance for the GalNAc selectivity of H1-CRD [7,49]. In addition, it has also been implicated as an important residue for the ligand release during endocytosis [96]. Mutagenesis studies have shown that His256 renders RHL-1 more sensitive to acidic pH and influences the ligand binding in a pH dependent manner. His256 is thought to act as a hydrogen bond (H-bond) acceptor for the amide group of asparagine (Asn) 264. Asn264 in turn interacts with the second calcium binding site and is also involved in the ligand binding. A drop in pH, as in the endosomes, will cause protonation of the histidine, making it unable to accept the hydrogen bond from Asn264. As a consequence, Asn264 is destabilized and will subsequently release the ligand [96,97]. The corresponding residue in H2-CRD is a glutamate (Glu), and as such it will be deprotonated in the endosomes. Glutamate is then capable of accepting H-bonds, stabilizing Asn264 and sustaining ligand binding. In order to confirm this hypothesis, His256 was exchanged for Glu as one of the binding site mutants created during this work. This subproject will serve two purposes. For one, it will further facilitate our understanding of the influence of pH on ligand binding to ASGP-R. Secondly, it also offers an intriguing insight into the differences in behavior and properties of the two subunits making up the receptor.

1.7.3 Isotope labeling of H1-CRD

The third project within this work has been to establish a method for isotope labeling of H1-CRD. It is the first step in an endeavor aiming to determine the structure of H1-CRD by NMR. The crystal structure of H1-CRD has been resolved by X-ray [5]. Still, X-ray crystallography does not exclude the use of NMR or vice versa. As previously noted, NMR can further improve the understanding of H1-CRD by studying the protein in solution. An additional benefit is the possibility to study protein-ligand interactions, and by doing so determine the mechanism for ligand binding to the receptor subunit.

A relatively large amount of isotope labeled H1-CRD is required for the NMR studies. As a general rule of thumb, 2-8 mg protein is needed for structure determination by NMR [9]. Hence, a reproducible method for high efficiency isotope labeling of H1-CRD is necessary, ensuring a sufficient supply of protein.

1.7.4 Aim

In summary, in order to further improve the knowledge on H1-CRD, three projects have been undertaken, all aiming to elucidate the relationship between structural elements and functional behavior of the subunit. The following goals have been set for the work:

- To determine the relative importance of disulfide bridge 153-164 and the odd cysteine 152 for the function of H1-CRD, as well as their involvement in dimer formation, by use of SDM.
- To investigate the contributions of residues in the binding site of H1-CRD to the functionality of the subunit using SDM. In addition, the influence of histidine vs. glutamate, in position 256, on the pH-dependent ligand binding should be determined.
- To establish a method for high efficiency isotope labeling of H1-CRD

2 Material and Methods

2.1 Material and reagents

Double-distilled Millipore water was used for the preparation of all buffers and solutions unless otherwise specified.

E.coli strains DH5 α and Top10 were originally from Invitrogen, while AD494(DE3) and BL21(DE3) originated from Novagen. *E.coli* XL-Blue 1 was purchased together with the ExSite PCR-based site-directed mutagenesis kit from Stratagene. Bacterial medium components such as Bacto-tryptone, Bacto-yeast extract and Bacto-agar were acquired from Becton Dickinson BD. Ampicillin sodium salt from Applichem, Carbenicillin disodium salt from Sigma and Kanamycin monosulfate from Fluka was used. IPTG was from Promega. ¹⁵N ammonium chloride (>99%) and ¹³C-D-Glucose (>98%) were purchased from Spectra Stable Isotopes. Thiamine hydrochloride and BME vitamins solution 100X were both purchased from Sigma.

Sepharose® 6B for chromatography was acquired from Sigma. Acrylamide 30% solution was purchased from Applichem. Sodium dodecylsulfate (SDS) from Biorad, N,N,N,N,'- Tetra-methylene-diamine (TEMED) from Serva and ammonium persulfate (APS) >98% for electrophoresis from Sigma was utilized for SDS polyacrylamide gel electrophoresis (PAGE). Coomassie Brilliant Blue G-250 was purchased from Biorad. Poinceau S solution and anti-chicken-IgG coupled with alkaline phosphatase (AP) were both purchased from Sigma. Nitro-blue tetrazoliumchloride, 5-bromo-4-chloro-3-indolylphosphate toluidine (NBT/BCIP), 67% in DMSO was from Fluka. Protein standard of bovine serum albumin (BSA) 1 mg/ml and SigmaMarker™ Low range were bought from Sigma.

The restriction enzymes *Bam* H1 and *Nde* 1 from New England Biolabs (NEB) were used together with 10X *Bam* H1 buffer and 100X BSA from the same supplier. λ -DNA was also acquired from NEB. DNA gel loading buffer 10X was purchased from Eppendorf. DNA molecular weight marker X and alkaline phosphatase enzyme with a 10X buffer was bought from Roche. The ExSite PCR-based site-directed mutagenesis kit (including rATP), *Pfu* turbo polymerase and REDTaq™ DNA polymerase were

purchased from Stratagene. *Pfu* polymerase and T4 DNA ligase with buffers were purchased from Promega. A 10 mM stock solution of dNTPs was made from dATP, dCTP, dGTP and dTTP from Roche. Big dye terminator v1.1. ready reaction kit was bought from Applied biosystems.

The β -GalNAC-polymer was purchased from Lectinity and streptavidin-peroxidase came from Roche Applied Sciences. ABTS peroxidase substrate kit was bought from Biorad. D-Galactose was from Senn Chemicals AG and *N*-acetyl-D-galactosamine from Acros.

2.2 Equipment

Incubation of liquid cultures was performed in Innova 4000 incubator Shaker from New Brunswick Scientific, while bacteria on plate were incubated in an incubator from Memmert. Bacteria cultures were prepared in a VS 120 AFY laminar flow from Skan. Cell cultures were centrifuged in a Sorvall RC-5B Refrigerated Superspeed Centrifuge using a GS 3 Rotor. Cultures of competent cells were harvested in centrifuge 5804R from Eppendorf. Finally, small culture samples (<2 ml) were centrifuged in Biofuge 13 from Heraeus Sepatech. Cell density was measured with SmartSpec™ 3000 spectrophotometer from Biorad.

Ultrasonication was carried out using a VibraCell sonication apparatus from Biorad unless otherwise specified. Ultracentrifugation of the protein suspensions during the purification procedure was done with a centrifuge T-1075 from Kontron instruments. ZelluTransRoth 6.0 dialysis tube with a 10 kDa cutoff was purchased from Roth. The FPLC-system BioLogic duo-flow version 3 and Bio-Scale MT2, MT10, MT20 columns from Biorad were used for affinity chromatography. The second chromatography system used within this work was the HPLC system Agilent 1100 purification system from Agilent. A Shodex IEC DEAE-825 column from Brechbühler AG and a C18 column from Vydac were utilized with this system. A mini-Protean 2 system from Biorad was employed for SDS-PAGE. The TransBlot semi-dry transfer cell used for immunoblotting also came from Biorad. A heatblock of model Thermomixer comfort from Eppendorf was used for heating and incubation of samples volumes <2 ml. The Spectramax 190 plate reader from Bucher Biotec was employed for absorbance measurements.

GFX microprep plasmid kit was purchased from Amersham Biosciences, Qiagen Plasmid Maxi kit from Qiagen and Perfect Prep Clean-up kit from Eppendorf. Electroporation was carried out with Gene/*E.coli* pulser® cuvettes and a micropulser electroporator, both from Biorad. PCR reactions were performed in the iCycler 1.259 thermal cycler from Biorad. Small sample volumes (≤ 2 ml) were centrifuged in a 5415D centrifuge from Eppendorf, while larger volumes (up to 50 ml) were centrifuged in a 5804R centrifuge also from Eppendorf. Special 96-well plates, Nunc-Immuno™ Plate MaxiSorp™ Surface from Nunc™ Brand products were used for the solid phase competition assay.

2.3 General methods for working with *E.coli* as expression system

2.3.1 *E.coli* strains used for cloning and protein expression

E.coli strains eligible for cloning are strains with background expression kept to a minimum. Consequently, the stability of the plasmid increases since there is no production of proteins, which could be of potential harm to the host cell [59].

In this work, *E.coli* strains DH5 α and Top 10 from Invitrogen and XL-Blue 1 from Stratagene were used for cloning of H1-CRD and mutants. All strains lack the T7 RNA polymerase gene and have the genotype endA1 recA1, *i.e.* reduced endonuclease activity and recombinant deficiency, facilitating plasmid stability.

E.coli strains suitable for protein expression are lysogens of the bacteriophage DE3, *i.e.* they contain the gene for T7 RNA polymerase under the control of the *lacUV5* promoter. Addition of IPTG will lead to transcription of the T7 *gene 1*, yielding T7 RNA polymerase, which in turn will transcribe the target DNA in the plasmid and the protein is expressed [53].

For expression of H1-CRD, *E.coli* AD494(DE3) from Novagen was used. The strain AD494(DE3) holds an advantage as it contains a mutation for the enzyme thioredoxin reductase (*trxB*) and thereby promotes formation of disulfide bonds in the *E.coli*

cytoplasm. AD494(DE3) cells also harbor an internal resistance to kanamycin at 15 µg/ml.

The *E.coli* strain BL21(DE3) from Novagen was used for expression of isotope labeled H1-CRD. It is a general purpose expression host, lacking the proteases *lon* and *ompT*, which can cause degradation of proteins during purification. Due to the protease deficiency, target proteins should be more stable in BL21(DE3) cells compared to other expression hosts [59].

2.3.2 Bacterial growth and expression medium

The bacterial growth media and supplements used for the cloning and protein expression procedures within this work are listed below [88]. All medias, additives and antibiotics were prepared using Millipore water. Medias were autoclaved at 121°C for 20 minutes prior to use, while the antibiotics, susceptible to heat degradation, were filter sterilized using a Millex®GP 0.22 µm filter (Millipore).

LB medium, 1 liter

Bacto-tryptone	10 g
Bacto-yeast extract	5 g
NaCl	10 g
Adjust pH to 7.5	

LB-agar medium, approximately 20 plates

LB medium	500 ml
Bacto-agar	7.5 g
Cool to 50°C before adding antibiotics and pouring plates, ca. 25 ml per plate	

TB medium, 1 liter

900 ml:	
Bacto-tryptone	12 g
Bacto-yeast extract	24 g
Glycerol	4 ml
100 ml:	
KH ₂ PO ₄	2.31 g
K ₂ HPO ₄	12.54 g

SOB medium, 1 liter

Bacto-tryptone	20 g
Bacto-yeast extract	5 g
NaCl	0.5 g
KCl 250 mM	10 ml
Add 5 ml 2 M MgCl ₂ prior to use	

SOC medium, 1 liter

SOB medium with addition of 20 ml 1 M glucose

Ampicillin 1000X, 5 ml

Ampicillin 0.5 g

Carbenicillin 1000X, 5 ml

Carbenicillin 0.25 g

Kanamycin 1000X, 5 ml

Kanamycin monosulfate 0.075 g

IPTG 0.4 M, 5 ml

IPTG 0.476 g

The bacteria cultures were prepared under sterile conditions in a laminar flow. Autoclaved Erlenmeyer flasks were used as vessels, filled maximum to half the volume to allow for good aeration of the bacteria. Typically the cells were grown at 37°C and 300 rpm, varying the choice of medium and the time depending on the purpose of the culture.

2.3.3 Competent cells

In order to clone and later express H1-CRD, competent *E.coli* cells of the previously mentioned strains were prepared. Competence refers to the ability of *E.coli* cells to take up foreign DNA, a property the bacteria normally lack, but which the cells can obtain through special treatments. One can distinguish between electrocompetent and calciumcompetent cells [66].

2.3.4 Electrocompetent cells

Uptake of free DNA can be induced in electrocompetent cells by subjecting the bacteria to a high-strength electric field. The mechanism by which the DNA is taken up into the cells is not fully known. However, it is believed that the electroshock leads to formation of transient pores in the cell wall, through which the DNA can pass into the bacteria [66,68].

Electrocompetent cells were made of the bacterial strains *E.coli* DH5 α and Top10. Non-transformed cells were streaked out on non-selective LB-agar plates and incubated at 37°C overnight (O/N). A single colony was picked and grown in 5 ml SOB medium

at 37°C and 300 rpm O/N. The preculture was inoculated into 250 ml SOB medium with addition of 10 mM MgCl₂. The culture was grown at 37°C and 300 rpm until the cell density reached an OD value of 0.5-0.6. The cells were centrifuged at 5000 rpm and 4°C for 10 minutes and the pellet resuspended in an equal volume 10% glycerol. Centrifugation was repeated twice, followed by resuspension in 50 ml and 1 ml of 10% glycerol respectively. The final cell suspension was aliquoted as 50 µl in Eppendorf tubes and frozen in liquid nitrogen. The electrocompetent cells were stored at -80°C [68].

2.3.5 Calciumcompetent cells

An alternative method to induce bacterial uptake of recombinant DNA is the use of calciumcompetent cells. The cells are initially treated with ice-cold CaCl₂ and then exposed to a high temperature, most often 42°C, for approximately 90 s. The mechanism of action is not fully known for this procedure either, but it is assumed that the bacterial cell wall become permeable in localized regions, allowing plasmid DNA to enter into the cells [66].

Calciumcompetent cells were made of bacterial strains *E.coli* AD494(DE3) and BL21(DE3). A single colony of non-transformed bacteria was picked from a freshly streaked plate and grown in 4 ml LB medium O/N at 37°C and 300 rpm. 1 ml preculture was inoculated into 100 ml LB medium and the culture grown at 37°C and 300 rpm until a cell density of 0.4-0.5 was reached. The cell culture was placed on ice for 20 minutes, followed by centrifugation at 5000 rpm and 4°C for 10 minutes. The cells were resuspended in 10 ml pre-chilled 50 mM CaCl₂ and left on ice for 40 minutes. Centrifugation was repeated once and the cells resuspended in 1 ml pre-chilled 50 mM CaCl₂/20% glycerol. The final cell suspension was aliquoted as 80 µl in Eppendorf tubes and frozen in liquid nitrogen. The calciumcompetent cells were stored at -80°C [68].

2.4 General methods for working with recombinant DNA

2.4.1 Plasmid

The vector pET-3b from Novagen was used within this work. It belongs to the pET system, derived from the plasmid pBR322. The pET-3b plasmid is 4639 bp in total with

a *Bam* H1 cloning site. It carries the T7 promotor and confers resistance to ampicillin [98].

2.4.2 Agarose gel electrophoresis analysis

Gel electrophoresis is a commonly used technique to separate DNA fragments as well as proteins. Samples are loaded onto a gel and an electric field is applied, causing e.g. the DNA fragments in the samples to migrate through the matrix toward the anode. The distance and the speed with which the fragments will move is dependent on the size of the macromolecules and the pore size of the gel [66,68].

Agarose gel electrophoresis was used in this work to analyze isolated plasmids, PCR products and digested DNA fragments. By varying the percentage of agarose in the gel, the separation resolution could be influenced. A 1.0% agarose gel was prepared by dissolving 1.0 g agarose in 100 ml TBE buffer (44.6 mM Tris, 44.6 mM boric acid, 1 mM EDTA) during heating in a microwave. Once completely dissolved, the liquid agarose was allowed to cool down to approximately 50°C before adding 2 µl ethidium bromide 1% (w/v) to 80 ml agarose solution. The agarose was poured into an electrophoresis apparatus and combs forming wells were inserted. Once polymerized, the combs were removed and the gel covered with TBE buffer. DNA marker X from Roche was loaded as a reference. The samples were prepared with DNA gel loading buffer 10X and diluted with water if necessary. The gel electrophoresis was run at 80 V.

2.4.3 Restriction digestion

A prerequisite for recombinant DNA technology is the ability to cut the target DNA and the cloning vector into precise and reproducible fragments to enable further cloning of the gene. Restriction endonucleases are enzymes, which possess the capability to recognize and cut DNA molecules internally at specific base pair sequences, and have proven to be crucial for molecular cloning [66].

The sequence of H1-CRD, inserted into the cloning region of pET-3b between the *Bam* H1 and *Nde* 1 restriction sites, was kindly provided by Prof. Spiess (Biocenter, University of Basel, Switzerland). Restriction digestion was carried out according to one of the following set-ups, the choice depending on the sample volume (Table 1).

Table 1. Two alternative set-ups were used for the restriction digestion of DNA, the choice depending on the volume of the sample to be digested. Water was added in an adequate amount to reach the final volume specified for each set-up.

Digest set-ups	Set-up 1 (μl)	Set-up 2 (μl)
DNA sample	8-12	30/50
<i>Nde</i> 1	1	2
<i>Bam</i> H1	1	2
10X <i>Bam</i> H1 buffer	2	4/6
BSA 100X (1 mg/ml)	2	-
Water (sterile)	<i>q.s.</i>	<i>q.s.</i>
Σ	20	40/60

The buffer 10X *Bam* H1 from NEB was used as both *Bam* H1 and *Nde* 1 have been reported to exhibit 100% activity under these conditions. The samples were prepared by adding the restriction enzymes last. Digestion was carried out at 37°C in a heatblock without shaking for 4-5 hours, alternatively O/N for set-up 1.

2.5 Expression, purification and analysis of ASGP-R H1-CRD

2.5.1 Confirming the sequence of H1-CRD DNA

Prior to expression and purification of H1-CRD, the accuracy of the protein encoding DNA sequence was confirmed. Small-volume cultures consisting of 3 ml LB medium supplemented with 50 μg/ml carbenicillin and 15 μg/ml kanamycin were inoculated with *E.coli* AD494(DE3) cells carrying the construct pET-3bH1. The cultures were grown O/N at 37°C and 300 rpm. The following day, cells were streaked onto a freshly prepared LB-ampicillin agar plate and sent to Microsynth for sequencing.

2.5.2 Expression of WT H1-CRD

AD494(DE3) cells, carrying the construct pET3bH1, were grown in 500 ml TB medium supplemented with 1% glucose, 50 μg/ml carbenicillin and 15 μg/ml kanamycin at 300 rpm and 37°C. The culture was initially inoculated with cells grown O/N under the same conditions, aiming at an OD of 0.1 to start the growth. Expression was induced at OD ~ 0.8 by addition of IPTG to a final concentration of 0.4 mM, and lasted five hours. Cells were harvested by centrifugation at 4°C and 5000 rpm for 10 minutes. The supernatant was discarded and the cell pellets stored either at 4°C when followed by

purification the next day, or at -20°C when purification was carried out two or more days later.

2.5.3 Material for the purification of WT H1-CRD

The purification procedure of H1-CRD requires material such as dialysis tubes and a Galactose-Sepharose-column (Gal-Sep-column). These were prepared beforehand as described in the sections below.

2.5.3.1 Dialysis tubes

Dialysis tubes with a 10 kDa cutoff were prepared from ZelluTransRoth 6.0 dialysis tube. The tubes, 20 cm long, were gently stirred in 500 ml 10mM NaHCO_3 (pH 8), 1 mM EDTA, preheated to 80°C , during 30 minutes. The buffer was exchanged with water, decreasing the temperature stepwise to 60°C , 40°C and 25°C every 10th minute. The tubes were transferred into 1 mM EDTA and stored at 4°C [99].

2.5.3.2 Preparation of a Galactose-Sepharose column

40 ml Sepharose® 6B was divided into two Falcon tubes and washed with water by centrifugation at 3000 rpm for 5 minutes three times. 20 ml 0.5 M carbonatbuffer (pH 11) and 2 ml divinylsulfon was added to each tube, which was subsequently incubated for 1 hour at room temperature (RT) during light shaking. The sepharose was rinsed with water by filtration and divided into new tubes in equal amounts. 20 ml 20% D-galactose in 0.5 M carbonatbuffer (pH 10) was added to each tube and thoroughly mixed with the sepharose prior to incubation at RT under light shaking O/N. The sepharose was centrifuged at 3000 rpm for 5 minutes and washed once with water as described above. 20 ml 0.5 M carbonatbuffer (pH 8.5) and 400 μl beta-mercaptoethanol was added to each tube, followed by incubation for 2 hours at RT under light shaking. The sepharose was centrifuged at 3000 rpm for 5 minutes and washed with water twice as previously described. Finally the sepharose was added stepwise to the column, allowing the slurry to settle and ensure proper packing up to the top, before the column was fully assembled and closed (Fornstedtet al 1975). Sterile filtered water was run through the column at a flow rate of 2 ml/min for a minimum of 40 minutes, before switching to 20% methanol, using the same settings. The column was stored in 20% methanol at 4°C [100].

2.5.4 Solubilization and renaturation of WT H1-CRD

A cell pellet originating from 500 ml expression culture was resuspended in 25 ml 20 mM Tris-HCl (pH 8) with addition of 8 M urea and 0.1% beta-mercaptoethanol. Complete cell lysis was achieved by ultrasonication on ice, in intervals of 20 seconds sonication and 10 seconds stop, for a total of 30 minutes. The protein suspension was centrifuged at 4°C and 19000 rpm for 20 minutes to separate soluble proteins and cell debris. The supernatant containing solubilized proteins was collected and stored on ice. The cell debris pellet was resuspended once more in 5 ml 20 mM Tris-HCl (pH 8) with addition of 8 M urea and 0.1% beta-mercaptoethanol, followed by 4 minutes ultrasonication and centrifugation as described above. The two supernatants were combined and 20 mM Tris-HCl (pH 8) added to reach a final volume of 50 ml with addition of 0.5 M NaCl, 25 mM CaCl₂ and 0.3% beta-mercaptoethanol. The protein suspension was incubated 1 hour on ice under light shaking followed by dilution with 25 ml 20mM Tris-HCl (pH 8), 0.5 M NaCl, 25 mM CaCl₂ and centrifugation at 4°C and 22000 rpm for 20 minutes. The supernatant was collected and partitioned into three dialysis tubes holding 25 ml protein suspension each. The dialysis tubes were immersed in 400 ml precooled (4°C) dialysis buffer, 20 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 25 mM CaCl₂, and allowed to equilibrate for 8-12 hours under light stirring. Five buffer changes were made in total, each step lasting 8-12 hours, to ensure complete refolding. The dialyzed protein solution was centrifuged at 4°C and 22000 rpm for 1 hour to remove any precipitate prior to purification by affinity chromatography.

2.5.5 Purification by FPLC affinity chromatography

Affinity chromatography was used to purify H1-CRD on an FPLC-system, as described below (Table 2). The dialyzed protein solution was loaded onto a 10 ml Galactose-Sepharose column, followed by washing with 20 mM Tris-HCl (pH 7.8), 0.5 M NaCl, 25 mM CaCl₂ to remove non-bound proteins. Elution of H1-CRD was carried out with 20 mM Tris-HCl (pH 7.8), 0.5 M NaCl, 2 mM EDTA containing buffer. The chromatography was monitored at 280 nm.

Table 2. FPLC program for the purification of WT H1-CRD by affinity chromatography.

Step	Volume (ml)	Flow rate (ml/min)
Load	100-200	1
Wash	25	1
Elution	30	1
Wash 2	5	2

Eluted fractions were stored at 4°C. The column was rinsed with water, before storage in 20% methanol at 4°C.

2.5.6 Protein analysis by SDS-PAGE

Protein samples, both after expression and purification as well as final products, were analyzed by SDS-PAGE electrophoresis according to the procedure by Laemmli [101]. SDS renders the proteins denatured and uniformly negatively charged. Similar to agarose gel electrophoresis, the proteins in a sample will migrate with different speed through the polyacrylamide gel towards the anode based on their size [102].

As the expected molecular mass of H1-CRD is 17 kDa, a 15% separating gel was prepared together with a stacking gel according to the set-up in table 3.

Table 3. Composition of a separating gel with 15% polyacrylamide and a stacking gel as used for SDS-PAGE analysis of H1-CRD.

Substance	Separating gel 15%	Stacking gel
Acrylamide solution 30%	7.5 ml	1.7 ml
Water	4.3 ml	7.2 ml
Separating buffer	3.0 ml	-
Stacking buffer	-	1.0 ml
SDS 10%	150 µl	100 µl
TEMED	7.5 µl	10 µl
APS 40%	50 µl	33 µl

The compositions of the buffers used for SDS-PAGE are listed in table 4.

Table 4. Compositions of the buffers used for SDS-PAGE analysis.

Buffers	Composition
Separating buffer	1.9 Tris-HCl (pH 8.8)
Stacking buffer	1.25 M Tris-HCl (pH 6.8)
Running buffer 10X	0.25 M Tris-HCl (pH 8.3), 2 M glycine, 1% SDS
SDS-PAGE sample buffer 5X - non-reducing	200 mM Tris-HCl (pH 6.8), 37.5% glycerol, 5% SDS, bromphenol blue
SDS-PAGE sample buffer 5X - reducing	200 mM Tris-HCl (pH 6.8), 37.5% glycerol, 5% SDS, bromphenol blue, 2 M beta-mercaptoethanol

The SDS-PAGE was run at 30 mA per gel during the stacking gel and 40 mA per gel during the separating gel. For visualization, the gels were stained either with Coomassie staining or silver staining.

2.5.6.1 Silver staining

Silver staining can detect proteins in the nanogram range [103]. It is the method of choice whenever high staining sensitivity is desirable. SDS-PAGE gels were stained with a silver staining kit from Amersham Biosciences, Pharmacia, according to the supplier's protocol.

2.5.6.2 Coomassie staining

Gels stained by Coomassie were initially incubated in a fixation solution containing 20% EtOH and 5% glacial acetic acid in water for 20 minutes under light shaking. The gels were washed with water 3x5 minutes before transfer to Enhanced Coomassie solution (0.02% Coomassie Brilliant Blue G-250, 2% phosphoric acid, 5% aluminum sulfate, 10% ethanol) for a minimum of 2 hours [104]. Destaining was done with water until desired contrast was reached.

2.5.7 Western Blot analysis

Western blot analysis [102] was used to confirm the identity of H1-CRD in protein samples after FPLC affinity chromatography and HPLC chromatography.

18 Whatman papers, typically of size 6x8.5 cm, were incubated as 6, 3 and 9 sheets in anode buffer I, anode buffer II and cathode buffer respectively for 5 minutes. For buffer compositions see table 5.

Table 5. Composition of the buffers used for Western Blot.

Buffers	Composition
Anode buffer I	300 mM Tris, 20% Methanol
Anode buffer II	25 mM Tris, 20% Methanol
Cathode buffer	40 mM Aminocaproic acid, 20% Methanol
Blocking buffer	20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 4% non-dried milk powder
TTBS	20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 0.05% Tween-20
TBS	20 mM Tris-HCl (pH 7.5), 500 mM NaCl
Substrate solution	100 mM Tris-HCl (pH 8.8), 100 mM NaCl, 5 mM MgCl ₂

A nitrocellulose membrane of the corresponding size was incubated 2 minutes in water followed by 5 minutes in anode buffer II. Finally, the gel was put in cathode buffer for 5 minutes. The protein transfer was accomplished by using TransBlot semi-dry transfer cell, stacking the papers on top of the anode in the following order: 6 sheets from anode buffer I, 3 sheets from anode buffer II, the nitrocellulose membrane, the gel and on the very top 9 sheets from cathode buffer. Immunoblotting was done at 14V for 1 hour.

The membrane was reversibly stained with ponceau S solution 0.1% for 3-5 minutes to confirm that the transfer had been successful. The molecular weight marker bands were marked with a pencil and the membrane was washed with water. The membrane was further incubated with blocking buffer at 4°C O/N. The following day, the membrane was washed 3x5 minutes with TTBS, followed by incubation with a primary antibody (polyclonal anti-H1-CRD-IgY produced in chicken [94] and diluted 1/5000 in blocking buffer) during light shaking at RT for 2 hours. Washing was repeated prior to incubation with the second antibody (polyvalent anti-chicken-IgG-AP produced in rabbit and diluted 1/5000 in blocking buffer) at RT for 2 hours. The membrane was washed twice with TTBS and once with TBS. 15 ml substrate solution supplemented with 0.4% NBT/BCIP was added for incubation under light shaking. Once bands became visible, developing was stopped by addition of water.

2.5.8 Separation of monomer and dimers by HPLC ion exchange chromatography

All buffers used for HPLC chromatography were prepared with gradient-grade water G chromasolv® from Sigma.

Ion exchange chromatography was used to separate monomers and dimers of H1-CRD on a HPLC system. A DEAE column, a weak positively charged anion exchanger, was employed for the separation. The column was preconditioned with running buffer A and B, containing 25 mM Tris-HCl (pH 8) and 25 mM Tris-HCl (pH 8), 250 mM CaCl₂ respectively. The buffers were run at 0.5 ml/min in a ratio of 85:15 (DR-thesis).

In theory, separation of monomers and dimers is achieved by slowly increasing the percentage of running buffer B. By changing the gradient, the concentration of CaCl₂ increases, gradually displacing the proteins from the column [105]. Monomers, which are less charged than the dimers, will elute at a lower CaCl₂ concentration. Therefore, an initially low salt content of the samples is crucial as the elution is ion concentration dependent. Since the H1-CRD samples contained a high salt molarity (500 mM NaCl) as a result of previous purification steps, a buffer change was necessary prior to IEC separation. Desalting was accomplished by loading 1.5 ml sample onto a HiTrap desalting column followed by elution with 2 ml running buffer A.

The protein samples were injected, typically 1400 µl at a time, on the DEAE column and separation carried out according to the program below (Table 6).

Table 6. Program for the separation of monomers and dimer by HPLC IEC.

Time (min)	% buffer B
2	15
25	40
27	40
29	15
32	Stop

Peaks were detected at 280 nm and collected for further analysis. While the column was kept at 20°C, both samples and collected fractions were stored at 5°C during the run.

2.5.9 Separation of monomers and dimers by HPLC reversed phase chromatography

An alternative method used to separate monomers and dimers of H1-CRD, developed by Dr. Said Rabbani at IMP, was HPLC reversed phase chromatography (RP). A C18-column was employed and run at 1 ml/min with a solvent gradient of 0.01 M TFA in water and acetonitrile respectively. Samples were loaded onto the column and eluted by increasing the acetonitrile proportion from 25% to 38% during 10 minutes. The acetonitrile was kept constant for an additional 10 minutes and then reduced again to 25% during 5 minutes. Collected fractions were frozen in liquid nitrogen and lyophilized. The column was washed and stored in water/acetonitrile 50/50.

2.5.10 Purification by HPLC affinity chromatography

Monomers purified by HPLC IEC were further run on a 2 ml GalNAc-Sep column (Bio-scale MT2 column) on HPLC, both to reconfirm the binding activity of the protein and to concentrate the samples. The column was preconditioned with running buffer containing 10 mM HEPES (pH 7.4), 10 mM CaCl₂ and kept at a constant flow-rate of 1 ml/min. The samples were loaded onto the column by repeated injections of 1400 µl, and eluted with 10 mM HEPES (pH 7.4), 2 mM EDTA, see program in table 7.

Table 7. Program used for final purification and pooling of H1-CRD monomers by HPLC affinity chromatography.

Time (min)	% buffer B
2	0
12	100
14	100
15	0
25	Stop

Peaks were detected at 280 nm and collected fractions stored at 5°C for further analysis.

2.5.11 Concentration determination by the Bradford assay

The Bradford assay was employed to measure the concentration of final protein samples [106]. A standard curve was established by using a dilution series of protein standard BSA, ranging from 100 ng/µl to 1000 ng/µl. The concentration of unknown

protein samples was then determined from the standard curve based on measured absorbance intensity.

A 96-well microplate was loaded with 200 μ l Bradford dye per well. 10 μ l protein solution was added to each well and mixed by gently tapping the plate. Standard solutions of BSA were added as triplicates and unknown samples as quadruplicates. The plate was incubated at RT for 5 minutes and the absorbance was measured at 595 nm. Final protein samples were stored at -20°C after determining the concentration.

2.6 Cloning, expression and characterization of mutants of H1-CRD

Site-directed mutagenesis can be performed with template DNA either in the form of an isolated DNA fragment or with the target sequence present in a plasmid. Template DNA isolated from cloning strains of *E.coli* shows higher quantities and quality when compared to DNA isolated from expression strains. [59,66] The construct pET-3bH1, carrying the sequence for WT H1-CRD, was therefore to be isolated from AD494(DE3) cells and transformed into a cloning strain prior to the planned mutagenesis studies.

2.6.1 Isolation of pET-3bH1

The plasmid pET-3bH1 was isolated from AD494(DE3) cells after confirmation that the DNA sequence encoding H1-CRD was correct. Plasmid isolation was carried out by alkaline lysis using the GFX microprep plasmid kit (Amersham Biosciences). 1.5 ml O/N cultures were processed according to the manufacturer's protocol with one exception, *i.e.* the final elution of the miniprep was done in 50 μ l instead of the recommended 100 μ l, in order to obtain samples of higher concentration. The isolated DNA was analyzed by agarose gel electrophoresis, prior to transformation into cloning strains of *E.coli*.

2.6.2 Preparation of template H1-CRD DNA

The plasmid pET-3bH1 was transformed into calciumcompetent DH5 α and Top10 cells, as well as electrocompetent DH5 α cells.

Transformation of the calciumcompetent cells was induced by heat shock. An aliquot of 80 μ l DH5 α or Top10 cells was thawed on ice and 1.5 μ l DNA was added. The cells

were incubated on ice for 30 minutes, followed by a heat shock at 42°C in a thermoblock for 90 seconds. The cells were placed on ice for another 2 minutes. LB medium was added to a total volume of 1 ml and the cell suspension was incubated at 37°C and 550 rpm for 1 hour. The cells were centrifuged at 15000 rpm for 1 minute, resuspended in 100 µl LB medium and plated on a LB-ampicillin agar plate. The plate was incubated at 37°C O/N.

Electroporation was carried out with 50 µl DH5α cells. The cells were transferred to a pre-chilled electroporation cuvette and 1.5 µl DNA added. The cuvette was placed in an electroporator with the settings 400 Ω, 1.75 kV and 25 µF. Following the electroporation, the cells were immediately mixed with 950 µl SOC medium and transferred to an Eppendorf tube. The cell suspension was incubated at 37°C and 550 rpm for 1 hour. 300 µl cell suspension was plated on a LB-ampicillin agar plate and incubated at 37°C O/N.

In order to obtain sufficient amounts of DNA template, plasmid isolation using both miniprep and maxiprep kits was performed.

Cultures consisting of 3 ml LB medium supplemented with 100 µg/ml ampicillin were inoculated with a single colony of the transformed DH5α or Top10 cells. The cells were grown at 37°C and 300 rpm for approximately 8 hours, followed by plasmid isolation using GFX microprep plasmid kit. 1.5 ml of each culture was processed according to the manufacturer's protocol, with the final elution in 70 µl. The miniprep samples were stored at -20°C.

Plasmid purification was also carried out using the Qiagen Plasmid Maxi kit. A 3 ml LB medium culture, supplemented with 100 µg/ml ampicillin, was inoculated with a single colony of transformed DH5α cells. The culture was incubated at 37°C and 300 rpm. After 8 hours the starter culture was diluted 1/100 into 100 ml selective LB medium and grown at 37°C and 300 rpm O/N. The cells were harvested by centrifugation at 4°C and 5500 rpm for 20 minutes. Plasmid isolation and purification was done according to the manufacturer's protocol using a QIAGEN-tip 500. The final DNA pellet was resuspended in 500 µl TE buffer consisting of 10 mM Tris-HCl (pH 8), 1 mM EDTA and stored at -20°C.

In order to confirm the presence of the H1-CRD sequence in the isolated plasmids, both the miniprep and maxiprep samples were digested with the restriction enzymes *Bam* H1 and *Nde* 1 O/N. Restriction digestion was done according to set-up 1 (see section 2.4.3) with 12 μ l miniprep and maxiprep sample respectively.

2.6.3 Quantification of DNA

The concentration of the isolated DNA was determined by agarose gel electrophoresis. A dilution series of known concentrations λ -DNA, ranging from 2.5 ng to maximum 600 ng, was prepared and loaded onto the gel together with the plasmid DNA. The gel was run at 80 V.

Quantification of the DNA was done partly by independent estimates, partly by using the software Quantity One from Biorad Laboratories, Inc. A calibration curve was established based on the λ -DNA series, and the concentration of the unknown samples calculated based on the standard curve.

2.6.4 Preparing plasmid pET-3b

The plasmid pET-3b was to be used for down-stream applications and therefore had to be prepared without insert(s). 50 μ l pET-3bH1 isolated from DH5 α -cells was digested with *Bam* H1 and *Nde* 1 according to set-up 2 (see section 2.4.3). To avoid self-ligation, the digested vector was dephosphorylated according to the set-up below, see table 8.

Table 8. Set-up for the dephosphorylation of a digested vector.

Substance	Volume (μ l)
Digested vector sample	60
Alkaline phosphatase enzyme	1.5
Alkaline phosphatase buffer 10X	1.5
Water (sterile)	7
Σ	70

Dephosphorylation was carried out at 37°C for 60 minutes, followed by incubation at 65°C for 10 minutes to deactivate the enzyme. The vector was loaded onto a 1.25% agarose gel for isolation and purification using the Perfect Prep Clean-up kit from Eppendorf.

2.7 Site-directed mutagenesis

Three different methods were employed to create point mutations in H1-CRD; conventional PCR with mutagenic primers, overlap extension PCR and ExSite PCR-based site-directed mutagenesis system. The choice of method was ultimately dependent on the location of the desired mutation in the template DNA sequence. Point mutations residing within 10 amino acids from either end of the sequence were attempted by conventional PCR with mutagenic primers, while overlap extension PCR was used for mutations further into the sequence. ExSite PCR-based site-directed mutagenesis system, a commercial kit available from Stratagene, could be used regardless of the position of the mutations.

2.7.1 Primers

Primers were designed according to the guidelines for each specific method. A requirement for the ExSite PCR-based mutagenesis system was the phosphorylation of one primer in a pair [107]. The primers used within this work for site-directed mutagenesis are summarized in table 9 below, with sequence and melting temperatures (T_m) specified. The mutations are highlighted with bold capitals in the sequences.

Table 9. Summary of the primers used in mutagenesis reactions of H1-CRD. The given names are specified along with the primer sequences, the method for which they were designed and utilized, as well as the melting temperatures as calculated by the manufacturer.

Name	Primer sequence	Method	T _m
H1-CRD FW	5'- GGA ATT CCA TAT GGG CTC AGA AAG GAC CTG CTG -3'	Amplification	66.1°C
H1-CRD RW	5'- CGC GGA TCC TTA TTA AAG GAG AGG TGG CTC CTG GCT -3'	Amplification	69.0°C
H1-CRD mut 152 FW	5'- GGA ATT CCA TAT GGG CTC AGA AAG GAC CTC CTG CCC G -3'	PCR with mut. primers	69.8°C
H1-CRD mut 153 FW	5'- GGA ATT CCA TAT GGG CTC AGA AAG GAC CTG CTC CCC G -3'	PCR with mut. primers	69.8°C
H1-CRD mut 152/153 FW	5'- GGA ATT CCA TAT GGG CTC AGA AAG GAC CTC CTC CCC G -3'	PCR with mut. primers	69.8°C

H1-CRD mut 164 FW	5'- GAC CGC AGC TCC TAC TGG TTC -3'	OE-PCR	61.5°C
H1-CRD mut 164 RW	5'- GAA CCA GTA GGA GCT GCG CTC -3'	OE-PCR	61.5°C
H1-CRD mut 237 FW	5'- AAC TGG AGG GCG GAG CAG CCG -3'	OE-PCR	65.2°C
H1-CRD mut 237 RW	5'- CGG CTG CTC CGC CCT CCA GTT -3'	OE-PCR	65.2°C
H1-CRD mut 238 FW	5'- TGG AGG CCG GGG CAG CCG GAC - 3'	OE-PCR	69.3°C
H1-CRD mut 238 RW	5'- GTC CGG CTG CCC CGG CCT CCA -3'	OE-PCR	69.3°C
H1-CRD mut 256 FW	5'- GAC TGT GCC GAG TTC ACC GAC -3'	OE-PCR	61.5°C
H1-CRD mut 256 RW	5'- GTC GGT GAA CTC GGC ACA GTC -3'	OE-PCR	61.5°C
H1-CRD mut 259 FW	5'- CAC TTC ACC GCC GAC GGC CGC -3'	OE-PCR	67.3°C
H1-CRD mut 259 RW	5'- GCG GCC GTC GGC GGT GAA GTG -3'	OE-PCR	67.3°C
H1-CRD mut 260 FW	5'- TTC ACC GAC GCC GGC CGC TGG -3'	OE-PCR	67.3°C
H1-CRD mut 260 RW	5'- CCA GCG GCC GGC GTC GGT GAA -3'	OE-PCR	67.3°C
H1-CRD mut 152/153 1/1	5'- TCC TCC CCG GTC AAC TGG GTG -3'	ExSite	63.6°C
H1-CRD mut 152/153 1/2	5'- GGT CCT TTC TGA GCC CAT ATG -3'	ExSite	57.4°C

All primers were ordered from and synthesized by Microsynth. Prior to use, the lyophilized primers were dissolved in sterilized water to make stock solutions of 200 μ M. The primers were briefly vortexed and incubated at 65°C for 5 minutes. Working solutions were prepared by diluting the stock solutions 1/10 with sterile water. All primers were stored at -20°C.

2.7.2 ExSite PCR-based site-directed mutagenesis system

The ExSite system can be used to introduce site-specific mutations at any position in double-stranded DNA [107]. Starting with plasmid DNA, the mutation(s) is introduced by a mutagenic primer simultaneously with amplification of the complimentary strand

by a reverse primer. ExSite utilizes a DNA polymerase blend consisting of both *Taq* and *Pfu* polymerase, increasing both the efficiency and reliability of the transcription. After completion of the PCR reaction two plasmid populations exist, the parental template plasmids and the produced linear DNA carrying the desired mutation. The restriction enzyme *Dpn* 1 is added, causing digestion of all methylated parental plasmids, leaving only the mutated DNA intact. The latter is polished with *Pfu* DNA polymerase and recircularized by T4 DNA ligase prior to transformation into competent *E.coli* cells [107].

The mutagenesis reaction was carried out in accordance with the producer's protocol and the supplied reagents. Different reaction set-ups were tested to define the optimal conditions. Primer H1-CRD mut 152/153 1/1, carrying the mutations C152S/C153S, was used together with primer H1-CRD mut 152/153 1/2 (see table 9). The reaction set-ups, which were tried, are summarized in table 10.

Table 10. Different set-ups attempted for mutation C152/153S using the ExSite PCR-based site-directed mutagenesis system.

Component	Reaction A	Reaction B	Reaction C	Reaction D
Template	1.68 µg	1.68 ng	16.8 ng	16.8 ng
Primer 1	15 pmol/95 ng	15 pmol/95 ng	15 pmol/95 ng	19 pmol/120 ng
Primer 2	15 pmol/95 ng	15 pmol/95 ng	15 pmol/95 ng	19 pmol/120 ng
dNTP (25 mM)	1 µl	1 µl	1 µl	1 µl
10X mutagenesis buffer	2.5 µl	2.5 µl	2.5 µl	5 µl
DNA polymerase blend	1 µl	1 µl	-	-
<i>Pfu</i> turbo polymerase	-	-	1 µl	1 µl
Water (sterile)	<i>q.s.</i>	<i>q.s.</i>	<i>q.s.</i>	<i>q.s.</i>
Σ	25 µl	25 µl	25 µl	50 µl

Reaction A is based on the recommendations for the ExSite system, differing from reaction B only in the amount of template used. Reaction C and D utilize *Pfu* turbo polymerase and differ in primer concentration.

The maxiprep DNA sample, purified from DH5 α cells, was used as template. The reaction samples were prepared on ice, adding the polymerase last. PCR was run according to the program in table 11.

Table 11. PCR program when using the ExSite PCR-based site-directed mutagenesis system. Reaction A and B were run with the first alternative times specified at 72°C, and reaction C and D with the second option.

Step	Cycles	Temperature	Time
Denaturation	1x	94°C	4 min
		55°C	2 min
		72°C	11/6 min
Annealing	12x	94°C	1 min
		57°C	2 min
		72°C	5.5/6 min
Elongation	1x	72°C	5/10 min
		4°C	∞

Reaction A and B were digested and polished by 1 μ l *Dpn* 1 and 0.5 μ l *Pfu* respectively at 37°C during 1 hour followed by incubation at 72°C for an additional 30 minutes. The DNA samples were diluted with water, 10X mutagenesis buffer and rATP according to the protocol prior to ligation of 10 μ l sample and 1 μ l T4 DNA ligase at 37°C for 1 hour. Small samples were taken from the reactions for subsequent agarose gel electrophoresis after completing the PCR reaction, digestion and dilution to analyze the integrity of the DNA product throughout the mutagenesis procedure. Finally, an 80 μ l aliquot of XL1-Blue supercompetent cells was incubated with 2 μ l ligation product on ice for 30 minutes. Transformation was carried out through heat pulse at 42°C for 60 seconds and the cells were plated on selective LB-ampicillin agar plates. Plates were then incubated at 37°C O/N.

In contrast, digestion of reaction C and D was carried out at 37°C for 3 hours before the samples were cooled to 4°C. The incubation time was prolonged to ensure that all parental plasmids were fully digested. No additional *Pfu* was added to reaction C and D, as the PCR was carried out exclusively with *Pfu* Turbo polymerase from the start. In

order to reduce the risk of contaminants interfering with downstream applications, the digested DNA was purified with the Perfectprep Gel CleanUp sample kit from Eppendorf following the manufacturer's protocol. Dilution of the samples, ligation and transformation was carried out according to the same procedures as applied for reaction A and B.

Single clones were randomly picked from the selective plates and grown in 3 ml LB cultures supplemented with 100 µg/ml ampicillin at 37°C and 300 rpm O/N. Plasmid purification was carried out on 1.5 ml from each culture using the GFX micro plasmid prep kit from Amersham Biosciences. The DNA was eluted in a final volume of 80 µl to achieve maximal recovery with high end-concentration. All samples were digested with the restriction enzymes *Bam* H1 and *Nde* 1 to verify the presence of H1-CRD in the isolated plasmid. Cultures of positive clones were re-streaked on plates and sent to Microsynth for sequencing.

2.7.3 Conventional PCR with mutagenic primers

Mutations residing within 10 amino acids from either end of the H1-CRD sequence were created by conventional PCR, each reaction consisting of one mutagenic primer and one reverse fully complimentary primer, *i.e.* H1-CRD RW in this case.

REDTaq™ DNA polymerase was initially used to optimize the PCR conditions in respect to template amount, primer and dNTP concentration, as well as optimal annealing temperature for the primers. While *Taq* polymerase is useful for screening of settings due to its high extension rate, the high-fidelity *Pfu* polymerase is better suited for PCR reactions where the products are to be used in downstream applications. The PCR conditions optimized with REDTaq™ DNA polymerase were used together with *Pfu* DNA polymerase to generate the mutated DNA sequence. The PCR reaction set-up, using either REDTaq™ or *Pfu* DNA polymerase, can be seen in table 12. A negative control was also included in the PCR run, exchanging the template for sterile water.

Table 12. PCR reaction set-ups used for mutations attempted by conventional PCR. REDTaq™ DNA polymerase was initially used to screen for optimized reaction conditions.

PCR reaction set-up	REDTaq™ DNA polymerase	Pfu DNA polymerase
Template (0.8 ng)	1 µl	1 µl
Primer 5' (20 µM)	1 µl	1 µl
Primer 3' (20 µM)	1 µl	1 µl
dNTP (10 mM)	1 µl	1 µl
10X polymerase buffer	5 µl	5 µl
RedTaq™ DNA polymerase	2.5 µl	-
Pfu DNA polymerase	-	1 µl
H ₂ O (sterile)	38.5 µl	40 µl
Σ	50 µl	50 µl

The samples were prepared on ice and the DNA polymerase added last. The PCR was run as described in table 13.

Table 13. PCR program for mutations created by conventional PCR.

Step	Cycles	Temperature	Time
Denaturation	1x	95°C	3 min
Annealing	30x	95°C	1 min
		Optimized A _{temp}	1 min
		72°C	2 min
Elongation	1x	72°C	10 min
		4°C	∞

The final PCR products were analyzed by agarose gel electrophoresis to confirm the purity as well as size of the DNA fragments. Typically, two PCR reactions run under the same conditions, showing the same results, were pooled and purified with the GenElute PCR Clean-Up kit from Sigma, to obtain samples with high end-concentration. The purified DNA was digested with *Bam* H1 and *Nde* 1 according to set-up 2 (see section 2.4.3) at 37°C for 4 hours. Next, the digested DNA was loaded onto a 1.25% agarose gel as one sample and run at 80 V for approximately 30-60 minutes. The gel was quickly observed under UV-light and the desired band excised,

followed by purification using the Perfect Prep Cleanup kit. Final elution was done with 30 µl elution buffer. Ligation was carried out as described in section 2.7.5.

2.7.4 Overlap extension PCR

Overlap extension PCR was attempted when the desired mutation(s) was located more than 60-70 nucleotides from either end of the H1-CRD sequence. In a first step, two similar reactions were run, differing only in the primer combination. In reaction A, the flanking primer H1-CRD FW was paired with e.g. the internal primer H1-CRD mut 164 RW, while reaction B combined flanking primer H1-CRD RW and e.g. internal primer H1-CRD mut 164 FW. REDTaq™ DNA polymerase was initially used to optimize the PCR conditions, as described above for conventional PCR with mutagenic primers (see section 2.7.3). Once the optimal PCR settings were determined, the reactions were carried out in duplicates with *Pfu* DNA polymerase. The DNA products were analyzed by agarose gel electrophoresis to confirm the success of the reactions. The fragments were further isolated and purified from a preparative agarose gel, using the Perfect Prep Clean-up kit. Final elution was done in 30 µl elution buffer.

In a second step, the two fragments were combined for the overlap extension PCR. Typically, a temperature gradient PCR reaction screening for the optimal annealing temperature was first performed. Simultaneously, the concentration of the fragments as well as the time point of primer addition was investigated. Three alternatives for primer addition were studied; addition from the start of the reaction, after 10 cycles and in a separate PCR reaction after completing the extension step. The reaction was set up according to table 14.

Table 14. Reaction set-up for the second step of overlap extension PCR. The dilution and amount of each fragment had to be determined individually based on the yields of the first reaction.

PCR set-up	Volume
Fragment A – dilution 1/X	1-5 μ l
Fragment B – dilution 1/X	1-5 μ l
5' primer (20 μ M)	1 μ l
3' primer (20 μ M)	1 μ l
dNTP (10 mM)	1 μ l
10X polymerase buffer	5 μ l
<i>Pfu</i> DNA polymerase	1 μ l
Water (sterile)	37 μ l
Σ	50 μl

The samples were prepared on ice and the *Pfu* polymerase added last. The PCR was run as described in table 15.

Table 15. Program for mutations created by overlap extension PCR.

Step	Cycles	Temperature	Time
Denaturation	1x	95°C	3 min
Annealing	30x	95°C	1 min
		Optimized A_{temp}	1 min
		72°C	2 min
Elongation	1x	72°C	10 min
		4°C	∞

Aliquots of the resulting PCR products were analyzed by agarose gel electrophoresis prior to purification of remaining products by the Perfect Prep Clean-Up kit without first isolating the DNA from an agarose gel. Final elution was done with 30 μ l elution buffer. The purified DNA was digested with *Bam* H1 and *Nde* 1 according to set-up 2 (see section 2.4.3) at 37°C for 4 hours. Following digestion, the mutated H1-CRD sequence was isolated and purified using agarose gel electrophoresis and the Perfect Prep Clean-Up kit. Final elution was done with 30 μ l elution buffer.

2.7.5 Ligation

Ligation was carried out with the digested and purified insert, using two different set-ups, see table 16. The samples were prepared by adding the T4 DNA ligase last.

Table 16. Two alternative ligation reaction set-ups used in parallel.

Ligation set-up	Set- up I	Set-up II
Insert	2 μ l	2 μ l
Vector	2 μ l	1 μ l
T4 DNA ligase	1 μ l	1 μ l
10X ligase buffer	2 μ l	1 μ l
rATP (10 mM)	2 μ l	1 μ l
Water (sterile)	11 μ l	4 μ l
Σ	20 μl	10 μl

The reaction tubes were placed in a RT-water bath and incubated at 4°C O/N. The optimal temperature for ligation is a balance between the optimal temperature for T4 DNA ligase enzyme activity and the temperature necessary to ensure annealing of the fragment ends. By placing the reaction tubes in RT-water at 4°C, a range of temperatures will be covered, passing the optimal temperature for the enzyme [108].

2.7.6 Transformation by electroporation

Prior to transformation of electrocompetent DH5 α -cells, the ligated DNA samples were desalted by dialysis. A petri dish was filled with sterile water and a Millipore nitrocellulose 0.025 μ m filter (Millipore) placed on the water surface. The ligation samples were positioned as individual drops on the filter and left at RT for 45 minutes. After dialysis, the samples were collected for electroporation.

Electroporation was performed with 50 μ l electrocompetent DH5 α cells and 5 μ l of each ligation sample. The procedure was carried out as described in section 2.6.2, using the same electroporation and incubation settings. Following incubation, 300 μ l of each cell suspension was plated on a LB-ampicillin agar plate. The plates were incubated at 37°C O/N.

O/N-cultures, 3 ml LB medium with 100 μ g/ml ampicillin, were inoculated with transformed DH5 α cells and grown at 37°C and 300 rpm O/N. Plasmid purification

was carried out on 1.5 ml from each culture using the GFX micro plasmid prep kit. The DNA was eluted in a final volume of 80 μ l and stored at -20°C . Aliquots of the miniprep samples were digested with the restriction enzymes *Bam* H1 and *Nde* 1 to verify the presence of H1-CRD using set-up 1 (see section 2.4.3). Cultures of positive clones were either re-streaked on plates and sent to Microsynth for sequencing, or sent to the in-house sequencing facility.

2.7.7 Big dye 1.1 terminator sequencing

Prior to in-house sequencing, samples were prepared by amplification of the template DNA with Big dye 1.1 [109] and subsequent precipitation and purification. For this purpose, primers encoding either the T7 promotor or the T7 terminator sequence were used. The primers were synthesized by Microsynth and stock solutions prepared as described in section 2.7.1.

T7 promotor primer: 5'- TAA TAC GAC TCA CTA TAG G -3' Tm 48.8°C
 T7 terminator primer: 5'- TAT GCT AGT TAT TGC TCA G -3' Tm 48.8°C

The PCR reaction was set up as outlined in table 17.

Table 17. Reaction set-up for labeling of samples prior to sequencing.

Sequencing PCR set-up	Volume
Template (miniprep sample)	5 μ l
T7 promotor <i>or</i> T7 terminator primer (20 nM)	1 μ l
Terminator ready reaction mix	3 μ l
Water (sterile)	1 μ l
Σ	10 μl

The sequencing sample was prepared on ice and the terminator ready reaction mix added just prior the start of the PCR, see table 18.

Table 18. PCR program for DNA amplification with Big dye 1.1 prior to sequencing.

Step	Cycles	Temperature	Time
Denaturation	1x	96°C	3 min
Annealing	25x	96°C	10 sec
		50°C	5 sec
		60°C	4 min
Elongation	1x	4°C	∞

Precipitation of the DNA was achieved by adding 10 µl sterile water, 2 µl 3 M NaAc (pH 4.6), 2 µl 125 mM EDTA and 50 µl pure ethanol. The mixture was placed on ice in the dark for 15-30 minutes, followed by centrifugation at 10000 rpm and 4°C for 20 minutes. The supernatant was removed with a fine tip Pasteur pipette. 50 µl ethanol 70% was added to the pellet and centrifugation repeated. After removing the supernatant, the pellet was allowed to dry at RT in the dark with open tubes for 30 minutes. The DNA sample was finally given to the sequencing facility at Biocenter (University of Basel, Switzerland).

2.7.8 Glycerol stocks

Clones containing the correct insert were stored as glycerol stocks. Cultures of 3 ml LB medium supplemented with 100 µg/ml ampicillin were inoculated with 20 µl O/N-culture and grown at 37°C and 300 rpm. Bacteria cultures prepared for storage should be as viable as possible, preferably mid log-phase and not yet entered the stationary phase. After 3-4 hours, 850 µl cell suspension was added to 150 µl sterile glycerol in cryovials (TPP) and briefly vortexed. The bacteria was quickly frozen in liquid nitrogen and stored at -80°C.

2.7.9 Transformation into calcium competent cells

Confirmed mutated constructs of H1-CRD were transformed into calcium competent AD494(DE3) cells for subsequent protein expression. An aliquot of 80 µl AD494(DE3) cells was thawed on ice and 1.5 µl mutant DNA added. The sample was incubated on ice for 30 minutes, followed by a heat-shock at 42°C for 90 seconds. The cells were transferred to ice for another 2 minutes, prior to addition of 900 µl LB medium with 0.4% glucose and 10 mM MgCl₂. The cell suspension was incubated at 37°C and 550

rpm for 1 hour, and then centrifuged at 13000 rpm for 1 minute. The supernatant was discarded and the pellet resuspended in 100 μ l LB medium, supplemented as specified above. The cells were plated on a selective LB-agar plate and incubated at 37°C O/N.

2.7.10 Small-scale expression analysis

Transformed AD494(DE3) cells were screened for high expression levels. O/N cultures of LB medium supplemented with appropriate antibiotics were inoculated with freshly transformed AD494(DE3) cells, randomly chosen. The cultures, typically 8-10, were grown at 37°C and 300 rpm O/N. Small expression cultures made up of 4 ml TB medium, 7.5% glucose, 15 μ g/ml kanamycin and 50 μ g/ml carbenicillin were inoculated with 50 μ l O/N-culture the next day and grown at 37°C and 300 rpm. One clone was prepared in duplicates, the additional culture used for OD measurements. As the cell density reached OD \sim 0.5, a 1 ml sample was taken from each culture and centrifuged at 13000 rpm for 2 minutes. Pellets were stored on ice, designated non-induced samples. Expression was induced by addition of 4 μ l 0.4 M IPTG. After 5 hours, another 1 ml sample was taken from every culture and centrifuged at 13000 rpm for 2 minutes. The pellets were stored on ice and designated expressed samples.

The expression levels were compared by SDS-PAGE analysis. Non-induced and induced samples were prepared by TCA precipitation. Pellets were resuspended in 1 ml PBS and 500 μ l cell suspension sonicated at RT for 5 minutes. 50 μ l TCA was added and samples vortexed for a minimum of 15 seconds. Samples were incubated on ice for 15 minutes and subsequently centrifuged at 13000 rpm for 10 minutes. After removal of the supernatant, pellets were washed with 50 μ l acetone and centrifuged at 13000 rpm for 5 minutes. Supernatant was removed and pellets allowed to dry, prior to resuspension in 50 μ l PBS. Reducing sample buffer 5X was added and samples sonicated for 5 minutes. Finally, the samples were incubated at 95°C and 1200 rpm for 5 minutes, before loading onto the SDS-PAGE (see section 2.5.6).

2.7.11 Determination of target protein solubility

The solubility of mutant proteins was determined according to the following protocol developed by Qiagen [110]. A culture of 10 ml LB medium, 100 μ g/ml ampicillin and 15 μ g/ml kanamycin was inoculated with freshly transformed AD494(DE3) cells, and grown at 37°C and 300 rpm O/N. The O/N-culture was diluted 1/20 into a 50 ml culture the following day and grown at 37°C and 300 rpm until an OD of 0.5 was

reached. 1 ml cell culture was centrifuged at 13000 rpm for 5 minutes and the pellet subsequently resuspended in 50 μ l SDS-PAGE sample buffer 1X (non-induced sample). Expression was induced by addition of IPTG to a final concentration of 1 mM. The culture was grown for another 5 hours. A second 1 ml cell culture sample was centrifuged at 13000 rpm for 5 min and resuspended in 100 μ l SDS-PAGE sample buffer 1X (induced sample). The cells were harvested by centrifugation at 4°C and 4000 rpm for 20 minutes.

The cell pellet was resuspended in 5 ml lysis buffer, *i.e.* 50 mM NaH₂PO₄ (pH 8), 300 mM NaCl, 10 mM imidazole, and subjected to freeze/thawing in a dry ice/ethanol bath. Complete cell lysis was achieved by sonication on ice for 6 x 10 seconds with 10 seconds pause between the steps. The cell lysate was centrifuged at 4°C and 5000 rpm for 25 minutes. The supernatant was decanted and stored on ice as crude extract A (soluble protein fraction). The pellet was resuspended in 5 ml lysis buffer and stored on ice as crude extract B (insoluble protein fraction). Both crude extracts were prepared for SDS-PAGE analysis by mixing 10 μ l sample with 10 μ l SDS-PAGE sample buffer 2X. All samples were heated at 95°C for 5 minutes and subsequently centrifuged at 13000 rpm for 1 minute. SDS-PAGE electrophoresis was prepared and run as previously described in section 2.5.6. The gel was loaded with 20 μ l of the non-induced and induced samples, and 10 μ l of both crude extracts.

2.8 Large-scale expression of mutant H1-CRD

Mutant forms of H1-CRD were expressed by *E.coli* AD494(DE3) cells as previously described in section 2.5.2. Typically, one or three 500 ml TB cultures were expressed for 5 hours.

2.8.1 Solubilization and renaturation of mutant H1-CRD

Purification of mutant forms of H1-CRD was carried out according to the protocol established for the wild-type (see section 2.5.4). In the case of culture volumes larger than 500 ml, the protocol was scaled up proportionally, but keeping the dialysis set-up constant, *i.e.* 3x25 ml protein solution in 400 ml dialysis buffer.

A different sonicator (Banson Sonifier 250, Skan AG Basel-Allschwil) was used for purification of the mutant forms of H1-CRD, due to prevailing circumstances. Sonication was performed with a 50% duty cycle and output control 5 for 15 minutes.

2.8.2 Purification by FPLC affinity chromatography

Mutants of H1-CRD were purified over a 20 ml Gal-Sep column on the FPLC system in a similar manner to WT H1-CRD (see section 2.5.5). The FPLC program can be seen in table 19.

Table 19. FPLC program for the purification of mutant forms of H1-CRD by affinity chromatography.

Step	Volume (ml)	Flow rate (ml/min)
Load	100-225	1
Wash	45	1
Elution	50	1
Wash 2	10	2

Eluted fractions were analyzed by SDS-PAGE and stored at 4°C until further purified.

2.8.3 Separation of monomers and dimers by HPLC IEC

An alternative desalting method was used for the mutant forms of H1-CRD prior to HPLC IEC. Amicon® Ultra-4 centrifugal filter device (Millipore) with a cutoff of 5 kDa was employed to concentrate and desalt protein samples after purification by FPLC. The centrifugal device was first prerinsed with 4 ml sterile water at 4°C and 4000 g for 20 minutes in accordance with the supplier's guidelines. Typically, 7 ml protein sample was concentrated in two centrifugation steps at 4°C and 4000 g for 20 minutes. The concentrated protein was then desalted with 2.8 ml running buffer A, *i.e.* 25 mM Tris-HCl (pH 8), three times using the same centrifugation settings, giving a final volume of approximately 1 ml.

Separation of monomers and dimers of mutant H1-CRD was carried out according to the protocol established for WT H1-CRD (see section 2.5.8).

2.8.4 Purification by HPLC affinity chromatography

Monomeric samples of mutant H1-CRD were purified and concentrated in a final step of affinity chromatography on HPLC as previously described for WT H1-CRD in section

2.5.10. Following affinity chromatography, the concentration was determined using the Bradford assay (see section 2.5.11) and the samples were stored at -20°C .

2.9 Evaluation of WT and mutant H1-CRD

2.9.1 The solid phase competition assay

The in-house developed target-based assay was used to evaluate ligand binding affinity of mutant H1-CRD compared to the wild-type [77]. β -GalNAc-polymer was initially prepared and precomplexed with streptavidin-peroxidase according to the set-up listed in table 20. After mixing the components, the reaction set-up was incubated at 37°C for 2 hours, followed by storage at 4°C .

Table 20. Set-up for formation of the β -GalNAc-polymer-streptavidin-peroxidase-complex.

Components	Volume (μl)
Biotinylated β -GalNAc-polymer (1 mg/ml)	20
Streptavidin-POD-conjugate (500 U/ml)	80
Fetal Call Serum	20
HEPES assay buffer	80
Σ	200

Two different approaches of the assay were undertaken:

- 1) Incubation with only GalNAc-polymer, indicating percentage binding relative to WT H1-CRD
- 2) Incubation with GalNAc-polymer and an inhibitor competing for immobilized H1-CRD, allowing for IC_{50} determination.

A 96-well plate was coated with $100\ \mu\text{l}$ protein solution per well. The protein solution, optimized for a concentration of $3\ \mu\text{g/ml}$, was prepared with HEPES assay buffer containing $20\ \text{mM}$ HEPES (pH 7.4), $150\ \text{mM}$ NaCl, $1\ \text{mM}$ CaCl_2 . 1-2 rows were filled with buffer only, serving as background and blank. In addition, the outer wells were filled with $150\ \mu\text{l}$ HEPES assay buffer to ensure that all wells were kept under equivalent conditions. The plate was placed in a humid chamber and incubated at 4°C O/N.

Following coating, the protein solution was discarded by flipping the plate upside down and tapping it on tissue paper. 150 μ l blocking buffer (1.5% BSA, 0.1% NaN₃ in HEPES assay buffer) was added to all wells, incubating at 4°C in the humid chamber for a minimum of one hour. The blocking buffer was discharged and the plate washed three times with 150 μ l HEPES assay buffer per well. In between the washing steps, the plate was dried by tapping it on tissue paper. Next, the plate was to be incubated either with only polymer or polymer competing with an inhibitor substance.

Incubation of the protein with only polymer was performed by adding 100 μ l β -GalNAc-polymer to the wells at a concentration of 0.5 μ g/ml in HEPES assay buffer. In the second approach, a dilution series of GalNAc, acting as inhibitor, was first prepared in HEPES assay buffer, ranging from 2 mM to 0.006 mM. 50 μ l GalNAc was then added to the wells, followed by 50 μ l 1 μ g/ml β -GalNAc-polymer. Consequently, the final polymer concentration was identical in both assay set-ups. The wells serving as blank were incubated with buffer only. The plate was incubated at 37°C for 2 hours followed by washing two times as described above.

In a final step, 100 μ l ABTS substrate solution was added to each well. The horseradish peroxidase substrate solution was freshly prepared by mixing solution A and B of the ABTS peroxidase substrate kit from Biorad in a ratio of 10:1. The reaction between the peroxidase and the substrate could be monitored through green color that was seen to develop over time. When the color was deemed strong enough, the reaction was interrupted by addition of 100 μ l 2% oxalic acid. OD was measured at 415 nm.

2.9.2 The Biacore assay

All Biacore studies within this work were carried out by Daniel Ricklin, who also developed the assay for ASGP-R H1-CRD [95]. A Biacore 3000 system and research grade CM5 sensor chips, all from Biacore AB, was employed for all SPR experiments.

Samples of H1-CRD, WT and mutants, were immobilized on separate flow cells of a CM5 sensor chip using standard amine coupling procedure [80]. The protein samples, originally in 10 mM HEPES (pH 7.4), 2 mM EDTA, were diluted in 10 mM acetate buffer (pH 4.5) to a final concentration of approximately 20 μ g/ml. After activating the surface for 5 to 10 minutes at a flow rate of 5 μ l/min, the protein was immobilized on

the sensor chip by injection for 5 to 15 minutes. Afterwards the surface was deactivated for a time corresponding to the activation phase.

Stock solutions of Gal and GalNAc were prepared by dissolving the monosaccharides in running buffer (10 mM HEPES (pH 7.4), 50 mM CaCl₂) to a final concentration of 100 mM. High-resolution screening was performed by injecting twofold serial dilutions between 5 mM and 5 μM as randomized triplicates. Each sample was injected for 30 seconds with an undisturbed dissociation phase of 20 seconds using the instruments kinject command at a flow rate of 50 μl/min. No regeneration or washing steps had to be applied. Five buffer blanks were injected at the beginning and one between the triplicate series. Signals of an untreated flow cell and averaged blank injections were subtracted from the sample sensograms. Since referenced sensograms showed negative SPR signals, the whole data set had to be mirrored by multiplication of each data point with -1. Mirrored steady state data were evaluated between 10 and 20 seconds of the injection period and were fitted to a single site binding model [95].

2.10 NMR studies of H1-CRD

All NMR experiments within this work were carried out by Dr. Brian Cutting at IMP. Furthermore, all NMR spectra were processed using the XWINNMR software package commercially available from Bruker AG. The samples were measured at 200 μl using a restricted volume, Shigemi, NMR tube [111]. The Shigemi tubes minimized the quantity of protein needed as the sample was contained only in the active volume of the NMR probe. The use of traditional NMR tubes requires approximately 500 μl for high magnetic field homogeneity. Shigemi tubes allowed the use of a reduced volume as the liquid is contained within magnetic susceptibility plugs optimized for H₂O both above and below.

2.10.1 T1rho measurements

To assess ligand binding to WT and mutant H1-CRD, transverse relaxation rates were measured. The measurement used an in-house sequence to quantitate T1rho, which is a measure of the transverse relaxation, but is less sensitive to magnetic field inhomogeneity. T1rho is advantageous since changing from the free ligand to the ligand with the protein might have different levels of magnetic field homogeneity. Transverse relaxation can be used to determine ligand binding because the rate of

relaxation increases with molecular weight [112]. Hence, the protein-ligand complex displays a greater rate of relaxation compared to the free ligand.

T1rho experiments were conducted with the triple mutant to assess binding affinity to GalNAc. An equivalent measurement with the wild-type protein served as reference. The protein concentrations were in the range 16 to 30 mM, while a GalNAc concentration of approximately 300 mM was used. For all T1rho experiments, the buffer consisted of 20 mM Tris-HCl (pH 7), 1 mM CaCl₂ in 99.5% D₂O.

To estimate the transverse relaxation rate, five decay times, beginning from 10 ms and ending at 200 ms, were measured. All T1rho measurements were performed on a 500 MHz Bruker DRX NMR spectrometer. The signal loss with increasing relaxation times was quantified and analyzed with the Prism curve-fitting software package from GraphPad Software, Inc.

2.10.2 Stability measured by T1rho

T1rho experiments were also performed with WT H1-CRD monomers purified by HPLC RP. The sample concentration was 45 mM H1-CRD and 500 mM GalNAc added. To better understand potential complications arising with the measurement of the binding, additional stability measurements were conducted. Eight measurements of the T1rho relaxation of the GalNAc mixed with the protein were recorded over a period of 56 days. In each case, the signal decay after 200 ms was compared with the decay after 10 ms. During the 56 days of observation, the sample was stored at RT.

2.10.3 Saturation transfer difference measurements

The saturation transfer difference (STD) NMR experiment was used as a complementary diagnosis of binding. The STD measures an intermolecular Overhauser transfer of magnetization. Due to the mechanism of the Overhauser process, the transfer is strongly distance dependent and limited to internuclear distances of approximately 4 Angstroms or less. This enables a qualitative assessment of binding as evidenced by a signal from the ligand in the STD spectrum upon irradiation of the protein [113].

STD was used to evaluate binding to WT H1-CRD purified either by HPLC RP or HPLC IEC. Samples of a protein concentration of 45 mM H1-CRD and 500 mM GalNAc added were analyzed. The NMR experiment was performed as described by Meyer *et*

al., with the exception that the protein was saturated with a train of 40 E-Burp-1 pulses [114] of 40 Hz root-mean-square intensity for 50 ms at 0 ppm [115]. All STD measurements were performed on the 500 MHz Bruker DRX NMR spectrometer.

2.11 Expression of isotope labeled H1-CRD

Isotope labeled protein is a prerequisite for 2D- and 3D heteronuclear NMR. The most straightforward way to obtain such protein is to grow the expression host in isotope enriched medium. For this purpose, different medias and set-ups are routinely being employed [86]. Here, two strategies applied within this work are presented.

- 1) Two-stage protocol, *i.e.* growth in rich medium and expression in minimal medium
- 2) Growth and expression in minimal medium

2.11.1 Minimal medium for bacterial growth and protein expression

The minimal media and supplements used for bacterial growth and expression of isotope labeled proteins are listed below. The M9 minimal medium is adapted from Sambrook and Mantias [88]. All medias and additives were prepared using sterile Millipore water. The M9 salts, NH₄Cl 1%, MgSO₄ and CaCl₂ were autoclaved at 121°C for 20 minutes, while the thiamin and D-glucose stock 20% solutions were filter sterilized.

M9 minimal medium, 1 liter

M9 salts 5X	200 ml
NH ₄ Cl 1%	100 ml
D-glucose stock 20%	20 ml
1M MgSO ₄	2.0 ml
1M CaCl ₂	0.1 ml
Supplement with appropriate antibiotics	

M9 minimal medium plus vitamins, 1 liter

M9 salts 5X	200 ml
NH ₄ Cl 1%	100 ml
D-glucose stock 20%	20 ml
1M MgSO ₄	2.0 ml
1M CaCl ₂	0.1 ml
BME Vitamins 100X	1ml
Thiamin 20 mg/ml	0.1ml
Supplement with appropriate antibiotics	

M9 salts 5X, 1 liter

KH ₂ PO ₄	15 g
Na ₂ HPO ₄	30 g
NaCl	2.5 g
adjust pH to 7.2	

NH₄Cl 1% stock, 0.5 liter

NH ₄ Cl	5 g
adjust pH to 7.2	

2.11.2 Expression of isotope labeled protein using a two-stage protocol

The two-stage protocol for expression of isotope labeled protein, presented by Marley *et al.*, generates cell mass using unlabeled rich media followed by exchange into labeled minimal media for protein expression [90]. It was initially applied for isotope labeling of H1-CRD in collaboration with an external colleague [116].

2.11.2.1 Growth and expression trial using the two-stage protocol

Growth and expression in minimal medium was controlled with a set of initial experiments. Cultures of 3 ml LB medium supplemented with appropriate antibiotics and inoculated with transformed AD494(DE3) or BL21(DE3) cells were grown at 37°C and 300 rpm O/N. The cultures were subsequently inoculated in 200 ml TB or LB medium with appropriate antibiotics and in the former case also supplemented with 1% glucose, aiming at an OD of 0.1. The cultures were grown at 37°C and 300 rpm until an OD of 1-2.0 was reached. The cells were collected at 4°C and 5000 rpm for 30 minutes, and pellets washed with 40 ml M9 salts. Centrifugation was repeated and the cells resuspended in 100 ml M9 minimal medium, concentrating the cells with a factor of 2X or 4X depending on the OD at which they were harvested.

The cultures were further grown at 30°C or 37°C at 300 rpm for 1.5-2 hours prior to induction. A 1 ml sample was removed from each culture, centrifuged at 13000 rpm for 3 minutes and the pellet stored on ice. Expression was induced by addition of IPTG to a final concentration of 0.4 mM, and growth continued under the same conditions. The growth was continuously controlled by measuring the OD. 1 ml samples were regularly removed at different time points from each culture and pellets stored on ice for subsequent analysis (*see section 2.11.4*).

2.11.2.2 Optimization of the expression set-up using experimental design

The expression set-up can be optimized in respect to a number of variables to achieve optimal yields, e.g. cell concentration, expression temperature, expression time, medium composition *etc.* Optimization of the conditions using a “one-factor-at-a-time”-approach can prove not only to be time-consuming, but may also lead to the wrong conclusions. [117,118]. A more efficient way to identify the most favorable settings is to apply experimental design, in which a small set of experiments is devised varying the factors systematically. In fractional factorial design, one class of experimental design set-ups, a number of experiments are chosen out of all possible options to given the maximum amount of variation and information with the smallest number of experimental runs. [119] The ideas behind this approach were adopted for the optimization of expression of isotope labeled H1-CRD according to the two-stage protocol. The growth and expression conditions that were to be optimized can be seen in table 21.

Table 21. Growth and expression conditions that were to be optimized for the two-stage protocol, specified with the alternatives chosen.

Factor	Options
Expression temperature	3: 25°C, 30°C, 37°C
Cell concentration factor	3: 2X, 4X, 6X
Growth culture medium	2: LB, TB
IPTG concentration	2: 0.4 mM, 1 mM

Investigating all possible combinations of the factors would require 36 different experiments. Instead 14 experimental runs were chosen in an attempt to reach as broad diversity as possible, while simultaneously attain maximum amount of information. The choices are illustrated in figure 16.

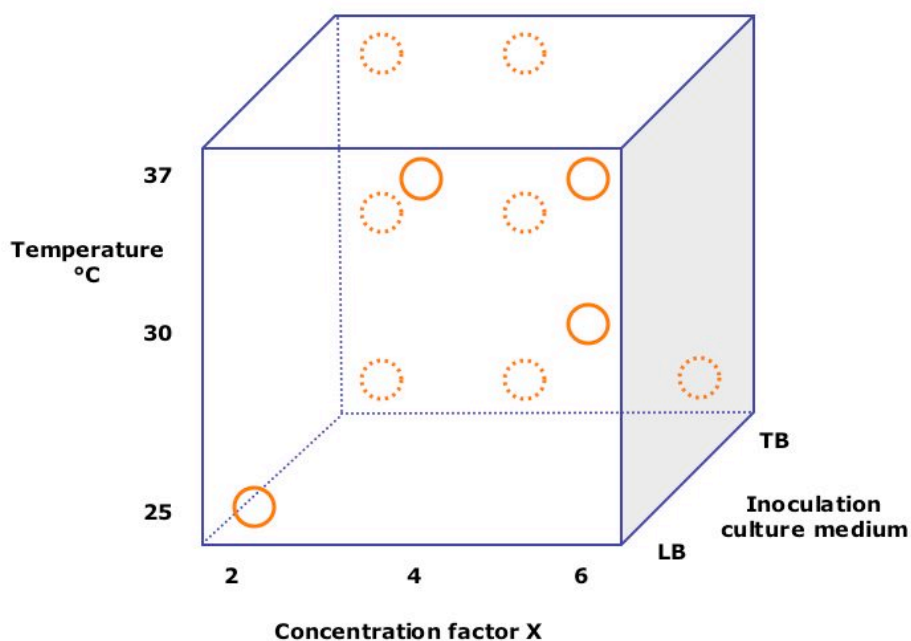


Figure 16. A graphic presentation of the options when optimizing the growth and expression conditions. Three choices are given for the expression temperature and concentration factor of cells, while two options are given for the inoculation culture medium, resulting in 18 possible set-ups. In addition, a fourth dimension is added when considering the two choices of IPTG concentration, doubling the number of expression set-ups. The experiments finally chosen and realized are marked with circles.

In addition, the optimal expression time was to be determined for each individual expression set-up. Also, addition of labeled full medium to the minimal medium was considered.

Transformed BL21(DE3) cells were grown in small cultures of LB medium supplemented with 100 µg/ml ampicillin at 37°C and 300 rpm O/N. The cultures were subsequently inoculated in 100, 200 or 300ml cultures of LB or TB medium supplemented with 1% glucose and 100 µg/ml ampicillin, aiming at an initial OD of 0.1. Cultures were grown at 37°C and 300 rpm until an OD of 1.0 was reached. The cells were collected at 4°C and 5000 rpm for 30 minutes, and pellets washed with 40 ml M9 salts. Centrifugation was repeated and the cells resuspended in 50 ml M9 minimal medium plus vitamins concentrating the cells with a factor of 2X, 4X or 6X. The cultures were further grown at 25°C, 30°C or 37°C at 300 rpm for 1 hour prior to induction. A 1 ml sample was removed from each culture, centrifuged at 13000rpm for 3 minutes and the pellet stored on ice. Expression was induced by addition of IPTG to a final concentration of 0.4 mM or 1 mM, and growth continued under the same

conditions. 1 ml samples were regularly removed at different time points and pellets stored on ice for subsequent analysis (see section 2.11.4).

2.11.2.3 Expression of H1-CRD according to the two-stage protocol under optimized conditions

Transformed BL21(DE3) cells were grown in small cultures of 3 ml LB medium with 100 µg/ml ampicillin at 37°C and 300 rpm O/N. The cultures were subsequently inoculated in 3x500 ml TB medium supplemented with 1% glucose and 100 µg/ml ampicillin, aiming at an initial OD of 0.1. Cultures were grown at 37°C and 300 rpm until an OD of 1.0 was reached. The cells were collected at 4°C and 5000 rpm for 30 minutes, and each pellet washed with 40 ml M9 salts. Centrifugation was repeated and the cells resuspended in 400 ml M9 minimal medium plus vitamins concentrating the cells with a factor of 4X. The cultures were further grown at 30°C and 300 rpm for 1 hour prior to induction. Expression was induced by addition of IPTG to a final concentration of 1 mM. After 5 hours, the cells were collected by centrifugation at 4°C and 5000 rpm for 15 minutes.

2.11.2.4 Alternative expression set-up of the two-stage protocol

Transformed BL21(DE3) cells were grown in small cultures of 3 ml LB medium with 50 µg/ml carbenicillin at 37°C and 300 rpm O/N. The cells were subsequently inoculated in 400 ml LB medium supplemented with 1% glucose and 50µg/ml carbenicillin, aiming at an initial OD of 0.1. The culture was grown at 37°C and 300rpm until an OD of 1.0 was reached. The cells were collected at 4°C and 5000rpm for 30 minutes, and the pellet(s) washed with 40 ml M9 salts. Centrifugation was repeated and the cells resuspended in 100 ml M9 minimal medium plus vitamins (carbenicillin substituting ampicillin), concentrating the cells with a factor of 4X. The culture was further grown at 30°C and 300 rpm for 2 hours prior to induction. Expression was induced by addition of IPTG to a final concentration of 0.4 mM. After 5 hours, the cells were collected by centrifugation at 4°C and 5000 rpm for 15 minutes.

2.11.3 Expression of isotope labeled protein by exclusive use of minimal medium

A second method routinely used for production of isotope labeled protein is the exclusive use of minimal medium for both generating cell mass and expressing protein. [120,121]. By varying the composition of the medium different labeling patterns can be attained, e.g. ¹⁵N and/or ¹³C labeling, selective amino acid labeling etc [86].

2.11.3.1 Expression of H1-CRD in minimal medium

Transformed BL21(DE3) cells were grown in small cultures of 3 ml LB medium with 50 µg/ml carbenicillin at 37°C and 300 rpm for approximately 8-9 hours. The cells were then collected as three individual pellets by centrifugation at 4°C and 5000 rpm for 10 minutes. Each pellet was subsequently resuspended in 50 ml **unlabeled** M9 minimal medium plus vitamins and incubated at 37°C and 300 rpm O/N. The cells were collected anew as three pellets by centrifugation, followed by washing with 40 ml M9 salts per pellet. Centrifugation was repeated and each pellet resuspended in 500 ml **labeled** M9 minimal medium plus vitamins. The cultures were further grown at 37°C at 300 rpm. When an OD of 1.0 was reached, expression was induced by addition of IPTG to a final concentration of 0.4 mM. After 5 hours, the cells were collected by centrifugation at 4°C and 5000 rpm for 15 minutes.

2.11.4 Sample preparation for SDS-PAGE electrophoresis

Pellets collected for SDS-PAGE analysis were well resuspended in lysis buffer (20 mM Tris-HCl (pH 8), 8 M urea and 0.01% beta-mercaptoethanol), briefly vortexed and sonicated for 10 minutes. The lysate was centrifuged at 13000 rpm for 10 minutes. Samples were prepared with reducing SDS-PAGE sample buffer 5X and incubated at 95°C and 1200 rpm for 5 minutes. SDS-PAGE electrophoresis was run as previously described (see *section 2.5.6*).

2.11.5 Solubilization and renaturation of isotope labeled H1-CRD

The isotope labeled H1-CRD was purified and refolded as previously described in section 2.5.4. Bacteria pellets collected after the two-stage method were lysed in proportions to the total volume of rich media that was used rather than the volume of minimal media. The reason for this was to ensure complete lysis of all bacteria and maximal extraction of protein.

Similar to purification of mutant forms of H1-CRD, a different sonicator (Banson Sonifier 250, Skan AG Basel-Allschwil) was used. Sonication was performed with a 50% duty cycle and output control 5 for 15 minutes.

2.11.6 Purification by FPLC affinity chromatography

Refolded protein was purified over a 10 ml Gal-Sepharose column on the FPLC system as described in section 2.5.5. Collected fractions were analyzed by SDS-PAGE electrophoresis.

2.11.7 Separation of monomers and dimers by HPLC IEC

HPLC IEC was used to separate monomers and dimers of the isotope labeled H1-CRD as described in section 2.5.8. Prior to the chromatography run, the protein samples were desalted using Amicon ultra-4 centrifugal filter devices.

Following the separation, collected monomer fractions were concentrated to approximately 500 μ l with Amicon ultra-4 centrifugal filter devices. A further concentration step with simultaneous buffer exchange to reduce the calcium content was done with Microcon ultracel YM-10 centrifugal filter device. 15 N labeled H1-CRD was kept in water, while double-labeled H1-CRD was exchanged into deuterium for future NMR experiments. Final protein sample in 25 mM Tris-HCl (pH 7.4), 1 mM CaCl_2 was stored at -20°C .

2.11.8 Separation of monomers and dimers by HPLC RP

HPLC RP was used to separate monomers and dimer for subsequent MS analysis. Separation was done according to the procedure described in section 2.5.9. Monomeric samples were submitted for MS analysis after lyophilization.

2.11.9 Mass spectrometry analysis

Isotope labeled H1-CRD was analyzed by MS to determine the level of isotopic enrichment. Protein mass analysis was performed using electrospray ionisation technology. All mass spectrometry experiments were carried out in the MS facility in the laboratory of Prof. Paul Jenö (Biocenter, University of Basel, Switzerland).

2.11.10 NMR studies of isotope labeled H1-CRD

As previously noted, the NMR experiments were carried out by Dr. Brian Cutting. Shigemi NMR tubes were used also for the studies of isotope labeled H1-CRD, as well as the XWINNMR software.

2.11.10.1 Heteronuclear Multiple Quantum Coherence (HMQC)

Heteronuclear Multiple Quantum Coherence experiments [122] were performed at 500 MHz ^1H resonance frequency (11.7 Tesla). The ^1H and ^{13}C pulses used for the experiments were calibrated using a standard sample containing 20 mM glucose enriched in ^{13}C . The spectral widths in the experiment were 70 ppm in the ^{13}C dimension, centered at 40 ppm, to ensure that all the C^α and side-chain carbons were measured in the protein. The ^1H spectral width was 5.5 ppm, centered at 2.6 ppm. Due to the uncertainty regarding the level of enrichment, more scans were recorded than were typically used for this experiment. 512 increments were recorded in the indirect ^{13}C dimension, each of which was constructed with 24 scans. The delay between scans was 1.5 seconds resulting in total experimental time of 10.2 hours. Decoupling in both dimensions was achieved by an 180° pulse and a GARP method.

2.11.10.2 HSQC at 800 MHz ^1H resonance frequency

Heteronuclear Single Quantum Coherence experiments [123] were performed at 800 MHz ^1H resonance frequency (18.8 Tesla) on a spectrometer equipped with a high-sensitivity cryoprobe. The spectrometer was kindly made available by Prof. S. Grzesiek. To increase the quality of the measurement, the sensitivity-enhanced version of the HSQC was performed [124]. The ^1H and ^{15}N pulses used for the experiments were calibrated using a standard sample containing 50 mM urea enriched in ^{15}N .

The spectral widths in the experiment were 25 ppm in the ^{15}N dimension, centered at 116 ppm, to ensure that all the backbone amide and side-chain nitrogens were measured in the protein. The ^1H spectral width was 6.0 ppm, centered at 9.0 ppm. 256 increments were recorded in the indirect ^{15}N dimension, each of which was constructed with 32 scans. The delay between scans was 2.3 seconds resulting in an approximate experimental time of 6 hours. Decoupling in both dimensions was achieved by an 180° pulse and a GARP method. No spectral referencing was performed.

2.11.10.3 HSQC at 600 MHz ^1H resonance frequency

HSQC experiments were performed at 600 MHz ^1H resonance frequency (14.1 Tesla). The spectrometer was kindly made available by Prof. S. Grzesiek. To increase the quality of the measurement, the sensitivity-enhanced version of the HSQC was performed [124]. The spectral widths in the experiment were 25 ppm in the ^{15}N dimension, centered at 110 ppm, to ensure that all the backbone amide and side-chain

nitrogens were measured in the protein. Due to the particular method of processing, the apparent width of the ^{15}N dimension appears twice as broad of the actual width. The ^1H spectral width was 6.0 ppm, centered at 9.0 ppm. 128 increments were recorded in the indirect ^{15}N dimension, each of which was constructed with 4 scans. The delay between scans was 0.8 seconds resulting in an approximate experimental time of less than 10 minutes. Decoupling in both dimensions was achieved by an 180° pulse and a GARP method. No spectral referencing was performed.

2.11.10.4 Ligand titration measured by HSQC at 600 MHz ^1H resonance frequency

HSQC experiments, using the same parameters described in section 2.11.10.3, were performed at 600 MHz. A series of different concentrations of GalNAc was added to the protein to determine the binding affinity. One HSQC was measured for each concentration and the chemical shift changes of the protein upon binding were monitored [125].

To minimize the volume added to the Shigemi NMR tube, two stock solutions of GalNAc, dissolved in D_2O , were used, 20 mM or 100 mM. The seven different HSQC experiments corresponded to measurement of the protein with GalNAc concentrations ranging from 0 mM to 1500 mM. The total volume added was 6.2 μl of GalNAc to the 200 μl solution of protein.

HSQC spectra were processed identically for each GalNAc concentration using XWINNMR. Each spectrum was then plotted with XWINPLOT, saved as postscript format, read into and superimposed with Adobe Illustrator.

2.11.10.5 ^{13}C isotopic enrichment experiments

Half-filter [126] experiments were created to estimate the level of isotopic enrichment in the protein samples. The experiment measures ^1H signals either from ^{13}C or from ^{12}C depending upon the experimental conditions. The ratio of the ^1H signals originating from ^{13}C to that of those originating from ^{12}C provided an estimate of the incorporation level. To validate the method, two control samples were measured prior to the protein. An in-house synthesized E-Selectin antagonist [127], with natural abundance ^{13}C (1.11%), and acetic acid fully enriched in ^{13}C . The experiments were performed on the same spectrometer used for the [^1H , ^{13}C] HMQC spectra. The delays used in the half-filter experiment were chosen for a value of a coupling constant of 145 Hz, resulting in a delay of 3.45 ms.

3 Results

3.1 Expression, purification and analysis of H1-CRD

Protein in preparative amounts is a necessity for structural and biochemical studies. A method for expression and purification of ASGP-R H1-CRD from *E.coli*, initially developed by Meier *et al.* [5], and further improved by Rita Born at IMP [94], was used within this work.

3.1.1 Expression and purification of H1-CRD

E.coli strain AD494(DE3) was used for overexpression of H1-CRD. Expression was induced by addition of IPTG and lasted five hours. As bacteria lack the ability to correctly fold recombinant proteins and do not support the formation of native disulfide bonds in the cytoplasm, expressed H1-CRD was accumulated as inclusion bodies. Lysis of the bacteria and complete denaturation of the protein content was carried out to solubilize the aggregated protein. The protein was subsequently refolded by stepwise dialysis and purified by affinity chromatography on a FPLC system. The protein could be seen in the chromatogram as a sharp peak during the elution phase, indicating correct folding and binding to galactose (*Figure 17*). The peak was collected as different fractions.

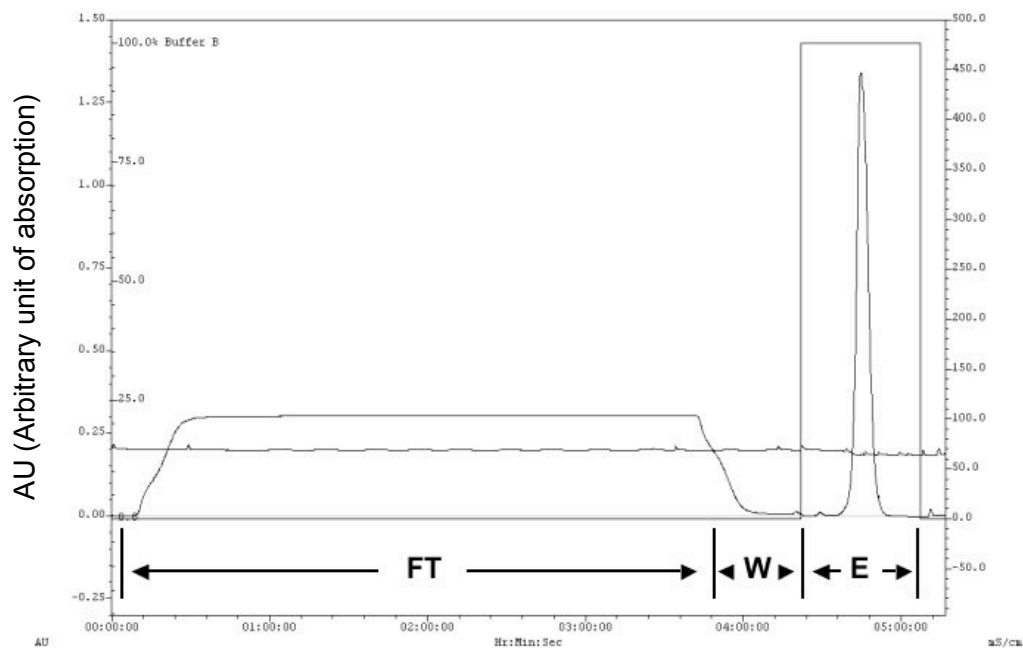


Figure 17. FPLC chromatogram of WT H1-CRD. FT corresponds to the flow-through, followed by the wash (W) to remove any unbound proteins. A distinct peak can be seen during the elution phase (E), representing H1-CRD.

Eluted fractions were analyzed by reducing SDS-PAGE (*Figure 18*). The low molecular weight marker in lane 1 was used as reference. H1-CRD was predominately observed in lane 8 and 9, corresponding to the elution peak in the chromatogram. The protein had clearly been enriched as seen when compared to the sample load (lane 3) before affinity chromatography. H1-CRD could also be spotted in the flow-through (lane 4), indicating either overloading of the column or alternatively misfolded protein.

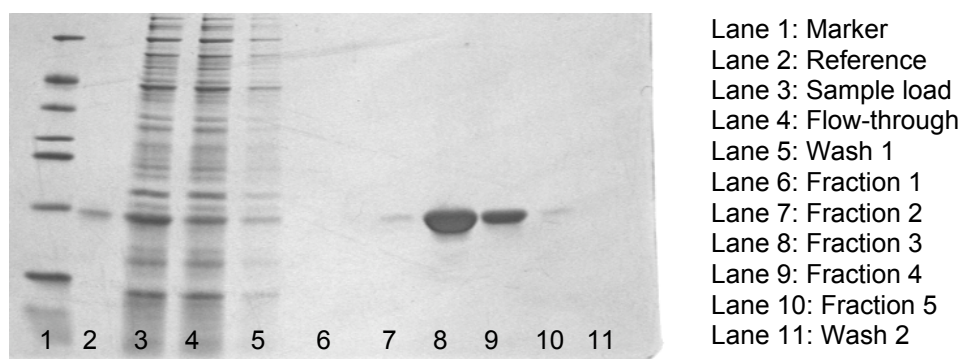


Figure 18. Reducing SDS-PAGE analysis of fractions collected during affinity chromatography of H1-CRD (15% gel, silver staining).

Western blot was used to verify the identity of H1-CRD after affinity chromatography on FPLC. A polyclonal anti-H1-CRD-IgY antibody produced in chicken [94] was utilized in the initial antigen-specific binding step, followed by a secondary antibody coupled with alkaline phosphatase. A thick band representing H1-CRD was clearly seen after development with NBT/BCIP (*Figure 19*).

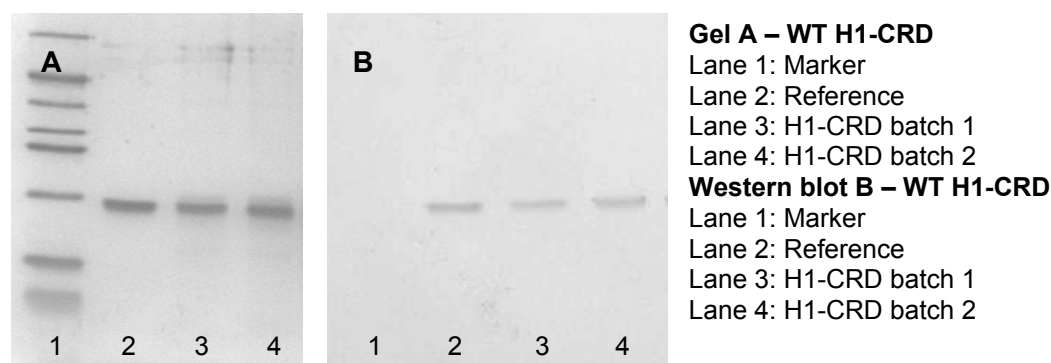


Figure 19. A) Reducing SDS-PAGE analysis of purified H1-CRD originating from different batches (15% gel, silver staining). B) Western blot of an identical gel as seen in figure 19A. A polyclonal anti-H1-CRD-IgY antibody produced in chicken was used for detection [94].

Analysis with non-reducing SDS-PAGE revealed occurrence of both monomeric and dimeric H1-CRD (*Figure 21*, lane 2). No consistent ratio of the two species could be determined as the fractions appeared to vary among different batches. To separate the monomers and dimers, HPLC IEC was employed. Protein samples were loaded onto a DEAE column and eluted with a Tris/CaCl₂ buffer gradient. Monomeric H1-CRD, less charged than the dimeric form, were found to elute at a lower CaCl₂ concentration, while the dimers emerged at higher CaCl₂ concentration. Repeated injections of H1-CRD showed highly similar chromatograms, confirming the reproducibility of the method. The monomers (peak 1) could be seen in the chromatogram as the first peak with a tailing shoulder. A small, unidentified peak (peak 2) was seen following the monomers and preceding the dimers, and is most likely an impurity. The final peak (peak 3) corresponds to dimers of H1-CRD (*Figure 20*).

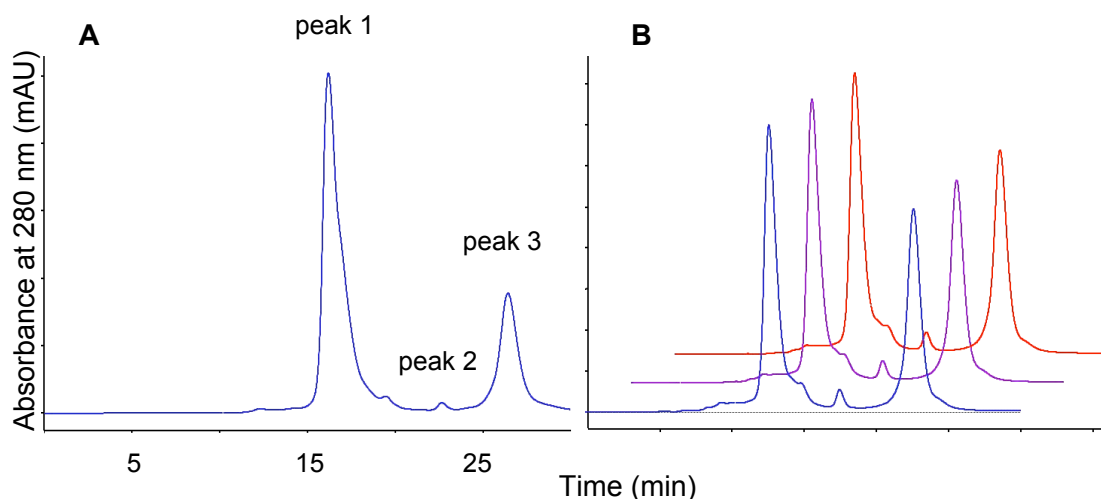
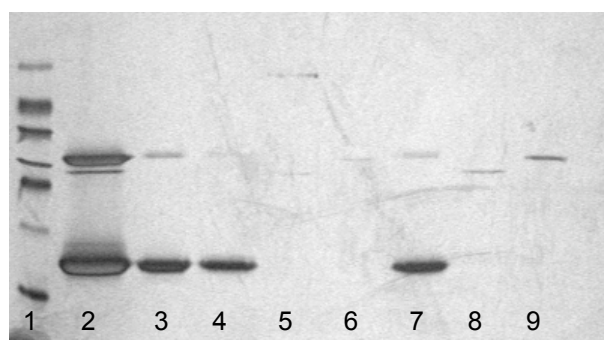


Figure 20. A) Separation of monomers and dimers by HPLC IEC. The monomers can be seen as peak 1 with a tailing shoulder, while the dimers elute in the end (peak 3). B) A second batch showing a higher content of dimer, but with a highly similar separation profile. Repeated injections show high reproducibility.

Collected fractions were analyzed by non-reducing SDS-PAGE to verify the separation of monomers and dimers of H1-CRD (*Figure 21*). The gel revealed traces of dimeric H1-CRD in the monomeric (lane 3 and 7) as well as the monomeric shoulder (lane 4) fraction. The assumed impurity (lane 5 and 8) could be seen to correspond to a band appearing closely below the dimers in the crude sample of H1-CRD (lane 2). Finally, the dimeric fraction (lane 6 and 9) exhibited pure dimers when analyzed, without contamination of either monomers or the impurity.



Lane 1: Marker
 Lane 2: H1-CRD eluted from FPLC
 Lane 3: Monomer fraction – IEC
 Lane 4: Shoulder fraction – IEC
 Lane 5: Impurity fraction - IEC
 Lane 6: Dimer fraction – IEC
 Lane 7: Monomer fraction – IEC
 Lane 8: Impurity fraction - IEC
 Lane 9: Dimer fraction – IEC

Figure 21. Non-reducing SDS-PAGE analysis of fractions collected during HPLC IEC. The monomer fractions (lane 3, 4 and 7) are clearly enriched in monomers, but also contain traces of dimers (15% gel, silver staining).

In an effort to concentrate the purified monomers and simultaneously confirm that the protein had maintained binding affinity for its ligand(s) throughout the purification procedure, affinity chromatography on HPLC was utilized. The sample was loaded

onto a 2 ml GalNAc-sepharose column and eluted with HEPES buffer containing EDTA. The protein could be observed in the chromatogram as a single peak, indicating that both folding and binding affinity remained intact (*Figure 22*). The peak was also seen to increase in size when increasing the sample load, but eluting at roughly the same retention time.

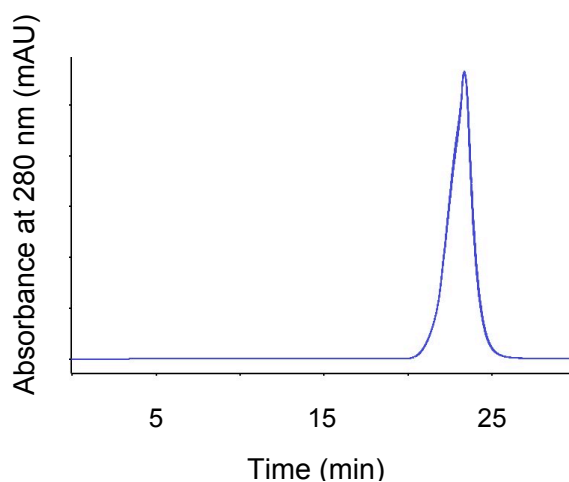


Figure 22. Monomers pooled and purified by HPLC affinity chromatography using a GalNAc-Sep column.

The peak was collected in three fractions, *i.e.* the rise of the peak, the central peak and the tail of the peak, in order to maximize the protein content in one final sample. Therefore, all fractions corresponding to the central peak after each HPLC affinity run were combined in a single sample, intended for future use in assays *etc.* The Bradford assay was used to determine the concentration of final samples, which were subsequently aliquoted and stored at -20°C .

The applied procedure for production of H1-CRD is summarized below in figure 23. Estimated yields are based on results obtained from different batches.

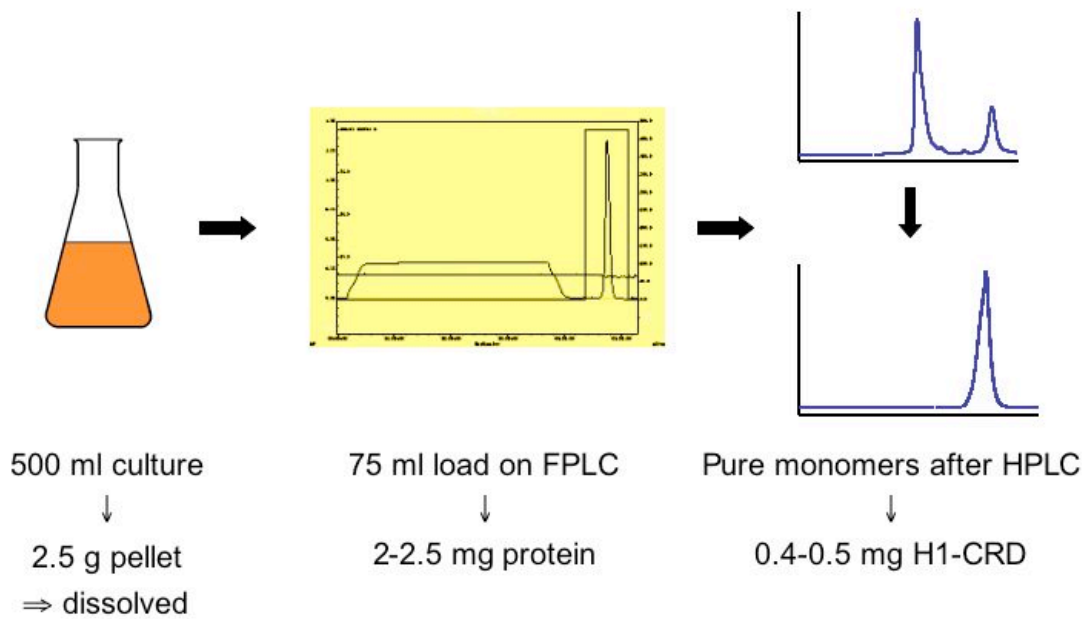


Figure 23. Flow-chart of the production scheme for H1-CRD. Estimated yields are specified for each step.

3.1.2 Reference batch of H1-CRD

In order to have a homogenous sample for assay development and referencing, a large batch of H1-CRD was produced following the established standard protocol. 4 liters expression culture were processed, resulting in a total of four samples of monomer and dimers, see table 22.

Table 22. Yields of the reference batch of H1-CRD after HPLC IEC and affinity chromatography respectively.

Final method	Monomers	Dimers
HPLC IEC	0.24 mg	0.07 mg
HPLC affinity chromatography	4.30 mg	1.23 mg

Final samples were stored as aliquots à 500µl at -20°C and used for development of the Biacore assay by Daniel Ricklin [95] and the solid-phase competition assay by Daniela Stokmaier [77]. The batch was also used as a reference in comparative studies with mutant forms of H1-CRD.

3.1.3 Separation of monomers and dimers using HPLC reversed phase chromatography

An alternative method, reversed phase chromatography, was also employed to separate monomers and dimers of H1-CRD on HPLC. RP chromatography separates proteins based on their differing hydrophobicity, eluting less hydrophobic proteins first. Samples of H1-CRD was loaded onto a C18 column and eluted by an increasing gradient of water/acetonitrile with 0.01 M TFA. Monomeric H1-CRD, less hydrophobic than the dimers, was seen to elute first in the chromatogram (*Figure 24*). The dimers followed closely in a second peak, barely separated from the first. The monomer peak was therefore collected as two samples to avoid contamination with dimers. Repeated injections of H1-CRD showed highly similar chromatograms. The peaks were collected separately and lyophilized.

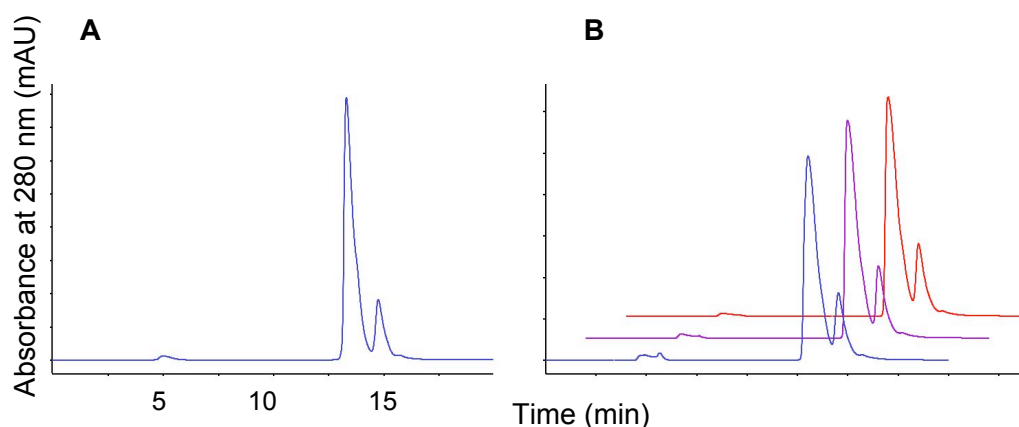


Figure 24. A) Separation of monomers and dimers by HPLC RP. The monomers can be seen as the first peak, closely followed by the smaller dimer peak. B) Repeated injections show reproducible results.

To verify the separation of monomers and dimers, samples were analyzed by non-reducing SDS-PAGE (prior to lyophilization) (*Figure 25*). In addition, a sample taken after FPLC without separating monomers and dimers was loaded onto the gel to serve as a reference of the initial ratio. Monomeric H1-CRD clearly dominated the two monomer samples (lane 4 and 5), but traces of dimers could also be observed, similar to the samples purified by HPLC IEC. There appeared to be slightly more dimers in the second monomer sample (lane 5), possibly due to the narrow separation profile. The second peak showed an exclusive presence of dimers without interfering monomers (lane 6).

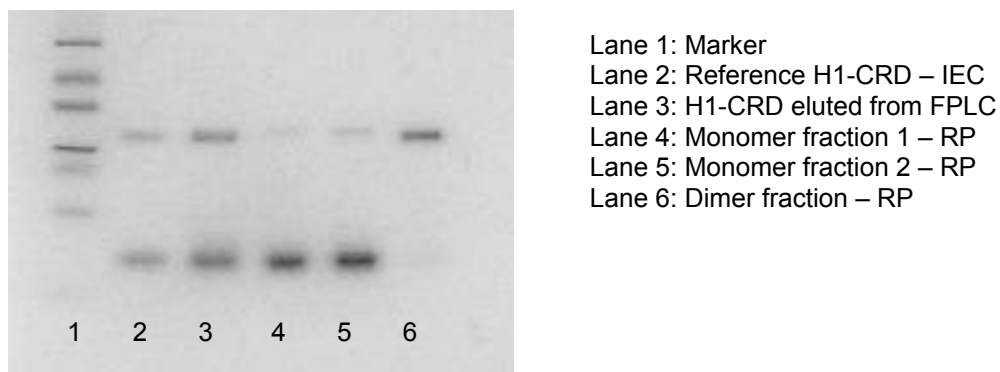


Figure 25. Non-reducing SDS-PAGE of monomers and dimers after separation by RP chromatography. The FPLC sample (lane 3) shows a 1:1 ratio of the species prior to separation. Fractions collected after RP show a clear enrichment of monomers and dimers respectively (15% gel, Coomassie staining).

3.1.4 Comparison of the binding affinity of monomers purified either by HPLC IEC or RP

RP chromatography is generally not recommended for protein purification. The use of organic solvents could harm and/or denature the proteins and compromise both activity and structure [128]. To ensure that H1-CRD monomers purified by HPLC RP showed equivalent affinity to that of monomers purified by HPLC IEC, a series of assays was carried out with the two samples for comparison.

3.1.4.1 Evaluation of WT H1-CRD using the solid-phase competition assay

Lyophilized H1-CRD monomers isolated by RP were initially reconstituted as 500 $\mu\text{g/ml}$ in HEPES assay buffer. WT H1-CRD monomers purified by IEC or RP were coated on 96-well plates at a concentration of 3 or 5 $\mu\text{g/ml}$ protein and incubated O/N. The plate was blocked with BSA prior to incubation with 0.5 $\mu\text{g/ml}$ GalNAc-polymer. No inhibitor was added as the purpose was to compare binding of the two protein samples to the GalNAc-polymer. As an estimate of the binding, the absorbance at 415 nm was measured after the reaction between ABTS substrate and peroxidase. The results are depicted in the graph below (*Figure 26*).

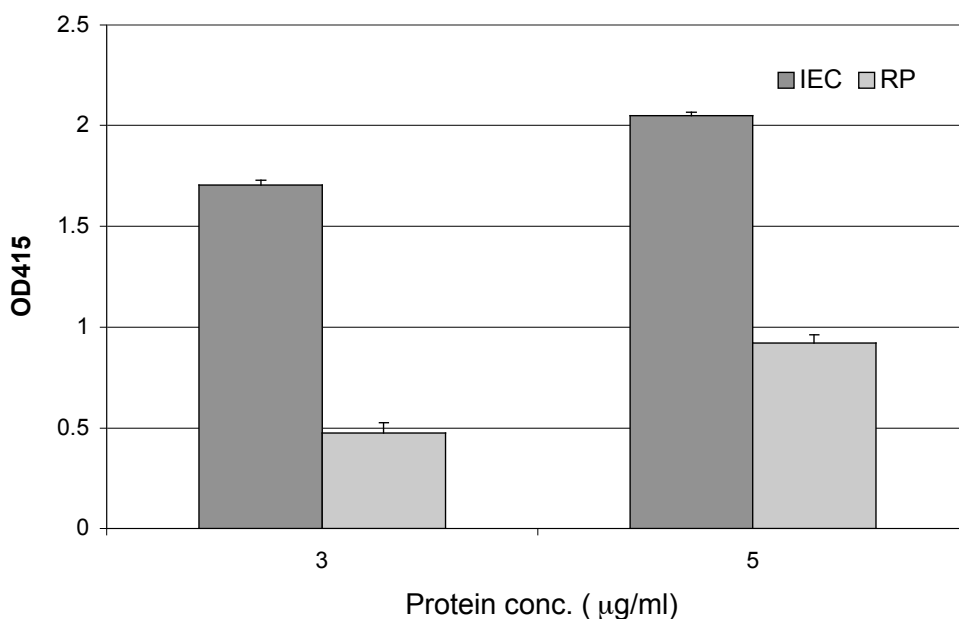


Figure 26. Binding of GalNAc-polymer to H1-CRD monomers purified by HPLC IEC (dark grey) or HPLC RP (light grey) as measured by the solid-phase competition assay.

A clear distinction could be observed between the two monomer samples. The IEC sample showed higher absorbance values at both protein concentrations tested, indicating a higher degree of binding. The RP monomers also showed binding, but to less extent. The measured values at 3 µg/ml, the optimal protein concentration for the assay [77], differed with a factor close to 4. In conclusion, the results generated by the solid-phase competition assay imply that there is a difference in binding affinity exerted by the two protein samples.

3.1.4.2 Evaluation of WT H1-CRD using the Biacore assay

WT H1-CRD monomers purified either by IEC or RP (reconstituted as in section 3.1.4.1) were immobilized on a Biacore sensor chip using standard amine coupling. Although the samples were of similar concentrations, the RP monomers proved more difficult to immobilize. Final protein densities used for the assay were 6000 and 4000 resonance units (RU) of the IEC and RP monomers, respectively. Such a difference does not constitute an obstacle as far as the assay is concerned, but should be kept in mind when evaluating the results.

Dilution series of GalNAc and Gal were injected and binding to the protein monitored. The resulting sensograms were fitted to a 1:1 binding model and K_D was determined (Figure 27).

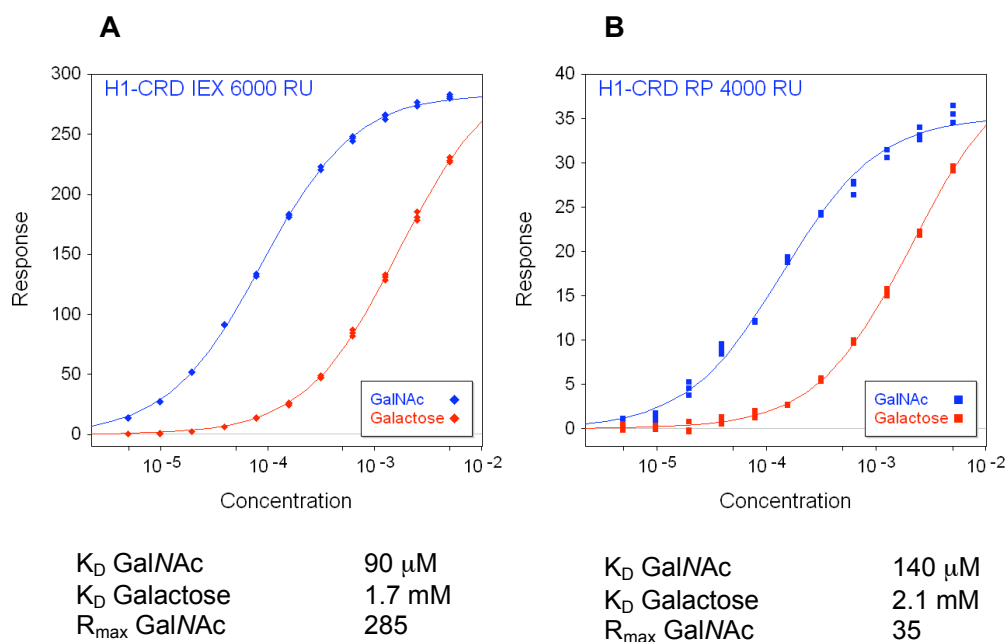


Figure 27. A) Binding of Gal (red) and GalNAc (blue) to WT H1-CRD monomers purified by HPLC IEC. B) Binding of Gal (red) and GalNAc (blue) to WT H1-CRD monomers purified by HPLC RP.

While a difference in K_D of GalNAc and Gal bound to the protein samples could be observed, the variation in the signal intensity was more striking. The signal response was considerably lower for the RP monomers. With the difference in protein density taken into account, the RP sample showed only roughly 20% of the signal intensity measured for the IEC monomers binding GalNAc. In summary, these results suggest that although the RP monomers are able to bind GalNAc, only a small fraction appears to be active. The lower signal response could be explained by damage or inactivation of part of the protein sample, caused by the RP method and/or the subsequent lyophilization. HPLC IEC is therefore better suited for separation of monomers and dimers of H1-CRD.

3.1.4.3 Evaluation of WT H1-CRD using NMR

Two independent NMR experiments were performed to assess binding of GalNAc to H1-CRD monomers purified either by HPLC RP or IEC. The T1rho experiment and the STD NMR experiment are subject to independent sources of experimental errors. Thus, a positive identification of ligand binding from both experiments would serve as cross-validation of the results.

T1rho experiments measure the transverse relaxation rate, which increases with the molecular weight. Hence, a ligand that binds to a large receptor displays a faster decay of magnetization than a free ligand. T1rho was measured at equivalent ratios of protein and GalNAc for the two monomer samples (*Figure 28*).

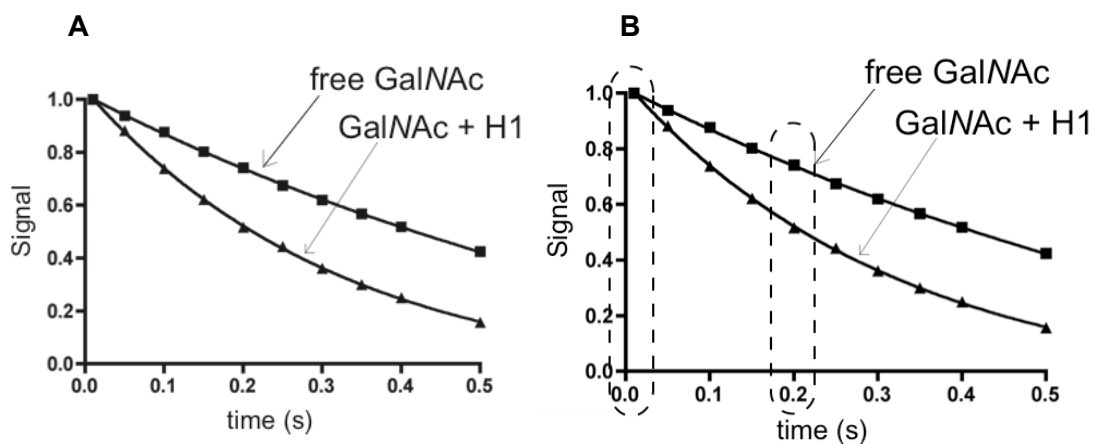


Figure 28. A) T1rho measurements of GalNAc in the absence (■) and presence (▲) of H1-CRD monomers purified by HPLC IEC. B) T1rho measurements of GalNAc in the absence (■) and presence (▲) of H1-CRD monomers purified by HPLC RP. For comparison of the samples, the ratio of signals measured after 200 and 10 ms (dashed lines) was calculated.

The T1rho spectra obtained for the two samples displayed increased relaxation in the presence of the protein, indicating binding. When viewed at identical relaxation time, for example 200 ms, the two samples showed comparable losses in magnetization. While primarily a qualitative assay, the T1rho experiments show that GalNAc binds to both proteins samples.

The two monomers samples, that were analyzed with T1rho, were also subjected to STD measurements (*Figure 29*). STD measures the transfer of magnetization to a bound ligand following irradiation of the protein. The experiment serves as an assay for binding since ligands that bind will give rise to a signal in the STD spectrum, while non-binding ligands will not. The experiment is primarily of a qualitative nature since the intensity of the STD spectrum is not trivially related to the affinity.

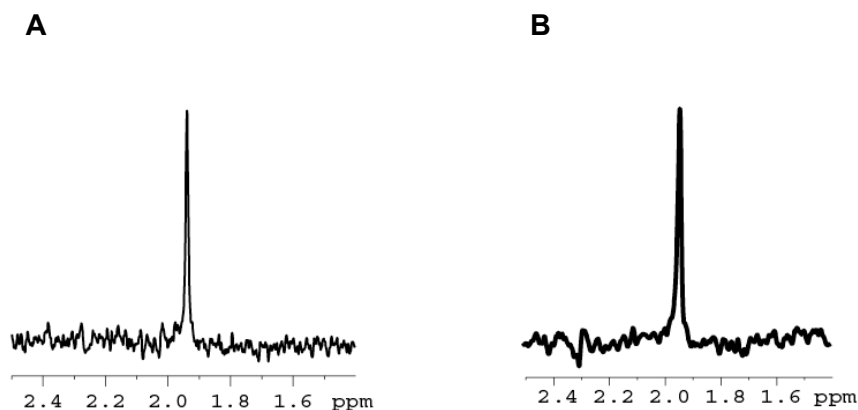


Figure 29. A) STD measurements of GalNAc in the presence of H1-CRD monomers purified by HPLC IEC. B) STD measurements of GalNAc in the presence of H1-CRD monomers purified by HPLC RP.

The major peak seen in the STD spectrum was found at 2.0 ppm, and is due to the methyl group of GalNAc. The signal could be observed for both proteins, indicating that they bind GalNAc. Control experiments were performed with each sample without irradiation of the protein, in which no STD signal was observed. These control experiments confirmed that the previous STD experiments with irradiation of the protein indeed signaled binding.

Positive results for binding from the T1rho as well as STD experiments were observed, adding to the certainty that both monomer samples retained GalNAc affinity regardless of either purification method.

3.1.4.4 Evaluation of the stability of WT H1-CRD using NMR

T1rho experiments with the RP monomer sample were also performed over a period of 56 days to assess protein stability. The signal at 200 ms was divided by the signal at 10 ms for each measurement (*Figure 30*). The protein sample was stored at RT between the measurements. The purpose of this analysis was to ascertain whether the structure of the protein would remain intact during the long-term period needed for three-dimensional NMR experiments required for structure elucidation.

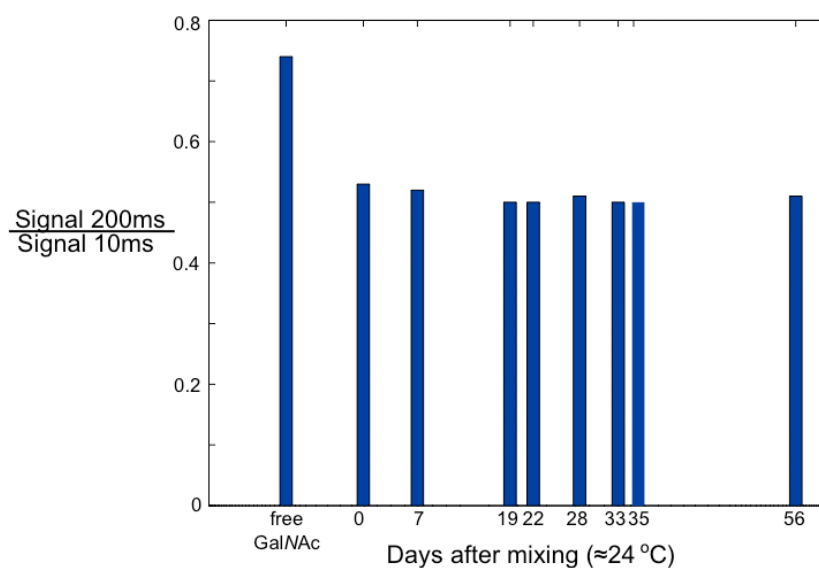


Figure 30. Repeated T1rho measurements of GalNAc and H1-CRD monomers purified by HPLC RP over time. A ratio of the signal measured after 200 ms to that after 10 ms was determined and compared in between the measurements.

The results showed a high degree of protein stability over time, as proved by the small fluctuation in the measured signals. The signal measured at the outset of the trial was nearly indistinguishable from that measured after almost two months. These preliminary results demonstrate that the protein can be expected to be stable during the time required for three-dimensional NMR experiments.

3.2 Site-directed mutagenesis of H1-CRD

3.2.1 Cloning of H1-CRD

A prerequisite for site-directed mutagenesis based on PCR is access to template DNA in the form of an isolated DNA sequence or the target sequence present in a plasmid. In the latter case, plasmid DNA isolated from cloning strains of *E.coli* is preferable as both yields and quality are higher [59,66]. Therefore, the construct pET-3bH1 was extracted from *E.coli* AD494(DE3) cells in a first step, subsequently transformed into *E.coli* DH5 α and Top10 cells and isolated after bacteria proliferation to serve as template. The accuracy of the cloned sequence was confirmed at the outset, showing 100% agreement with the DNA sequence of H1-CRD as determined 1985 [129] and posted in the ExPASy protein database [130].

The plasmid was isolated by alkaline lysis using commercial kits. Following extraction, the DNA was digested with restriction enzymes and analyzed by electrophoresis (Figure 31).

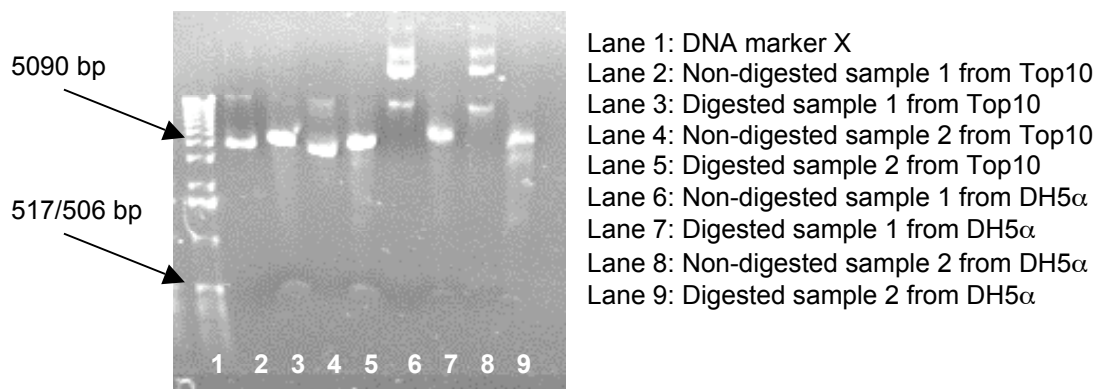


Figure 31. Non-digested and digested plasmids carrying the sequence of H1-CRD, isolated from DH5 α and Top10 cells.

Digested DNA samples originating either from DH5 α or Top10 cells displayed two fragments of expected molecular weight, *i.e.* the vector (4639 bp) below 5090 bp and the insert (435 bp) at 517/506 bp. In contrast, non-digested samples were seen at different molecular weights depending on the host strain. DNA extracted from Top10 cells were seen at 4072 bp as one dominant band, while the DH5 α samples exhibited three bands all above 5090 bp. This difference is likely due to various forms of the extracted DNA, which migrate through the gel at different velocities [68].

Quantification of DNA was carried out using agarose gel electrophoresis. The concentration of the plasmid samples were estimated based on comparison with a dilution series of known concentrations λ -DNA (Figure 32).

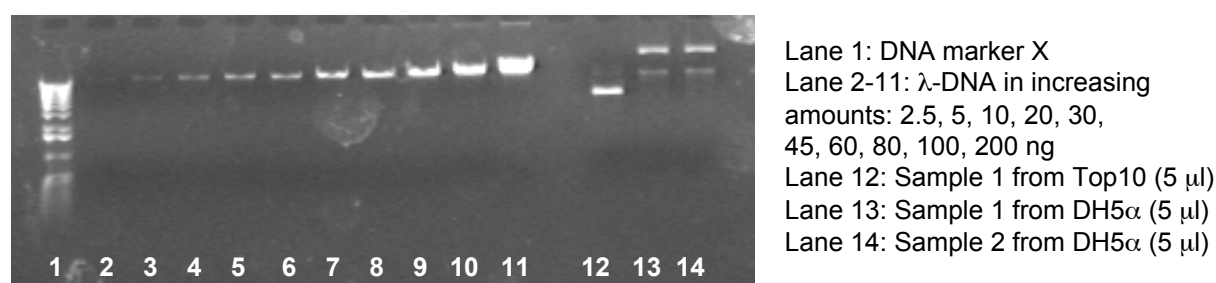


Figure 32. Quantification of miniprep samples isolated from DH5 α and Top10 cells compared to standardized amounts of λ -DNA.

The concentration of miniprep samples T1, D1 and D2 were determined by visual assessment, while quantification of the maxiprep sample was done using the software Quantity One from Biorad Laboratories, Inc. A calibration curve was initially established, from which the concentration of the maxiprep sample could be derived. The results are summarized in table 23.

Table 23. Concentration of the DNA samples as determined by agarose gel electrophoresis analysis.

DNA sample	Conc. (ng/ μ l)
T1	12
D1	8
D2	8
Maxiprep	547

3.2.2 Site-directed mutagenesis

There exist numerous methods for introducing site-specific mutations in a DNA sequence. Three methods were attempted within this work to create point mutations in H1-CRD, namely ExSite PCR-based site-directed mutagenesis system, conventional PCR with mutagenic primers and overlap extension PCR.

3.2.2.1 ExSite PCR-based site-directed mutagenesis system

The commercial kit ExSite PCR-based site-directed mutagenesis system from Stratagene was employed to generate a double substitution of Cys152 and Cys153 to serines. The mutagenesis reaction was carried out in accordance with the recommended protocol, varying the concentration of template and primers, as well as using different DNA polymerases. A total of 4 reaction set-ups (A, B, C and D, see section 2.7.2) were tested and analyzed by agarose gel electrophoresis (Figure 33).

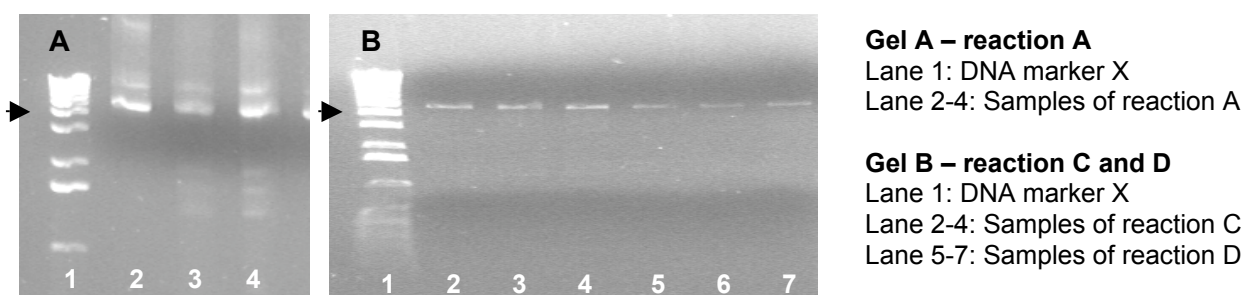


Figure 33. A) Agarose gel analysis of reaction A. Samples taken after PCR (lane 2), *Dpn* 1 digestion (lane 3) and dilution (lane 4). The arrow indicates the 4072 bp band in the marker. B) Agarose gel analysis of

reaction C (lane 2-4) and D (lane 5-7), after PCR, *Dpn* 1 digestion and dilution respectively. The arrow indicates the 4072 bp band in the marker.

DNA fragments around 4072 bp, thought to correspond to the PCR product, could be spotted in all samples after completing the initial PCR run. Reaction A clearly exhibited the most intense band, with additional bands residing at higher molecular weight. Reaction B (not shown), C and D looked essentially the same, showing single bands of decreasing intensity in the same order. Aliquots from the reaction set-ups analyzed after digestion with *Dpn* 1 and dilution highly resembled the initial PCR samples. It was decided to proceed with ligation and transformation of reaction B, C and D. Reaction A was omitted as it failed to give a single clean band, most likely due to excess template, present also after digestion.

Calcium-competent XL1-blue cells were transformed with samples of reaction B, C and D, and grown on selective plates. Randomly chosen single clones were cultivated for plasmid isolation followed by restriction digestion by *Bam* H1 and *Nde* 1, and agarose gel analysis (Figure 34).

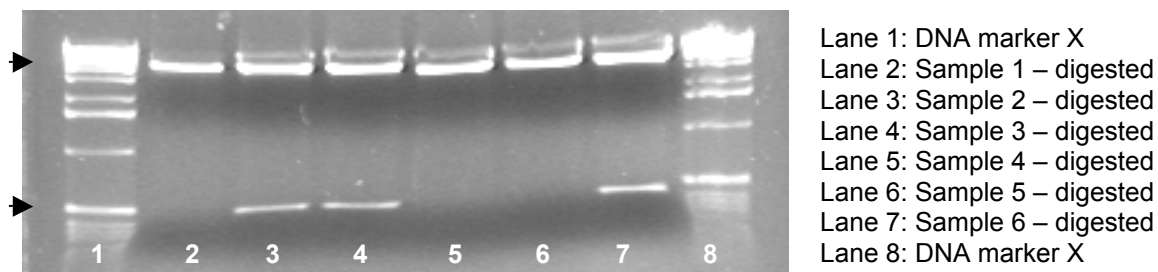


Figure 34. Agarose gel analysis of digested plasmid samples purified from XL1-Blue cells transformed with ExSite mutagenesis reaction B. The upper arrow denotes the 4072 bp band in the marker, while the lower arrow marks the 517/506 bp band.

Digested samples were seen to show two fragments of expected molecular weight, *i.e.* a band around 4072bp corresponding to the vector and a smaller fragment close to 517/506bp matching the insert of 435bp. Samples showing only the higher molecular weight fragment were also observed. Clones such as these, growing on selective plates and containing a plasmid but lacking the target sequence are referred to as false positives (see section 3.2.3).

Digested samples showing the expected fragments were chosen for sequencing, carried out by Microsynth GmbH sequencing group (Balgach). Samples originating from all three reactions were submitted, but without successful results. Out of twelve samples, six retained the wild-type sequence. One sample showed a severe frame-shift at the site of the intended mutations. Four of the samples were observed to partially have incorporated the mutations, but simultaneously caused a frame-shift. Finally, one sample had successfully integrated the mutations, but lacked two preceding amino acids and was therefore not perceived as eligible for downstream applications.

In summary, the ExSite PCR-based site-directed mutagenesis system did not meet the expectations and was therefore excluded as an option from further work.

3.2.2.2 Conventional PCR with mutagenic primers

A traditional approach to create site-specific mutations in H1-CRD was attempted, entailing a single PCR reaction with mutagenic primers followed by ligation of the mutated insert and a predigested vector. The method was used for mutations residing within 10 amino acids from either end of the H1-CRD sequence, *i.e.* single and double substitutions of Cys152 and Cys153 to serines.

The PCR conditions were initially optimized in respect to template amount, primer and dNTP concentration, as well as optimal annealing temperature for the primers, using REDTaq™ polymerase. A set-up utilizing a final concentration of 0.4 μM primer, 0.2 mM dNTP and 16 pg/μl template gave the best results, as proved by agarose gel analysis, and was used throughout the following mutagenesis reactions. The optimal annealing temperature for primer pair H1-CRD mut 152/153 and H1-CRD RW was found to be 61.3°C. This temperature was also used for mutagenesis reactions with primer H1-CRD mut 152 and H1-CRD mut 153, as all three primers had identical melting points, 69.8°C, and were therefore expected to behave in a similar fashion.

Two PCR reactions à 50 μl were run for each intended mutation, with an additional negative control exchanging the template for sterile water. Aliquots of the final PCR products were observed by agarose gel analysis to confirm the purity as well as the size of the DNA fragments prior to pooling of the samples and purification. The purified DNA was digested with restriction enzymes *Bam* H1 and *Nde* 1, and isolated through preparative agarose gel electrophoresis. Recovery of the digested mutated sequence

was confirmed by agarose analysis, while simultaneously comparing it with the vector DNA before ligation (Figure 35).

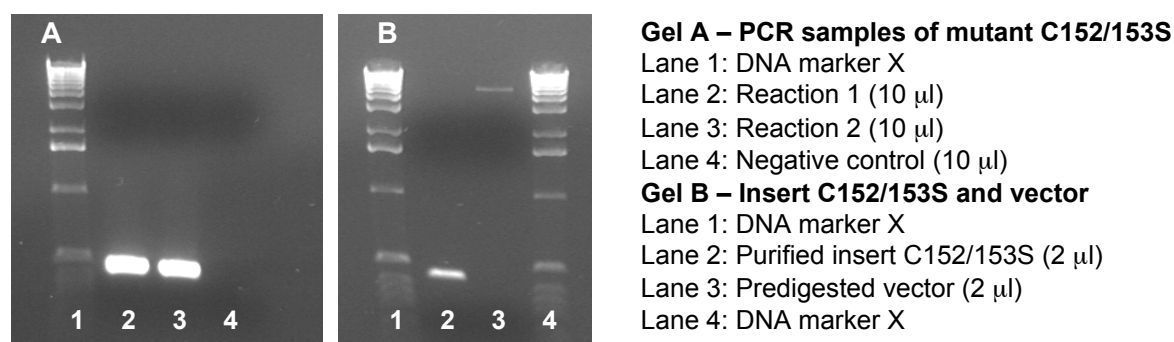
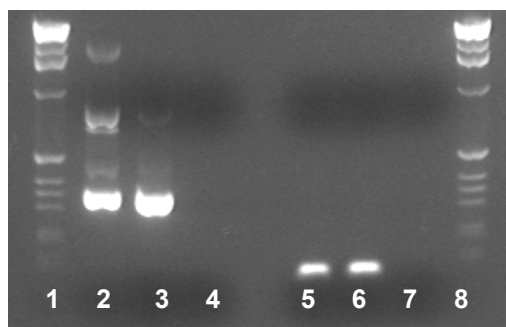


Figure 35. A) Agarose gel analysis of the mutated sequence C152/153S after a PCR reaction with *Pfu* polymerase at 61.3°C. B) Analysis of the gel purified insert (mutant C152/153S) and pET3b vector after restriction digestion.

3.2.2.3 Overlap extension PCR

Overlap extension PCR was used to create mutations located more than 60-70 nucleotides from either end of the H1-CRD sequence. Conventional PCR is not eligible for such mutations as it would require very long primers, risking unspecific binding and inaccurate DNA synthesis. As a result substitution of Cys164, Pro237, Glu238, His256, Asp259 and Asp260, all fulfilling mentioned criteria, was performed by overlap extension PCR.

In a first step, two reactions were carried out in parallel, generating overlapping halves of the target sequence, both carrying the desired mutation(s). Similarly to the conventional PCR set-up, the conditions were initially optimized with REDTaq™ polymerase, showing that a final concentration of 0.4 µM primer, 0.2 mM dNTP and 16 pg/µl template yielded the best results. The optimal annealing temperatures are summarized in table 24. PCR products intended for downstream applications were ultimately generated using *Pfu* polymerase and analyzed to confirm that the fragments showed the expected size (Figure 36). Occasionally, additional bands of higher molecular weight were observed, but these were considered to be of less importance as the desired fragments were isolated and purified by preparative agarose gel electrophoresis. Hence, contamination with PCR products larger or smaller than expected was avoided.



Lane 1: DNA marker X
 Lane 2: Fragment A – reaction 1
 Lane 3: Fragment A – reaction 2
 Lane 4: Fragment A – negative control
 Lane 5: Fragment B – reaction 1
 Lane 6: Fragment B – reaction 2
 Lane 7: Fragment B – negative control
 Lane 8: DNA marker X

Figure 36 Agarose gel analysis of fragment A and B of mutant H256E after the first step of overlap extension PCR. The expected size of the fragments is estimated to 351 and 114 nucleotides, respectively.

The second step of overlap extension PCR combines the two fragments, allowing for annealing and extension at the overlapping sequences, with simultaneous amplification accomplished by the flanking primers. In a first trial attempting to generate mutant C164S, the overlap and extension was run as one individual PCR reaction, excluding the flanking primers and only comprising 10 cycles of annealing and elongation. The annealing temperature was set to 61°C, corresponding to the optimized conditions of the first PCR reaction. However, the set-up failed to achieve an overlap extension, resulting only in amplification of fragment B, the larger half produced in the first step. Agarose gel analysis clearly showed the final PCR product and fragment B to be of equal size, while the positive control could be spotted at higher molecular weight (*Figure 37A*).

The second step was repeated to optimize the conditions, running 30 cycles of annealing and elongation. A temperature gradient ranging from 55°C to 65°C was tested in different rounds. Reactions with or without (w/o) addition of the flanking primers were run in parallel. In addition, the ratio of fragment A and B was also varied, trying 1:1 and 3:1.

Agarose gel analysis could show that none of the reactions run without flanking primers were able to produce a visible product (*Figure 37B*). In contrast, reactions including the primers showed a very intense and broad band between 517/506 and 369 bp. Due to the position, the PCR product was deemed likely to contain the expected full-length fragment, but it also appeared to be heavily contaminated with fragment B. The success of the isolation and purification of the desired product from such a sample seemed improbable. Bands of lower molecular weight and lower intensity could also be observed and were believed to represent non-completed products.

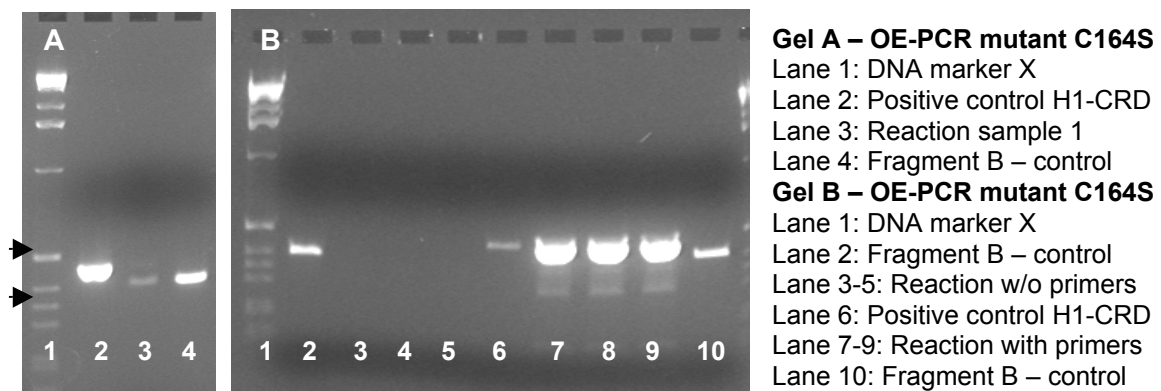


Figure 37. A) Agarose gel analysis of the second step of the overlap extension PCR of mutant C164S without addition of flanking primers. The reaction sample (lane 3) can be seen to correspond to the size of fragment B (lane 4) rather than the expected full-length fragment (lane 2). The upper arrow marks the 517/506 bp band in the reference, while the lower arrow points at the 396 bp band. B) Agarose gel analysis of the second step of the overlap extension PCR of mutant C164S. The reactions in lane 3-5 were run without addition of flanking primers at 59.9°C, 57.2°C and 55.0°C respectively. In contrast, the reactions in lane 7-9 were run with addition of the flanking primers at the same temperatures as mentioned above.

Yet another strategy was tested with addition of the flanking primers after 10 cycles of annealing and elongation and then continuing to run the PCR for another 20 cycles. A temperature gradient was applied, running the reactions at 60°C, 56.2°C, 52°C and 50°C respectively. Furthermore, fragment B was diluted 1/10 from which 3 μ l was added to the reaction set-up. Fragment A was not diluted and added as 1 μ l to the sample mix. Following the PCR, the samples were analyzed by agarose gel electrophoresis together with a positive control and fragment B. This strategy proved to be very successful and also an exceptionally illustrative example of the importance of optimal annealing temperature as seen in figure 38.

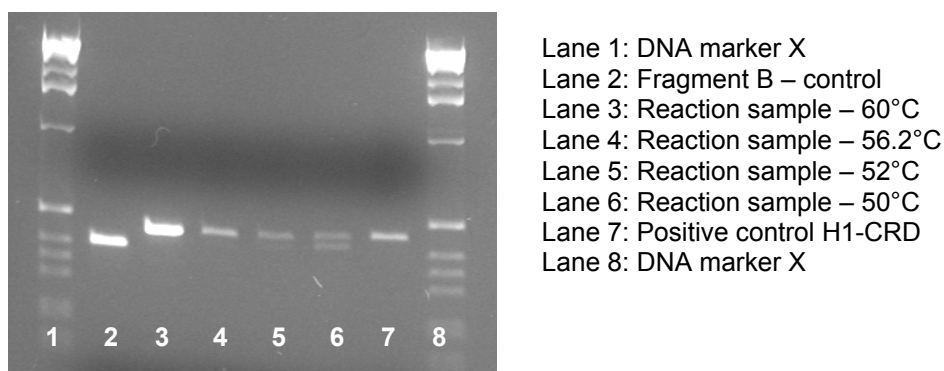


Figure 38. Agarose gel analysis of the second step of the overlap extension PCR of mutant C164S. The reactions were run at different temperatures, adding the flanking primers after 10 cycles of annealing and elongation, followed by another 20 cycles.

While all four reactions (lane 3-6) clearly demonstrated the ability to produce the expected full-length fragment, the intensity of the product(s) varied significantly. A clear correlation could be seen between increased band intensity and increased annealing temperature. Moreover, at the lowest temperature tested, an additional band was seen, corresponding well to the size of fragment B. This second band further supports the necessity to optimize the annealing temperature, in order to avoid contamination with incorrect products or unspecific primer binding. The second vital issue contributing to the success of the set-up is the time point at which the flanking primers are added. By excluding the primers during the first 10 cycles, the two fragments are allowed to anneal and extend without any external interference. However, it is crucial to add the primers before completing the PCR run to achieve amplification and consequently a sufficient amount of product for detection and downstream applications. Due to the most satisfying results obtained when using this set-up for mutant C164S, it was decided to apply it in further mutagenesis reactions carried out by overlap extension PCR. Nonetheless, optimal annealing temperature and fragment concentration had to be determined empirically for each individual reaction (*Figure 39*).

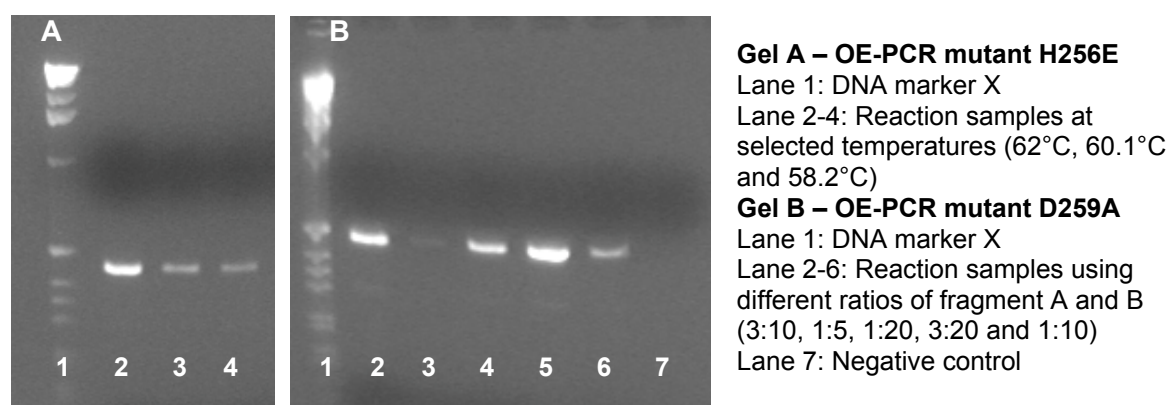


Figure 39. A) Agarose gel analysis of the second step of the overlap extension PCR of mutant H256E. A temperature gradient was applied, showing highest yields at 62°C (lane 2). B) Agarose gel analysis of the second step of the overlap extension PCR of mutant D259A. The temperature was kept constant at 61°C, while the ratio of fragment A and B from the first step was varied. Best result is observed at a ratio of 3:20 (lane 5).

The annealing temperatures used in the first and second step of the overlap extension PCR reactions can be found summarized in table 24.

Table 24. Optimized annealing temperatures used in the first and second step of overlap extension PCR reactions.

Mutation	Annealing temp. 1st step	Annealing temp. 2nd step
C164S	61°C	60°C
P237A	60°C	62°C
E238G	58.5°C	61.4°C
H256E	59.5°C	62°C
D259A	59.9°C	61.8°C
D260A	59.9°C	61.8°C

Duplicate reactions were run under optimized conditions and pooled after confirming presence of pure products by agarose gel analysis. The DNA was purified and digested by restriction enzymes, followed by isolation and purification by preparative agarose gel electrophoresis prior to ligation.

3.2.3 Transformation by electroporation

Purified inserts, originating either from conventional PCR or overlap extension PCR, were ligated with predigested and dephosphorylated pET3b vector O/N. Two set-ups were used, varying the insert-vector ratio from 1:1 to 2:1.

An aliquot of electrocompetent DH5 α -cells were transformed with 5 μ l desalted ligation sample and plated on selective LB-agar plates. Colony formation could be observed after incubation O/N, seemingly independent of the ligation set-up of the sample used for transformation. Randomly chosen single colonies were grown for subsequent plasmid isolation and restriction digestion. Agarose gel analysis was used to distinguish between false and true positives (*Figure 40*).

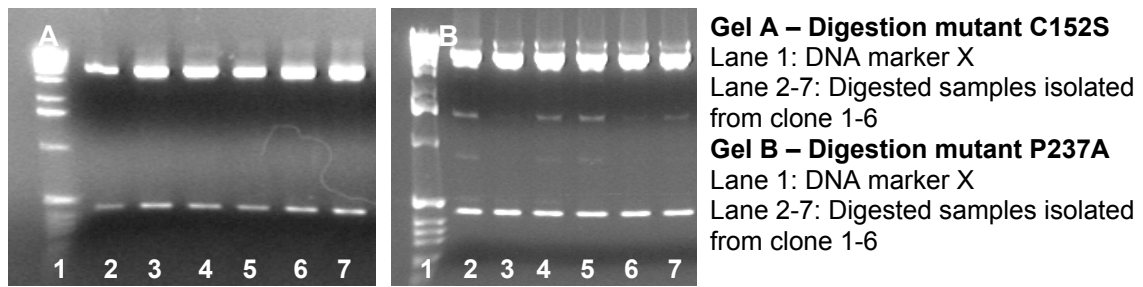


Figure 40. A) Agarose gel analysis of digested miniprep samples of mutant C152S, isolated from DH5 α cells. The samples show clean bands of expected molecular weight corresponding to that of the insert and the vector. B) Agarose gel analysis of digested miniprep samples of mutant P237A, isolated from DH5 α cells. Bands of intermediate molecular weight can be observed in some samples in addition to those of the insert and vector (lane 2, 4, 5 and 7).

Ideally, restriction digestion of the isolated plasmid should result in two bands corresponding to the vector and the cloned insert. Such a sample signals a true positive, *i.e.* the clone pass the selection as a result of the internalized vector, which in turn succeed to fuse with the insert during the ligation. In contrast, a false positive is a clone, which has taken up a self-ligated vector and consequently survives the selection, but fails to present the cloned insert after digestion. A third alternative is that of samples presenting additional bands. Unexpected bands can be a sign of supplementary cleavage sites or star activity. The latter can occur when *e.g.* excess glycerol or enzyme is present [131]. Use of *Bam* H1 is known to cause star activity [132], but can be avoided when adhering to the suppliers' recommendations and including BSA in the digestion set-up [133].

Positive clones were sequenced either by Microsynth or the in-house facility in Biocenter.

3.2.4 Sequencing results

The sequencing results were obtained as nucleotide sequences and converted into protein sequences using the Translate tool [134] available through the ExPASy proteomics server [130]. Thus, correct mutations could easily be confirmed while simultaneously ascertaining the accuracy of the rest of the sequence. In addition, the obtained sequences were aligned with the WT H1-CRD sequence utilizing the program LALIGN, also accessible through ExPASy [135]. Alignment serves as a great aid in highlighting potential differences and was used to validate the correctness of the mutated sequence (*Figure 41*).

WT	M	G	S	E	R	T	C	C	P	V	N	W	V	E	H	E	R	S	C	19	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Mc7	M	G	S	E	R	T	S	S	P	V	N	W	V	E	H	E	R	S	S	19	
WT	Y	W	F	S	R	S	G	K	A	W	A	D	A	D	N	Y	C	R	L	38	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Mc7	Y	W	F	S	R	S	G	K	A	W	A	D	A	D	N	Y	C	R	L	38	
WT	E	D	A	H	L	V	V	V	T	S	W	E	E	Q	K	F	V	Q	H	57	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Mc7	E	D	A	H	L	V	V	V	T	S	W	E	E	Q	K	F	V	Q	H	57	
WT	H	I	G	P	V	N	T	W	M	G	L	H	D	Q	N	G	P	W	K	76	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Mc7	H	I	G	P	V	N	T	W	M	G	L	H	D	Q	N	G	P	W	K	76	
WT	W	V	D	G	T	D	Y	E	T	G	F	K	N	W	R	P	E	Q	P	95	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Mc7	W	V	D	G	T	D	Y	E	T	G	F	K	N	W	R	P	E	Q	P	95	
WT	D	D	W	Y	G	H	G	L	G	G	G	E	D	C	A	H	F	T	D	114	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Mc7	D	D	W	Y	G	H	G	L	G	G	G	E	D	C	A	H	F	T	D	114	
WT	D	G	R	W	N	D	D	V	C	Q	R	P	Y	R	W	V	C	E	T	133	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Mc7	D	G	R	W	N	D	D	V	C	Q	R	P	Y	R	W	V	C	E	T	133	
WT	E	L	D	K	A	S	Q	E	P	P	L	L	-	-							
	:	:	:	:	:	:	:	:	:	:	:	:	:	:							
Mc7	E	L	D	K	A	S	Q	E	P	P	L	L	-	-							

Figure 41. Alignment of the WT H1-CRD sequence with that of clone 7 of mutant C152/153/164S (here denoted Mc7). Identical residues in the two sequences are marked with a semicolon, while lack thereof signals a difference. In addition, the residues differing in the two sequences are denoted by a box.

All planned and attempted mutations could be verified through translation and alignment, demonstrating the capability of the two applied methods and their usefulness in site-directed mutagenesis.

3.2.5 Transformation into calcium competent cells

DNA sequences shown to contain the correct mutation(s) were transformed into calcium competent AD494(DE3) cells and plated on selective LB-agar plates. Colony formation could be observed after incubation O/N. Randomly chosen clones were cultivated and screened for high expression levels, comparing non-induced and induced samples by reducing SDS-PAGE electrophoresis (*Figure 42*).

Important aspects taken into consideration when selecting a clone intended for large-scale protein production were 1) markedly high expression levels of the mutant H1-

CRD, and 2) low expression levels of foreign proteins showing minor variation in the presence of IPTG. One or two clones meeting posed criteria were identified for each mutant protein and conserved as glycerol stocks.

3.2.6 Determination of target protein solubility

The solubility of the mutant proteins was probed according to a protocol established by Qiagen [110]. Samples collected from non-induced and induced cells were compared with soluble and insoluble fractions derived from induced cultures using SDS-PAGE electrophoresis.

Overexpression of mutant H1-CRD could be confirmed by comparing the non-induced and induced samples, supporting previous findings from the small-scale expression analysis. No mutant H1-CRD could be spotted in the soluble fraction, while a distinct band was observed in the insoluble matter (Figure 42). Similar to WT H1-CRD, the mutants of H1-CRD fail to fold correctly and accumulate as insoluble aggregates in the bacteria cells. Consequently, the protein must be denatured in order to appear soluble in solution, followed by renaturation to restore the native structure.

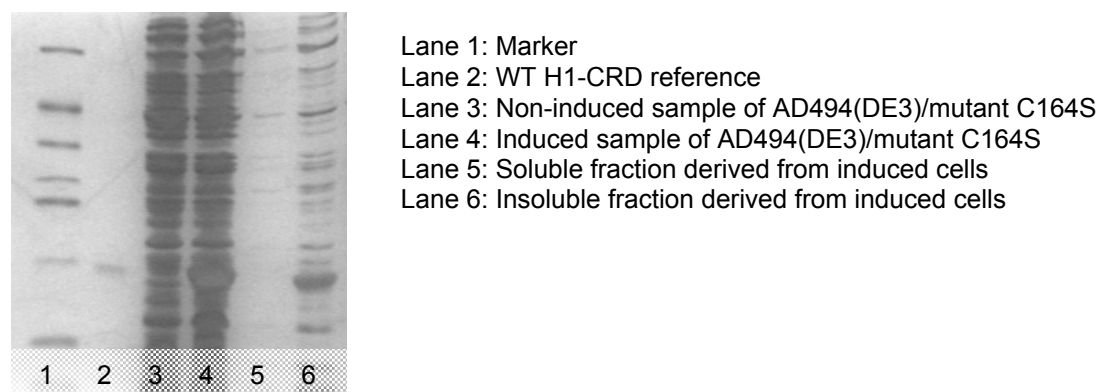


Figure 42. SDS-PAGE analysis of non-induced, induced, soluble and insoluble matter derived from *E.coli* cells transformed with the mutant sequence C164S. The mutant protein, corresponding well to the reference sample of WT H1-CRD, can be spotted in the induced sample, as well as the insoluble fraction. In contrast, no H1-CRD is seen in the soluble fraction (15% gel, silver staining).

3.3 Expression and characterization of cysteine mutants of the ASGP-R H1-CRD

Mutant forms of H1-CRD created to investigate the functional importance of the three cysteines (152, 153, 164) closest to the *N*-terminus, were expressed and purified

according to the method established for the WT protein. Non-reducing SDS-PAGE electrophoresis was used to detect dimer formation, while the functionality was analyzed by the solid-phase competition assay, Biacore and NMR.

3.3.1 Expression and purification of cysteine mutants of H1-CRD

The cysteine mutants of H1-CRD were expressed in *E.coli* and subsequently purified by complete denaturation, refolding and affinity chromatography using galactose as a ligand. The chromatogram profile closely resembled that of WT H1-CRD, exhibiting a distinct, sharp peak during the elution phase. All seven cysteine mutants could be purified through this method, indicating that they all maintained binding affinity for galactose. Eluted fractions were analyzed by SDS-PAGE analysis, showing an apparent enrichment of mutant protein in the fractions corresponding to the peak.

3.3.2 Analysis of dimer content

Non-reducing SDS-PAGE was used to analyze the dimer content of the cysteine mutants after purification by affinity chromatography. The gels were developed with silver staining (Figure 43).

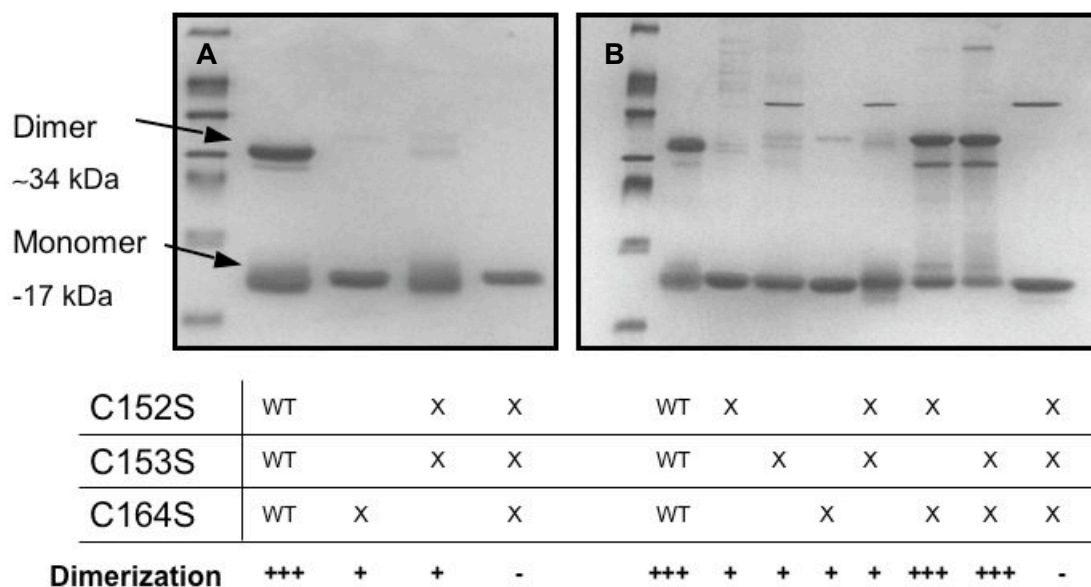


Figure 43. Dimer content of samples analyzed by non-reducing SDS-PAGE after affinity chromatography. A) WT H1-CRD is compared to the three initially created cysteine mutants. The single mutant C164S and double mutant C152/153S only show traces of dimer compared to the WT, while no dimers are seen at all for the triple mutant (15% gel, silver staining). B) Comparison of WT H1-CRD and all seven cysteine mutants. Dimers can be spotted to a higher or lower degree for all protein samples except the triple mutant. Bands of higher molecular weight than expected can be observed in several samples (mutant

C153S, C152/153S and C152/153/164S), most likely the result of contamination with bacterial proteins (15% gel, silver staining).

Three main conclusions can be made from observing the resulting SDS-PAGEs. First, the single mutants show a significant reduction in dimer formation compared to the WT. These proteins with an even number of cysteine will preferentially form intramolecular disulfide bonds. However, intermolecular disulfide bonds can also occur occasionally as seen from the traces of dimers present in the samples. Second, the double mutants show dimerization similar to the WT, except when Cys164 remains as the odd fifth cysteine. It is possible that Cys164 is positioned in such a way that it cannot easily give rise to dimers, which in turn would suggest that it is in fact engaged in an intramolecular disulfide bond in native H1-CRD. No obvious difference can be observed between the C152/164S and C153/164S mutants, indicating that both Cys152 and Cys153 are capable of reacting with a free thiol in a neighboring subunit. Finally, replacement of all three cysteines completely abolishes dimer formation as proved by the triple mutant. Hence, the dimer formation seen upon expression of H1-CRD *in vitro* is indeed caused by the odd cysteine and intermolecular disulfide bonds. Furthermore, the cysteines taking part in the third disulfide bridge, Cys153-Cys164, can also give rise to dimers, as shown by mutant C152S (and the other two single mutants). This observation together with the successful affinity chromatography purification of the double and triple mutants suggests that the formation of the third disulfide bond is not crucial for the binding affinity for Gal. In contrast, the remaining four cysteines in the subunit do not take part in intermolecular disulfide bonds, indicating that the bridges they form are stable and an absolute necessity for active H1-CRD.

3.3.3 Separation of monomers and dimers

Separation of monomers and dimers was carried out with ion exchange chromatography for all cysteine mutants, also the ones showing only traces of dimers or even none, in order to treat all samples equal (*Figure 44*). The samples with a high dimer content, as confirmed by SDS-PAGE analysis, showed a separation profile closely resembling the WT (see mutant C152/164S in *figure 44*). The monomers eluted first (peak 1) with a tailing shoulder peak (peak 2), followed by the impurity peak (peak 3) and finally the dimer peak (peak 4). On the contrary, samples containing merely traces of dimers were seen to display a single peak corresponding to the monomers (see

mutant C164S in figure 44). Eluted fractions were controlled with non-reducing SDS-PAGE (Figure 45A).

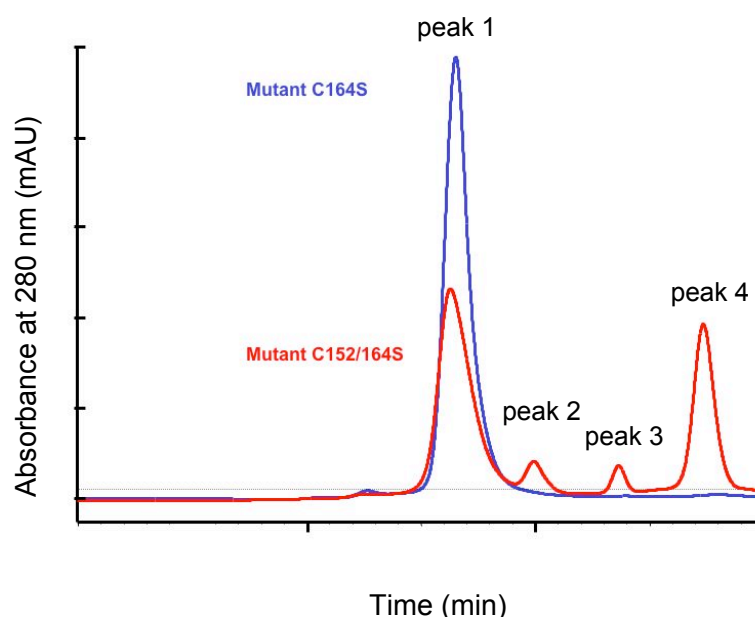
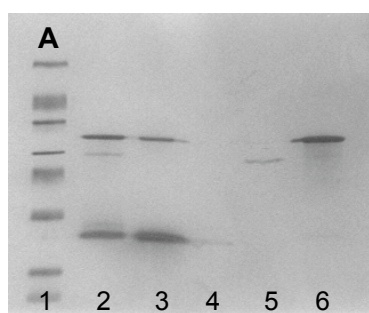


Figure 44. Superimposed chromatograms of mutant C164S in blue and mutant C152/164S in red. The single mutant exhibits only one peak corresponding to the monomers, while the double mutant shows a distinct separation between its monomers and dimers.

In a final step, the monomers were purified and pooled over a GalNAc-Sep-column on HPLC. All mutant proteins could be seen to elute as an individual peak, indicating that they all maintained binding affinity for GalNAc. Samples were analyzed by non-reducing SDS-PAGE to verify the purity of the proteins (Figure 45B).



Lane 1: Marker
Lane 2: Mutant C152/164S after FPLC
Lane 3: Monomer fraction – major peak
Lane 4: Monomeric shoulder fraction
Lane 5: Impurity fraction
Lane 6: Dimer fraction

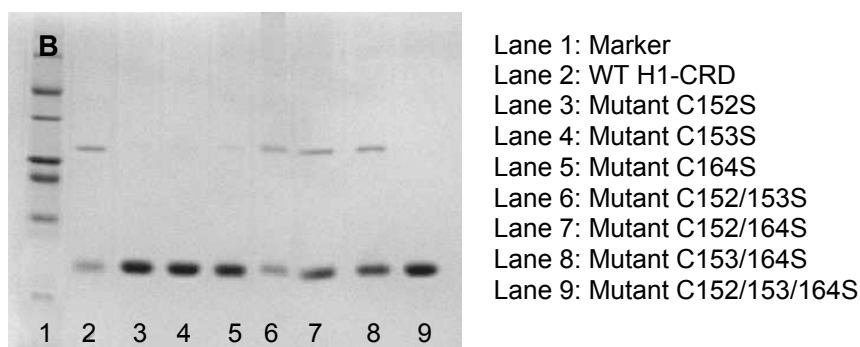


Figure 45. A) Non-reducing SDS-PAGE analysis of fractions collected during HPLC IEC of mutant C152/164S (15% gel, silver staining). B) Non-reducing SDS-PAGE analysis of WT H1-CRD and all seven cysteine mutants after final purification by HPLC affinity chromatography (15% gel, coomassie staining).

3.3.4 Evaluation of the binding affinity of the cysteine mutants

The issue concerning the binding affinity of the cysteine mutants was also addressed within this work. For the proteins to prove useful in further studies, apart from preventing dimerization, it is equally important that the mutations do not affect the binding properties of H1-CRD. Therefore, a non-negotiable demand set for the mutants was that binding affinity and native behavior of the WT protein would be preserved after the alterations. To ensure that this was the case, *i.e.* that the cysteine mutants and WT H1-CRD showed equivalent affinity, a series of assays were carried out for comparison. The triple cysteine mutant was the prime candidate for these studies as it was the only mutant successfully escaping dimer formation. Nevertheless, the other six substituted proteins were also included in some of the experiments to elucidate the role of the third disulfide bridge.

3.3.4.1 Evaluation of the cysteine mutants of H1-CRD using the solid-phase competition assay

Binding affinity of the cysteine mutants for GalNAc was tested using the solid-phase competition assay and compared with that of the WT. Two approaches were adopted, the first measuring relative binding of a GalNAc-polymer to the proteins and the second using a competitive binding set-up allowing for determination of IC_{50} values.

In the first approach, the cysteine mutants and WT H1-CRD were coated onto plates O/N at a concentration of 3 $\mu\text{g/ml}$ protein. The plate was then blocked with BSA prior to incubation with 0.5 $\mu\text{g/ml}$ GalNAc-polymer. No inhibitor was added. The degree of binding was determined by measuring the absorbance at 415 nm after addition of ABTS substrate. Relative binding of GalNAc to the mutants was calculated using WT H1-CRD

as reference for maximum binding. Two independent measurements, carried out within four months, are depicted in the graph below (Figure 46).

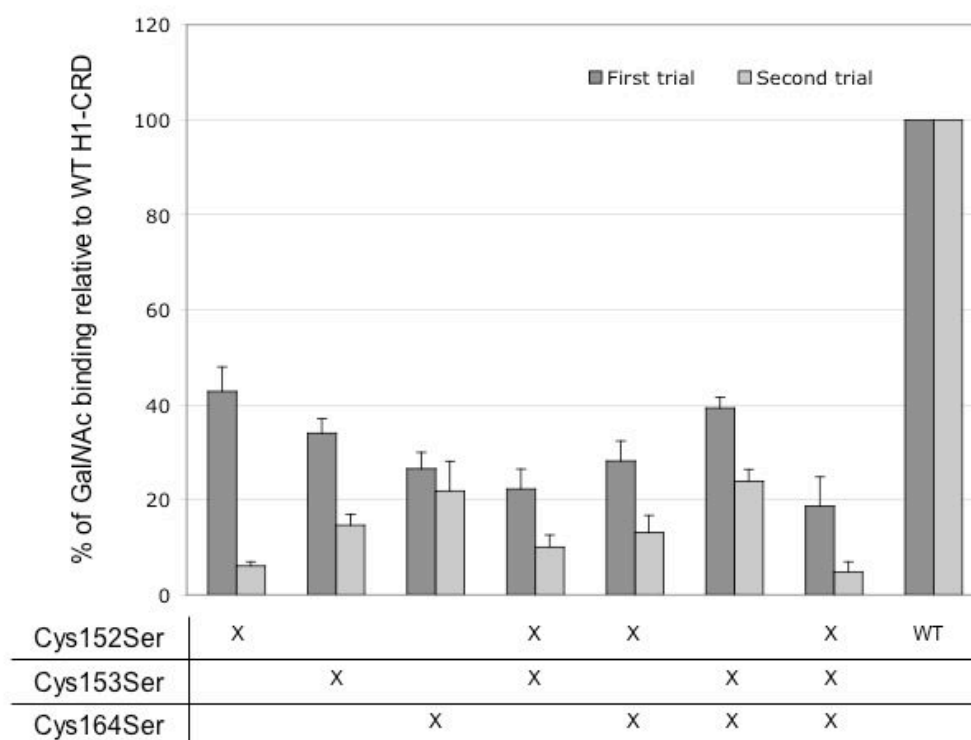


Figure 46. Results of the solid-phase competition assay expressed as percentage of GalNAc-polymer binding relative to that of WT H1-CRD. The dark grey columns represent the results obtained from the first trial. The assay was repeated after approximately four months, the results of which are depicted in light gray.

An obvious distinction in GalNAc binding can be observed when comparing the cysteine mutants and WT H1-CRD. GalNAc clearly binds with lower affinity to the cysteine-substituted proteins, exhibiting less than 50% of the signal intensity of the wild-type protein. Minimum binding is demonstrated by the triple mutant. The single mutants and the double mutants show similar binding affinities, but without revealing a possible dependence on one cysteine or another. Still, it can be concluded that not only is the third disulfide bridge important for maintaining native binding properties, but also the seventh odd cysteine. Moreover, comparison of assays carried out at different occasions demonstrates comparable results, but with a further decrease in signal intensity and binding of the cysteine mutants. Ranking of the mutant proteins remains roughly the same, with the triple mutant consistently showing the lowest degree of binding of all proteins tested. The single and double mutants also display highly similar results, with minor individual changes. The seemingly time-dependent

decrease in binding affinity could originate from reduced stability of the cysteine mutants due to structural alterations, further discussed in section 4.2.2 Furthermore, the mutant proteins were of rather low concentration, less than 200 $\mu\text{g/ml}$, which could also have a negative influence on the protein stability. Taken together, the results suggest that the cysteine mutants are susceptible to inactivation and it is possible that only a fraction of the protein samples are actually active and able to bind GalNAc.

The second approach was carried out similarly to the first, but with two exceptions. The cysteine mutants and WT H1-CRD were coated onto plates O/N at a concentration of 3 $\mu\text{g/ml}$ protein and 20 mM CaCl_2 instead of the previous 1 mM. The calcium concentration was increased as it had been shown to affect the signal intensity (data not shown). The plate was blocked with BSA prior to incubation with 0.5 $\mu\text{g/ml}$ GalNAc-polymer. Monovalent GalNAc, acting as inhibitor, was added in concentrations ranging from 2 mM to 0.06 mM simultaneously with the polymer. Binding was determined by measuring the absorbance at 415 nm after addition of ABTS substrate. The results were plotted as percentage polymer binding against concentration of monovalent GalNAc as seen in figure 47. Maximum binding was represented by polymer binding in the absence of monosaccharide.

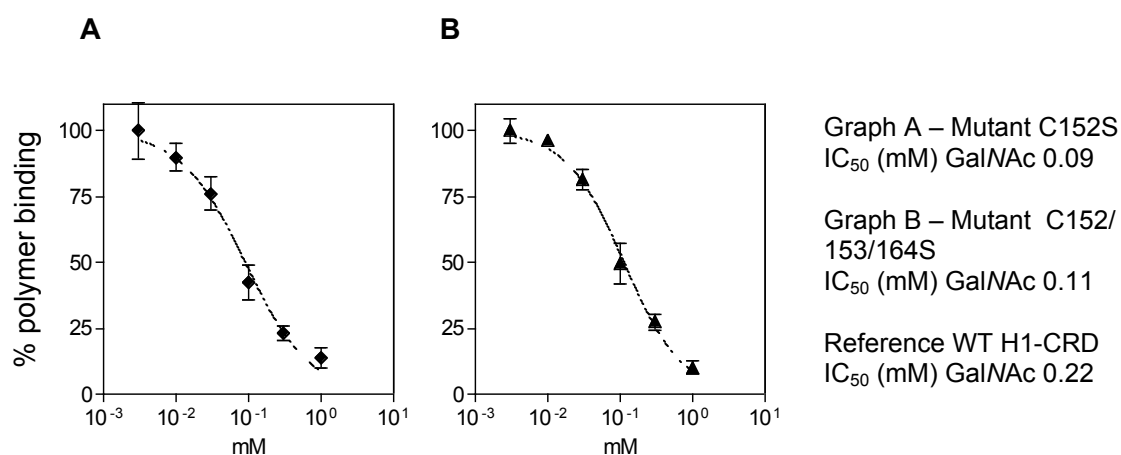


Figure 47. GalNAc-polymer competing with monovalent GalNAc for binding to the cysteine mutants. A) Percentage polymer binding to mutant C152S in the presence of increasing concentrations of monovalent GalNAc. B) Percentage polymer binding to mutant C152/153/164S in the presence of increasing concentrations of monovalent GalNAc.

The resulting graphs, showing a sigmoid dependency curve, were used to determine the IC_{50} of GalNAc binding to the different forms of H1-CRD. To facilitate the

evaluation of the cysteine mutants, relative IC₅₀ (rIC₅₀) values were calculated using WT H1-CRD as reference. The results are summarized in table 25.

Table 25. Summary of rIC₅₀ values derived from the solid-phase competition assay performed with monovalent GalNAc competing with GalNAc-polymer for binding to the cysteine mutants.

Assay (rIC ₅₀)	C152S	C164S	C152/153S	C152/164S	C152/153/164S	WT
1	0.32	0.50	0.53	0.31	0.58	1
2	0.40	0.59	0.52	0.42	0.50	1
3	0.32	0.55	0.46	0.38	0.45	1
X±Std.	0.35±0.05	0.55±0.05	0.50±0.04	0.37±0.06	0.51±0.07	1

All cysteine mutants show rIC₅₀ values below zero, indicating that the monovalent GalNAc is able to inhibit polymer binding to the mutants at lower concentrations than compared to the WT. The GalNAc-polymer, which is expected to bind stronger due to its polyvalent nature, clearly exhibits higher affinity for the wild-type protein. The mutant proteins could experience folding or stability problems, disturbing their structure and as a result affect their capacity for ligand binding negatively. Comparison of the mutants reveals rather consistent values around 0.5, with the exception of a single mutant showing a rIC₅₀ closer to 0.3. Somewhat surprisingly, the triple mutant does neither show the lowest rIC₅₀ nor does it deviate from the other mutants. Conversely, it can be argued that it ought to behave similar to the double mutants, also missing a third disulfide bridge, and not suffer from lack of the odd cysteine. Instead, a marked divergence between the single mutants and the other altered proteins would be expected and is further discussed in section 4.2.2. As an additional remark it should be noted that the cysteine mutants generally gave rise to lower signal intensities than the WT. One explanation for this observation is the lower polymer binding, but it could also be attributed to a lesser fraction of active protein present in the mutant samples.

In summary, the cysteine mutants and WT H1-CRD do indeed differ in binding affinity for GalNAc as proved by the solid-phase competition assay. The ligand consistently shows higher binding to the wild-type protein, while the mutants repeatedly fall short in signaling binding. In addition, the stability of the mutated proteins is questioned as suggested by the reduction in ligand binding over time.

3.3.4.2 Evaluation of the triple cysteine mutant of H1-CRD using the Biacore assay

The triple cysteine mutant was immobilized alongside WT H1-CRD on a Biacore sensor chip using standard amine coupling. Final protein densities obtained were 3400 and 5100 RU of the triple mutant and the wild-type monomers respectively. Dilution series of GalNAc and Gal were injected and binding to the proteins monitored. The resulting sensograms were fitted to a 1:1 binding model and K_D determined (Figure 48).

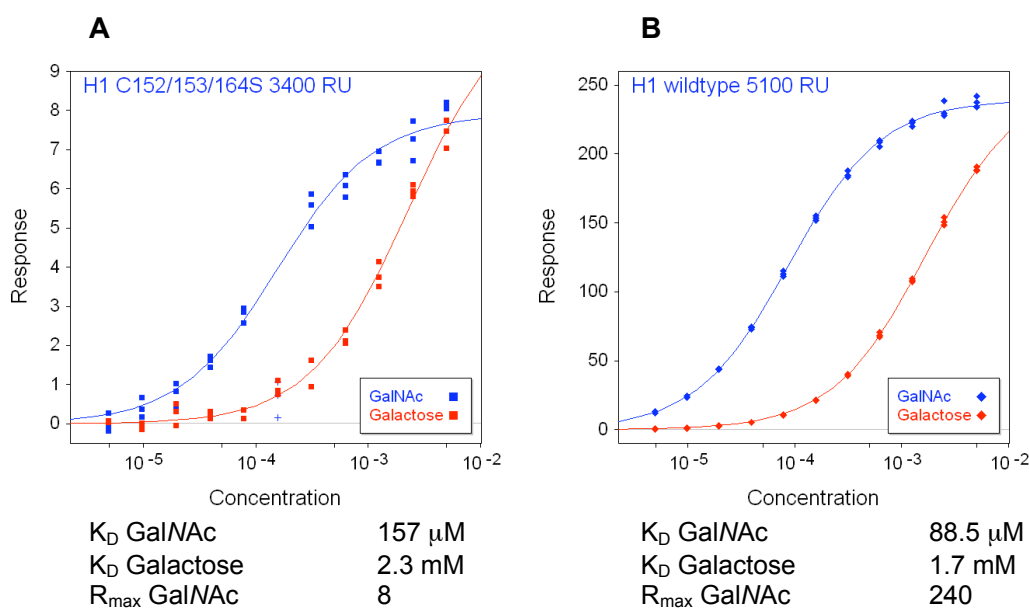


Figure 48. A) Binding of Gal (red) and GalNAc (blue) to the triple cysteine mutant of H1-CRD measured by Biacore. B) Binding of Gal (red) and GalNAc (blue) to WT H1-CRD measured by Biacore.

A conspicuous difference can be observed between the two resulting graphs. The signal intensity of the triple cysteine mutant is significantly lower compared with that of WT H1-CRD. While the maximum response recorded for GalNAc binding to the native protein is around 240, the substituted protein gives rise to a mere 8. The experimentally determined K_D values of GalNAc also differs with 88.5 μ M and 157 μ M for WT H1-CRD and the mutant respectively. However, the accuracy of the latter value is difficult to conclude due to the low signal response. Nevertheless, even though complicated to quantify, the triple cysteine mutant clearly suffers from problems hampering its activity, e.g. decreased affinity, deactivation and/or stability problems *etc.*

3.3.4.3 Evaluation of the triple cysteine mutant of H1-CRD using NMR

T1rho experiments were performed with GalNAc interacting with the triple cysteine mutant and compared to reference spectra of GalNAc with the wild-type protein (see section 3.1.4.3). Binding was observed on a qualitative level with a faster decay of magnetization for the GalNAc in the presence of the mutant protein than in its absence (Figure 49).

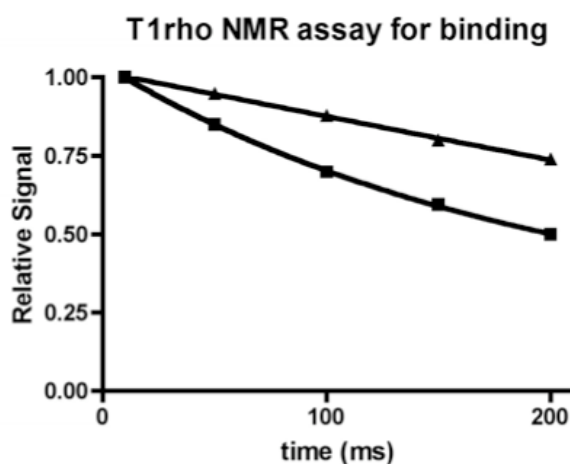


Figure 49. T1rho experiments with GalNAc binding to the triple mutant. The squares (■) represent GalNAc bound to the mutant protein, while the triangles (▲) symbolizes the free GalNAc.

Although the spectra obtained for the mutant showed good agreement with the reference spectra (Figure 28), the results cannot be used for a quantitative measure of the binding.

3.4 Investigation of the binding site of the ASGP-R H1-CRD

Mutant forms of H1-CRD, created to investigate the functional importance of individual amino acid residues in the binding site, were expressed and purified according to the method established for the wild-type protein. The functionality of the mutant proteins was analyzed by the solid-phase competition assay and Biacore.

3.4.1 Expression and purification of binding site mutants of H1-CRD

The binding site mutants of H1-CRD were expressed in *E.coli* and subsequently purified by complete denaturation, refolding and affinity chromatography using galactose as a ligand. The chromatographic profile closely resembled that of the wild-type, exhibiting a distinct, sharp peak during the elution phase. All five binding site mutants, *i.e.* P237A,

E238E, H256E, D259A and D260A, could be purified through this method, indicating that they all maintained binding affinity for galactose. Eluted fractions were analyzed by SDS-PAGE analysis, showing an apparent enrichment of mutant protein in the fractions corresponding to the peak. For mutant H256E, the flow-through and the eluted peak was seen to be of similar height, indicating a high fraction of non-bound proteins. In addition, SDS-PAGE analysis showed a protein band corresponding to the mutant protein in the flow-through, presumably misfolded. In an attempt to recover the lost protein, the flow-through fraction was denatured and refolded once more. Affinity chromatography was repeated, showing a small peak during the elution phase. SDS-PAGE analysis confirmed that the peak consisted of mutant H256E, but the amount was estimated to be rather low. Therefore, a second denaturation and refolding procedure was deemed to be of minor significance and contribution, and excluded from further purification procedures.

Separation of monomers and dimers was carried out with HPLC IEC, showing good correlation with the separation profile seen for WT H1-CRD. Monomeric protein was seen to elute first, followed by dimers. In a final step, the monomers were purified and pooled over a GalNAc-Sep-column on HPLC. All mutant proteins could be seen to elute as an individual peak, indicating that they all maintained binding affinity for GalNAc.

3.4.2 Evaluation of the binding site mutants of H1-CRD using the solid-phase competition assay

The solid-phase competition assay was initially carried out with binding site mutant P237A and H256E to estimate the relative binding of GalNAc-polymer to the proteins. The two mutant proteins and WT H1-CRD were coated onto plates O/N at a concentration of 3 µg/ml protein. The plate was subsequently blocked with BSA, followed by incubation with 0.5 µg/ml GalNAc-polymer. No inhibitor was added. The absorbance at 415 nm was measured after addition of ABTS substrate. Relative binding of GalNAc to the mutants was calculated using WT H1-CRD as reference of maximum binding. Two independent measurements are depicted in the graph below (*Figure 50*).

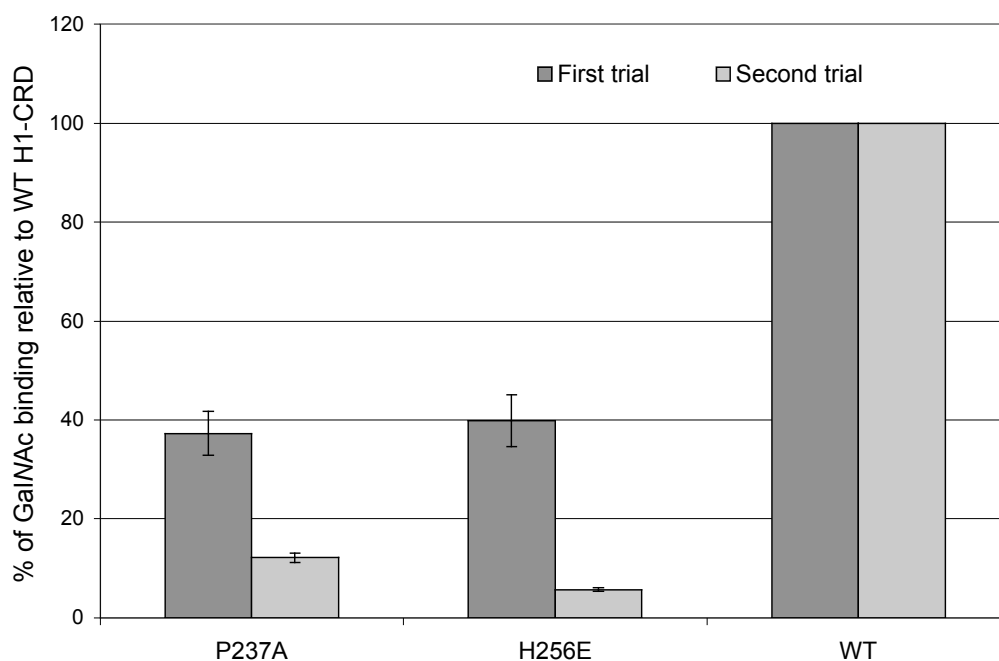


Figure 50. Results of the solid-phase competition assay performed with mutant P237A and H256E, expressed as percentage of GalNAc-polymer binding relative to that of WT H1-CRD. The dark grey columns represents the results obtained from a first trial, while the light grey values are derived from a second attempt.

Similar to the cysteine mutants, GalNAc binds with decreased affinity to the binding site mutants compared to the wild-type protein. The first assay showed a relative binding of 37% and 40% for P237A and H256E respectively. In a repeated assay six months later, the binding was significantly lower for both mutants. P237A displayed a mere 12% binding, while H256E was even lower at 5%. The substantial variation in the experimental data makes it difficult to accurately determine the degree of GalNAc binding. Nonetheless, while unable to provide a reliable measure of the binding, these results suggest that the mutant proteins do differ from the wild-type in respect to ligand binding and/or protein stability. As previously remarked for both the WT monomers purified by RP and the cysteine mutants, the binding site mutants might suffer from inactivation, leaving only part of the protein sample active.

The second assay approach was undertaken for all five binding site mutants and performed in an identical manner to that of the cysteine mutants. The proteins were coated O/N at a concentration of 3 $\mu\text{g/ml}$ protein and 20 mM CaCl_2 . Blocking with BSA was followed by simultaneous incubation of 0.5 $\mu\text{g/ml}$ GalNAc-polymer and monovalent GalNAc. The latter was added in concentrations ranging from 2 mM to

0.06 mM. The absorbance at 415 nm was measured after addition of ABTS substrate. The results were plotted as percentage binding of the polymer against concentration of monovalent GalNAc (Figure 51).

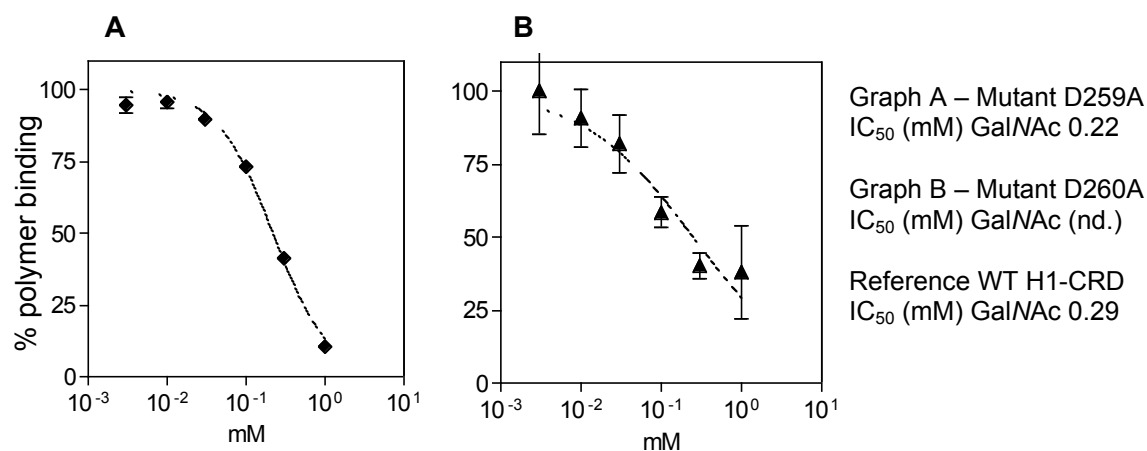


Figure 51. GalNAc-polymer competing with monovalent GalNAc for binding to binding site mutants of H1-CRD. A) Percentage polymer binding to mutant D259A in the presence of increasing concentrations monovalent GalNAc. B) Percentage polymer binding to mutant D260A in the presence of increasing concentrations of monovalent GalNAc. A reliable IC₅₀ value could not be determined (nd) from the graph.

As for the cysteine mutants, IC₅₀ of monovalent GalNAc binding to the mutants was determined and expressed as rIC₅₀ using WT H1-CRD as reference. The results are summarized in table 26.

Table 26. Summary of rIC₅₀ values derived from the solid-phase competition assay performed with monovalent GalNAc competing with GalNAc-polymer for binding to the binding site mutants.

Assay (rIC ₅₀)	P237A	E238G	H256E	D259A	D260A	WT
1	0.75	2.89	0.57	0.71	nd.	1
2	0.72	4.19	0.54	0.58	nd.	1
3	0.87	5.85	nd.	0.74	nd.	1
χ±Std.	0.78±0.08	4.31±1.48	0.56±0.02	0.68±0.09	(-)	1

Of the five binding site mutants, four display an average rIC₅₀ below one, indicating that the GalNAc polymer binds the mutant proteins with lower affinity compared to WT H1-CRD. Two mutants, P237A and D259A exhibit values fairly close to one, suggesting that the effect of the mutations is minor. The residues in question have not previously been implicated in ligand binding and could be of structural rather than

functional importance, further discussed in section 4.3.1. Mutant H256E also show a rIC_{50} below one, as can be expected. Previous studies have identified His256 as a critical residue in the selectivity and binding of GalNAc [7]. Binding site mutant D260A failed to produce convincing results. The signal intensity was very low despite prolonged detection time, complicating the evaluation. The resulting graphs would barely show a sigmoid curve. Adding to the uncertainty is the variability in absorbance measured for the replicates of different concentrations of monovalent GalNAc which were tested. Although the experimental data suggest that the binding affinity of mutant D260A for GalNAc is significantly decreased, a quantitative measure could not be determined. Finally, one of the five binding mutants appeared to bind the GalNAc polymer with a higher affinity than WT H1-CRD. Mutant E238G displayed exceptionally strong signals in the assay, resulting in high IC_{50} values and rIC_{50} s above one. However, the rIC_{50} s showed a rather high variation and while it can be said that GalNAc binds the mutant protein roughly three times better than the wild-type, a second assay is needed to corroborate these findings.

3.4.3 Evaluation of the binding site mutants of H1-CRD using the Biacore assay

Mutant P237A was immobilized alongside WT H1-CRD on a Biacore sensor chip, resulting in protein densities of 5300 and 5100 RU for the mutant and wild-type monomers, respectively. Dilution series of GalNAc and Gal were injected and the obtained sensograms fitted to a 1:1 binding model (*Figure 48B, Figure 52*).

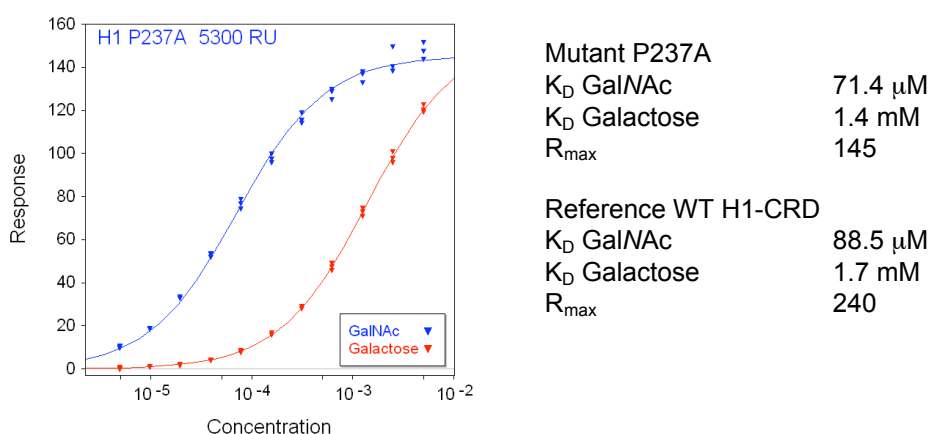


Figure 52. Binding of Gal (red) and GalNAc (blue) to mutant P237A as measured by Biacore. The reference data are derived from the graph presented in section 3.3.4.2.

Mutant P237A can be seen to show binding affinities in the same range as WT H1-CRD, even though the signal intensity admittedly is lower. The mutant protein displays a maximum response of 145 compared to 240 of the wild-type. Despite the difference in signal intensity, the K_D values for GalNAc showed good agreement with 88.5 μM and 71.4 μM for WT H1-CRD and the mutant respectively.

3.4.4 Probing pH-dependent ligand binding of mutant H256E

Ligand binding to mutant H256E was also studied at different pHs to deduce the pH-dependency. His256 in H1-CRD has been implicated as an important residue for ligand release at endosomal pH [96]. Ligand binding is likely to be interrupted due to protonation of the histidine and subsequent destabilization of a calcium- and ligand binding residue in the binding site. A glutamate, the corresponding residue in H2-CRD, is expected to be less influenced by the endosomal pH 5.4 as it would remain deprotonated and able to accept stabilizing hydrogen bonds. To verify this assumption, mutant H256E was analyzed by the solid-phase assay and Biacore at different pHs.

3.4.4.1 Probing pH-dependent binding affinity with the solid-phase competition assay

The solid-phase competition assay was carried out according to the first approach, only adding polymer to the protein. In order to investigate the effect of pH on binding, the polymer was prepared in buffer of different pHs, ranging from 7.4 to 5. Each pH was tested with four protein replicates. Following incubation and washing, the absorbance at 415 nm was measured after addition of ABTS substrate. The results were depicted in separate graphs for the mutant protein and the WT (*Figure 53*).

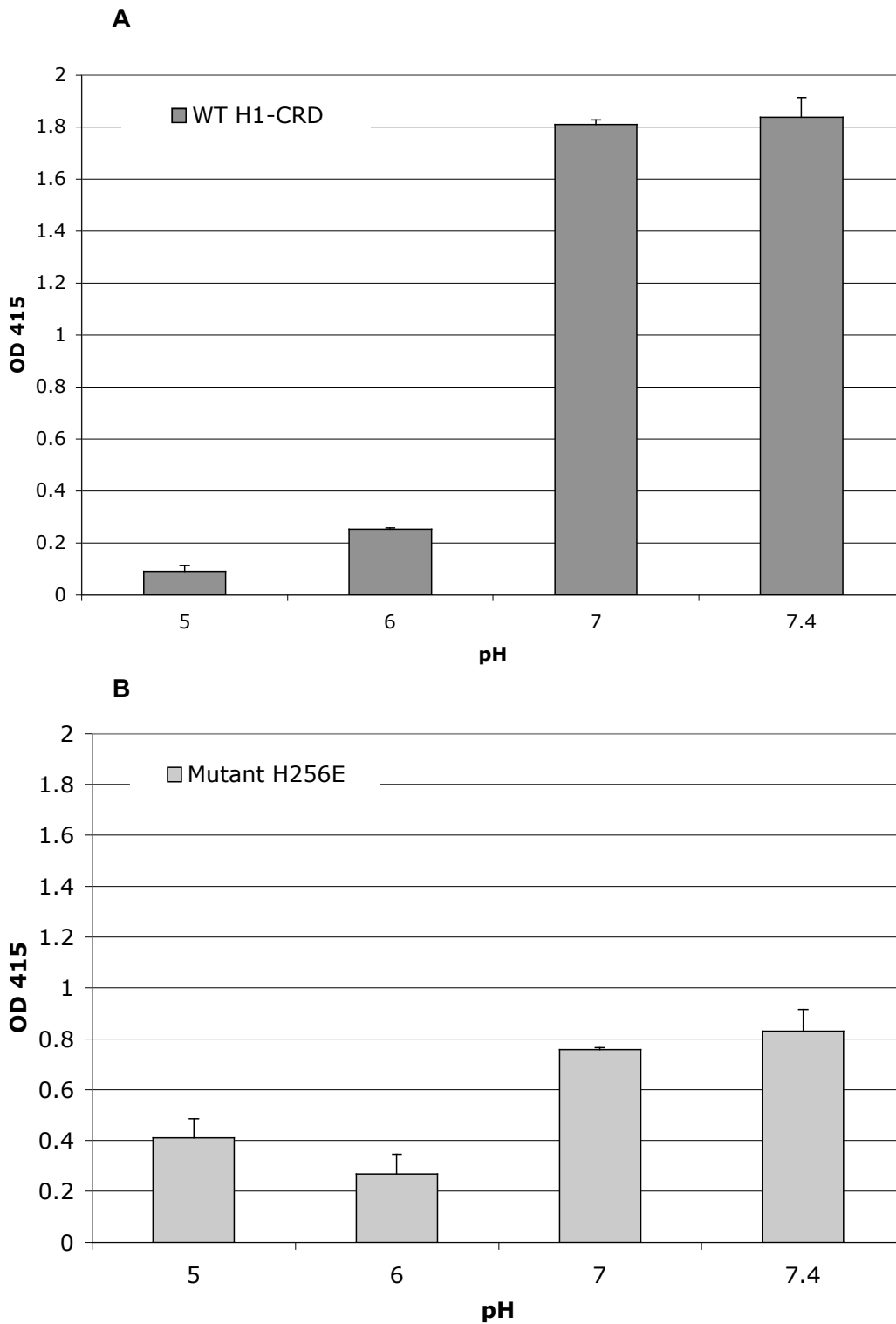


Figure 53. A) Binding of GalNAc polymer to WT H1-CRD measured at different pHs. B) Binding of GalNAc polymer to mutant H256E measured at different pHs. Note that the graphs are only qualitatively comparable, and not quantitatively. The mutant protein required longer detection time than the WT, which in turn would have reached saturation during the corresponding time period.

The results obtained for the WT showed a significant decrease in signal intensity accompanying the decrease in pH. Binding of the GalNAc polymer was reduced as the

pH was lowered beyond 7. At pH 5, only a tenth of the binding at physiological pH could be detected. A decrease in binding could also be observed for mutant H256E, but not to the same extent. The lowest signal intensity was detected at pH 6, approximately a third of that seen at pH 7.4. In this context it should also be remarked that the results of the two proteins cannot be directly compared. Since mutant H256E suffered from low signal response, the development time had been prolonged. In contrast, the WT gave rise to strong signals within a fairly short time, threatening to reach saturation if extended further. Still, even though the absolute signal values are not comparable, the profiles are.

3.4.4.2 Probing pH-dependent binding affinity using Biacore

WT H1-CRD and two independently produced batches of mutant H256E were immobilized on a Biacore sensor chip. Even though all three protein samples were of similar concentrations, both mutants required prolonged immobilization times, 20 vs. 5 minutes. Still, the final protein density obtained was lower for the mutants, 3000 RU compared to 5300 RU for the wild-type. GalNAc was screened in buffers at different pHs, ranging from 7.4 to 5.5, and both affinity and intensity was monitored (Figure 54). K_D of GalNAc was determined for each individual pH tested.

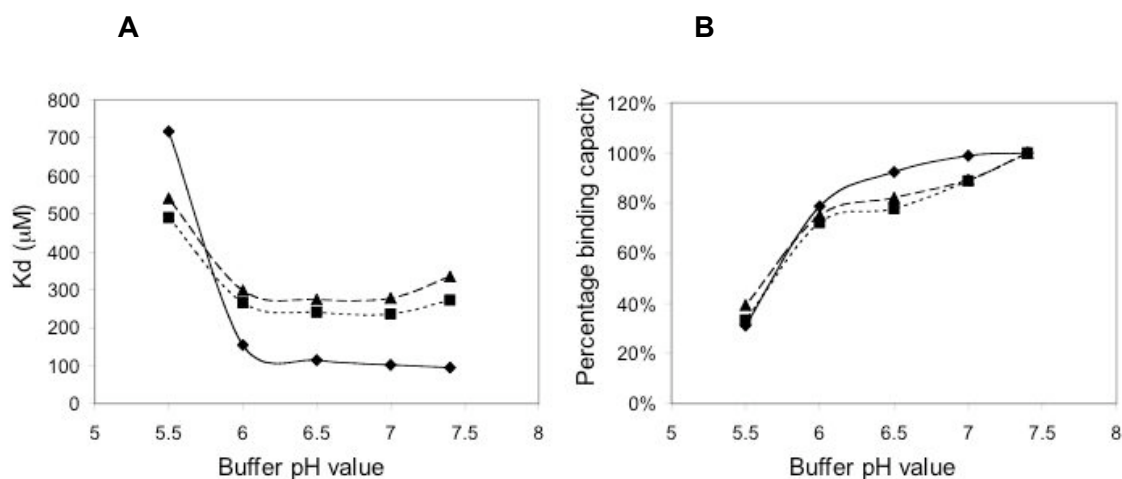


Figure 54. A) GalNAc binding affinity of WT H1-CRD (full line) and mutant H256E at pH ranging from 7.4 to 5.5. Two separately prepared samples of mutant H256E were included, represented by the two dashed lines. B) Percentage GalNAc binding capacity of WT H1-CRD and mutant H256E at pH ranging from 7.4 to 5.5.

All H1-CRD samples could be seen to show a dramatic loss of affinity below pH 6.0. The K_D is stable between pH 7.4 and 6.0, but deviates noticeably once the pH is

lowered further. Comparison of the mutant and the wild-type revealed that, although affected, the former appeared less sensitive to the drop in pH. This observation is further supported by the fact that the two mutant samples showed close to identical results. However, similar to the triple cysteine mutant, mutant H256E also suffers from the drawback of low signal intensity with a maximum response around 23 compared to 263 of WT H1-CRD.

3.5 Expression and purification of isotope labeled H1-CRD

Isotope labeled protein has shown to be invaluable for structural and dynamic studies with biomolecular NMR. By incorporating isotopes such as ^{13}C and ^{15}N , heteronuclear, multidimensional NMR experiments can be performed, permitting routinely analysis of proteins on the order of 20 kDa in molecular weight [83]. In this project, a method to generate uniformly isotope labeled H1-CRD was developed, with the further aim to study protein-ligand interactions and solve the protein structure by NMR.

3.5.1 Expression of isotope labeled H1-CRD using a two-stage protocol

Expression of isotope labeled H1-CRD was initially attempted by using the two-stage protocol developed by Marley *et al.*, [90]. *E.coli* AD494(DE3) and BL21(DE3) cells, transformed with pET-3bH1, were grown to high density in rich medium and subsequently exchanged into minimal medium. After 1.5-2.0 hours' incubation, allowing for depletion of remaining unlabeled nutrients, expression was induced by addition of IPTG. Growth and expression levels were monitored throughout the experiment. Non-induced and induced samples were lysed and denatured prior to analysis by reducing SDS-PAGE (Figure 55).

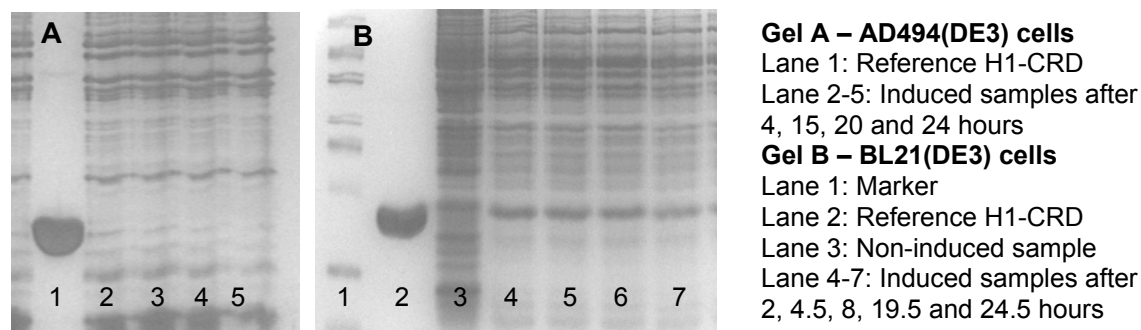


Figure 55. A) Reducing SDS-PAGE analysis of induced fractions collected from AD494(DE3) cells grown in minimal medium (15% gel, silver staining). Only traces of H1-CRD can be spotted in the expressed samples (lane 2-5). B) Reducing SDS-PAGE analysis of non-induced and induced fractions collected from

BL21(DE3) cells grown in minimal medium (15% gel, silver staining). H1-CRD can be observed in all induced samples (lane 4-7), already after 2 hours expression.

Induced samples of AD494(DE3) failed to show expression of H1-CRD in minimal medium. At best, traces of protein corresponding to the size of monomeric H1-CRD were seen, though the actual identity of the protein remained disputable. In contrast, BL21(DE3) clearly exhibited overexpression of H1-CRD in minimal medium. Distinct bands of H1-CRD could be seen expressed over time in analyzed samples.

At further comparison of the two bacteria strains, a difference in growth rate was observed (Figure 56). AD494(DE3) showed an increase in cell number of maximum 25% after inoculation in minimal medium, while corresponding measure for BL21(DE3) was roughly 66%. In this context it should be remarked that the two expression attempts were not completely equivalent as expression temperature and inoculation OD varied between the set-ups. Still, repeated expression attempts with AD494(DE3) grown at different temperatures indicated the same, the strain failed to grow in minimal medium. Hence, the distinction in growth rate between the two strains should be kept in mind.

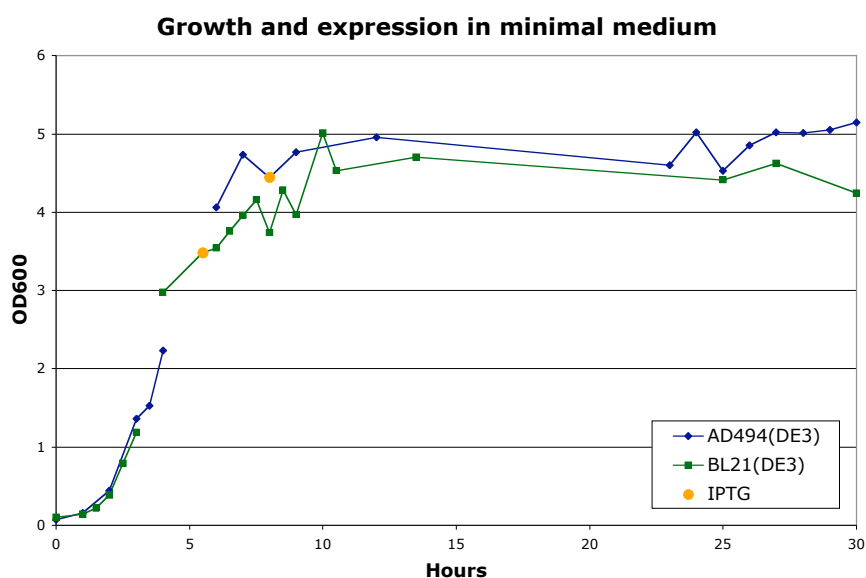


Figure 56. Superimposed growth curves of *E.coli* AD494(DE3) (blue) and BL21(DE) (green). Cells were initially grown in rich medium, followed by an exchange into minimal medium (represented by the gaps in the curves). Induction by IPTG is marked by an orange dot.

Based on these results, it was decided to use BL21(DE3) for expression of isotope labeled H1-CRD.

3.5.1.1 Optimization of the expression set-up using experimental design

The set-up for expression of isotope labeled H1-CRD from *E.coli* BL21(DE3) according to the two-stage protocol was optimized in terms of expression temperature, cell concentration factor, inoculation culture medium, IPTG concentration and expression time. An experimental design approach was undertaken to determine the most favorable conditions, *i.e.* a number of experiments was run in which the factors were varied systematically. Following the expression attempts, non-induced and induced samples from each culture were initially compared at each selected temperature level by SDS-PAGE analysis (*Figure 57*).

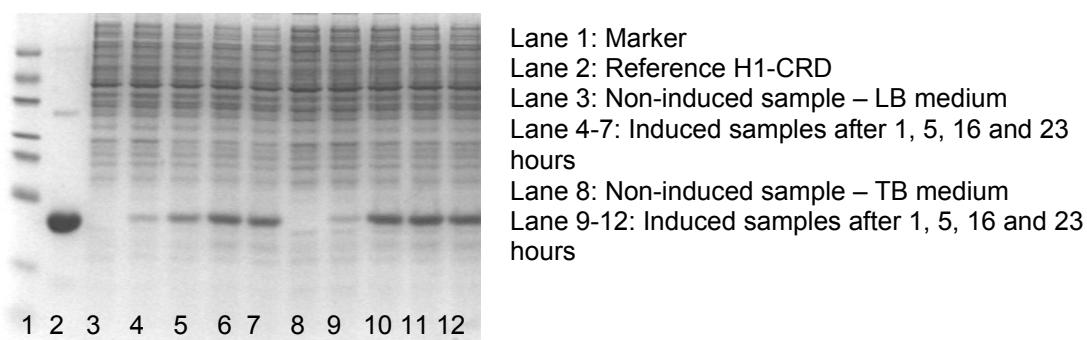


Figure 57. Reducing SDS-PAGE analysis of samples collected after H1-CRD expression at 25°C in minimal medium. After growth in rich medium, either LB (lane 3-7) or TB (lane 8-12), the BL21(DE3) cells were inoculated in minimal medium and concentrated with a factor of 2. Expression levels over time were monitored (15% gel, coomassie staining).

Samples showing the highest expression levels were identified, indicating which conditions were most favorable at each specific expression temperature. The results were summarized and displayed as shown in figure 58 below.

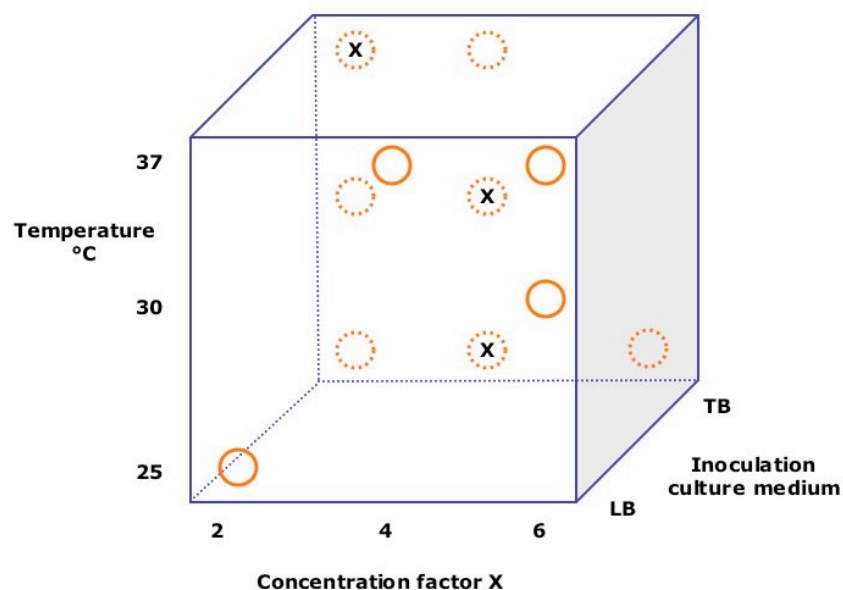


Figure 58. Optional growth and expression conditions, when using the two-stage protocol, graphically presented. The performed experiments are marked with circles. In addition, the set-up showing best yields at each selected temperature level is denoted with a cross.

Based on the experiments and the summary, a number of conclusions could be made. First, use of TB medium as growth culture medium gave the highest expression yields regardless of the expression temperature. It appears that preconditioning of the cells, to provide them with excess nutrients, have a positive influence on the protein expression in minimal medium. Second, to concentrate the cells as much as up to 6 times did not show to be optimal at any of the temperature levels tested. A concentration factor of 4 exhibited best outcome at the lower temperatures, *i.e.* 25°C and 30°C, while a concentration factor of 2 was most favorable at 37°C. In short, a considerable increase in cell concentration will result in higher protein yields at low temperatures, but it is not beneficial to increase the cell number above an OD of 2 at 37°C. Third, experiments carried out at 37°C showed highest protein yields during the first 10 hours, while samples taken at later stages exhibited less H1-CRD. In contrast, expression carried out at 25°C displayed a relatively constant protein amount after 5, 16 and 23 hours. Taken together these results indicate that, when expressed in *E.coli*, H1-CRD is more stable at lower temperatures, possibly due to decreased or slower protein degradation. Fourth, a conclusion about the optimal IPTG concentration proved difficult to make as no obvious difference could be discerned between cultures induced by the lower or higher concentration during the experiments. The expression yield appeared to increase slightly with a high IPTG concentration at the lower temperatures.

In order to determine the optimal expression temperature, the samples presenting the best protein yields for each level were analyzed together by reducing SDS-PAGE electrophoresis (Figure 59).

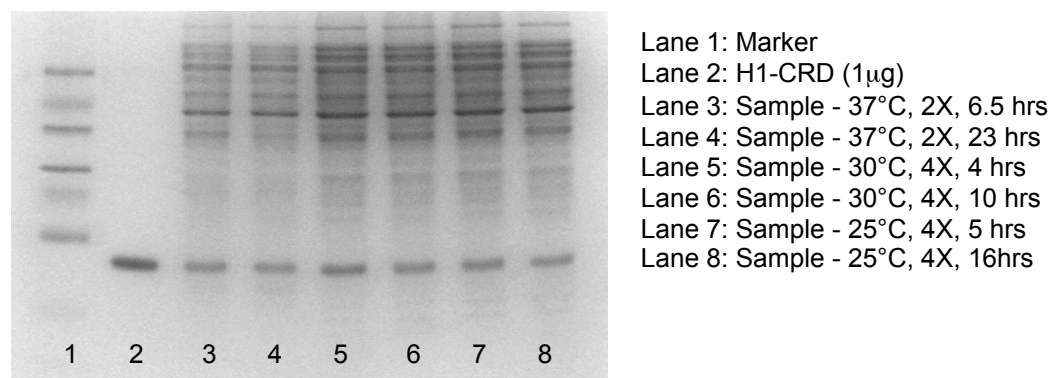


Figure 59. Reducing SDS-PAGE analysis of induced samples presenting the best protein yields at each temperature level tested (15% gel, coomassie staining).

Expression at 30°C could be seen to result in the highest yields of H1-CRD (lane 5). Compared to the expression set-up at 25°C (lane 7 and 8), which was identical apart from the temperature, more intense bands of H1-CRD were observed at 30°C (lane 5 and 6). Looking at the samples induced at 37°C (lane 3 and 4), a significant difference in protein yields could be seen. Expression of both H1-CRD and other proteins was substantially lower at 37°C. In this case, it should be considered that the expression set-up differed also in regard to the concentration factor of cells. At 37°C, a concentration factor of 2 was used, while the cells were concentrated up to 4 times at 30°C. The lower concentration of cells could explain why the samples at 37°C, in particular the bacterial proteins, differ so much in intensity. Nonetheless, it was concluded that the optimal expression temperature was 30°C. As previously mentioned, TB medium and a concentration factor of 4 had been determined as the most favorable conditions at that temperature.

Optimal expression time and IPTG concentration were determined in a separate experiment. The expression trial was run according to the optimized conditions, using two cultures differing only in IPTG concentration, *i.e.* 0.4 mM and 1 mM IPTG, respectively. Samples were collected regularly and analyzed with reducing SDS-PAGE (Figure 60).

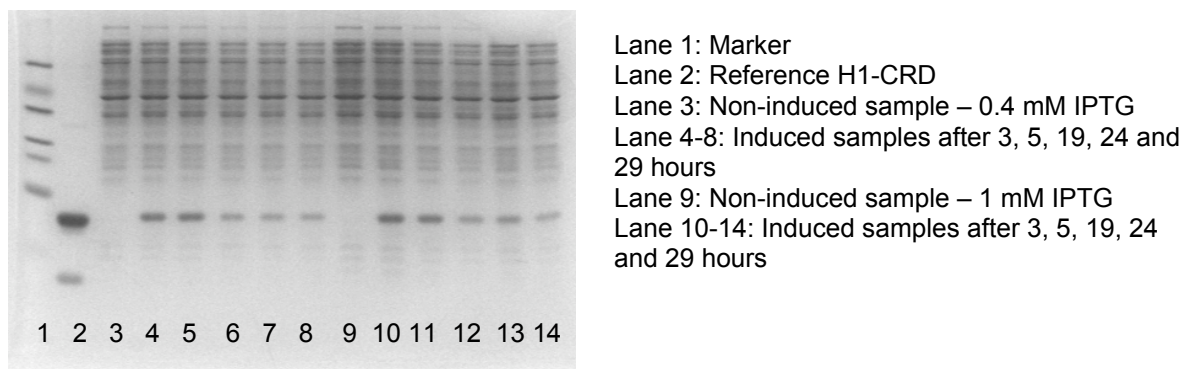


Figure 60. Reducing SDS-PAGE analysis of culture samples induced either by 0.4 mM (lane 3-8) or 1.0 mM (lane 9-14) IPTG (15% gel, coomassie staining).

The two cultures showed highly similar expression patterns. Most H1-CRD could be spotted after 3 and 5 hours, while samples taken later showed decreased levels, an observation true for both set-ups. Due to this and previous evaluations, it was decided that expression should be kept to maximum 5 hours. At further comparison, a small difference could be detected in the intensity of the protein bands between the samples, indicating a higher yield at induction with 1 mM IPTG. It was concluded that the higher IPTG concentration tested was most beneficial and was to be used in further experiments.

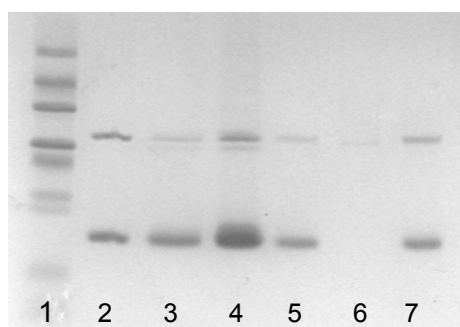
Possible benefits of adding labeled rich medium to the minimal medium were evaluated in a final experiment in the optimization trial. The expression trial was run according to the optimized conditions, using two cultures differing only in addition of rich medium, *i.e.* one was grown according to the standard protocol and one supplemented with 20% *E.coli* OD2 Silantes [136]. Samples were collected regularly and analyzed with reducing SDS-PAGE. However, addition of rich medium was seen to result only in a modest increase of expressed H1-CRD. The culture supplemented with *E.coli* OD2 Silantes exhibited bands of H1-CRD slightly more intense than the control culture. However, since the difference appeared to be minor, it was not deemed as significant enough to include addition of rich medium in the procedure. Labeled rich media, such as *E.coli* OD2 Silantes, are more expensive than minimal medium and to use the former as a supplement would not be cost effective in the long run.

3.5.1.2 Expression of H1-CRD according to the two-stage protocol under optimized conditions

Uniform labeling with only ^{15}N as well as combined with ^{13}C was carried out according to the two-stage protocol. *E.coli* BL21(DE3) cells were grown to high density in TB medium and subsequently exchanged into minimal medium according to the optimized conditions. After 1 hour's incubation, expression was induced by addition of 1 mM IPTG and left for 5 hours. The bacteria were shown to increase in cell numbers after inoculation in the minimal medium. Following addition of IPTG, the growth rate slowed down and reached steady state. A 400 ml culture was seen to result in approximately 6 gram of bacteria pellet after centrifugation.

Collected bacteria pellets were lysed in proportion to the initial volume of rich media used to ensure complete lysis of all bacteria and maximal extraction of protein. Hence, a pellet from a 400 ml minimal medium culture was treated according to its initial growth conditions, *i.e.* 3x525 ml, during the purification. The protein was subsequently refolded by stepwise dialysis and purified by affinity chromatography on a FPLC system. The dialysed protein solution was divided into two runs à 115 ml load each to avoid overloading the column. H1-CRD could be observed in the chromatogram as a sharp peak, indicating correct folding and binding to galactose. Eluted fractions were analyzed by reducing SDS-PAGE, in which H1-CRD was found in the fractions corresponding to the peak. Traces of H1-CRD could also be spotted in the flow-through, but to such a low extent it was not deemed as significant.

HPLC IEC was used to separate monomers and dimers of the isotope labeled H1-CRD prior to studies by NMR. Monomeric H1-CRD indisputably dominated the isotope labeled samples, seen as a high peak with a tailing end during IEC. In contrast, the dimeric peak was considerably lower, best seen in the double-labeled sample of H1-CRD. The difference could also be observed in the ^{15}N labeled sample, although it was less pronounced. Collected fractions were analyzed with non-reducing SDS-PAGE electrophoresis (*Figure 61*).



Lane 1: Marker
 Lane 2: Reference H1-CRD
 Lane 3: $^{13}\text{C}/^{15}\text{N}$ sample – FPLC
 Lane 4: Sample desalted and concentrated prior to HPLC IEC
 Lane 5: Monomer sample – IEC run 1
 Lane 6: Dimer sample – IEC run 1
 Lane 7: Monomer sample – IEC run 2

Figure 61. Reducing SDS-PAGE analysis providing an overview of the purification of $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD (15% gel, coomassie staining).

The analysis confirmed previous findings, *i.e.* the monomeric fraction exhibits traces of dimers even after the separation step. The dimeric fraction on the other hand appears homogenous with one single band. Furthermore, the monomers could be shown to exceed the dimers in numbers prior to the separation, supporting the results of the HPLC IEC.

Collected monomers were concentrated with a simultaneous buffer exchange prior to concentration determination by Bradford. Samples were stored at -20°C until use for NMR studies (Table 27).

Table 27. Samples of isotope labeled H1-CRD generated by the two-stage protocol.

Sample	Culture volume	Conc. ($\mu\text{g}/\text{ml}$)	Amount (μg)
^{15}N H1-CRD	400 ml	465	930
$^{13}\text{C}/^{15}\text{N}$ H1-CRD	400 ml	182	182

3.5.2 Analysis of the isotopic enrichment using NMR

The double labeled H1-CRD, expressed according to the two-stage protocol, was analyzed by 1D NMR half-filter experiment to assess the extent of ^{13}C enrichment. Prior to its analysis, two control experiments were performed to establish reference spectra for a sample at natural abundance versus one with 100% ^{13}C enrichment.

The half-filter experiment separates the data into two spectra, one resulting from hydrogens directly bound to ^{12}C , and the other from hydrogens directly bound to ^{13}C (Figure 62). For the sample at natural abundance, an E-selectin antagonist [127] was analyzed. The result derived from the ^{12}C bound hydrogens displayed a spectrum with

several intense peaks, while the corresponding spectrum for the ^{13}C bound hydrogens showed significantly diminished (if any) peaks. These results are in good agreement with the known 1.1% natural occurrence of ^{13}C . For acetic acid, the sample fully enriched with ^{13}C , the resulting spectra showed a reversed pattern. Nearly all of the signals were filtered into the spectrum corresponding to ^{13}C bound hydrogens while only a minor response was detected for the ^{12}C bound hydrogens. These two control samples indicated that the experiment was optimized sufficiently to test for an unknown level of ^{13}C enrichment.

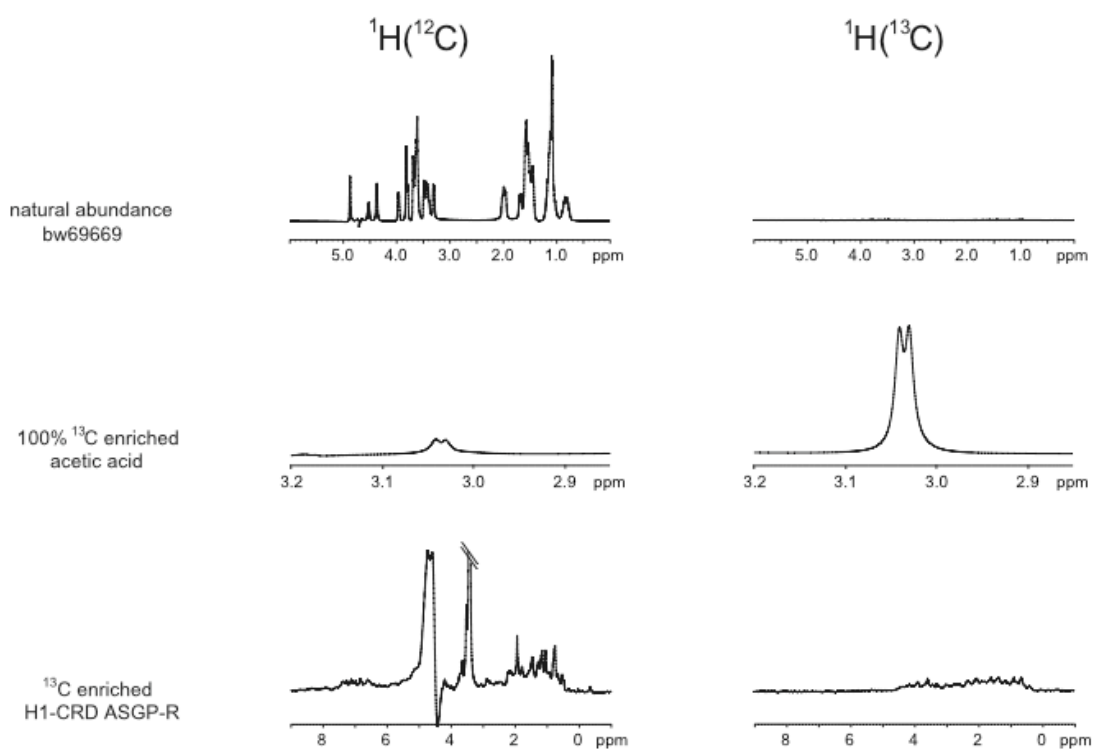


Figure 62. NMR half-filter experiments designed to estimate isotopic incorporation. The upper two rows correspond to control samples used to verify the method with an expected result. The lowest row is the actual test performed on the double labeled H1-CRD.

The half-filter experiment applied to the double labeled H1-CRD demonstrated that the level of ^{13}C enrichment was much less than expected. The majority of signals appear in the ^{12}C filtered spectrum with the peaks in the range of 0 to about 3 ppm corresponding to side-chain hydrogens. Peaks at similar frequencies can also be seen in the equivalent ^{13}C spectrum. The peaks in the ^{13}C spectrum differ from those in the ^{12}C spectrum as the former are composed of doublets, dividing each peak in two and subsequently decreasing signal height. The intense peaks found at 3.8 and 4.7 ppm can be attributed

to residual buffer and water, respectively. The absence of these intense peaks in the ^{13}C filtered spectrum is to be anticipated as they are not bound to ^{13}C nuclei. To accurately determine the extent of isotopic incorporation, the frequency range needed to be integrated. Upon integration, the relative distribution of signals indicated that approximately 60% of hydrogens were bound to ^{12}C and 40% to ^{13}C . A low level of isotopic enrichment (>90%) would require prolonged experimental measurement time, which would not be possible given finite magnetic field and protein stability. In addition, a high incorporation rate is more critical for ^{13}C than for ^{15}N [85].

3.5.3 Mass spectrometry analysis of isotope labeled H1-CRD produced by the two-stage protocol

HPLC RP chromatography was used to separate monomers and dimers prior to MS analysis, aiming to determine the incorporation of ^{15}N and ^{13}C . Even though the RP chromatography method was shown to negatively influence the activity of H1-CRD (see section 3.1.4), it holds an advantage when preparing samples for MS. The resulting protein samples can easily be lyophilized, unlike samples prepared by HPLC IEC. Furthermore, MS analysis can be performed regardless of whether the protein is correctly folded or not, and therefore the activity is of less concern. The monomers could be seen in the chromatogram as a sharp peak, close but clearly separated from the following dimer peak. The monomeric fraction was lyophilized and analyzed by MS. An equivalent sample of unlabeled monomeric H1-CRD was used as reference. Deconvolution of the resulting MS spectra revealed an unexpected high degree of similarity between the samples (*Figure 63*).

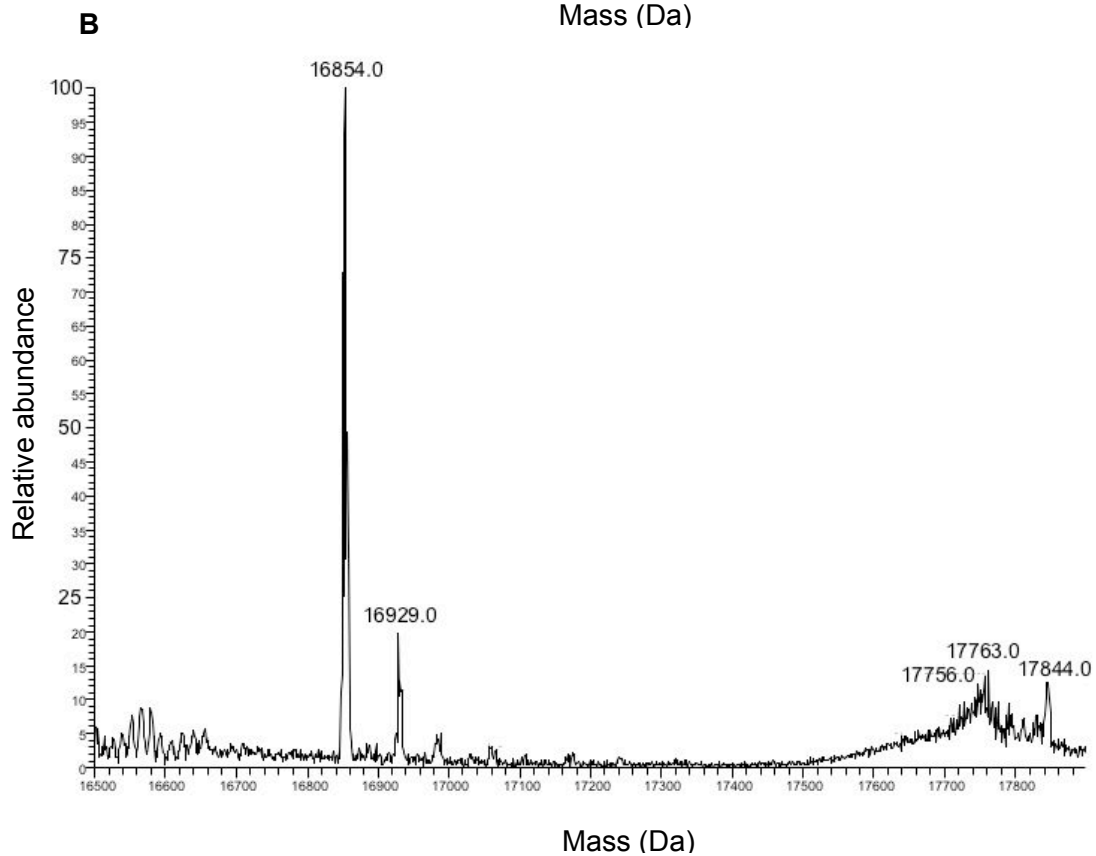
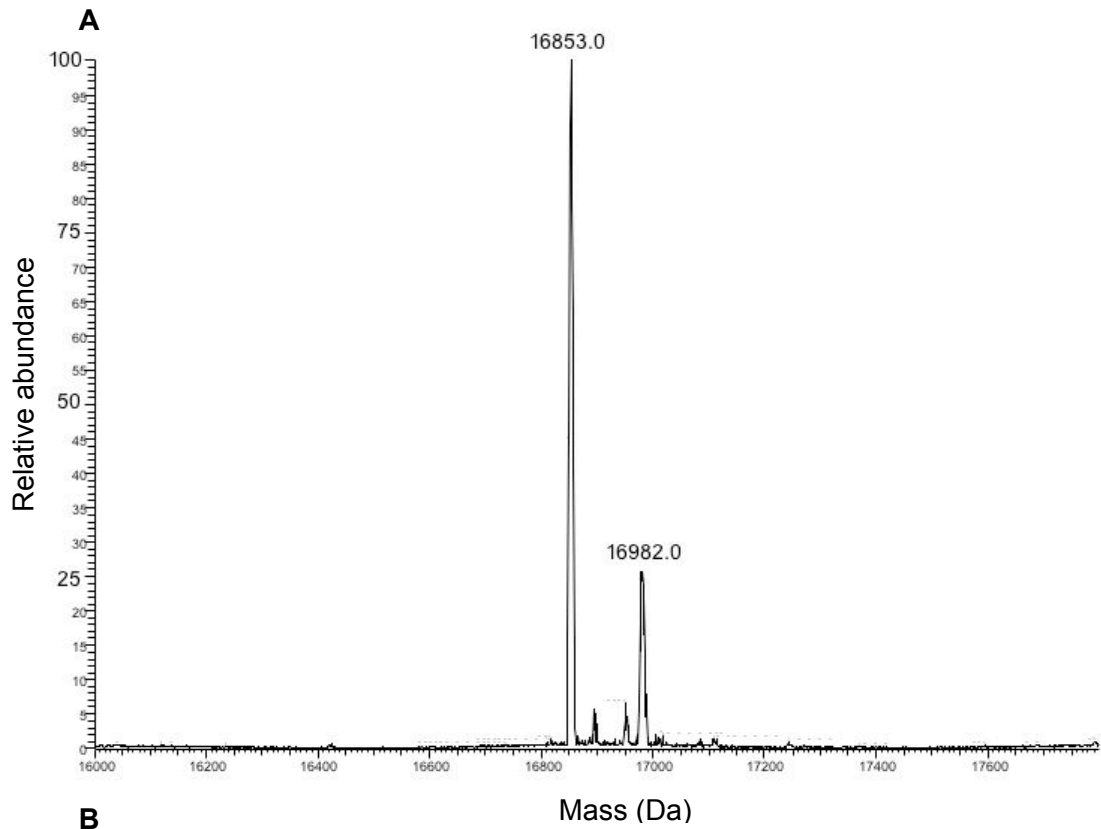


Figure 63. A) MS spectra of unlabeled H1-CRD. The major fragment shows a mass of 16853 Da, explained by cleavage of the *N*-terminal methionine. B) MS spectra of $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD. The major fragment shows a mass of 16854 Da.

The average mass of unlabeled H1-CRD is estimated to 16991 Da, while corresponding masses for ^{15}N and $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD are 17201 and 17953 Da, respectively. The major peak seen in the reference spectra of unlabeled H1-CRD showed a mass of 16853Da. The difference of 138 Da can be deduced to cleavage of the *N*-terminal Methionine (-131 Da) and reduction of the disulfide bridges (+2 Da per broken bond). Unexpectedly, the major peak shown in the spectra of $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD has a mass of 16854 Da, closely resembling the unlabeled sample. A minor peak can be seen at 16929 and lots of small, merged peaks around 17750 Da. The latter peaks correspond well with masses expected for double-labeled H1-CRD, but constitute an inferior part of the protein sample. Similar results were found for the ^{15}N labeled H1-CRD with a major peak at 16853Da. In summary, the MS results proved that primarily unlabeled H1-CRD was expressed during the experiments and only a minor population of labeled H1-CRD was present in the final samples. Multidimensional NMR experiments rely on usage of highly enriched isotope labeled proteins and these samples, produced according to the two-stage protocol, would therefore not be of satisfactory quality. The method either needs to be improved or exchanged in order to fulfill the demands.

3.5.4 HSQC of ^{15}N labeled H1-CRD expressed according to the two-stage protocol

^{15}N labeled H1-CRD, expressed according to the two-stage protocol, was recorded using a 2D HSQC experiment on a 800 MHz NMR equipped with a low-temperature cryoprobe. The HSQC spectrum, although weak in intensity, displays numerous signals that are significantly stronger than the noise and therefore correspond to ^{15}N enriched amino acids (*Figure 64*).

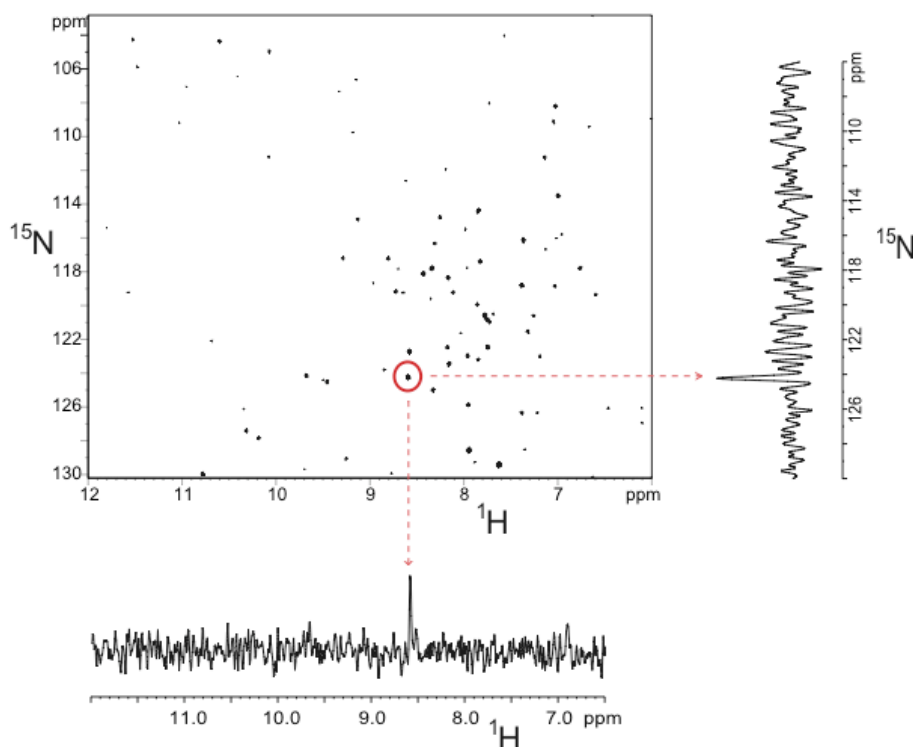


Figure 64. HSQC spectrum of ^{15}N labeled H1-CRD. The vertical and horizontal insets are taken from the frequencies of the circled peak, which represents the most intense signal.

The peaks are well dispersed in the ^1H dimension, indicative of a stable tertiary structure. One factor determining the distribution of amide hydrogen chemical shifts in a folded protein is the various degrees of hydrogen bonding in different structural elements. If the protein was not folded, the amide hydrogens would experience an averaged chemical shift due to a high degree of flexibility. Consequently, the amide hydrogen peaks would be expected to be confined within a narrower range of ^1H frequencies [137,138].

Even though the amide hydrogens of the protein could be observed, a measurement time longer than expected was required to obtain clear signals with a reasonable degree of confidence. At the current level of technology, an 800 MHz NMR using a cryoprobe is among the most sensitive spectrometers available [139]. Assuming a high ^{15}N incorporation rate and high protein concentration, an adequate measurement time of 10 minutes would have been anticipated, rather than of 6 hours. Three-dimensional NMR experiments, aiming to determine the structure, would need at least one month measurement time each with this sample. Due to the drift of the magnetic field during this time, and the availability of the instrument, such experiments would be impossible.

Therefore, either the sample quality or quantity or both ought to be substantially improved before attempting structural analysis.

3.5.5 HMQC of $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD expressed according to the two-stage protocol

$^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD, expressed according to the two-stage protocol, was recorded using a 2D HMQC experiment on a 500 MHz NMR. The spectrum shows the characteristic diagonal distribution of alpha and beta carbons associated with a protein (Figure 65). The downfield region represents the alpha carbons, while the beta carbons can be seen in the upfield region.

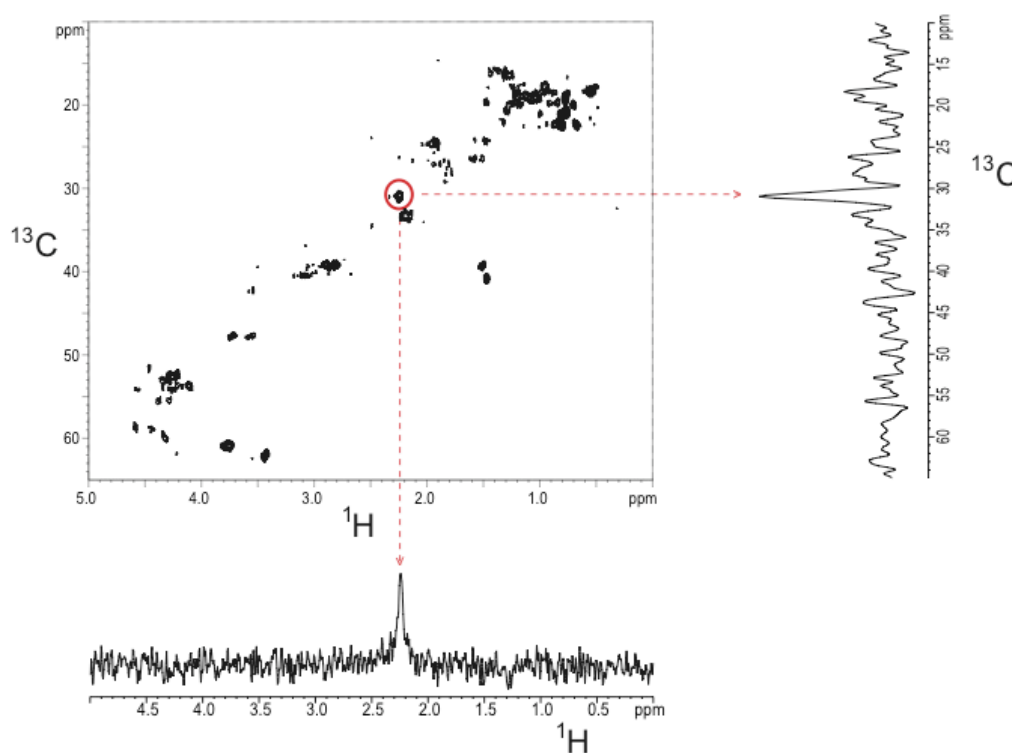


Figure 65. HMQC spectrum of $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD. The vertical and horizontal insets are taken from the frequencies of a representative peak.

The observed signals were weak given the extended measurement time. Similar to the ^{15}N H1-CRD sample measured by HSQC at 800 MHz, the double enriched sample did not show satisfying results to fulfill the prerequisites for structural analysis.

3.5.6 Alternative expression set-up of the two-stage protocol

Uniform labeling of H1-CRD with ^{15}N was attempted with the alternative two-stage protocol, aiming to improve the labeling efficiency. *E.coli* BL21(DE3) cells were grown

to high density in LB medium and subsequently exchanged into minimal medium, concentrating the cells with a factor of 4. After 2 hours' incubation, expression was induced by addition of 0.4 mM IPTG and left for 5 hours. The cell numbers were fairly constant, showing a small decrease, during the first 2 hours. Following induction, an increase in cell density could be observed. A 100 ml minimal medium culture gave a final bacteria pellet of 1.75 gram.

H1-CRD was purified by the previously described and applied procedure, comprising denaturation, refolding and purification by affinity chromatography. HPLC RP chromatography was used to separate monomers and dimers prior to MS analysis (see *section 3.5.3*). The monomers could be seen in the RP chromatogram as a peak with a tailing shoulder, closely separated from the following smaller dimer peak.

The monomeric fraction was lyophilized and analyzed by MS, aiming to determine the incorporation of ^{15}N . As stated earlier, the expected molecular weight of H1-CRD completely labeled with ^{15}N is 17201 Da. This most recent sample showed a major fragment with an average molecular weight of 17053 Da. The level of isotope incorporation can be calculated by dividing the difference of the experimental labeled and unlabeled masses with the difference of the theoretical labeled and unlabeled masses. Using the data from the MS analysis of unlabeled H1-CRD (see *section 3.5.5*) gives an experimental difference of 200, compared to a theoretical difference of 211. Consequently, the isotope enrichment could be calculated to 95%.

3.5.7 Expression of isotope labeled H1-CRD by exclusive use of minimal medium

As the two-stage protocol initially failed to produce isotope labeled H1-CRD, a second method was attempted. Uniform labeling with only ^{15}N as well as combined with ^{13}C was carried out according to the protocol for expression by exclusive use of minimal medium. A small preculture was grown in **unlabeled** minimal medium O/N and subsequently diluted 10X in **labeled** minimal medium the following day. Expression was attempted at 30°C and 37°C, favoring the latter temperature as the protein yields appeared slightly higher. Expression was induced at an OD of 1.0 and lasted 5 hours. Growth and expression levels were monitored throughout the experiment. Cell numbers could be seen to increase steadily after addition of IPTG. A batch of 3 cultures à 500 ml would typically yield around 14 gram bacteria pellet. The protein was

purified according to the standard method established for H1-CRD. Following solubilization and refolding, the protein was purified by affinity chromatography on a FPLC system (Figure 66).

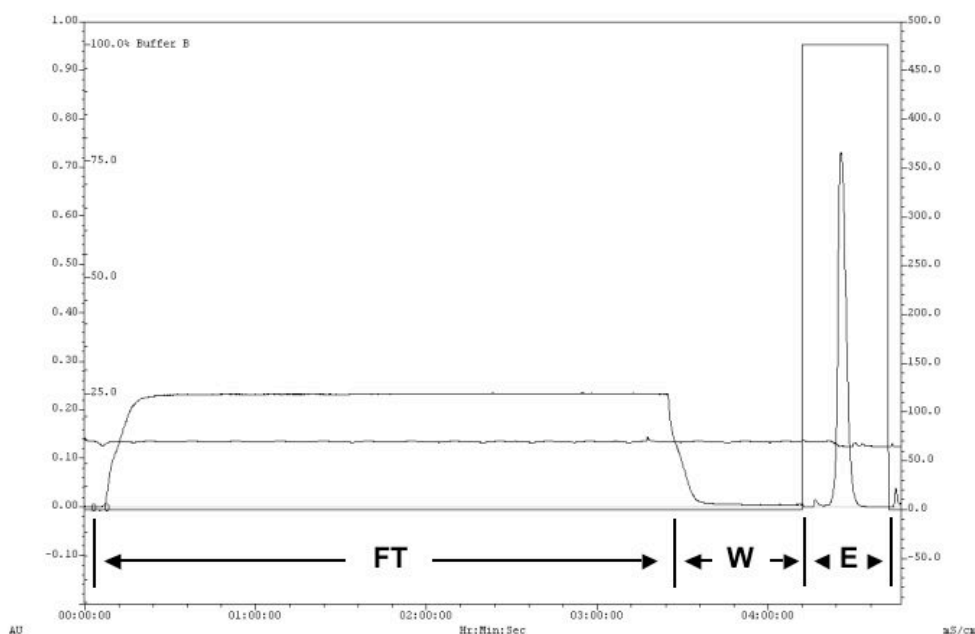


Figure 66. FPLC affinity chromatography of ^{15}N labeled H1-CRD expressed by exclusive use of minimal medium. A sharp peak can be seen during the elution phase, representing the purified protein.

H1-CRD could be seen in the chromatogram as a sharp peak during the elution phase, indicating correct folding and binding to galactose. Eluted fractions were analyzed by reducing SDS-PAGE, showing good agreement with the affinity chromatography and indicating a high yield of purified H1-CRD.

HPLC IEC was used to separate monomers and dimers of the isotope labeled H1-CRD prior to studies by NMR (Figure 67). The resulting chromatograms showed one dominant peak with a tailing shoulder at the time expected for elution of monomers. The impurity could not be detected, while the presence of dimers was negligible. The separation profile looked the same for both ^{15}N and $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD.

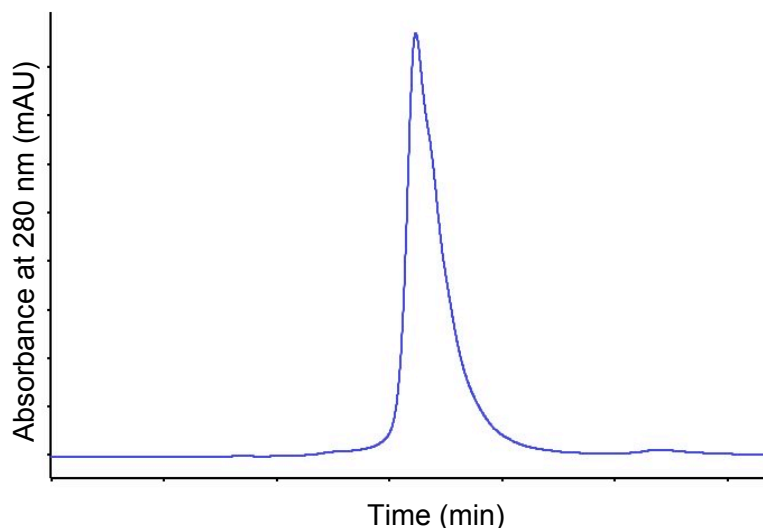


Figure 67. Separation of monomers and dimers by HPLC ion exchange chromatography of $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD. A dominant peak of monomers can be seen, while the presence of dimers is negligible.

Collected monomers were concentrated with a simultaneous buffer exchange prior to concentration determination by Bradford. Samples were stored at -20°C until use for NMR studies (Table 28).

Table 28. Samples of isotope labeled H1-CRD generated by exclusive use of minimal medium.

Sample	Culture volume	Conc. (mg/ml)	Amount (μg)
^{15}N H1-CRD	3 x 500 ml	2.88	720
$^{13}\text{C}/^{15}\text{N}$ H1-CRD	3 x 500 ml	2.43	608

HPLC RP chromatography was used for separation of monomers and dimers prior to MS analysis (see section 3.5.3). The monomers could be seen as a peak with a tailing shoulder, while the dimeric fraction was minor. The monomers were lyophilized and submitted for MS (Figure 68).

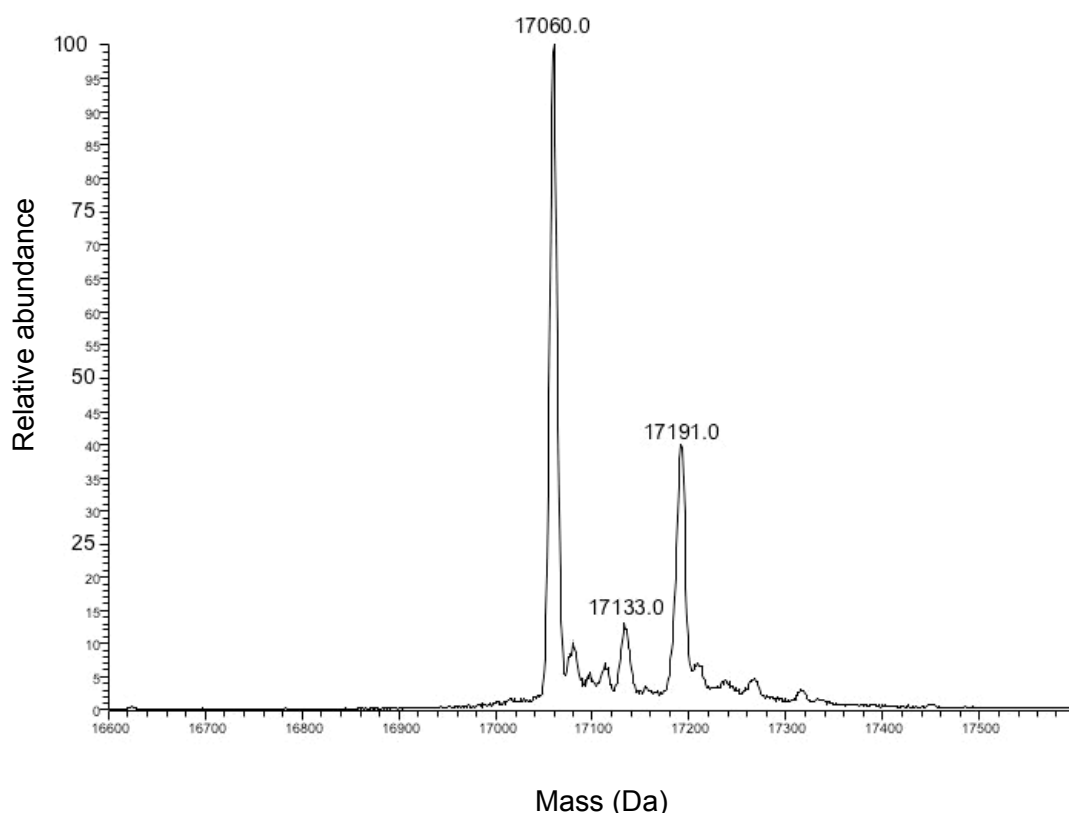


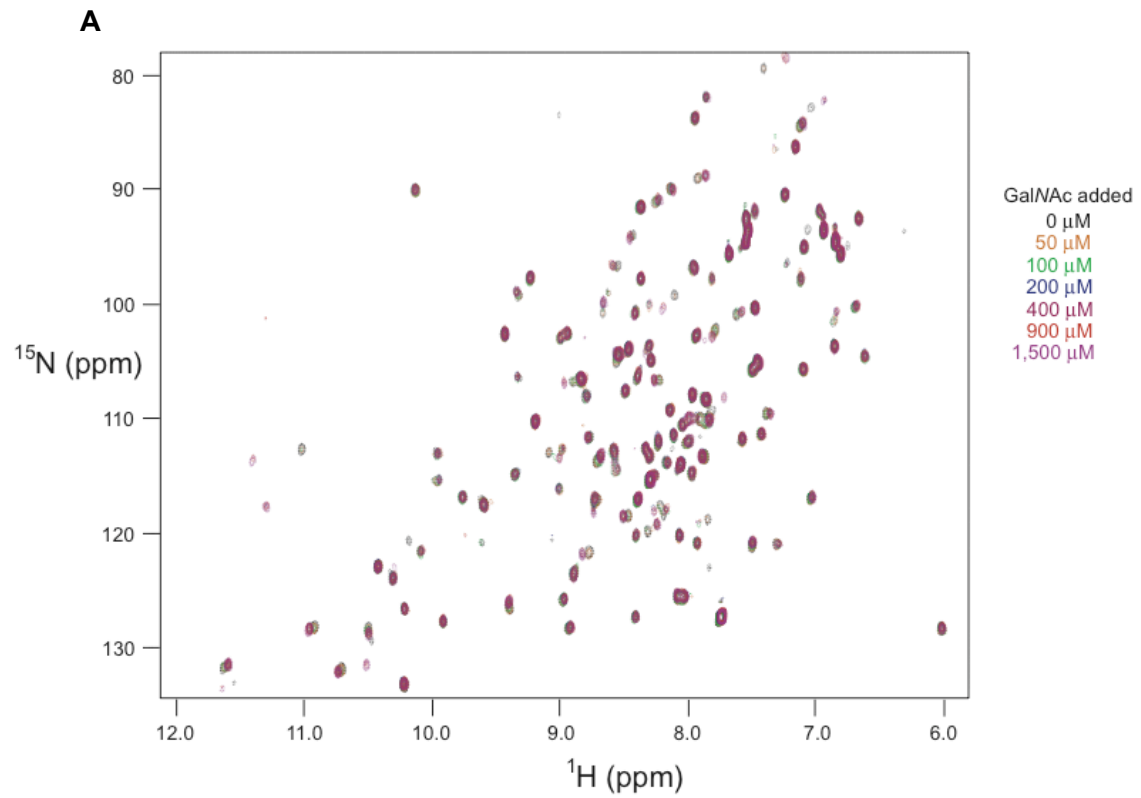
Figure 68. MS spectra of ^{15}N labeled H1-CRD. The major fragment shows a mass of 17060 Da.

Analysis of ^{15}N labeled H1-CRD showed a major fragment with an estimated average molecular weight of 17060 Da, *i.e.* a difference of 207 Da compared to the reference spectra of unlabeled H1-CRD. Dividing the experimental mass difference with the theoretical difference of 211 Da gives an isotopic enrichment of 98%. Corresponding results for the $^{13}\text{C}/^{15}\text{N}$ labeled H1-CRD displayed a deviation of 935 Da when comparing the experimental data. With an expected mass difference of 962 Da, the isotope incorporation can be determined to 97%.

3.5.8 HSQC of ^{15}N labeled H1-CRD produced by exclusive use of minimal medium

^{15}N labeled H1-CRD was recorded using a 2D HSQC experiment on a 600 MHz NMR. The HSQC spectrum acquired within 10 minutes measurement time showed strong signals, which indicated that a significant amount of isotopically enriched protein was present (*Figure 69*). Similar to the measurements with the sample used on the 800 MHz, the peaks were well dispersed in the ^1H dimension, indicative of a stable tertiary structure. The greater signal-to-noise of the peaks measured at 600 MHz increased the

reliability that the observed signals actually corresponded to amide resonances in the protein. The high quality of the sample, as proved by this initial NMR experiment and the MS analysis, makes it eligible to attempt more challenging experiments required for protein NMR assignment and structural analysis.



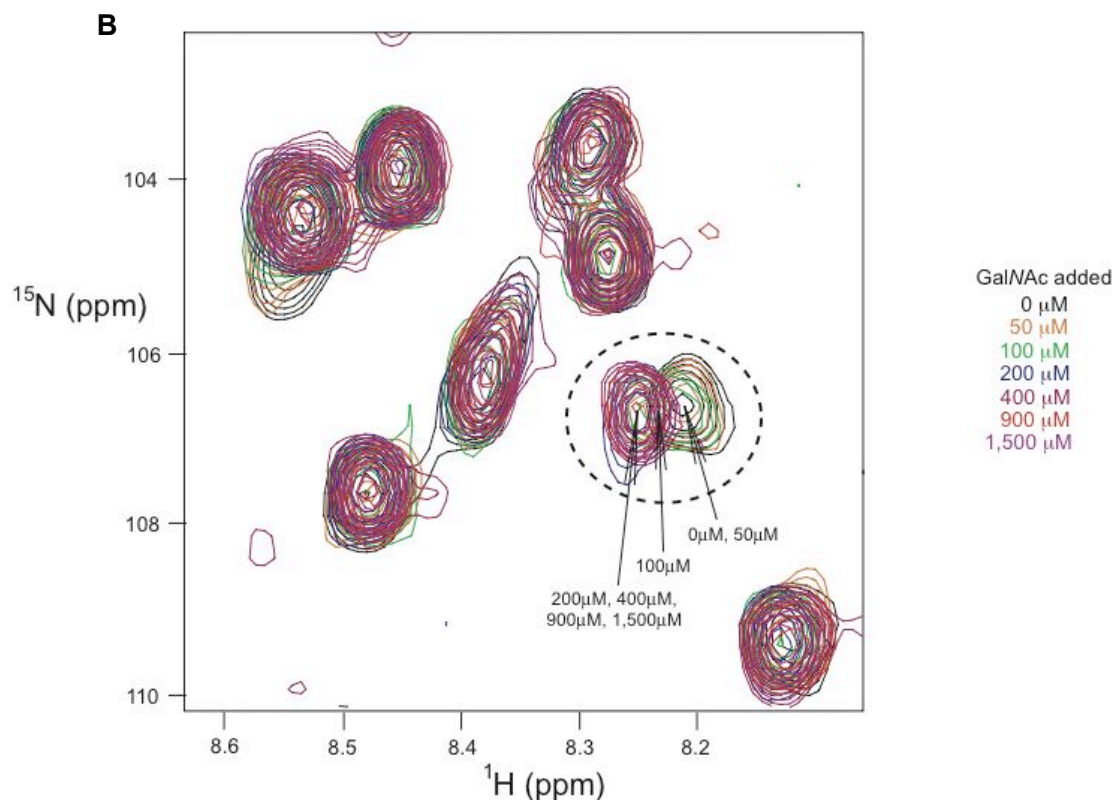


Figure 69. HSQC spectrum of ^{15}N labeled H1-CRD. A) The entire spectral width of the protein is shown with a series of different GalNAc concentrations. B) To aid in visualization, one of the amide residues showing a perturbed resonance in the presence of GalNAc is expanded. In addition to the aforementioned residue, several other residues are shown which remain at their location and are therefore assumed not to be involved in binding.

Binding to the protein was investigated by adding a series of different concentrations of GalNAc and monitoring changes in HSQC peak positions (Figure 69) [93]. The observed peaks in the spectrum represent the average chemical shift of the free and bound form of the protein at each concentration of GalNAc. With no GalNAc added, this average is purely due to the free form. In contrast, at a saturating concentration of GalNAc, all protein will exist in a bound state. At intermittent concentrations, both free and bound forms of the protein are present, shifting the position of the peak in between the two extremes. Analysis of the changes in the peak position with differing amounts of ligand allows the determination of the affinity constant. Superimposing HSQC spectra of H1-CRD with different GalNAc concentrations demonstrated that the majority of the peaks were unaffected, however, a number of peaks could be seen to change location. The relocated peaks moved in a uniform manner, reaching their final position at the same GalNAc concentration, proving the specificity of the binding. The preliminary results indicated that the K_D of GalNAc to H1-CRD was 100 μM . This value

was determined from the concentration of GalNAc required to shift the peaks halfway from their initial position to their saturated location.

4 Discussion

4.1 Expression and purification of H1-CRD – protein production interrupted

Access to protein in milligram amounts is a prerequisite for structural and functional protein studies. Recombinant DNA technology combined with heterologous expression systems have been of immense importance to facilitate protein production and to fulfill these needs. A protocol for the expression of ASGP-R H1-CRD in *E.coli* has previously been established and enabled determination of the crystal structure of the subunit in 2000 [5]. The original protocol was further optimized in a doctoral thesis by Rita Born at IMP, greatly improving the yields [94]. The optimized expression and purification method was employed within this PhD-thesis. The protein, present in inclusion bodies after expression, is denatured and reduced during the purification procedure. To recover the native structure, it is refolded through step-wise dialysis and finally purified by affinity chromatography. The procedure is fairly straightforward and only utilizes conventional reagents. However, it suffers from a major setback as the final protein sample consists of two protein populations, namely monomers and dimers of H1-CRD. The dimers present a predicament due to their abnormal formation. Although ASGP-R certainly is an oligomeric receptor, the subunits associate via the extracellular stalk segment and not at the CRD. The dimerization *in vitro* is further problematic as it cannot be controlled and its incidence is variable. Consequently, the samples become largely heterogeneous, both within and in between batches of H1-CRD.

At the outset of this PhD thesis, little was known about the dimers of H1-CRD. An odd cysteine was suggested as the cause of the dimer formation, supported by the absence of dimers under reducing conditions during SDS-PAGE analysis. Yet, the theory had not been verified, nor had the odd cysteine been properly identified. Furthermore, the behavior of dimeric H1-CRD compared to pure monomers had not been determined, leaving the possibility that the two protein species exhibit different binding affinities. Until such a difference, or the opposite, had been established, it was resolved to perform further studies with samples of pure monomeric H1-CRD.

A method to separate monomers and dimers using HPLC IEC, developed by Daniel Ricklin at IMP [95], was used in this work. The two protein species were successfully divided during the run of the program and collected as separate fractions. The large batch that was produced during this PhD thesis, and that subsequently aided the development of the Biacore assay and the solid-phase competition assay, was purified by HPLC IEC and resulted in 4 mg monomers of H1-CRD. However, SDS-PAGE analysis revealed that a small fraction of dimers was present in the monomer samples even after the separation step. It can reasonably be assumed that these dimers formed after the IEC as a consequence of the high concentration of monomers. Aliquots of monomeric H1-CRD were stored either at -20°C or at 4°C after being thawed and opened. Once in liquid state, monomers are theoretically able to associate in dimeric complexes until an equilibrium is reached. While SDS-PAGE analysis could demonstrate that the monomers still constituted a majority of the samples even after storage at 4°C , the presumed eradicated dimers were also clearly present. Under these conditions, a new aliquot of H1-CRD would need to be thawed for each new experiment to ensure a high content of monomers, but even then the homogeneity of the sample would remain uncertain.

An alternative method, based on HPLC RP and developed by Dr. Said Rabbani at IMP, was also attempted for separation of monomers and dimers of H1-CRD. The major appeal of the method was the possibility to store the final protein samples as powder after the chromatography and lyophilization. It was theorized that the stability of the protein, over time, would improve in the solid state. In addition, liquid protein samples of fixed concentrations could be prepared on demand. However, the aptness of the method was questioned due to the relatively harsh conditions used during the chromatography step. Monomers purified by HPLC RP were therefore compared to monomers purified by HPLC IEC in a series of assays determining binding affinity to GalNAc. Binding of GalNAc could be detected in three independent studies to both protein samples. While the NMR data could only provide a qualitative statement, a clear difference in the binding could be observed in the Biacore experiments and the solid-phase competition assay. Binding of GalNAc occurs to less extent to the RP purified monomers when equivalent amounts of protein are used. While the determined K_D values were in a similar range, the signal intensity between the protein samples differed substantially in both assays. Taken together the results suggest that only a small fraction of protein is fully active in the RP sample, as indicated by the K_D .

The majority of protein, however, appears inactive and incapable of GalNAc binding, resulting in low signal response. As the two monomer samples only differ in the final purification step, it lies close at hand to assume that the RP step and/or the lyophilization harms the protein and renders it inactive. Consequently, IEC remains as the method of choice for separation of monomers and dimers of H1-CRD.

Although a symptomatic relief, the separation method(s) cannot be considered an optimal solution as the dimer formation is in fact neither cured nor abolished. The production of H1-CRD suffers from a significant protein loss on behalf of the dimerization as monomeric H1-CRD is trapped in dimers. Moreover, even after going to the lengths required for separating the two protein species, a stable and invariable sample of monomers cannot be guaranteed. Finally, additional purification steps are known to decrease the final protein yields [128], which is also true for the production of H1-CRD. In summary, a method able to prevent dimer formation, either through complete inhibition or at an early point of the protein purification, would be highly advantageous for the production of H1-CRD.

Protein dimers arising as a result of intermolecular covalent disulfide bonds can be inhibited by modification of the cysteine(s) involved. An effective way to abolish dimer formation caused by covalent disulfide bonds is to substitute the involved cysteine for serine or alanine through site-directed mutagenesis [140,141]. This approach was undertaken within this work and is discussed further in section 4.2. An alternative method is to chemically modify the cysteines by thiol reagents such as *N*-ethylmaleimide (NEM), 5,5'-dithiobis-(2-nitrobenzoate) (DTNB) or iodoacetamide [142,143]. As an example, NEM was utilized in a study investigating activity and oligomeric state of the lactose transport protein (LacS). By incubating a cysteine mutant of the protein with NEM, the activity of LacS was considerably reduced. In contrast, a second mutant lacking a cysteine was left unaffected by the NEM treatment [144].

Initial experiments combining H2-CRD with NEM carried out within a master thesis project by Daniela Abgottspon at IMP have shown promising results [145]. Following expression and purification by affinity chromatography on FPLC, samples of H2-CRD were incubated with different concentrations of NEM. Non-reducing SDS-PAGE analysis could reveal that the dimer content in the NEM treated samples was significantly lower compared to the negative controls after one day as well as one

week. A small fraction of dimers was present in the NEM samples, but is supposedly the result of dimerization occurring already during the refolding and purification of the protein. If, as the preliminary results suggest, NEM can prevent dimer formation of the ASGP-R subunits, and it is combined with the HPLC separation method, samples of pure monomers can be obtained. Furthermore, if added at an early stage of the protein purification, protein loss due to dimerization could be minimized as NEM will block the free cysteine. Similarly, purified monomers, modified by NEM, would also remain stable after separation from the dimers.

4.2 Expression and characterization of cysteine mutants of ASGP-R H1-CRD

Site-directed mutagenesis is often resorted to when investigating the functional importance of cysteine residues [146,147], and also as a means of inhibiting formation of protein dimers or higher oligomers [141]. In one study, the technique was used to prevent non-native disulfide bonding during expression of human fibroblast β -interferon (IFN β) in *E.coli* [148]. The initially expressed wild-type protein was revealed to show only a tenth of the expected activity. In addition, most of the IFN β was found to exist as dimers or higher oligomers that were inactive. As the protein contains three cysteine residues, it was theorized that one or more of them could be involved in non-native intermolecular disulfide bridging. Comparison with the structurally similar protein IFN α indicated Cys17 as a likely candidate for dimer formation and the residue was therefore substituted for a serine. No multimeric complexes were formed during expression of the mutant IFN β -C17S, proving the assumption to be right. Moreover, the mutated protein showed a specific activity similar to that of purified native fibroblast interferon. Similar to the study of IFN β , SDM was applied in this work to prevent the formation of non-native dimers of H1-CRD formed upon expression and purification from *E.coli*.

H1-CRD contains seven cysteines, six of which are engaged in disulfide bridges, Cys153-Cys164, Cys181-Cys276 and Cys254-Cys268. The odd cysteine, Cys152 as proposed by the crystal structure, has no counterpart to react with within the subunit, leaving it free to form intermolecular disulfide bonds. Cys164 has also been suggested as the odd cysteine in earlier work at IMP [94]. As the odd cysteine remained unidentified at the outset of this thesis, it was decided to mutate the three cysteines

152, 153 and 164, individually and collectively, to determine their potential involvement in dimerization of H1-CRD. In addition, the functional relevance of the odd cysteine and the third disulfide bridge, either Cys152-Cys153 or Cys153-Cys164, was to be concluded. Even though the primary goal of the mutagenesis studies was to inhibit the dimerization, it is of outmost importance that the activity and ligand binding affinity of the subunit is sustained in order for the mutated protein to be of further use.

4.2.1 Dimer formation can be stopped by site-directed mutagenesis...

The three cysteines closest to the *N*-terminus in H1-CRD, *i.e.* C152, C153 and C164 were exchanged for serines, either as single, double or triple mutations, generating seven mutant proteins in total. The mutations and the accuracy of the DNA sequence were verified by sequencing, prior to protein expression in *E.coli*. Similar to WT H1-CRD, the mutant protein(s) were found to accumulate in inclusion bodies and the protein content was therefore denatured and reduced. The native structure was restored by stepwise dialysis, followed by purification by affinity chromatography. All seven mutants showed to both express and purify well by the procedures established for the wild-type protein. Hence, it could be concluded that, qualitatively, the cysteine mutants of H1-CRD maintained binding affinity for galactose.

The dimer content of the purified mutant samples was investigated by non-reducing SDS-PAGE analysis and compared with WT H1-CRD. Out of the seven protein samples only one showed to be completely devoid of dimers, namely the triple mutant. The single mutants, even though containing an even number of cysteines, were all seen to show traces of dimers. Although the cysteines at the *N*-terminus preferentially form an intramolecular disulfide bridge, regardless which of the three cysteines it is formed between, they can also engage in intermolecular bonding. It can therefore be argued that a single point mutation would not be sufficient to eliminate the dimer formation. The double mutants show a high content of dimers, as can be expected. An odd number of cysteines, as in the wild-type, leaves one free to react with a cysteine in a neighboring subunit. However, an exception to this theory could also be observed in the case of mutant C152/153S, in which Cys164 was the fifth unpaired cysteine. Mutant C152/153S only displayed traces of dimers, more resembling the single mutants than WT H1-CRD. The lack of dimers suggests that Cys164 is indeed engaged in an intramolecular disulfide bond in native H1-CRD, as suggested by the crystal structure. However, whether it is paired with Cys152 or Cys153 cannot be determined by these

results. Finally, as previously mentioned, the triple mutant was found to exist exclusively as monomeric protein. Based on these observations it can be concluded that the dimer formation of H1-CRD *in vitro* is caused by the three cysteines closest to the *N*-terminus. The odd cysteine is predominately responsible for the dimerization, but the other two cysteines can also contribute. Concurrently, it can be concluded that the remaining four cysteines in H1-CRD, *i.e.* Cys181, Cys254, Cys268 and Cys276, do not take part in the dimer formation, as no dimers were detected in the triple mutant sample. Considering the crystal structure, the intramolecular disulfide bonds formed by these four cysteines are expected to be of crucial importance for the functionality of the subunit.

In summary, site-directed mutagenesis has proved to be a most powerful tool for preventing the dimer formation of H1-CRD when expressed and purified from *E.coli*.

4.2.2 ...but at the expense of protein affinity and stability

Following the purification by affinity chromatography on FPLC, all mutant samples were loaded onto the anion-exchange column on HPLC. In a final step, the mutant monomers were purified and pooled over the GalNAc-Sep-column on HPLC. The samples were used to determine ligand binding affinity of the mutant proteins. Emphasis was put on the triple mutant, as it was the only sample showing a complete lack of dimers.

The results derived from the solid-phase competition assay were unambiguous. The cysteine mutants all showed lower affinity for GalNAc compared to the wild-type. The signal intensity of the mutant proteins was constantly lower, indicating weaker binding of the GalNAc polymer. Further support for this observation was seen in the competitive approach, where the calculated rIC_{50} values of the mutants compared to the wild-type were all below zero. That is to say, a lower concentration of monovalent GalNAc is needed to inhibit binding of the GalNAc polymer to the mutants.

Further studies were undertaken with the triple cysteine mutant and included NMR and Biacore experiments. While NMR clearly could show binding of GalNAc to the mutant protein, it could not provide a quantitative measure. In contrast, the Biacore experiment resulted in a K_D value for comparison with WT H1-CRD. The K_D values differed significantly, with a factor of almost two, in favor of the wild-type. In

agreement with previous results, the Biacore experiments indicated that the mutant protein exhibit a lower affinity for GalNAc. However, it should also be noted that the signal intensity of the triple mutant was exceptionally low, only slightly higher than the background, and could therefore limit the accuracy of the final results.

A second concern to be addressed is that of the stability of the mutant proteins. In the first approach of the solid-phase competition assay, only binding of the polymer without a competing inhibitor substance was measured. Repeated attempts of the assay set-up showed significantly reduced ligand binding over time. All seven mutants displayed a reduction in signal intensity, seemingly in similar proportions. The protein samples used in the different assays originated from the same batches, but new aliquots were thawed for the assays performed in the second run. In contrast, the wild-type did not appear to lose affinity during the equivalent time period, but showed consistent results when examined in the assay.

The decrease in stability, and indirectly also in affinity, can partly be explained by the loss of the third disulfide bridge. Disulfide bonds are known to improve the stability of proteins, making them more resistant to e.g. high temperatures, acidic or basic pH, high concentrations of organic solvent, or to increase the half-life of protein therapeutics [149]. There are also examples of protein engineering where disulfide bridges have been introduced to increase the stability of specific proteins. In one study, T4 lysozyme was subjected to SDM and mutants with one or more non-native disulfide bonds were constructed. One double-disulfide mutant turned out especially successful as it was not only fully active, but also retained activity at temperatures substantially higher than the wild-type enzyme [150]. In a study more closely related to H1-CRD, disulfide mutants were created of the H2b subunit [151]. The extracellular domain of H2b, as H1, has eight cysteines, proposed to form four disulfide bonds. The importance of each individual bond was investigated by converting corresponding cysteines to alanines and in doing so disrupting the disulfide bridge. The mutant proteins were subsequently expressed in 3T3 fibroblasts. The results revealed that one mutant, lacking bond Cys157-Cys171 nearest to the transmembrane region, was able to fold correctly and was found expressed on the cell surface. In contrast, mutations of the other three disulfide bonds prevented proper folding of the subunit, and all three mutant proteins were degraded in the endoplasmic reticulum. To conclude, the study showed that three of the four disulfide bridges in H2b are essential for proper folding of the protein.

Corresponding bridges in H1 are Cys153-164, Cys181-Cys276 and Cys254-Cys268. Hence, it can be assumed that the bridge formed between Cys153 and Cys164 is important for the structure of H1-CRD. Consequently, the lack of the bond could explain the reduced protein stability observed in the studies with the cysteine mutants and it could also reflect negatively on the affinity. It does, however, not explain the results obtained for the single mutants, or at the every least mutant C152S, in which the third bridge ought to be intact. If the bridge were indeed important for the subunit stability and affinity, a marked difference between the single mutants and the other mutated proteins would be expected. However, as all seven cysteine mutants did show similar signal intensities and $rI_{C_{50}}$ values, the altered properties cannot be attributed exclusively to an absence of the third bridge.

In summary, even though dimer formation of H1-CRD can be inhibited by substitution of three cysteines, the mutant protein does not constitute an optimal solution. The affinity as well as the stability of the subunit is compromised by the mutations, making it behave differently compared to the WT. Therefore, SDM cannot be considered an eligible method in this case.

4.3 Investigation of the binding site of ASGP-R H1-CRD

Another part of this PhD thesis has been dedicated to investigate the relative importance of specific amino acid residue located in the binding site of H1-CRD. While the majority of the residues in the binding site, stretching from Arg236 to Cys268, have been studied and their contributions to the functionality of the subunit have been deduced, a few residues have remained unexamined. This subproject has aimed to further expand the overview of residues in the binding site.

4.3.1 Identification of residues important for the functionality of ASGP-R H1-CRD

Five mutant proteins were created, namely P237A, E238G, H256E, D259A and D260A. The mutations were confirmed by sequencing prior to protein expression in *E.coli*. The mutant proteins were purified according to the procedure established for WT H1-CRD. After the initial purification by affinity chromatography, the proteins were subjected to HPLC IEC for separation of monomers and dimers. Final samples were obtained after running the monomeric fractions over the GalNAc-Sep-column on

HPLC. The solid-phase competition assay and Biacore was used to evaluate binding affinity of the binding site mutants of H1-CRD.

Mutant P237A was investigated with both the solid-phase competition assay and Biacore. Taken together, the results demonstrated binding affinity for GalNAc in the same range as the wild-type. The K_D derived from the Biacore data was slightly lower than that for the WT. Concurrently, IC_{50} values obtained from the solid-phase competition assay resulted in rIC_{50} s just below one. Although the mutant displayed a lower signal response compared to the WT in both assays, it was clearly sufficient for evaluation of the data. Based on the results, it can be concluded that Pro237, located in loop four within the binding site of H1-CRD, plays a minor role in the functionality of the subunit. In general, proline is considered a rigid structure, which reduces the flexibility of the protein whenever it occurs. It is often found in β -turns, due to its ability to assume cis configuration [55]. It can be theorized that Pro237 is of importance for the local structure and stability of H1-CRD, but its conversion has little influence on the binding affinity.

The neighboring residue, Glu238 was converted to a glycine. Previous studies have shown that simultaneous insertion of Arg236 and Gly238 into corresponding positions of MGR increased the selectivity for GalNAc twofold [7]. Glycine is the native residue found in RHL-1, in which Arg236, as in H1-CRD, is also present. Mutant E238G was created to investigate potential differences in affinity caused by the two amino acids. The solid-phase competition assay showed a convincing increase in binding affinity for GalNAc of the mutant compared to the wild-type. Even though the standard deviation calculated for the three determined rIC_{50} values was fairly large, the results were unanimous. As suggested by earlier studies, residue 238 in H1-CRD can influence the binding affinity for GalNAc. Based on the results obtained within this work, binding is clearly favored by glycine compared to glutamate. While glycine is unlikely to interact with the ligand itself, it may influence the structural flexibility of the protein and by doing so facilitate binding. In contrast, glutamate carries a larger side chain, limiting its flexibility. Examination of the wild-type crystal structure reveals that Glu238 is positioned at the surface of the protein in a solvated area. Glutamate can be found deprotonated under physiological conditions and as it can form hydrogen bonds with surrounding water molecules, it is likely to have a positive influence on the protein solubility.

Histidine 256 in H1-CRD has previously been revealed as a critical residue involved in GalNAc binding to the subunit [7,49]. It has also been implicated in the ligand release during endocytosis [96,97]. In this work, His256 was converted to a glutamate to further investigate pH dependent ligand binding associated with the residue. Purification of mutant H256E proved to be more complicated compared to the other mutant proteins. Affinity chromatography resulted in only a minor elution of the mutant, while the protein could also be detected in the flow-through. Repeated denaturation and refolding of the flow-through was able to recover a small fraction of the initially misfolded H256E protein. The mutant was analyzed by the solid-phase competition assay and Biacore, confirming previous results. K_D of GalNAc determined by Biacore was higher for the mutant compared to WT H1-CRD, while the rIC_{50} based on the solid-phase competition assay was found to be below 1. Hence, His256 is important for binding of GalNAc and exchanging the residue will result in a decrease in binding affinity. The influence of pH on the ligand binding is discussed separately in section 4.3.2 below.

Mutant D259A was investigated with the solid-phase competition assay. The calculated rIC_{50} values, just below one, showed a slight decrease in binding of the polymer and hence affinity for GalNAc. However, the resulting rIC_{50} values closely resemble those determined for mutant P237A. It is possible that a second method would show binding similar to the WT, as was seen for mutant P237A. The corresponding residue in RHL-1 is a threonine. Although predicted as too distant for direct contact with a bound ligand, it has been shown to influence binding of GalNAc, possibly through positioning of His256 [7]. Therefore, any differences in binding affinity arising as a consequence of varying the residue at position 259 are likely to depend on the residue's ability to modify the protein structure. Converting Asp259 to alanine appears to be of minor importance for the binding of GalNAc. Observation of the crystal structure of H1-CRD locates Asp259 at the surface of H1-CRD (*Figure 70*).

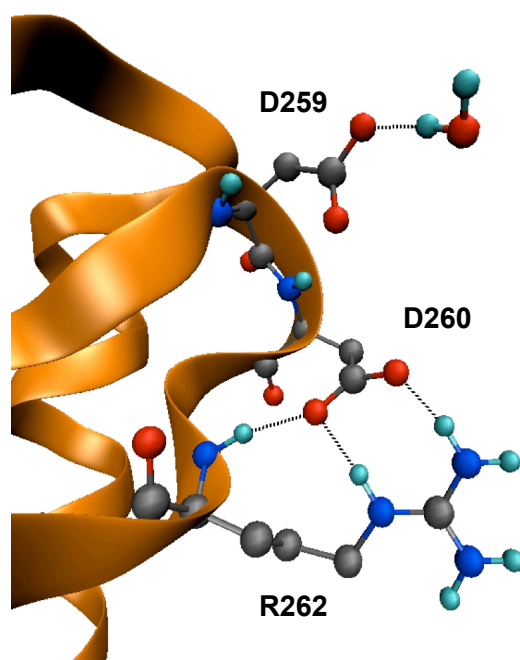


Figure 70. Close-up of the crystal structure of H1-CRD, focusing on D259 and D260. The residues are found at the surface of the protein. Asp259 can form hydrogen bonds with surrounding water molecules, while Asp260 can be seen to stabilize Arg262 [50].

The side chain is directed towards the surroundings, forming a hydrogen bond with a proximate water. Substitution with alanine causes the residue to lose its ability to act as a hydrogen acceptor and the solubility and the structure of the protein could be affected.

Aspartate 260, the final residue to be mutated, is conserved in both H1- and H2-CRD of human, mouse and rat. It was converted into alanine, to explore its contribution to the functionality of H1-CRD. The mutant protein was analyzed by the solid-phase competition assay. Detection of GalNAc binding was very limited due to the low signal response, also for protein replicates incubated exclusively with polymer. Even though the development time after addition of the substrate was prolonged, the results did not improve much. Taken together, the results suggest that the mutation D260A causes a major loss in binding affinity for GalNAc. The crystal structure shows Asp260 at the surface of the protein, closely interacting with a second residue, arginine 262 (*Figure 70*). The two residues are connected by three hydrogen bonds, stabilizing the loop in which they are situated and also the local protein structure. In contrast, alanine is unable to act as a hydrogen bond acceptor. The mutation is therefore likely to result in a destabilization of the protein as well as a change in the structure, reflecting negatively on the binding affinity of the subunit.

4.3.2 Histidine 256 – responsible for pH dependent ligand binding

Previous mutagenesis studies of RHL-1 have shown that histidine 256 plays an important role in the pH dependent ligand binding exhibited by the subunit [96,97]. In this work, His256 of H1-CRD was substituted by glutamate, the corresponding residue in H2-CRD. Binding of GalNAc to mutant H256E and WT H1-CRD was investigated by the solid-phase competition assay and Biacore at different pH ranging from 7.4 to 5. The solid-phase competition assay showed that binding of the GalNAc-polymer decreased at pH 6 or lower. Similar results were obtained from the Biacore experiments. K_D of GalNAc was seen to increase at pH lower than 6, signaling a reduction in affinity. While these observations were true for both the mutant and WT protein, a more dramatic change in ligand binding and K_D was detected for the latter. Histidine 256 clearly renders H1-CRD more sensitive to low pH compared to a glutamate in the same position.

His256 has been proposed to interact with Asn264, which in turn is involved in both calcium- and ligand binding. The histidine stabilizes Asn264 by a hydrogen bond formed between the imidazole and the amide (*Figure 71*).

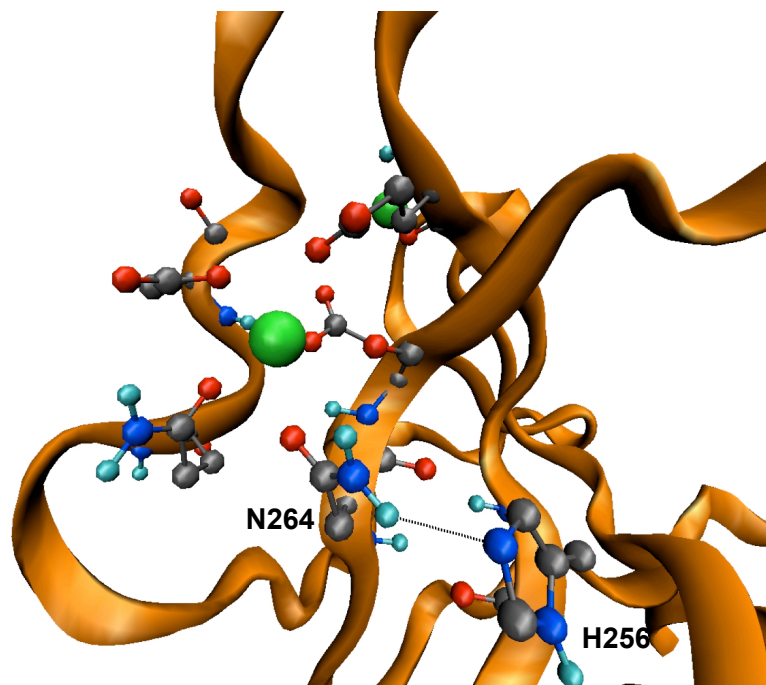


Figure 71. Close-up of the crystal structure of H1-CRD, focusing on N264 and H256 (drawn as ball-and-sticks). A stabilizing hydrogen bond with an approximate length of 2.43 Å is marked between the residues. Ligands for the calcium ion (green) are also highlighted as ball-and-sticks [50].

Protonation of the imidazole, caused by a drop in pH during endocytosis, will disrupt the hydrogen bond. As a consequence, the asparagine is destabilized and both ligand and calcium binding is disturbed. In contrast, glutamate is deprotonated at endosomal pH and theoretically able to accept a hydrogen bond from asparagine. Ligand binding can therefore be sustained at lower pH compared to the wild-type protein. The results obtained in this study could also show that mutant H256E was less affected by a decrease in pH. As H2-CRD contains a glutamate instead of a histidine, it can be expected that H2-CRD is less sensitive to acid pH than H1-CRD. However, it can also be anticipated that H2-CRD exhibits lower affinity for GalNAc due to the lack of His256. This theory is supported by preliminary results obtained within the ASGP-R project group at IMP [152].

4.4 Isotope labeling of ASGP-R H1-CRD

NMR holds a central position in protein studies. The method is often used to analyze protein-ligand interactions as a qualitative measure. It can also be applied in more elaborate studies to determine the 3D structure of proteins [92]. The first step when attempting structural elucidation by NMR is to generate protein enriched by rare isotopes such as ^{15}N and/or ^{13}C . As protein in milligram quantities is usually required, the expression system has to be both efficient and reproducible. The establishment of a method for isotope labeling of H1-CRD and enable future structure determination by NMR has been the aim of a third project carried out within this thesis.

Heteronuclear labeled protein is commonly produced by growing *E.coli* in minimal medium supplemented with a ^{15}N - and /or ^{13}C source. In order to facilitate bacterial growth and protein expression, the culture media as well as expression conditions are routinely optimized. One alternative expression set-up, here referred to as the two-stage protocol, utilizes unlabeled rich medium for growth of the *E.coli*, following by exchange of the bacteria into labeled minimal medium for protein expression [90]. During this work, both the two-stage protocol and exclusive use of minimal medium was attempted for labeling of H1-CRD.

4.4.1 H1-CRD expressed by the two-stage protocol – a vanishing act of ^{13}C and ^{15}N

The two-stage protocol was initially applied for expression of isotope labeled H1-CRD. Growth and expression conditions were optimized in terms of starter culture medium, concentration factor of cells, expression temperature and time as well as IPTG concentration. After determining the conditions resulting in the highest protein yields, H1-CRD labeled with ^{15}N , or both ^{15}N and ^{13}C , was expressed. The purification procedure established for WT H1-CRD was employed with good results as evidenced by sharp peaks seen during both FPLC affinity chromatography and HPLC IEC. NMR analysis of the two protein samples could confirm that the protein was well folded. However, the signals were fairly weak and the measurement time had to be prolonged beyond what is considered standard procedure. It was suspected that the isotopic incorporation level might be lower than initially expected, a suspicion that proved to be true. MS analysis and a NMR half-filter experiment showed that the majority of the protein sample was indeed unlabeled. Isotope labeled H1-CRD could also be observed, but to less extent. Due to the nature of multidimensional NMR experiments, a protein sample has to be isotopically enriched to a minimum of 90% in order to prove useful. Hence, the current method was evidently not optimal for the task it was intended and therefore needed to be improved. Furthermore, the obvious lack of labeled H1-CRD was found most peculiar as the protein had been expressed in minimal medium, in which the only nitrogen- and/or carbon source(s) were isotopically labeled. Similarly, the dominant population of unlabeled H1-CRD could initially not be explained. It can be argued that some unlabeled metabolites might remain in the bacteria also after inoculation in minimal medium and could be incorporated into the protein during expression, but it does not account for the clear majority of unlabeled protein. Moreover, the protocol includes one hour incubation in minimal medium prior to induction, partly to allow clearance of such unlabeled metabolites. Since expression is only induced after transferring the bacteria into the labeled minimal medium, the fraction of unlabeled protein should be kept to a minimum.

An explanation for the expressed unlabeled protein was proposed by external fellow scientists working with isotope labeling [153]. It was suggested that the problem arose due to unintended induction during the growth phase in rich medium. Rich medium, such as LB or TB, can contain traces of lactose, which in turn can induce the *lacUV5*

promotor and expression of T7 RNA polymerase. Hence, the bacteria would initiate expression already in the rich medium, explaining the high content of unlabeled protein.

The phenomenon of unintended induction of the T7 expression system was addressed in a recent paper [154]. Two different batches of N-Z-amine were used for protein expression and while the use of one was observed to cause unintended induction, the other one did not. A sample of Bacto tryptone from Difco (the brand used within this work) was also shown to cause unintended induction. It was concluded that the differences in inducing activity were likely to be dependent on varying amounts of lactose in the media components. N-Z-amine and tryptone are derived from enzymatic digestion of casein, a milk protein. Even though the casein would have been purified prior to digestion, trace amounts of lactose could remain in the final product.

TB medium contains more tryptone and yeast extract than LB, and consequently runs a higher risk of unintended induction. As TB medium had been used during the expression of H1-CRD, it was suggested to repeat the experiment using LB medium as starter culture medium instead. In addition, it was proposed to prolong the incubation time in minimal medium prior to induction. As the growth temperature was decreased following the medium exchange, the rate of bacterial metabolism could be affected. To ensure complete clearance of unlabeled metabolites it was therefore advised to extend the time between media change and induction. Finally, it was suggested to use carbenicillin instead of ampicillin, as ampicillin is more sensitive to culture conditions and degradation. Use of carbenicillin decreases the risk of possible contamination of non-transformed bacteria. The proposals were implemented in the method and referred to as an alternative expression set-up of the two-stage protocol. To examine whether the suggestions would improve the labeling efficiency, a small batch of ^{15}N labeled H1-CRD was prepared. Following expression, the protein was purified according to the standard procedure and HPLC RP was used to prepare a monomeric sample for MS analysis. The results showed a clear improvement in isotope enrichment with an incorporation level of 95%. However, while improving the first protocol, a second method using only minimal medium had simultaneously been attempted. The latter proved to result in an even higher labeling efficiency and it was chosen as the method of choice. It can be argued that a method based solely on minimal medium is more reliable as it will never risk incorporation of unlabeled metabolites.

4.4.2 Expression of isotope labeled H1-CRD

The second method made use of minimal medium for both bacterial growth and protein expression. To adapt the bacteria to medium poor in nutrients, the cells were initially grown in unlabeled minimal medium O/N. The starter culture was subsequently diluted 10X in labeled minimal medium and expression induced at an appropriate OD. High efficiency labeling was achieved for both ^{15}N and $^{13}\text{C}/^{15}\text{N}$, resulting in incorporation levels of 98% and 97%, respectively. The yield varied between 400 and 500 μg per liter expression culture. Initial NMR experiments with the ^{15}N labeled H1-CRD could confirm that the protein was well folded. A HSQC spectrum was acquired within 10 minutes and strong signals recorded, supporting the results of the MS analysis, *i.e.* the protein was highly enriched by ^{15}N . Binding of GalNAc was also investigated by adding different concentrations of the monosaccharide to the protein sample and monitoring changes in the peak positions. Taken together, the results indicate that the protein sample and hence, the method by which it was produced, provide a good basis for a future structure determination of H1-CRD by NMR. A high level of isotopic incorporation can be achieved and the final protein is both pure and well-behaved in terms of folding and ligand binding.

4.5 Summary and outlook

The work carried out within this thesis has strived to improve the knowledge on the H1-CRD subunit of the ASGP-R. Three subprojects were undertaken with the common goal to deduce the functionality of different structural elements of the subunit. Both single residues as well as the whole subunit have been targeted in the different projects. The individual aim of each project, as posted in the introduction, can be found summarized together with the obtained results and an outlook below.

4.5.1 Role of the three N-terminal cysteines

The goal of the project was to determine whether dimer formation of H1-CRD is caused by one of the cysteines closest to the *N*-terminus. Furthermore, it aimed to clarify how the cysteines are relevant for the functionality of the subunit.

The results showed that the dimerization seen upon expression of H1-CRD in *E.coli* is indeed dependent on the cysteines closest to the *N*-terminus, primarily Cys152 and/or Cys153. A single cysteine-to-serine point mutation was seen to reduce the dimer

content substantially. Merely traces of dimers were spotted compared to the wild-type protein. Double mutations did not improve the situation further, but substitution of all three cysteines abolished the dimer formation completely.

The cysteine mutants were further analyzed in respect to binding affinity for GalNAc. Disappointingly, all seven mutant proteins showed a decrease in affinity compared with the WT. No apparent difference could be observed between the single, double or triple mutants. The undisputable decrease in affinity suggested that not only the third disulfide bridge is important for the functionality of the subunit, but also the seventh odd cysteine.

In summary, although able to prevent dimer formation of H1-CRD, site-directed mutagenesis cannot be regarded as a suitable method due to the loss in GalNAc affinity. Alternative strategies, which do not influence the binding behavior of the subunit, ought to be considered. Chemical modification of the odd cysteine, using e.g. NEM or iodoacetamide, constitutes one such alternative and interesting approach.

4.5.2 Investigation of the binding site of H1-CRD

The importance of individual residues in the binding site of H1-CRD was to be determined by SDM. In addition, the project aimed to elucidate the role of histidine vs. glutamate at position 256 in pH-dependent ligand binding

Five mutant proteins were created containing single point mutations, namely P237A, E238G, H256E, D259A and D260A. Two of the mutant proteins, P237A and D259A, showed GalNAc affinity in the same range, or slightly lower, as WT H1-CRD. Hence, it can be concluded that these residues are of minor relevance for the functionality of the subunit. Conversion of Glu238 to glycine resulted in an increase in affinity for GalNAc. While previous studies exclude interaction between glycine and a ligand, the residue may influence the local structure of the protein and by doing so facilitate ligand binding. The fourth mutant, D260A, resulted in a substantial loss in binding affinity for GalNAc. Observation of the crystal structure of H1-CRD revealed Asp260 to form close interactions with a neighboring residue. The residue is expected to be important for the structural stability and indirectly also the ligand binding.

In summary, these results provide indications of the relevance of the investigated residues for the functionality of H1-CRD. An additional assay would be beneficial to accurately quantify the contributions of the residues to ligand binding.

Finally, mutant H256E was created to investigate the influence of histidine vs. glutamate on pH dependent ligand binding. While WT H1-CRD showed a dramatic loss in binding affinity for GalNAc at low pH, the change observed for the mutant protein was substantially less. It should be remarked, however, that the mutant protein showed lower binding affinity for GalNAc at physiological pH as a result of the substitution. In summary, His256 confers pH sensitivity as well as high ligand binding affinity to H1-CRD. Although H2-CRD theoretically would be able to sustain ligand binding at acidic pH, it lacks high affinity for GalNAc as glutamate, instead of histidine, is present in position 256.

4.5.3 Isotope labeling of H1-CRD

The aim of the project was to establish a method for high efficiency isotope labeling of H1-CRD to enable future structure determination by NMR.

Two methods were attempted. The first, referred to as the two-stage protocol, combined growth in unlabeled rich medium with expression in labeled minimal medium to reduce costs and increase yields. Despite much effort aimed to optimize the expression set-up, the final protein sample showed an unexpected high degree of unlabeled H1-CRD. Presence of the unlabeled H1-CRD was suggested to be due to unintended induction caused by lactose traces in the rich medium. After modification of the protocol, a protein sample with a ^{15}N incorporation level of 95% was obtained.

A second method, relying solely on use of minimal medium, was employed as the two-stage protocol initially failed to produce isotope labeled H1-CRD. The second method was immediately successful as a high enrichment of ^{15}N , 98%, could be achieved. Repeated expression with ^{13}C and ^{15}N showed reproducible results with an overall incorporation level of 97%. In addition, preliminary NMR experiments with the ^{15}N labeled H1-CRD could confirm that the protein maintained a stable tertiary structure and that GalNAc affinity was preserved.

In summary, a method for high efficiency isotope labeling of H1-CRD has been established. The method makes use of only minimal medium, thereby minimizing the risk of unintended induction and expression of unlabeled protein. Obtained protein samples show high isotope incorporation levels and high quality, making them suitable for structural elucidation by NMR.

5 References

1. V. P. Torchilin, *Eur J Pharm Sci* **2000**, *11*, 81-91.
2. F. Winau, O. Westphal, R. Winau, *Micorbes Infect* **2004**, *6*, 786-789.
3. J. Wu. M. H. Nantz, M. A. Zern, *Front Biosci* **2002**, *7*, 717-725.
4. P. H. Weigel, J. H. N. Yik, *Biochim Biophys Acta* **2002**, *1572*, 341-363.
5. M. Meier, M. D. Bider, V. N. Malashkevich, M. Spiess, P. Burkhard, *J Mol Biol* **2000**, *300*, 857-865.
6. S. T. Iobst, K. Drickamer, *J Biol Chem* **1994**, *269*, 15512-15519.
7. S. T. Iobst, K. Drickamer, *J Biol Chem* **1996**, *271*, 6686-6693.
8. A. R. Kolatkar, W. I. Weis, *J Biol Chem* **1996**, *271*, 6679-6685.
9. W. Jahnke, H. Widmer, *Cell Mol Life Sci* **2004**, *61*, 580-599.
10. A. G. Morell, R. A. Irvine, I. Sternlieb, I. H. Scheinberg, G. Ashwell, *J Biol Chem* **1968**, *243*, 155-159.
11. G. Ashwell, J. Harford, *Annu Rev Biochem* **1982**, *51*, 531-554.
12. K. Drickamer, *J Biol Chem* **1988**, *263*, 9557-9560.
13. D. C. Kilpatrick, *Biochim Biophys Acta* **2002**, *1572*, 187-197.
14. A. N. Zelensky, J. E. Gready, *FEBS J* **2005**, *272*, 6179-6217.
15. R. J. Stockert, *Physiol Rev* **1995**, *75*, 591-609.
16. K. B. Chiacchia, K. Drickamer, *J Biol Chem* **1984**, *259*, 15440-15446.
17. M. Spiess, H. F. Lodish, *Proc Natl Acad Sci* **1985**, *82*, 6465-6469.
18. J. H. N. Yik, A. Saxena, P. H. Weigel, *J Biol Chem* **2002**, *277*, 23076-23083.
19. Picture courtesy of D. Abgottspon, Institute of Molecular Pharmacy, University of Basel, Switzerland, **2007**
20. M. R. Hardy, R. R. Townsend, S. M. Parkhurst, Y. C. Lee, *Biochemistry* **1985**, *24*, 22-28.
21. A. L. Schwartz, C. J. Steer, E. S. Kempner, *J Biol Chem* **1984**, *259*, 12025-12029..
22. M. D. Bider, J. M. Wahlberg, R. A. Kammerer, M. Spiess, *J Biol Chem* **1996**, *271*, 31996-32001.

23. M. D. Bider, R. Cescato, P. Jenö, M. Spiess, *Eur J Biochem* **1995**, *230*, 207-212.
24. U. Westerlind, J. Westman, E. Törnquist, C. I. E. Smith, S. Oscarson, M. Lahmann, T. Norberg, *Glycoconj J* **2004**, *21*, 227-241.
25. P. C. N. Rensen, L. A. J. M. Sliedregt, M. Ferns, E. Kieviet, S. M. W. van Rossenberg, S. H. van Leeuwen, T. J. C van Berjel, E. A. L. Biessen, *J Biol Chem* **2001**, *276*, 37577-37584.
26. J. U. Baenziger, Y. Maynard, *J Biol Chem* **1980**, *255*, 4607-4613.
27. M. D. Bider, M. Spiess, *FEBS Lett* **1998**, *434*, 37-41.
28. R. Tozawa, S. Ishibashi, J. Osuga, K. Yamamoto, H. Yagyu, K. Ohashi, Y. Tamura, N. Yahagi, Y. Iizuka, H. Okazaki, K. Harada, T. Gotoda, H. Shimano, S. Kimura, R. Nagai, N. Yamada, *J Bio Chem* **2001**, *276*, 12624-12628.
29. S. Ishibashi, R. E. Hammer, J. Herz, *J Biol Chem* **1994**, *269*, 27803-27806.
30. E. Windler, J. Greeve, B. Levkau, V. Kolb-Bachofen, W. Daerr, H. Greten, *Biochem J* **1991**, *276*, 79-87.
31. R. F. Rotundo, R. A. Rebres, P. J. Mckeown-Longo, F. A. Blumenstock, T. M. Saba, *Hepatology* **1998**, *28*, 475-485.
32. L. Dini, F. Auturi, A. Lentini, S. Oliverio, M. Piacentini, *FEBS* **1992**, *296*, 174-178.
33. A. Rifai, K. Fadden, S. L. Morrison, K. R. Chintalacharuvu, *J Exp Med* **2000**, *191*, 2171-2182.
34. S. Becker, M. Spiess, H. D. Klenk, *J Gen Virol* **1995**, *76*, 393-399.
35. T. Owada, K. Matsubayashi, H. Sakata, H. Ihara, S. Sato, K. Ikebuchi, T. Kato, H. Azuma, H. Ikeda, *J Viral Hepat* **2006**, *13*, 11-18.
36. U. Treichel, K. H. Meyer zum Buschenfelde, H. P. Dienes, G. Gerken, *Arch Virol* **1997**, *142*, 493-498.
37. B. Saunier, M. Triyatni, L. Ulianich, P. Maruvada, P. Yen, L. D. Kohn, *J Virol* **2003**, *77*, 546-599.
38. I. Geffen, M. Spiess, *Int Rev Cytol* **1992**, *137*, 181-219.
39. M. Spiess, *Biochemistry* **1990**, *29*, 10009-10018.
40. Picture courtesy of Dr. D. Ricklin, Institute of Molecular Pharmacy, University of Basel, Switzerland, **2005**
41. A. L. Schwartz, D. Rup, *J Biol Chem* **1983**, *258*, 11249-11255.

42. C. Fuhrer, I. Geffen, M. Spiess, *J Cell Biol* **1991**, *114*, 423-431.
43. C. Fuhrer, I. Geffen, K. Huggel, M. Spiess, *J Biol Chem* **1994**, *269*, 3277-3282.
44. D. A. Wall, A. L. Hubbard, *J Cell Biol* **1981**, *90*, 687-696.
45. S. Matsuura, H. Nakada, T. Sawamura, Y. Tashiro, *J Cell Biol* **1982**, *95*, 864-875.
46. J-H. Park, E-W. Cho, S. Y. Shin, Y-J. Lee, K. L. Kim, *Biochem Biophys Res Comm* **1998**, *244*, 304-311.
47. J-H. Park, K. L. Kim, E-W. Cho, *Biotechnol Lett* **2006**, *28*, 1061-1069.
48. Y. Y. Seow, M. G. Tan, K. T. Woo, *Nephron* **2002**, *91*, 431-438.
49. A. R. Kolatkar, A. K. Leung, R. Isecker, R. Brossmert, K. Drickamer, W. I. Weis, *J Biol Chem* **1998**, *273*, 19502-19508.
50. Picture courtesy of M. Spreafico, Institute of Molecular Pharmacy, University of Basel, Switzerland, **2007**
51. J. Wu, P. Liu, J-L. Zhu, S. Maddukuri, M. A. Zern, *Hepatology* **1998**, *27*, 772-778.
52. L. A. J. M. Sliedregt, P. C. N. Rensen, E. T. Rump, P. J. van Santbrink, M. K. Bijsterbosch, A. R. P. M. Valentijn, G. A. van der Marel, J. H. van Boom, T. J. C. van Berkel, E. A. L. Biessen, *J Med Chem* **1999**, *42*, 609-618.
53. H. P. Sorensen, K. K. Mortensen, *J Biotechnol* **2005**, *115*, 113-128.
54. W. Schumann, L. C. S. Ferreira, *Genet Mol Biol* **2004**, *27*, 442-453.
55. D. L. Nelson, M. M. Cox, in *Lehninger Principles of Biochemistry*, 3rd edition, Worth Publishers, New York, **2000**
56. F. Baneyx, *Curr Opin Biotechnol* **1999**, *10*, 411-421.
57. K. Terpe, *Appl Microbiol Biotechnol* **2006**, *72*, 211-222.
58. S. C. Makrides, *Microbiol Rev* **1996**, *60*, 512-538.
59. Novagen, pET system manual 10th edition, **2003**
60. F. W. Studier, B. A. Moffat, *J Mol Biol* **1986**, *189*, 113-130.
61. W. E. Volk, B. A. Gebhardt, M-L. Hammar-sköld, R. J. Kadner, in *Essentials of Medical Microbiology*, 5th edition, Lippincott-Raven Publishers, Philadelphia, **1996**
62. H. P. Sorensen, K. K. Mortensen, *Microb Cell Fact* **2005**, *4*, 1-8.

63. P. Jonasson, S. Liljeqvist, P-Å. Nygren, S. Ståhl, *Biotechnol Appl Biochem* **2002**, 35, 91-105.
64. E. De Bernardez Clark, *Curr Opin Biotechnol* **1998**, 9, 157-163.
65. E. De Bernardez Clark, *Curr Opin Biotechnol* **2001**, 12, 202-207.
66. B. R. Glick, J. J. Pasternak, in *Molecular Biotechnology, Principles and Applications of Recombinant DNA*, 2nd edition, ASM Press, Washington, D. C., **1998**
67. M. M. Ling, B. H. Robinson, *Anal Biochem* **1997**, 254, 157-178.
68. P.C. Turner, A. G. McLennan, A.D. Bates, M. R. H. White, in *Instant Notes in Molecular Biology*, BIOS Scientific Publishers, Oxford, **1997**
69. F. Vallette, E. Mege, A. Reiss, M. Adesnik, *Nucleic Acids Res* **1989**, 17, 723-733.
70. R. Higuchi, B. Krummel, R. K. Saiki, *Nucleic Acids Res* **1988**, 16, 7351-7367.
71. A. Hemsley, N. Atnheim, M. D. Toney, G. Cortopassi, D. Galas, *Nucleic Acids Res* **1989**, 17, 6545-6551.
72. D. J. Scholl, J. N. Wells, *Biochem Pharmacol* **2000**, 60, 1647-1654.
73. J. N. Tinguely, B. Wermuth, *Eur J Biochem* **1999**, 260, 9-14
74. K. L. Morrison, G. A. Weiss, *Curr Opin Chem Biol* **2001**, 5, 302-307.
75. F. Lefevre, M-H. Remy, J-M. Masson, *Nucleic Acids Res* **1997**, 25, 447-447.
76. L. A. de Jong, D. R. Uges, J. P. Franke, R. Bischoff, *J Chromatogr B Analyt Technol Biomed Life Sci* **2005**, 829, 1-25.
77. D. Stokmaier, *Receptor-mediated endocytosis in human hepatocytes*, diploma thesis, University of Basel (Basel, Switzerland), **2004**
78. G. Weitz-Schmidt, D. Stokmaier, G. Scheel, N. E. Nifant'ev, A. B. Tuzikov, N. V. Bovin, *Anal Biochem* **1996**, 238, 184-190.
79. N. V. Bovin, *Glycoconj J* **1998**, 15, 431-446.
80. U. Jonsson, L. Fagerstam, B. Ivarsson, B. Johnsson, R. Karlsson, K. Lundh, S. Lofas, B. Persson, H. Roos, I. Ronnberg, *et al.*, *Biotechniques* **1991**, 11, 620-7
81. P. Torreri, M. Ceccarini, P. Macioce, T. C. Petrucci, *Ann 1st Super Sanita* **2005**, 41, 437-441.
82. Picture courtesy of A. Vögtli, Institute of Molecular Pharmacy, University of Basel, Switzerland, **2006**

83. N. K. Goto, L. E. Kay, *Curr Opin Struct Biol* **2000**, *10*, 585-592.
84. J. Keeler, in *Understanding NMR Spectroscopy*, John Wiley and Sons Ltd, West Sussex, **2005**
85. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, N. J. Skelton, in *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego, **1996**
86. H. Patzelt, N. Goto, H. Iwai, K. Lundstrom, E. Fernholz, in *BioNMR Techniques in Drug Research*, O. Zerbe, ed., Wiley-VCH, **2002**
87. G. Gottschalk, in *Bacterial Metabolism*, 2nd edition, Springer, **1986**
88. J. Sambrook, E.F. Fritsch, T. Maniatis, in *Molecular Cloning, A laboratory manual*, 2nd edition, Cold Spring Harbor Laboratory Press, **1989**
89. M. Cai, Y. Huang, K. Sakaguchi, G. M. Clore, A. M. Gronenborn, R. Craigie, *J Biomol NMR* **1998**, *11*, 97-102.
90. J. Marley, M. Lu, C. Bracken, *J Biomol NMR* **2001**, *20*, 71-75.
91. D. A. Lindhout, A. Thiessen, D. Schieve, B. D. Sykes, *Protein Sci* **2003**, *12*, 1786-1791.
92. G. C. K. Roberts, *DDT* **2000**, *5*, 230-240.
93. T. de Beer, R. E. Carter, K. E. Lobel-Rice, A. Sorkin, M. Overduin, *Science* **1998**, *281*, 1357-1360.
94. R. Born, *Benefit and Application of Antibodies against the H1 Carbohydrate Recognition Domain of the Human Hepatic Asialoglycoprotein Receptor*, PhD thesis, University of Basel (Basel, Switzerland), **2005**
95. D. Ricklin, *Surface Plasmon Resonance Applications in Drug Discovery with an Emphasis on Small Molecule and Low Affinity Systems*, PhD thesis, University of Basel (Basel, Switzerland), **2005**
96. S. Wragg, K. Drickamer, *J Biol Chem* **1999**, *274*, 35400-35406.
97. H. Feinberg, D. Torgersen, K. Drickamer, W. I. Weis, *J Biol Chem* **2000**, *275*, 35176-35184.
98. Novagen, pET-3a-d vectors, TB026, 12/98
99. M. P. Deutscher, in *Methods in enzymology, a guide to protein purification*, volume 182, Academic press, **1990**
100. N. Fornstedt, J. Porath, *FEBS Lett* **1975**, *57*, 187-191.
101. U. K. Laemmli, *Nature* **1970**, *227*, 680-685.

102. B. D. Hames, N. M. Hooper, J. D. Houghton, in *Instant Notes in Biochemistry*, BIOS Scientific Publishers, Oxford, **1997**
103. Amersham Biosciences, Silver staining kit, protein, data file 18-1122-21
104. D. Kang, Y. S. Gho, M. Suh, C. Kang, *Bull Korean Chem Soc* **2002**, 23, 1511-1512.
105. Amersham biosciences, *Ion Exchange Chromatography and Chromatofocusing, Principles and Methods*, Handbook 11-0004-21, edition AA
106. J. M. Walker, in *The Protein Protocol Handbook*, 2nd edition, Humana Press, **2002**
107. Stratagene, ExSite™ PCR-Based Site-Directed Mutagenesis kit instruction manual, Catalog 200502, Revision 073006f
108. Promega, T4 DNA Ligase, Technical Bulletin, Catalog M1801
109. Applied Biosystems, BigDye® Terminator v3.1 Cycle Sequencing kit, Protocol 4337035, 09/2002
110. Qiagen, *The QIAexpressionist*, 5th edition, **2003**
111. <http://www.geocities.com/~shigemi/>
112. J. W. Peng, C. A. Lepre, J. Fejzo, N. Abdul-Manan, J. M. Moore, *Methods Enzymol* **2001**, 338, 202-230.
113. M. Mayer, B. Meyer, *Angew Chem Int Ed* **1999**, 38, 1784-1788.
114. H. Geen, S. Wimperis, R. Freeman, *J Magn Reson* **1989**, 85, 620-627.
115. B. Cutting, S. V. Shelke, Z. Dragic, B. Wagner, H. Gathje, S. Kelm, B. Ernst, *Mag. Reson. Chem* **2007**, in press
116. H. Peters, Institute of Chemistry, University of Luebeck, Germany
117. C. Li, J. Bai, Z. Cai, F. Ouyang, *J Biotechnol* **2002**, 93, 27-34.
118. <http://www.acc.umu.se/%7Etnkjtj/chemometrics/editorial/aug2002.html>
119. J. Gabrielsson, N-O. Lindberg, T. Lundstedt, *J Chemometrics* **2002**, 16, 141-160.
120. A. Dutta, B. J. Rao, K. V. R. Chary, *Protein Expr Purif* **2003**, 29, 252-258.
121. M. Devany, N. P. Kotharu, H. Matsuo, *Protein Expr Purif* **2005**, 40, 244-247.
122. A. Bax, M. Ikura, L.E. Kay, D.A. Torchia, R. Tschudin, *J Magn Reson* **1990**, 86, 304-318.

123. G. Bodenhausen and D.J. Ruben, *Chem Phys Lett* **1980**, 69, 185-189.
124. J. Cavanagh, A.G Palmer, P.E. Wright, and M. Rance, *J Magn Reson* **1990**, 91, 429-436.
125. P. E. Johnson, P. Tomme, M. D. Joshi, L. P. McIntosh, *Biochemistry* **1996**, 35, 13895-13906.
126. G. Otting, K. Wuethrich, *J Magn Reson* **1989**, 85, 586-594.
127. H.C. Kolb, B. Ernst, *Eur J Chem* **1997**, 3, 1571-1578.
128. Amersham biosciences, *The Recombinant Protein Handbook, Protein Amplification and Simple Purification*, Handbook 18-1142-75
129. M. Spiess, A. L. Schwartz, H. F. Lodish, *J Biol Chem* **1985**, 260, 1979-1982.
130. <http://www.expasy.org/>
131. M. Nasri, D. Thomas, *Nucleic Acids Res* **1986**, 14, 811-821.
132. J. George, R. W. Blakesley, J. G. Chirikjian, *J Biol Chem* **1980**, 255, 6521-6524.
133. <http://www.neb.com/nebecomm/products/faqproductR0136.asp#1>
134. <http://www.expasy.org/tools/dna.html>
135. http://www.ch.embnet.org/software/LALIGN_form.html
136. Silantes, *E.coli* OD media, Technical Bulletin, **2002**
137. S. Honnappa, B. Cutting, W. Jahnke, J. Seelig, M. O. Steinmetz, *J Biol Chem* **2003**, 278, 38926-38924.
138. P. E. Wright, H. J. Dyson, *J Mol Biol* **1999**, 293, 321-331.
139. <http://www.bruker-biospin.com/cryoprobes.html>
140. C. Stover, M. P. Mayhew, M. J. Holden, A. Howard, D. T. Gallagher, *J Struct Biol* **2000**, 129, 96-99.
141. M. Lechmann, N. Kozor, E. Zinser, A. T. Prechtel, H. Sticht, A. Steinkasserer, *Biochem Biophys Res Commun* **2005**, 329, 132-139.
142. R. R. Gali, P. G. Board, *Biochem J* **1997**, 321, 207-210.
143. T. Betakova, B. Moss, *J Virol* **2000**, 74, 2438-2442.
144. L. M. Veenhoff, E. H. Heuberger, B. Poolman, *EMBO J* **2001**, 20, 3056-3062.
145. D. Abgottspon, B. Ernst, Institute of Molecular Pharmacy, University of Basel, Switzerland, *unpublished results*, **2007**

146. S. J. Ennion, R. J. Evans, *Mol Pharmacol* **2002**, *61*, 303-311.
147. U. Lathinen, K. Svensson, R. F. Petterson, *Eur J Biochem* **1999**, *260*, 392-397.
148. D. F. Mark, S. D. Lu, A. A. Creasey, R. Yamamoto, L. S. Lin, *PNAS* **1984**, *81*, 5662-5666.
149. G. Bulaj, *Biotechnol Adv* **2005**, *23*, 87-92
150. M. Matsumura, G. Signor, B. W. Matthews, *Nature* **1989**, *342*, 291-293.
151. M. H. Yuk, H. F. Lodish, *J Biol Chem* **1995**, *270*, 20169-20176.
152. S. Rabbani, D. Stokmaier, B. Cutting, D. Abgottspon, B. Ernst, Institute of Molecular Pharmacy, University of Basel, Switzerland, *unpublished results*, **2007**
153. EMBL Workshop, *Protein Expression, RNA Transcription and Isotope Labeling for NMR*, Heidelberg, Germany, **2006**
154. F. W. Studier, *Protein Expr Purif* **2005**, *41*, 207-234.

Karin Johansson

Date of birth 1979-01-06
Nationality Swedish
Current residence Muespacherstrasse 64, 4055 Basel, Switzerland
E-mail Karin.Johansson@unibas.ch

Academic studies

2007 – 2003 PhD project, Institute of Molecular Pharmacy, University of Basel, Switzerland
“Linking structure and function of the Asialoglycoprotein Receptor H1-CRD using Site-directed Mutagenesis and Isotope Labeling”

2003 Approved Pharmacist License, Uppsala University, Sweden

2002 – 1998 Master of Science in Pharmacy, Uppsala University, Sweden

2002 Master thesis project, King’s College London, United Kingdom
“Cytotoxic properties of a tropical black fungus and *Penicillium polonicum*”

Work experience

2003 Licensed Pharmacist, Stockholm, Sweden

Scientific endeavors

2006 EMBL workshop, Heidelberg, Germany
“Protein expression, RNA transcription and Isotope labeling for NMR”

2005 Oral presentation, Swiss Chemical Society Fall Meeting, Lausanne, Switzerland
“Cloning and Characterization of Cysteine Mutants of the ASGP-R H1-CRD”

Educational experience

2007 – 2004 Supervision of the student practical course “Modern Drug Design”, Institute of Molecular Pharmacy, University of Basel

2007 – 2005 Supervision of diploma work, Institute of Molecular Pharmacy, University of Basel

- Investigation of the biological function of H2-CRD in comparison to H1-CRD of the asialoglycoprotein receptor
- Site-directed mutagenesis of the binding site of the ASGP-R
- Optimizing an assay to evaluate bradykinin receptor antagonists for the bioassay practicum
- Characterization of the binding site of the H1-CRD subunit of the ASGP-R

Research areas

Cloning, site-directed mutagenesis, protein expression and purification, isotope labeling