Using biophysical characterization to begin to understand the onset, diagnosis, and treatment of cancer

by

Natalie Jean Krahn

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

University of Manitoba

Winnipeg

Copyright © 2017 by Natalie Jean Krahn

Abstract

Cancer is the leading cause of death among Canadians with about one third of the population at risk of being diagnosed with cancer in their lifetime. Efforts are in place to develop novel or improved ways to diagnose and treat cancer, in hopes to reduce these statistics. This thesis focuses on the role of two important proteins in cancer: high mobility group AT-hook 2a (HMGA2a) and a chimeric heavy-chain antibody (EG2-hFc).

HMGA2a (HMGA2) is a transcription activating factor which is found to be bound to chromatin in the nucleus. The overexpression of HMGA2 leads to increased cell proliferation, a crucial role in embryonic stem cells. However, the presence of significant amounts of HMGA2 in adult somatic cells leads to the formation of tumours. HMGA2 is made up of three AT-hook binding domains through which it interacts with various forms of DNA including dsDNA, and distorted DNA segments such as replication forks (RFs), and Holliday junctions. It is the interaction of HMGA2 with RFs that result in the observed high cell proliferation. HMGA2 was recombinantly produced and studied using a biophysical approach to investigate its behaviour in solution both in the bound and unbound states. We determined that HMGA2 is intrinsically disordered and behaves as a monomer up to concentrations of 10 mg/mL. Furthermore, upon interaction with RF DNA in a 1:1 ratio, HMGA2 was found to retain its flexibility and bind in a side-by-side orientation to the RF. We also confirmed the important role of HMGA2 in cell viability through cell culture knockdown studies. From this information we can begin to elucidate the interaction mechanism of HMGA2 with RF DNA and understand its biological significance.

EG2-hFc is a chimeric heavy-chain mAb against the epidermal growth factor receptor (EGFR). It was designed to be smaller in size than traditional mAb therapeutics for deeper

tumour penetration, but large enough such that it would not be easily cleared through the renal system. The composition and heterogeneity of the sugars attached at Asn297 in the Fc region can have a great impact on the efficacy of the mAb. Therefore I focused on implementing a glycoengineering approach to alter the N-glycosylation on EG2-hFc and furthermore analyze the effect that changing the glycosylation pattern had on its hydrodynamic and binding properties. Glycoprotein processing inhibitors added to the cell growth media generated mAbs with complex, high mannose (with or without glucose), or hybrid glycans. These altered EG2-hFc mAbs were observed to have similar secondary structure and hydrodynamic properties to one another. However, the presence and composition of sugars at Asn297 had an effect on the binding affinity (K_D) of EG2-hFc to FcγRI. Specifically, hybrid glycans reduced the K_D two-fold when compared to the complex glycans while removal of the N-glycan drastically (> 100-fold) reduced binding to a full-sized IgG1 but only marginally (~ 8-fold) reduced binding to EG2-hFc. These results suggested that the design of EG2-hFc stabilizes the Fc pocket such that complete removal of the N-glycan does not have as significant an effect on binding as would be observed with full-sized mAbs. Furthermore, as a potential therapeutic candidate for cancer treatment, this increased stability makes it attractive for industry.

Acknowledgements

I would like to acknowledge all those who have supported me these past four years in my graduate research. Specifically my direct supervisor, Dr. Jörg Stetefeld, who supported me over the years not only financially but as a researcher, wife and most recently a mother. Without his guidance and flexibility, I would not be the researcher I am today. Dr. Stetefeld encouraged me to maintain my collaborations with Dr. Michael Butler and MabNet from which I was able to attend multiple conferences and workshops in addition to publishing a number of peer-reviewed publications. His flexibility gave me the opportunity to continue with my research without compromising my growing family of which I am truly grateful.

I would also like to acknowledge Dr. Butler for his financial and moral support in not only my research but through attending conferences. I believe this experience of presenting at national and international conferences has broadened my scientific knowledge, provided me a multitude of scientific contacts and given me the confidence to speak publicly about my research. These opportunities were invaluable to my scientific career and greatly appreciated.

I would also like to thank Drs. Trushar Patel and Markus Meier for their time and patience in teaching me new techniques in the lab as well as the corresponding analyses. My research took on a whole new perspective with their guidance and expertise, and for that alone I am grateful. Additionally, all former and current students who I have encountered along the way both in Dr. Stetefeld's and Dr. Butler's labs as well as the Chemistry and Microbiology department. They have made my experience as a graduate student a truly enjoyable and unforgettable experience.

Finally I would like to thank my advisory committee Dr. Butler, Dr. Sean McKenna, and Dr. Francis Lin who were always quick to respond to emails and made themselves available for

meetings. Their efforts in reviewing my thesis proposal, candidacy exam and thesis as well as any guidance they have instilled upon me are greatly appreciated.

My research was supported by the Faculty of Science and Graduate Enhancement of Tri-Council Stipends (GETS) fellowships provided by the University of Manitoba. Travel support was provided by the Department of Chemistry and Faculty of Science from the University of Manitoba.

Table of Contents

Acknowledgements ..... **iv**

List of Tables ..... **ix**

List of Figures ..... **x**

Contributions of Authors ..... **xi**

List of Copyrighted Material for which Permission has been Obtained ..... **xii**

List of Abbreviations ..... **xiv**

Chapter 1 : Introduction **17**

1.1 Organization of the dissertation 17

1.2 Cancer in Canada 18

1.3 Detecting cancer 20

1.4 Antibodies as therapeutics 22

1.5 Resistance to cancer treatment 24

1.6 Research objectives and hypotheses 26

Chapter 2 : Theory behind the techniques **32**

2.1 Size exclusion chromatography 32

2.2 Dynamic light scattering 34

2.3 Analytical ultracentrifugation 37

2.3.1 Sedimentation velocity 38

2.3.2 Sedimentation equilibrium 39

2.3.3 Absorbance optics 40

2.3.4 Interference optics 41

2.3.5 Fluorescence optics 42

2.3.6 Data analysis 43

2.4 Circular dichroism 44

2.4.1 Principle 44

2.4.2 CD of proteins 45

2.4.3 CD of nucleic acids 48

2.5 Nuclear magnetic resonance 52

2.5.1 Theory 52

2.5.2 Sample preparation 52

2.5.3 Secondary structure determination 53

2.5.4 Chemical shift mapping 55
2.6 Small angle x-ray scattering 55
2.6.1 Data collection and analysis 55
2.6.2 Ab-inito modelling 58
Chapter 3 : Nano-scale assembly of high-mobility group AT-hook 2 protein with DNA replication fork **67**
3.1 Introduction 68
3.2 Materials and Methods 70
3.2.1 Protein expression and purification 70
3.2.2 Formation of RF DNA with and without HMGA2 71
3.2.3 Coomassie SDS-PAGE 71
3.2.4 DNA gel shift assays 72
3.2.5 Hydrodynamic characterization 72
3.2.6 Spectropolarimetry 73
3.2.7 Chemical shift mapping of amide backbone 73
3.2.8 Low resolution solution scattering 74
3.2.9 Calculation of hydrodynamic parameters from ab initio models. 74
3.2.10 Knockdown of HMGA2 75
3.3 Results 75
3.3.1 HMGA2, RF and the complex can be purified to homogeneity 75
3.3.2 HMGA2 and RF are monomeric in solution and form an equimolar complex 78
3.3.3 HMGA2 is disordered in solution 82
3.3.4 CD spectropolarimetry confirms assembly of RF 85
3.3.5 HMGA2 retains flexibility when in complex with RF 85
3.3.6 HMGA2 interacts with RF in a side-by-side manner 87
3.3.7 HMGA2 plays a role in cell proliferation 89
3.4 Discussion 91
Chapter 4 : Inhibition of glycosylation on a camelid antibody uniquely affects its FcγRI binding activity **106**
4.1 Introduction 107
4.2 Materials and Methods 111
4.2.1 Mammalian cell culture 111
4.2.2 FcγRI production and purification 111

4.2.3 Purification and concentration determination of mAbs 112
4.2.4 N-glycosylation analysis 113
4.2.5 Spectropolarimetry 114
4.2.6 Hydrodynamic studies 114
4.2.7 FcγRI receptor interaction 115
4.2.8 Homology modeling 116
4.3 Results 116
4.3.1 Growth and productivity 116
4.3.2 *MAb glycan microheterogeneity* 118
4.3.3 Secondary structure analysis 122
4.3.4 Homogeneity of mAbs 125
4.3.5 *Stability of mAbs* 129
4.3.6 *FcγRI binding affinity* 129
4.4 Discussion 132
Chapter 5: Conclusion 150
5.1 Thesis summary 150
5.2 Future perspectives 153

List of Tables

Table 3.1 Summary of Hydrodynamic Parameters for HMGA2, RF, and complex 81
Table 4.1 Summary of the composition of N-glycans attached to N297 for EG2-hFc and DP-12 samples treated with glycoprotein processing inhibitors..... 121
Table 4.2 Summary of the hydrodynamic and physical properties determined for EG2-hFc and DP-12 samples treated with glycoprotein processing inhibitors 125
Table 4.3 Effect of N-glycan on affinity (K_D) of EG2-hFc and DP-12 to FcγRI determined using SPR 130

List of Figures

Figure 1.1 Natural killer cells mediate ADCC response in cancer cells 23
Figure 1.2 Hypothetical scheme for involvement of HMG proteins in DNA damage recognition and repair 26
Figure 2.1 Size-exclusion chromatography calibration curve 33
Figure 2.2 Far UV CD spectra associated with various types of secondary structure 47
Figure 2.3 Circular dichroism spectra of B-A and B-Z transitions of DNA 51
Figure 2.4 Chemical shift dispersion of structured and intrinsically disordered proteins 54
Figure 3.1 Schematic of the domains of human HMGA2 highlighting the AT-hook domains ... 70
Figure 3.2 Purity of HMGA2, RF and complex 76
Figure 3.3 Coomassie SDS-PAGE of purified and cleaved HMGA2 77
Figure 3.4 Electrophoretic mobility shift assay of RF interaction with HMGA2 78
Figure 3.5 Monodisperse hydrodynamic radius and sedimentation coefficient distributions of HMGA2, RF, and complex 80
Figure 3.6 HMGA2 sequence prediction 82
Figure 3.7 Secondary structure analysis of HMGA2, RF, and complex 83
Figure 3.8 Secondary structure analysis of bound and unbound HMGA2 84
Figure 3.9 SAXS plots of HMGA2, RF, and complex merged from multiple concentrations 85
Figure 3.10 Low resolution models of RF and complex suggest RF interacts with HMGA2 in a side-by-side orientation 88
Figure 3.11 Knockdown of HMGA2 suggests role in cellular proliferation 90
Figure 4.1 Antibody domain schematic of IgG1 compared to EG2-hFc highlighting the position of their N-glycosylation sites on the C_H2 domains (red oval). 110
Figure 4.2 Effect of glycoprotein processing inhibitors on the cell growth and mAb yield 118
Figure 4.3 Gel electrophoresis analysis of glycosylated and non-glycosylated EG2-hFc and DP-12 mAbs 119
Figure 4.4 HILIC-HPLC glycan analysis of mAb glycans 120
Figure 4.5 Glycan profiles of exoglycosidase digests for swainsonine treated mAbs 122
Figure 4.6 Effect of glycoprotein processing inhibitors on the secondary structure of mAbs determined by CD 124
Figure 4.7 Monodispersed hydrodynamic radius (R_h) distribution by volume of variously glycosylated mAbs 126
Figure 4.8 Monodisperse sedimentation coefficient distribution in water of variously glycosylated mAbs 128
Figure 4.9 Effect of glycoprotein processing inhibitors on the binding of mAbs to FcγRI 131
Figure 4.10 Structural interpretation of N-glycosylated EG2-hFc using homology modeling .. 135

Contributions of Authors

The experimental work performed in Chapter 3 was under the supervision of Drs. T.R. Patel and J. Stetefeld. The author performed majority of the experiments with assistance of V. To and Dr. J. O'Neil for data collection and analysis of the nuclear magnetic resonance experiments and expertise of Drs. E.P. Booy and S.A. McKenna for the cell biology assays. Drs. K. McEleney and T.R. Patel assisted with the small-angle x-ray scattering data collection and analysis. The manuscript for Chapter 3 was prepared by the author with contributions and revisions from Drs. T.R. Patel, M. Meier, J. O'Neil, E.P. Booy, and J. Stetefeld.

The experimental work performed in Chapter 4 was under the supervision of Dr. J. Stetefeld in close collaboration with Dr. M. Butler. The mammalian cell culture experiments and N-glycan analysis were performed by Dr. M. Spearman and the author. All hydrodynamic and secondary structure characterization was performed by the author assisted by M. McDougall while the surface Plasmon resonance interaction studies were done by J. Dorion-Thibaudeau under the supervision of Drs. G. DeCrescenzo and Y. Durocher. The manuscript for Chapter 4 was prepared by the author and Dr. M. Spearman, with contributions and revisions from M. McDougall, Drs. T. R. Patel, M. Meier, G. DeCrescenzo, M. Butler, and J. Stetefeld.

List of Copyrighted Material for which Permission has been Obtained

Figure 1.1 – Reproduced with permission from Springer. License # 4093350053979

Iannello, A., and Ahmad, A. (2005) Role of antibody-dependent cell-mediated cytotoxicity in the efficacy of therapeutic anti-cancer monoclonal antibodies. *Cancer Metastasis Rev.* **24**, 487–499

Figure 1.2 – Reproduced with permission from Elsevier. License # 4093351507317

Reeves, R., and Adair, J. E. (2005) Role of high mobility group (HMG) chromatin proteins in DNA repair. *DNA Repair (Amst).* **4**, 926–938

Figure 2.1– Reproduced with permission from Elsevier. License # 4092790866525

Vander Heyden, Y., Popovici, S.T., Schoenmakers, P.J. (2002). Evaluation of size-exclusion chromatography and size-exclusion electrochromatography calibration curves. *J. Chromatogr. A.* **957** (2), 127-137.

Figure 2.2– Reproduced with permission from Elsevier. License # 4083400372579

Kelly, S.M., Jess, T.J., Price, N.P. (2005). How to study proteins by circular dichroism. *Biochim. Biophys. Acta*. **1751**, 119–139

Figure 2.4 – Reproduced with permission from Springer. License # 4093281163374

Brutscher, B., Felli, I. C., Gil-Caballero, S., Hošek, T., Kümmerle, R., Piai, A., Pierattelli, R., and Sólyom, Z. (2015) NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. in *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*, pp. 49–122, Springer

International Publishing Switzerland, Switzerland, Doi: 10.1007/978-3-319-20164-1_3

Chapter 4 – Reproduced with permission from Elsevier. License # 4083391095836

Krahn, N., Spearman, M., Meier, M., Dorion-Thibaudeau, J., McDougall, M., DeCrescenzo,G., Durocher, Y., Stetefeld, J., Butler, M. (2017). Inhibition of glycosylation on a camelid monoclonal antibody uniquely affects its FcγRI binding activity. *Eur. J. Pharm. Sci.* **96**,428-439

List of Abbreviations

°C	degrees Celsius
2D	two-dimensional
ΔG^0	standard free energy
ΔH^0	standard free enthalpy
ΔS^0	standard free entropy
η_0	viscosity
µM	micromolar
$\bar{\nu}$	partial specific volume
ρ_0	density
χ	Chi (goodness of fit)
A	adenine
AB	aminobenzamide
ADCC	antibody-dependent cellular cytotoxicity
Ala (A)	alanine
AP	apurinic/apyrimidinic
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
ATCC®	American type culture collection
AUC	analytical ultra-centrifugation
BAP	biotin acceptor peptide
BER	base excision repair
ccRCC	clear cell renal cell carcinoma
C	cytosine
Cas	castanospermine
CD	circular dichroism
CDC	complement-dependent cytotoxicity
CHO	Chinese hamster ovary
DLS	dynamic light scattering
D_{max}	maximal particle dimension
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
dsDNA	double-stranded deoxyribonucleic acid
dRP	5'-deoxyribosyl phosphate
E. coli	*Escherichia coli*
EBNA1	Epstein-Barr virus nuclear antigen 1
EGFR	epidermal growth factor receptor
ELISA	enzyme-linked immunosorbent assay
EMSA	electrophoretic mobility shift assay
ER	endoplasmic reticulum
ESC	embryonic stem cell
Fab	fragment, antigen-binding
Fc	fragment crystallisable
FcγR	fragment crystallisable gamma receptor

FDA	Food and Drug Administration
Fuc (F)	fucose
G	guanine
GFP	green fluorescent protein
GI	galactosylation index
Gly (G)	glycine
H	hydrogen
hcAb	heavy-chain antibody
HCl	hydrochloric acid
HEK293T	human embryonic kidney 293 expressing mutant version of SV40 large T antigen
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HILIC-HPLC	hydrophobic interaction liquid chromatography – high performance liquid chromatography
His (H)	histidine
HIV-1	human immunodeficiency virus -1
HMGA1	high mobility group AT-hook 1
HMGA2	high mobility group AT-hook 2
HSQC	heteronuclear single quantum coherence
IC	immune complex
IgG	immunoglobulin G
IMAC	immobilised metal ion affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
k_a	association kinetic constant
k_d	dissociation kinetic constant
K_D	dissociation constant of interaction
kDa	kilodalton
Kif	kifunensine
Lys (K)	lysine
M	molar
mAb	monoclonal antibody
Man (M)	mannose
mg	milligram
MHz	mega Hertz
mL	millilitre
mm	millimeter
mM	millimolar
mRNA	messenger ribonucleic acid
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
M_w	molecular weight
N	nitrogen
N_A	Avogadro's number
NaCl	sodium chloride
NaF	sodium fluoride
NaOH	sodium hydroxide
NH_4Cl	ammonium chloride

NK	natural killer
nm	nanometer
nM	nanomolar
NMR	nuclear magnetic resonance
NSD	normalized spatial discrepancy
OD_{600}	optical density at 600 nm
PAGE	polyacrylamide gel electrophoresis
PDB	protein data bank
Phe (F)	phenylalanine
ppm	parts per million
P(r)	pair-distance distribution function
Pro (P)	proline
ps	picosecond
PYLL	potential years of life left
R	gas constant
RF	replication fork
R_g	radius of gyration
R_h	hydrodynamic radius
RNA	ribonucleic acid
rpm	revolutions per minute
s	second
S	Svedberg
$s^0_{20,w}$	Svedberg constant at 20°C in water at infinite dilution
$s_{20,b}$	Svedberg constant at 20°C in buffer
$s_{20,w}$	Svedberg constant at 20°C in water
SAXS	small angle X-ray scattering
scFv	single-chain fragment variable
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
Ser (S)	serine
siRNA	small (short) interfering ribonucleic acid
SPR	Surface Plasmon Resonance
ssDNA	single-stranded deoxyribonucleic acid
SV	sedimentation velocity
Swa	swainsonine
T	thymine
TBE	Tris-borate ethylenediaminetetraacetic acid
TRF2	telomeric repeat-binding factor 2
Trp (W)	tryptophan
Tyr (Y)	tyrosine
UV	ultra-violet
Val (V)	valine
VCD	viable cell density
VEGF	vascular endothelial growth factor
W	watts

Chapter 1 : Introduction

1.1 Organization of the dissertation

This dissertation is organized as a sandwich thesis and is comprised of two first author papers published or accepted for publication in peer-reviewed scientific journals. An in depth discussion of the biochemistry of each protein is presented in the introduction of subsequent chapters to avoid duplication. Each manuscript contributes to the theme of this thesis which is the biochemical and biophysical analysis of proteins involved in the diagnosis, treatment, and chemoresistance of cancer.

Chapter 1 is an introduction to the theme of cancer in Canada and how the proteins HMGA2a and EG2-hFc fit into this theme. It delves into the details of how HMGA2 is a potential biomarker and has also been found to cause chemoresistance in chemotherapy treatment of cancer and the novel design of EG2-hFc as a new antibody-based therapeutic.

Chapter 2 takes an in-depth look into the biophysical and biochemical techniques used for the research in this thesis. It provides the theory behind the techniques as well as some background information into the experimental set up. The techniques are discussed in the order in which they appear in the thesis.

Chapter 3 is entitled *"Nano-scale assembly of high-mobility group AT-hook 2 protein with DNA replication fork"*. This is a manuscript accepted to the peer-reviewed journal *"Biophysical Journal"*. In this manuscript, an in-depth biophysical study was performed of HMGA2 unbound and in complex with a RF. The study concludes with a low-resolution model from solution scattering data which suggests the mechanism by which HMGA2 stabilizes RFs. It is this stabilization that aids in the increased cellular proliferation observed in embryonic stem

cells (ESCs) and cancer cells, and the DNA interaction which provides HMGA2 the capability to play a role in the BER mechanism that imparts chemoresistance on cancer cells.

Chapter 4 is entitled *"Inhibition of glycosylation on a camelid antibody uniquely affects its FcγRI binding activity"*. This manuscript is published in the peer-review journal *"European Journal of Pharmaceutical Sciences"*. It investigated the effect of N-glycosylation on the overall hydrodynamic properties of EG2-hFc in addition to its binding affinity to FcγRI when compared to a full-sized IgG1. These studies propose that EG2-hFc has unique binding characteristics which could be harnessed as a potential antibody-based therapeutic for cancer treatment.

Chapter 5 finally takes the information from both manuscripts and discusses how they relate to one another and the theme of the thesis. It proposes a framework of new techniques and experimentation which could be implemented to further the current research.

1.2 Cancer in Canada

Cancer is the number one cause of death in Canadians with one third of Canadians expected to develop cancer in their lifetime and one quarter expected to die from it. Additionally, cancer is the leading cause of premature mortality which is measured by potential years of life lost (PYLL). Between 2010 and 2012, the PYLL for all cancers combined outweighed all other leading causes of death by 1,500,000 (1). Furthermore, in 2008 the Public Health Agency of Canada estimated that cancer was the costliest illness in terms of lost productivity due to death than any other disease (2).

On a brighter note, since 1990 mortality rates have been decreasing which is in part due to the increasing knowledge we have acquired about the cause of cancer, its development, and treatment methods (1). Cancer research is ongoing but it remains a struggle to develop novel

therapeutics with clinical success. A study performed for the timeframe of 2006 – 2015 investigated the likelihood of drug approval by Food and Drug Administration (FDA) from Phase I trials for 15 categories of major diseases. Oncology was rated the lowest likelihood of approval with a mere 5.1% expected success (one third the success of 7 other categories of diseases). It was also at the bottom of the charts for the likelihood of approval from Phase II trials (8.1%) and Phase III trials (33.0%). Delving deeper into the Phase III success rates for oncology, drugs for chronic/small lymphocytic leukemia – non-Hodgkin's lymphoma had the highest success (close to 70%) while pancreatic cancer had the lowest success (~15%) (3). With that being said, continual efforts are being made to overcome this hurdle and develop successful therapeutics.

Tumours are considered a growth of abnormal cells and can be benign (non-cancerous) or malignant (cancerous). A malignant tumour is composed of cancerous cells which do not contain the cell adhesion molecules necessary to constrain the tumour to a specific location. This is how cancer spreads as the cells are unstable and can enter the bloodstream, split off and reside in various locations in your body. Cancer cells have been publicized as "good cells gone bad" which comes from the information that any normal body cell type can become cancerous. This occurs from the cell possessing multiple mutations in their cellular pathways or cellular division mechanisms which results in the cells not behaving as they should (4). Since multiple mutations must occur to transform a healthy cell into a cancerous one, cancer is found to affect those over the age of 50 four times more than between the ages of 30-49 (1). Individuals that are at a higher risk for certain cancers are those that have inherited mutations in tumor-supressor genes or germ-line mutations in one allele of a gene (adenomatous polyposis coli, *APC*, for example). This does not mean the individual has inherited cancer, but rather inherited mutations which with

additional mutations could lead to tumour formation (4). For simplicity, if three mutations result in tumour formation and you have already inherited one of those mutations, then statistically speaking the chances of you getting the other two mutations, with all things being equal, is a lot higher than someone who still needs to obtain all three.

1.3 Detecting cancer

As mentioned previously, cancer can easily spread throughout the body and therefore the greatest chance of survival is detecting it and treating it in its earliest stage. This is however not a trivial task, and not all cancers can be detected using the same methods. Breast, prostate, colon, lung, and cervical cancers are among the most common and therefore important for physicians to accurately detect in the early stages (5). Many of the tests performed to diagnose cancer such as blood work, biopsies, and pap tests are procedures which obtain samples for biomarker analysis. A biomarker is defined as a biological molecule found in blood, other body fluids or tissue which is a sign of a condition, disease, normal or abnormal process. These biomarkers are what differentiate a healthy person from one with the specific disease and include, but not limited to, proteins, nucleic acids, antibodies, and peptides. Additionally, they can be a collection of alterations found both in gene expression, and metabolomic or proteomic signatures (6). In the past few decades, numerous publications have investigated cancer biomarkers as a means for assessment, screening, diagnosis as well as monitoring the disease (7, 8). Just as there is a wide variety of biomarkers, their applicability can be highly specific or have broad specificity.

A potential broad range biomarker has been identified as the protein high mobility group AT-hook 2 (HMGA2). It is defined as a transcriptional activating factor (9), but further research

determined that abnormal expression of HMGA2 has been identified in both benign and malignant tumours. The degree of tumour malignancy and metasticity is directly correlated to the expression levels of HMGA2 (10, 11). Studies have investigated a broad range of cancers and have found HMGA2 to play a role in the diagnosis, prognosis, or clinical outcome of almost all of them.

A recent study of clear cell renal cell carcinoma (ccRCC) found that a positive correlation existed between the expression of HMGA2 and tumour size. Using univariate and multivariate model analysis, HMGA2 was found to be an independent prognostic factor that could lead to the overall survival in patients with ccRCC. Moreover, it has the potential to monitor the progress of treatments (12). A similar correlation was observed for colorectal cancer, in which they specifically looked at the expression levels of let-7a-5p (microRNA targeting HMGA2) in the tumour tissues. Implementing HMGA2 as a biomarker before and after colon surgery would provide suggestive information on the need for chemotherapy treatment (13). Separately, in testicular germ cell tumours HMGA2 was found to represent a novel marker in pathological subtyping (14). Information gathered from this will be beneficial for implementing the appropriate course of treatment.

Furthermore, in breast cancer, HMGA2 was found to have a strong positive correlation to specimens with advanced tumour grade and poor survival. This is in part a result from the role it plays in protecting cancer cells against agents which stimulate p53 (Ser15) phosphorylation (15). This control over maintaining cell growth causes tumours to grow, suggesting that it may also be a therapeutic target (16). Similar findings were observed in tongue (17), bladder (18), gastric (19), and prostate cancer (20), among others. This extensive breadth of cancers for which HMGA2 has been detected, suggests its role as a broad range biomarker.

1.4 Antibodies as therapeutics

Once cancer has been diagnosed in a patient, an appropriate course of treatment must be implemented. A recent strategy for treating tumours that has been established over the past 20 years is antibody-based therapy. It is based fundamentally on the knowledge that tumor cells express their own antigens. Antibody-based therapeutics are then designed to target these antigens and block their interaction (21). Although this may seem obvious, identifying an antigen which is suitable for antibody-based therapy can be the first key challenge to overcome. However, the process of designing an antibody with high antigen-binding affinity and high cytotoxicity is also not trivial.

The mechanism by which antibodies kill tumours can vary. They can be designed to act directly through blocking receptors, inducing apoptosis or delivering a cytotoxic agent; inducing immune-mediated cell killing mechanisms [complement-dependent cytotoxicity (CDC) or antibody-dependent cellular cytotoxicity (ADCC)]; or have a specific effect on the tumour vasculature or stroma. Most commonly, mAbs are designed to act as a scaffold bringing together a cell that will induce killing towards the tumour cell. The Fab region binds to the tumour antigen while the Fc region of the antibody binds to a cell-surface receptor to activate the CDC and ADCC activity (**Figure 1.1)**. As mentioned previously it is a challenge to choose the tumour antigen by which to target. Much research has been conducted to identify antigens which are now exploited in designing new antibody-based therapeutics. These include cell surface differentiation antigens, growth factors, and stromal or extracellular matrix antigens (22–25).

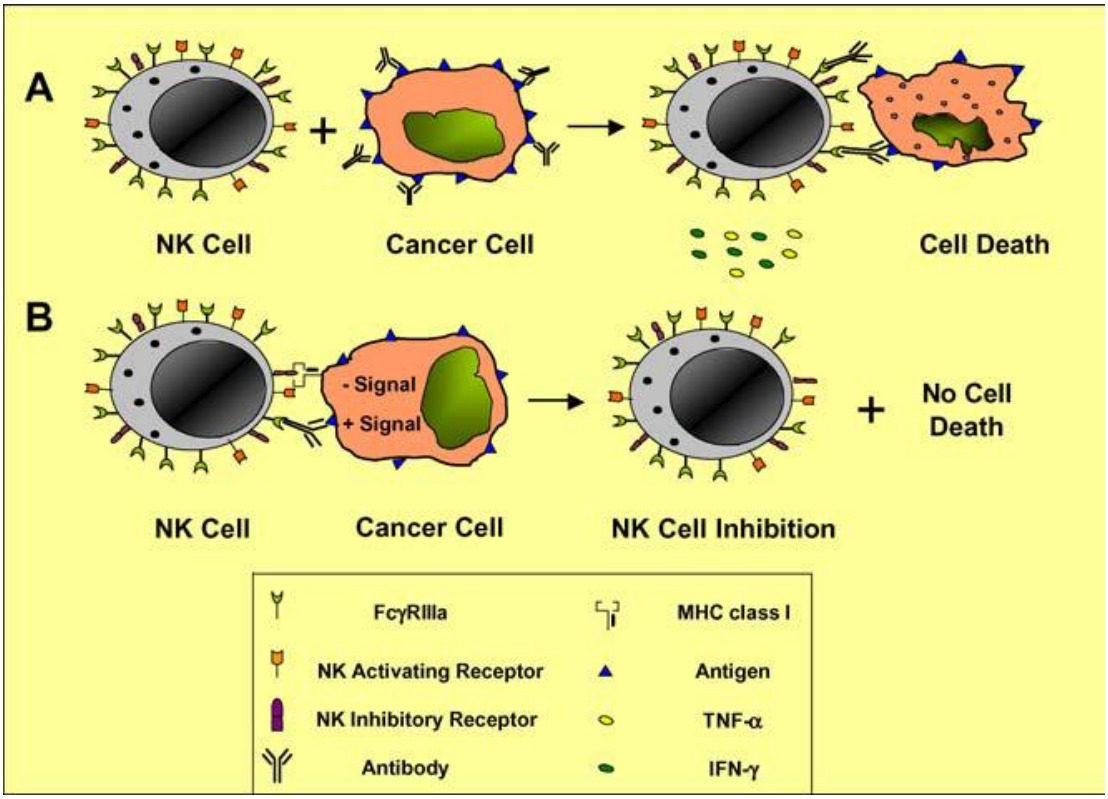

Figure 1.1 Natural killer cells mediate ADCC response in cancer cells

(A) The antigen-specific region of the antibody binds to the cancer cell while the Fc region binds to the Fc receptor (FcγRIIIa or CD16a) on NK cells. This cross-linking triggers NK cell cytotoxicity and the release of soluble mediators.
(B) Simultaneous co-ligation of an inhibitory receptor on the surface of the effector cell sends negative signals and inhibits NK cell activation and killing of cancer cells by ADCC (25).

Among the most successful antibody therapies in treating solid tumours are those that target the ErbB family (which includes epidermal growth factor receptor, EGFR) and vascular endothelial growth factor (VEGF). A recent study of colorectal cancer patients found that patients containing wild-type KRAS (a gene for the protein K-Ras involved in cellular regulation) had improved response, disease control and survival when treated with EGFR-specific antibodies (21). This is a prime example of how biomarkers can provide the information necessary to determine a successful course of treatment. Current FDA approved antibody-therapeutics for cancer are full IgG molecules (typically IgG1) which have been humanized,

chimerized, or conjugated (26). However, antibody engineering has advanced to bispecific antibodies, single-chain fragment variable (scFv) regions, Fabs, and heavy chain antibodies (hcAbs) as alternative therapeutic antibody molecules.

As such, a strategic network was initiated and subsequently funded through NSERC to focus on the development of a novel mAb for cancer therapy. EG2-hFc, a hcAb, is designed with a camelid scFv against EGFR fused to a humanized IgG1 Fc region for a molecular weight of ~80 kDa. The absence of the C_H1 domain and light chains creates a smaller antibody that has the advantage to penetrate tumours which would otherwise be inaccessible for current antibody therapeutics (~150 kDa) (27, 28). Moreover, with the presence of a human Fc region, Fc receptors will bind with similar affinity to current therapeutics to induce cytotoxicity through ADCC or CDC without posing immunogenic problems that would be associated if the mAb was instead designed with a camelid Fc region. As a result, it is suggested that EG2-hFc has the potential to have higher efficacy due to its enhanced tumour penetration and low immunogenic properties when compared to the current biopharmaceuticals on the market. However, in order for EG2-hFc to be a viable candidate for biotherapeutics, further research must investigate the upstream and downstream processing of the mAb, including glycoengineering, to produce the mAb in high quality and quantity.

1.5 Resistance to cancer treatment

Although immense efforts are implemented to further the treatment of cancer, it should not be overlooked that some patients can be resistant to treatment, specifically chemotherapy. As a result, research has also focused on how and why these patients develop chemoresistance. The

role of chemotherapy is to use various drugs to control cancer by preferentially targeting the rapidly growing cells. In a lot of situations, patients who initially respond well to chemotherapy, over time show a reduction in response resulting in tumor regrowth. This results from either the tumour cells having an inherent resistance due to a genetic characteristic they contain or they may acquire resistance after being exposed to the drug. Typically, 1 in $10^6 – 10^7$ cancer cells in a tumour are likely to have inherent resistance against a particular drug. Therefore in a clinically detectable tumour, which contains at least 10^9 cells, it would be expected to contain 10 – 1000 cells which contain some drug resistance. Moreover, these cells then have the capability to repopulate the tumour regardless of the destruction of their neighbouring sensitive cells (29).

A number of hypotheses exist as to how drug resistance occurs including altered membrane transport, genetic responses, alterations in target molecules, metabolic effects, growth factors, and enhanced DNA repair (29). In addition to the biomarker potential of HMGA2, it is proposed to impart resistance on chemotherapy treatment due to the base excision repair (BER) mechanism it is involved in. HMGA2 has been found to be efficient in cleaving apurinic/apyrimidinic (AP) sites in DNA through a process of β-elimination in addition to the related 5'-deoxyribosyl phosphate (dRP) lyase activity. Furthermore, upon exposure of cells expressing HMGA2 with hydroxyurea (a common chemotherapeutic agent for tumours) a 2-8 fold increase in cell survival was observed over the parental cells (expressing undetectable amounts of HMGA2) (9, 30). Therefore, the presence of HMGA2 in tumour cells not only enhances the rate of cellular division, but imparts an additional layer of protection from certain DNA-damaging chemotherapeutic agents. **Figure 1.2** suggests two different methods by which HMG proteins bind to induced DNA damage such as UV-radiation. HMG proteins either directly

bind to the lesions blocking repair, or they bind to the DNA itself, recruiting other repair proteins (31).

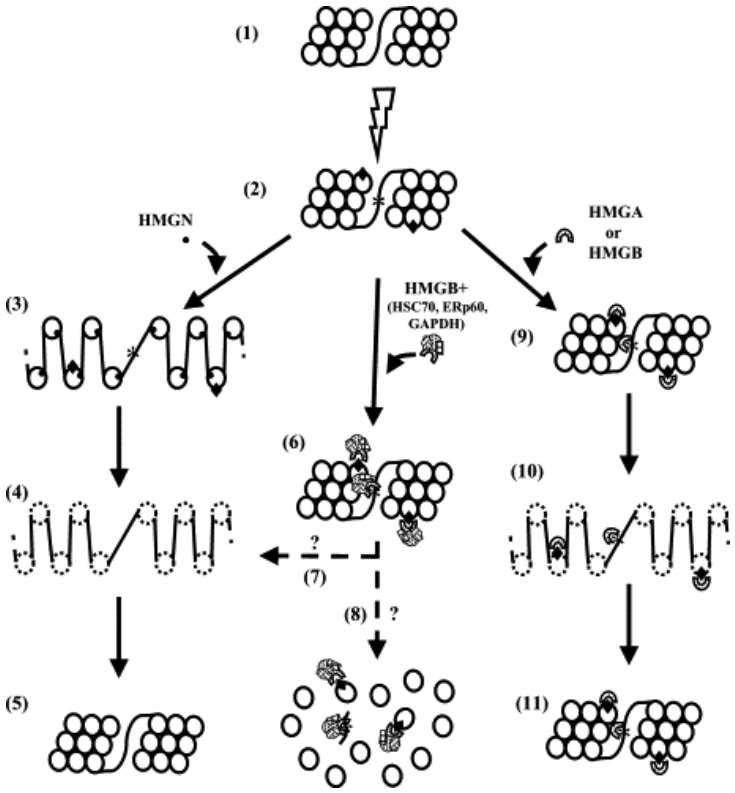

Figure 1.2 Hypothetical scheme for involvement of HMG proteins in DNA damage recognition and repair

(1) Nucleosome-free DNA or DNA wrapped around nucleosomes and folded into chromatin are susceptible to (2) damage induced by UV-radiation. HMG proteins recognize these lesions and can (9-11) directly bind to them inhibiting DNA repair, or (3-5) bind to the DNA recruiting other repair proteins to form a multiprotein complex. It is this complex that then is able to repair the DNA and (5) return it to its original state (31).

1.6 Research objectives and hypotheses

This research project focused on using a range of biophysical and biochemical techniques to investigate two unique proteins, HMGA2 and EG2-hFc, and their interactions for use as a biomarker, for cancer treatment or understanding the chemoresistance in cancer. In chapter 3, I

sought to elucidate the mechanism by which HMGA2 enhances cellular replication and protects cancer cells from DNA damage. Using recombinant HMGA2 and synthesized DNA I would characterize the biomolecules individually and in complex with one another. Secondary structure analysis using circular dichroism (CD) and nuclear magnetic resonance (NMR) will distinguish whether complex formation induces any secondary structural changes in the DNA RF and HMGA2, respectively. Furthermore, biophysical techniques such as dynamic light scattering (DLS) and sedimentation velocity (SV) will be implemented to understand how the hydrodynamic properties of the biomolecules change when they bind to one another. This will provide the information necessary to compare to the low resolution models generated from small angle X-ray scattering (SAXS) data. From this, we can begin to visualize the formed complex and distinguish the positioning of HMGA2 upon binding to the RF. This indepth analysis will determine that one HMGA2 binds in a side-by-side orientation to one RF molecule containing three AT-hooks. Through this interaction, we can begin to hypothesize the mechanism by which HMGA2 is able to stabilize stalled RFs furthering the research into HMGA2 as a potential biomarker or as a target for cancer therapy.

Chapter 4 follows that by delving into the glycoengineering of a novel mAb, EG2-hFc, to enhance its cytotoxicity. I decided to take a unique approach to this biotechnology related question by implementing the same biophysical and biochemical techniques which were previously used to analyze HMGA2 and the RF. Through altering the N-glycosylation pattern using glycoprotein processing inhibitors in the growth media, I wanted to observe any hydrodynamic or secondary structural changes of EG2-hFc. From this, I would investigate how these glycan changes affected mAb binding to its Fc receptor (FcγRI) and try to correlate the results. However, in order to get a full understanding of the N-glycan effect on EG2-hFc, all

analyses would be repeated with a full-sized IgG1 mAb (DP-12). From these results we would be able to determine that the EG2-hFc mAb design imparts unique properties onto the antibody which makes it suitable for antibody-based therapy in cancer. Although EG2-hFc has a lower affinity to FcγRI than a full-sized IgG1, the ability of EG2-hFc to bind to its receptor when aglycosylated makes it unique and worth further investigation for therapeutic use.

With cancer as the leading cause of death in Canada, this thesis aims to provide research which will get Canada as a whole closer to understanding how to reduce the number of cancer related deaths. Gaining knowledge about the solution behaviour of complexes and how sugars affect the structure-function relationship in a novel mAb-based therapy is a stepping stone in the right direction.

References

1. *Canadian Cancer Statistics 2016* (2016) Toronto, ON

2. Economic Burden of Illness in Canada *Public Heal. Agency Canada 2014*. [online] http://www.phac-aspc.gc.ca/ebic-femc/index-eng.php (Accessed March 22, 2017)

3. Thomas, D. W., Burns, J., Audette, J., Carroll, A., Dow-Hygelund, C., and Hay, M. (2016) *Clinical Development Success Rates 2006-2015*

4. Lodish, H., Berk, A., and Zipursky, S. L. (2000) Tumor Cells and the Onset of Cancer. in *Mol. Cell. Biol.*, 4th Ed., W.H. Freeman, New York

5. Zeichner, S. B. (2016) Detecting cancer: Pearls for the primary care physician. *Cleve. Clin. J. Med.* **83**, 515–523

6. Henry, N. L., and Hayes, D. F. (2012) Cancer biomarkers. *Mol. Oncol.* **6**, 140–146

7. Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M. R., Hayes, D. F., and Bast Jr, R. C. (2007) American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J. Clin. Oncol.* **25**, 5287–5312

8. Moslehi, R., Chu, W., Karlan, B., Fishman, D., Risch, H., Fields, A., Smotkin, D., Ben-David, Y., Rosenblatt, J., Russo, D., Schwartz, P., Tung, N., Warner, E., Rosen, B., Friedman, J., Brunet, J.-S., and Narod, S. A. (2000) BRCA1 and BRCA2 Mutation Analysis of 208 Ashkenazi Jewish Women with Ovarian Cancer. *Am. J. Hum. Genet.* **66**, 1259–1272

9. Summer, H., Li, O., Bao, Q., Zhan, L., Peter, S., Sathiyanathan, P., Henderson, D., Klonisch, T., Goodman, S. D., and Dröge, P. (2009) HMGA2 exhibits dRP/AP site cleavage activity and protects cancer cells from DNA-damage-induced cytotoxicity during chemotherapy. *Nucleic Acids Res.* **37**, 4371–84

10. Cui, T., and Leng, F. (2007) Specific recognition of AT-rich DNA sequences by the mammalian high mobility group protein AT-hook 2: A SELEX study. *Biochemistry*. **46**, 13059–13066

11. Galande, S. (2002) Chromatin (dis)organization and cancer: BUR-binding proteins as biomarkers for cancer. *Curr. Cancer Drug Targets*. **2**, 157–90

12. Na, N., Si, T., Huang, Z., Miao, B., Hong, L., Li, H., Qiu, J., and Qui, J. (2016) High expression of HMGA2 predicts poor survival in patients with clear cell renal cell carcinoma. *Onco. Targets Ther.* **9**, 7199–7205

13. Liu, T.-P., Huang, C.-C., Yeh, K.-T., Ke, T.-W., Wei, P.-L., Yang, J.-R., and Cheng, Y.-W. (2016) Down-regulation of let-7a-5p predicts lymph node metastasis and prognosis in colorectal cancer: Implications for chemotherapy. *Surg. Oncol.* **25**, 429–434

14. Kloth, L., Gottlieb, A., Helmke, B., Wosniok, W., Löning, T., Burchardt, K., Belge, G.,

Günther, K., and Bullerdiek, J. (2015) HMGA2 expression distinguishes between different types of postpubertal testicular germ cell tumour. *J. Pathol. Clin. Res.* **1**, 239–51

15. Wu, J., Zhang, S., Shan, J., Hu, Z., Liu, X., Chen, L., Ren, X., Yao, L., Sheng, H., Li, L., Ann, D., Wang, J., and Wang, X. (2016) Elevated HMGA2 expression is associated with cancer aggressiveness and predicts poor outcome in breast cancer. *Cancer Lett.* **376**, 284–92

16. Loughery, J., Cox, M., Smith, L. M., and Meek, D. W. (2014) Critical role for p53-serine 15 phosphorylation in stimulating transactivation at p53-responsive promoters. *Nucleic Acids Res.* **42**, 7666–7680

17. Zhao, X.-P., Zhang, H., Jiao, J.-Y., Tang, D.-X., Wu, Y., and Pan, C.-B. (2016) Overexpression of HMGA2 promotes tongue cancer metastasis through EMT pathway. *J. Transl. Med.* **14**, 26

18. Shi, Z., Li, X., Wu, D., Tang, R., Chen, R., Xue, S., and Sun, X. (2016) Silencing of HMGA2 suppresses cellular proliferation, migration, invasion, and epithelial–mesenchymal transition in bladder cancer. *Tumor Biol.* **37**, 7515–7523

19. Jun, K.-H., Jung, J.-H., Choi, H.-J., Shin, E.-Y., and Chin, H.-M. (2015) HMGA1/HMGA2 protein expression and prognostic implications in gastric cancer. *Int. J. Surg.* **24**, 39–44

20. Cai, J., Shen, G., Liu, S., and Meng, Q. (2016) Downregulation of HMGA2 inhibits cellular proliferation and invasion, improves cellular apoptosis in prostate cancer. *Tumor Biol.* **37**, 699–707

21. Scott, A. M., Wolchok, J. D., and Old, L. J. (2012) Antibody therapy of cancer. *Nat. Rev. Cancer*. **12**, 278–87

22. Van Cutsem, E., Köhne, C.-H., Hitre, E., Zaluski, J., Chang Chien, C.-R., Makhson, A., D'Haens, G., Pintér, T., Lim, R., Bodoky, G., Roh, J. K., Folprecht, G., Ruff, P., Stroh, C., Tejpar, S., Schlichting, M., Nippgen, J., and Rougier, P. (2009) Cetuximab and Chemotherapy as Initial Treatment for Metastatic Colorectal Cancer. *N. Engl. J. Med.* **360**, 1408–1417

23. Repp, R., Kellner, C., Muskulus, A., Staudinger, M., Mohseni Nodehi, S., Glorius, P., Akramiene, D., Dechant, M., Fey, G. H., van Berkel, P. H. C., van de Winkel, J. G. J., Parren, P. W. H. I., Valerius, T., Gramatzki, M., and Peipp, M. (2011) Combined Fc-protein- and Fc-glyco-engineering of scFv-Fc fusion proteins synergistically enhances CD16a binding but does not further enhance NK-cell mediated ADCC. *J. Immunol. Methods*. **373**, 67–78

24. Weiner, L. M., Surana, R., and Wang, S. (2010) Antibodies and cancer therapy: versatile platforms for cancer immunotherapy. *Nat. Rev. Immunol.* **10**, 317–327

25. Iannello, A., and Ahmad, A. (2005) Role of antibody-dependent cell-mediated cytotoxicity in the efficacy of therapeutic anti-cancer monoclonal antibodies. *Cancer*

Metastasis Rev. **24**, 487–499

26. FDA Approved Antibody-based Therapeutics (2013) *Immunol. Link.* [online] http://www.immunologylink.com/FDA-APP-Abs.html (Accessed March 30, 2017)

27. Butler, M. (2009) Strategies for the production of single-type glycoform monoclonal antibodies

28. Zhang, J., Liu, X., Bell, A., To, R., Baral, T. N., Azizi, A., Li, J., Cass, B., and Durocher, Y. (2009) Transient expression and purification of chimeric heavy chain antibodies. *Protein Expr. Purif.* **65**, 77–82

29. Luqmani, Y. A. (2005) Mechanisms of drug resistance in cancer chemotherapy. *Med. Princ. Pract.* **14**, 35–48

30. Sakano, K., Oikawa, S., Hasegawa, K., and Kawanishi, S. (2001) Hydroxyurea induces site-specific DNA damage via formation of hydrogen peroxide and nitric oxide. *Jpn. J. Cancer Res.* **92**, 1166–1174

31. Reeves, R., and Adair, J. E. (2005) Role of high mobility group (HMG) chromatin proteins in DNA repair. *DNA Repair (Amst).* **4**, 926–938

Chapter 2 : Theory behind the techniques

2.1 Size exclusion chromatography

SEC is the standard technique for the separation and quantification of dimers, trimers, multimers, and oligomers based off of their R_h. This is accomplished through the diffusion of biomolecules in an aqueous mobile phase along a stationary phase of spherical porous particles. These particles have a controlled pore size which allow for separation of a defined molecular size range. Alteration of the pore size will result in a change in the separation capacity and, as such, various SEC columns have been manufactured to suit different molecular sizes. The stationary phase in these columns is typically stable with the use of a wide range of pH, chemicals, and denaturants; providing flexibility in the composition of buffer (mobile phase) (1). In a chromatographic process the free energy can be described using Gibbs free-energy:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = RT \ln k \qquad (2.1)$$

where ΔG^0, ΔH^0, and ΔS^0 are the standard free energy, enthalpy, and entropy, respectively; R is the gas constant, T is the absolute temperature and k is the partition coefficient (2). SEC is typically an entropic process provided there is no protein adsorption to the column (ie. $\Delta H = 0$), and therefore the Gibbs free-energy equation reduces to:

$$\mathbf{ln}(\boldsymbol{K_D}) = -\frac{\boldsymbol{\Delta S^0}}{\boldsymbol{R}} \qquad (2.2)$$

where K_D is the thermodynamic retention factor, with the implication that temperature should have no effect on the retention of the protein (2).

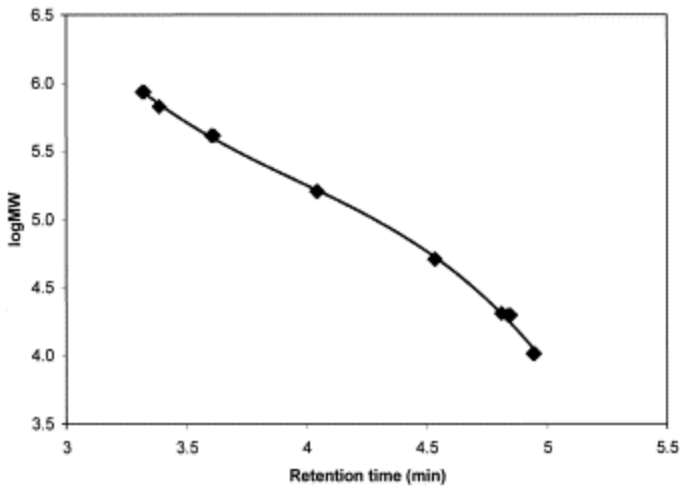

Figure 2.1 Size-exclusion chromatography calibration curve

An example of a size-based calibration curve for SEC. Log(M_W) is plotted as a function of the experimental retention time. A third order model was fit through the calibration points (3).

In SEC, a size-based calibration curve (**Figure 2.1**) can be generated from a set of proteins or polymers with known M_w to estimate the M_w of an unknown. The largest proteins elute first from the column, avoiding all the pores, while the remaining proteins elute in order of decreasing size. Plotting log M_w vs. the retention volume of the known proteins produces a third order polynomial with a linear region. It is this linear portion that describes the highest resolution and M_w accuracy of the select column (2) and can be defined by this relationship:

$$\log \boldsymbol{M_w = mK_D + b} \qquad (2.3)$$

where *m* and *b* are the slope and intercept of the line, respectively. To achieve higher resolution in a column, the size distribution of the pores has to be narrow such that the slope of this line is less steep.

The SEC calibration curve is most accurate for non-adsorbing globular proteins. For rod-like proteins or flexible chains, the Stokes radii do not correlate with the M_w and therefore cannot be accurately estimated from the calibration curve. Additionally, if the protein becomes adsorbed into the stationary phase, then the retention volume is altered and would not correlate with the

appropriate protein size (4). Adsorption can be reduced by limiting the non-specific electrostatic interactions through increasing the ionic strength (≥100 mM NaCl) (5) or using additives such as arginine (6).

Typically, SEC columns are composed of 3 – 20 µm particles with different pore sizes. They are typically 30 cm in length with an internal diameter of 4.6 – 8 mm to provide an analysis time between 15 and 50 mins (7). SEC columns are routinely packed with either silica (with or without modification) or cross-linked polymers which possess hydrophobic, hydrophilic or ionic characteristics (7). The SEC performed in my thesis implemented Superdex columns (GE Healthcare) which are packed with a spherical composite of cross-linked agarose and dextran (13 µm particle size). The Superdex 75 10/300 GL is most useful for separation of smaller proteins (3,000 – 70,000 Da) and therefore was used for purification of HMGA2 while the Superdex 200 10/300 GL has a larger range (10,000 – 600,000 Da) and was used when purifying the complexes.

2.2 Dynamic light scattering

DLS, also referred to as photon correlation spectroscopy or quasi-elastic light scattering, is a technique that uses a monochromatic light source to detect scattered light from macromolecules in solution. The Brownian motion of these macromolecules due to bombardment with solvent molecules causes fluctuations in the intensity of scattered light and results in obtaining the diffusion coefficient (D_τ) and furthermore the hydrodynamic radii (R_h) and polydispersity of the particles (8). D_τ is equal to the mean square displacement of the particles divided by two-fold time and is inversely proportional to the particle size. Particle size

can be determined by monitoring the Brownian motion over time since larger particles will move slower (maintain the same position for a longer period of time) and small particles will move faster (constantly changing position) (9). This motion is affected by factors such as temperature and solvent viscosity, and since solvent viscosity depends on temperature, measurements must be performed in a temperature controlled environment (10, 11).

The intensity of scattered light fluctuates in a time-dependent manner due to the continual movement of particles. The resulting constructive or destructive interference is collected by the detector and fitted with an autocorrelation function. The decay of this function can directly be related to the motion of the particles, their D_τ, and furthermore their dimensions (8, 9). The Zetasizer Nano S® machine (4 mW He-Ne laser with $\lambda_0 = 633$ nm) from Malvern Instruments (used in the described experiments) has a backscatter detection system at 173° which has the added advantage of allowing measurement of the D_τ for highly concentrated molecules since the setup avoids the multiple scattering phenomenon (scattering of a photon by more than one particle). With that being said, standard practice should measure the D_τ at multiple concentrations and extrapolate to infinite dilution (D^0_τ). The D_τ is an important parameter which can be related to the R_h using the following equation:

$$\boldsymbol{D_\tau = \frac{k_B T}{6\pi\eta R_h}} \qquad (2.4)$$

Where k_B is the Boltzmann coefficient (1.380×10^{-23} kg·m^2·s^{-2}·K^{-1}), T is the absolute temperature, and η is the solvent viscosity (12).

Additionally, the translational frictional coefficient, f, can be calculated from the D_τ:

$$\boldsymbol{f = \frac{RT}{N_A D_\tau}} \qquad (2.5)$$

Where R is the gas constant (8.314×10^{-7} erg/mol·K), T is the absolute temperature and N_A is Avogadro's number (6.022137×10^{23} mol). The f can be further used to determine the frictional

ratio (f/f_0, where $f_0 = 6\pi\eta R_0$ and R_0 is the radius of a sphere). The frictional ratio relates the R_h to the radius of a sphere with the volume of an unsolvated molecule, and therefore can provide information on the solution conformation of the macromolecules ($f/f_0 = 1$ for a compact sphere) (13).

The Zetasizer Nano S® machine is equipped with software that analyzes the scattering data to evaluate the size and homogeneity of macromolecules. The first information gained is an intensity-weighted distribution which results in obtaining the *z*-average R_h. This parameter requires the input of the refractive index of the macromolecules and solvent viscosity to be calculated accurately. Furthermore, the *z*-average R_h should only be compared if determined from a monodisperse distribution in a suitable buffer (8). The intensity distribution for a solution containing N_a and N_b molecules with size *a* and *b*, respectively, is described below.

$$\%\boldsymbol{I_a} = \frac{\boldsymbol{a^6 N_a \cdot 100}}{\boldsymbol{N_a a^6 + N_b b^6}} \tag{2.6}$$

Where $\%I_a$ represents the intensity-weighted distribution for particles with size *a* (14). This distribution can be converted to a volume-weighted distribution by using the Rayleigh approximation that the mass of a spherical molecule is proportional to its $(\text{size})^3$. Therefore, if the density of a system is uniform, then the mass can be approximated to the size of the molecule.

$$\%\boldsymbol{V_a} = \frac{\boldsymbol{a^3 N_a \cdot 100}}{\boldsymbol{N_a a^3 + N_b b^3}} \tag{2.7}$$

Where $\%V_a$ is the volume-weighted distribution for molecules with size *a* for a solution containing N_a and N_b molecules of size *a* and *b*, respectively (14). Based off of equation 2.7 it explains why the R_h obtained from a volume-weighted distribution is smaller than that for an intensity-weighted distribution. The volume-weighted distribution also requires knowledge of the viscosity and refractive index but in addition assumes that all the molecules are homogeneous

and spherical in shape. This distribution is found to be more practical because it presents a distribution of sizes based on the volume/mass of the molecules.

Additionally, the intensity distribution can be converted into a number-weighted distribution which represents the number of molecules of a given size. For a system containing two molecules the expression for the number-weighted distribution for molecules of size *a* is:

$$\%\boldsymbol{N_a} = \frac{\boldsymbol{N_a \cdot 100}}{\boldsymbol{N_a + N_b}} \quad (2.8)$$

Where N_a and N_b are the number of particles of size *a* and *b*, respectively (14).

By analyzing each weighted distribution, one can have a more complete view on the sample behaviour. For example, in an intensity-weighted distribution, a single large particle will scatter light more intensely than multiple small particles. However, in a volume-weighted distribution, the signal will be largely composed of multiple small particles rather than the single large one.

2.3 Analytical ultracentrifugation

AUC is an invaluable technique for characterizing the solution behaviour of macromolecules and can be used as a quantitative analysis of macromolecular interactions. Contrary to other methods, this technique is non-destructive and allows analysis of macromolecules in their native buffer. AUC works on the principle that mass will redistribute in a gravitational field until the gravitational potential energy exactly matches the chemical potential energy at each radial position. There are two main methodologies for which the AUC instrument is employed; sedimentation velocity and sedimentation equilibrium experiments (15).

2.3.1 Sedimentation velocity

Sedimentation velocity (SV) - AUC is a fundamental hydrodynamic method that allows the determination of purity or homogeneity of a macromolecular species with resolving power of 0.1 – 1 % of contaminants or aggregates (16). Through the use of centrifugation, SV experiments monitor the rate at which the boundaries of molecules move as mass redistributes from the applied gravitational field. This can be understood when one looks at the forces that are acting on a molecule in an experiment. The force on a particle due to a gravitational field is:

$$M_p\omega^2 r \tag{2.9}$$

where M_p is the mass of the particle, ω is the rotor speed in radians per second ($\omega=2\pi\cdot rpm/60$), and r is the distance from the center of the rotor. The solvent that is displaced then exerts a counterforce as the particle sediments, resulting in a net force of:

$$(M_p - M_s)\omega^2 r \tag{2.10}$$

And since:

$$M_s = M_p\bar{\nu}\rho \tag{2.11}$$

where $\bar{\nu}$ is the partial specific volume of the particle in cm^3/g and ρ is the solvent density in g/cm^3, then the effective (or buoyant) mass of the particle is:

$$M_b = M_p(1 - \bar{\nu}\rho) \tag{2.12}$$

Finally, there is a frictional force that develops from the particle moving through the solvent that is given by:

$$F_f = fv \tag{2.13}$$

where f is the frictional coefficient and v is the velocity of the particle through the solvent. Combining these forces together, we can define the sedimentation coefficient:

$$s \equiv \frac{v}{\omega^2 r} = \frac{M_b}{f} \tag{2.14}$$

Since only electrically neutral particles will sediment, proteins and nucleic acids should be in a buffer containing near physiological salt concentrations (> 100 mM). This provides sufficient counter ions in the immediate surroundings of the macromolecule that the sedimentation coefficient is relatively insensitive to the salt concentration. However, at low salt concentrations (< 10 mM) a greater region of solution is required to produce a neutral particle capable of sedimenting. In such a situation the apparent radius of the protein or nucleic acid will increase resulting in an larger sedimentation coefficient (17). The SV experiments described previously were performed in buffer conditions with a salt concentration ≥ 154 mM to avoid this ambiguity with the sedimentation coefficient.

2.3.2 Sedimentation equilibrium

Sedimentation equilibrium (SE) experiments determine the concentration distribution after equilibrium has been reached. This is accomplished by using a small centrifugal force such that the flux from the sedimentation is exactly balanced by the flux due to diffusion. SE experiments provide a very accurate way to determine M_w and consequently the oligomeric states for a biomolecule at low concentrations. For an ideal single non-interacting species, the equilibrium radial concentration, $c(r)$, is given by:

$$\boldsymbol{c(r) = c_0 \exp\left[\frac{M_b \omega^2}{RT}\left(\frac{r^2 - {r_0}^2}{2}\right)\right]} \qquad (2.15)$$

Where c_0 is the concentration at an arbitrary reference distance r_0. Deviations can occur for interacting species or multiple non-interacting species. SE can also be implemented at high concentrations to investigate non-ideal sedimentation which is typically applied to weakly attractive or repulsive interactions. This caliber of experiment requires additional fractionation and quantification after the samples have been centrifuged (18).

2.3.3 Absorbance optics

Absorbance optics is the most frequently used detector for AUC experiments (19) as it is the easiest to use, working as a double-beam spectrophotometer from a wavelength range of 190 to 800 nm. As a result, where the Beer-Lambert law holds, the absorbance is proportional to the concentration of the sample, c ($A=\varepsilon cl$, where A is the absorbance, ε is the molar extinction coefficient, and l is the sample pathlength) with a precision of ± 0.10 OD. The noise associated with absorbance measurements is primarily stochastic which appears as high-frequency "fuzz" around the signal. Little systematic noise is present which is either radially-independent or time-independent (15).

When planning an absorbance experiment, one should prepare samples with an absorbance reading between 0.2 and 1.0 OD. It is possible to scan different samples at different wavelengths, however the XLA detector is relatively imprecise (± 3 nm) when switching between wavelengths. Therefore in such experiments, it is suggested that the absorbance region being detected is flat, such as in peaks and valleys and not quickly changing over the wavelength. If this is not possible, it is suggested to maintain the same wavelength throughout the experiment for accurate readings. A potential drawback to using absorbance optics is the length necessary to complete a scan. Of all the three optical systems, absorbance optics takes the longest (15). In SE experiments this is not a problem, however for SV experiments, this limits the amount of data that can be collected over the duration of the experiment.

2.3.4 Interference optics

The Rayleigh interference optical system produces a signal with equally-spaced horizontal fringes whose vertical displacement (ΔY) is directly proportional to the optical path difference between the light beams passing through the sample and reference solutions. Any refractive index difference (n) will contribute to the path length in the following expression:

$$\Delta Y = \frac{\Delta n l}{\lambda} \tag{2.16}$$

Where l is the optical path length and λ is the wavelength of the light source (20). For proteins, the refractive index difference is proportional to the refractive index increment

$$\Delta n = c\left(\frac{dn}{dc}\right) \tag{2.17}$$

Where *dn/dc* has an average value of 0.186 mL/g (ranging from 0.173 mL/g to 0.215 mL/g) (21), and for the XLA λ = ~670 nm and l = 1.2 cm (19). Therefore, any material that has a difference in refractive index from the reference solution can be detected. This can be advantageous and problematic when it comes to detergents and lipid micelles, etc. While the molecular weight and refractive index can be characterized using interference optics, samples containing detergents should be analyzed using absorbance optics (22).

Contrary to the absorbance optics, interference experiments have very little stochastic noise. However, any pathlength difference between the sample and reference solutions, including optical imperfections such as dust, oil, scratches, etc. are visible in the signal which results in significant time-independent systematic noise. The images obtained from an interference experiment must be converted to fringe displacement measurements using a Fourier analysis. This involves determining the fractional fringe displacement (23) for which the first radial position is arbitrarily assigned a zero fringe displacement. Since these fringe displacements

cannot be traced through the menisci, they contain a radially-independent systematic noise and must be removed prior to data analysis (24).

Moreover, due to the nature of the interference optics, sapphire windows must be used with careful alignment and focussing. Any variation in the window or centerpiece flatness (>0.1λ) will cause a vertical shift in the data (20). Quartz windows should be avoided with interference experiments since at high speeds the interference fringe pattern begins to blur.

2.3.5 Fluorescence optics

The most recent addition to the XLI is fluorescence optics which are not as well characterized as the absorbance or interference optics. In the AU-FDS system, a 488 nm laser is used such that sufficient radial resolution can be achieved (~20 – 50 μm). As a result, molecules must be extrinsically labelled with labels including but not limited to fluorescein, BODIPY, NBD, green fluorescent protein (GFP), and their many derivatives (15). After being excited with the laser, emitted light passes through a pair of long (> 505 nm) dichroic filters which capture the maximum amount of light to provide good sensitivity. As a result, it is not possible to select a label based off of its emission characteristics and therefore the fluorescent optical system is not suited for multiple labels.

Fluorescence optics have a combination of the high-frequency stochastic noise found in the absorbance detector and the low-frequency systematic noise observed in interference optics. The stochastic noise tends to be about 1% of the signal intensity and uses a similar detector to that of absorbance optics. The systematic noise tends to arise from two sources; fluorescent material stuck to the window and background fluorescence. Fluorescent material can stick to the window when it is being handled, particularly in the areas where there was once an air-liquid

boundary. The severity is largely based on the nature of the sample being used; some proteins have little sticking while others leave an uneven coating over the window. The background fluorescence is mainly a result of the cell components and tends to be lower in magnitude and more uniform than the former. This allows for background noise to be removed in the data analysis.

Provided that the fluorescence signal is directly proportional to the concentration, one can calculate the sedimentation coefficient, diffusion coefficient, and molecular weight without having to do any conversions. However, if one wants concentration-dependent information, fluorescence detection does pose some problems and so will not be discussed further.

2.3.6 Data analysis

There are various methods for analyzing SV-AUC data which include many alternative approaches and software packages. For simplicity, I will discuss the c(s) method which was used in all the SV-AUC data analysis presented.

SEDFIT and SEDPHAT implements the c(s) method to obtain a sedimentation coefficient distribution function from a direct fit to the data (25). The program initially creates a grid of sedimentation coefficients which cover the range of interest. A scaling relationship is created between *s* and *D* by assuming a constant shape and an equal frictional ratio for all species. The program then simulates the sedimentation boundaries for each point by using a numerical solution of the Lamm equation.

$$\frac{\partial c}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[rD\frac{\partial c}{\partial r} - s\omega^2 r^2 c\right] \quad (2.18)$$

Where c is the concentration, s is the sedimentation coefficient, D is the diffusion coefficient, ω is the rotor speed, and r is the distance from the center of rotation (26). Following that, data are

fit to a sum of the Lamm equations using a least-squares fitting procedure to define the concentration of each species in the grid. At this time, the systematic noise of the baseline and vertical displacements are treated as additional linear fitting parameters and therefore removed. Finally, the program regularizes the result to provide a smooth distribution rather than the "spiky" initial result (15).

The c(s) method has excellent resolution and sensitivity, no limit on the number of scans that can be analyzed, and quite useful in characterizing homogeneity and quantifying impurities and aggregates in samples. However, it does require that all association reactions are at equilibrium before sedimentation begins. This method assumes a non-interacting mixture or one that associates/dissociates rapidly such that the distributions appear similar to those expected for limiting models. If the c(s) distribution is a good fit to the data, thermodynamic parameters can be extracted by integrating the distribution to correlate the dependence of weight average sedimentation coefficients on loading concentration (25).

2.4 Circular dichroism

2.4.1 Principle

The technique of circular dichroism (CD) is based on the principle that plane polarised light can be split up into two that are circularly polarised (CP) and of equal magnitude; one rotating counter-clockwise (left) and the other rotating clockwise (right). The phenomenon of CD results from the differential absorption of LCPL and RCPL. The recombination of light that is equally absorbed by LCPL and RCPL generates polarised radiation. However, if each component is absorbed to different extents, then the resulting radiation will be elliptically

polarised. To obtain elliptical polarisation (ie. a CD signal), the sample must be optically active as a result of intrinsic chirality, covalently linked to a chiral center, or placed in an asymmetric environment (27). The actual CD instruments measure the difference in absorbance of CPL ($\Delta A = A_L - A_R$), however generally report the detector result in terms of ellipticity (θ) in degrees which is related by:

$$\boldsymbol{\theta = 32.98\,\Delta A} \tag{2.19}$$

To generate a CD spectrum that can be further analyzed, the ellipticity (or ΔA) is recorded as a function of wavelength.

As with other techniques, there are various methods to observe the CD effect including (a) direct subtraction, (b) ellipsometric, and (c) modulation. The technique of direct subtraction involves measuring the radiation of each component separately and then subtracting them from one another. Using the ellipsometric method, it is the transmitted ellipticity that is directly measured. However, the most common method of modulation involves continually switching the incident radiation between L and R components with the resulting radiation detected by a photomultiplier (28). Spectropolarimeters that implement this method have a modulator subjected to an alternating electric field (typically 50kHz) which splits plane polarised light into its L and R components. The modulator is a piezoelectric quartz crystal with a thin plate of isotropic material tightly coupled to the crystal. As the electric field alternates, it induces structural changes in the quartz crystal resulting in the transmission of CPL at the field extremes (29).

2.4.2 CD of proteins

CD spectroscopy is a non-destructive low resolution technique in which structural features of proteins can be described. It is a much less demanding technique both in terms of

sample amount and time spent than high-resolution techniques such as NMR and XRD. Good quality data can be achieved with ≤ 0.1 mg of protein (far UV) or 1 mg (near UV and visible) in as little as 30 minutes. Additionally, CD is versatile such that proteins can be analysed in a wide range of experimental conditions and is often used as a method to establish stable buffer conditions for NMR studies. It should be noted, however, that proteins have a low CD signal (10 mdeg) and therefore experimental conditions should be chosen accordingly that the protein signal is not masked (27).

The CD technique is advantageous when studying proteins because there are spectral regions which can be easily assigned to a distinct structural feature of a molecule. Main chromophores of interest include the peptide bond (absorption below 240 nm), aromatic amino side chains (absorption in the range 260 – 320 nm), and disulphide bonds (weak broad absorption bands centered at 260 nm). Additionally, other non-protein cofactors can give rise to a CD signal such as pyridoxal-5'-phosphate (~330 nm), flavins (300 – 500 nm depending on oxidation state), haem groups (strongly at 410 nm with other bands in the range of 350 to 650 nm), and chlorophyll molecules (visible to near IR regions) (30). Moreover, when a number of these chromophores are in the same vicinity, they can give rise to a single absorbing unit resulting in characteristic features of the CD spectra.

When studying proteins, CD can be used to generate information on the secondary structural composition, tertiary structure fingerprint, and integrity of cofactor binding sites. In the far UV region, the secondary structure elements of proteins have characteristic CD spectra (**Figure 2.2**). As a result, a number of algorithms exist to elucidate the proportion of each secondary structure element present in a given sample. These algorithms include SELCON (self-consistent) (31), VARSLC (variable selection) (32), CDSSTR (33), K2d (34), and CONTIN

(35). DICHROWEB is an online server which was developed at Birbeck College, University of London, U.K. where it is currently hosted (36). This server allows for the input of data in a variety of formats which can then be analysed using the above mentioned algorithms. Each of these algorithms are implemented based off a database designated in the DICHROWEB program. Conventional CD instruments have a Xe arc light source which allows measurements to 180 nm. At wavelengths below 180 nm the intensity of the radiation drops off and significant absorption is observed with the N_2 used for purging the system and the H_2O solvent. To achieve a more reliable estimate of secondary structure it is beneficial to go further into the far UV region, below 170 nm, which can be accomplished using a synchrotron light source (33, 37). Furthermore, for reliable secondary structure analysis, the protein concentration should be accurately known and the CD instrument should be properly calibrated.

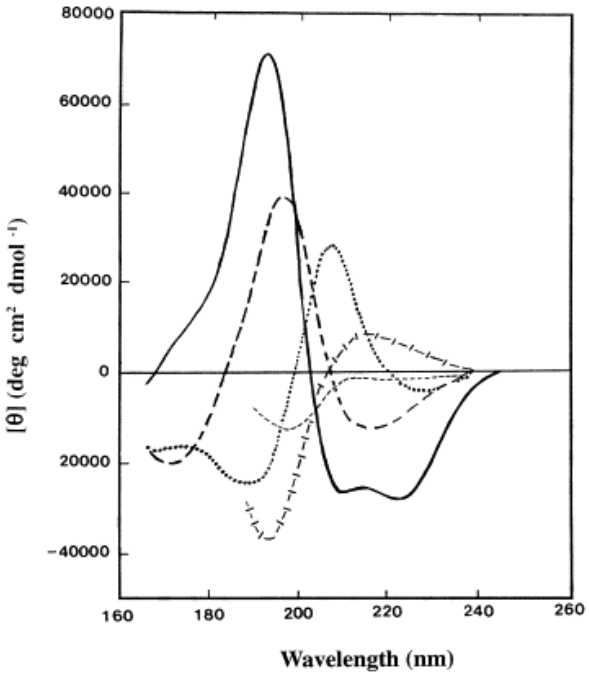

Figure 2.2 Far UV CD spectra associated with various types of secondary structure

CD spectra of the following secondary structures: Solid line, α-helix; long dashed line, anti-parallel β-sheet; dotted line, type I β-turn; cross dashed line, extended 31-helix or poly (Pro) II helix; short dashed line, irregular structure (27).

For analysis of the tertiary structure, the 260 – 320 nm range is investigated which arises from the characteristic profiles of the aromatic amino acids. Trp has a peak close to 290 nm with fine structure between 290 nm and 305 nm. Tyr has a peak at 275 nm and 282 nm with a shoulder at longer wavelengths that is typically obscured by the Trp profile. Phe has weaker but sharper peaks with fine structure between 255 nm and 270 nm. The fine structure results from the vibronic transitions in different vibrational levels of the excited state that are involved (38). The actual shape and magnitude of the near UV region of the CD spectra will depend on the number of each type of aromatic amino acid, their mobility, the nature of their environment, and their relative position in the protein. The tertiary structure analysis is still quite limited and therefore this region is typically used as a footprint to compare, for example, wild-type and mutant forms of a protein (27).

A final way to take advantage of the CD technique for proteins is investigating the integrity of cofactor binding. Typically, organic cofactors show little or no CD signal when they are free in solution but once bound to its protein partner in a site that imparts chirality, a signal is produced. Through a comparative analysis of CD spectra, signal in the appropriate spectral region for the cofactor is an excellent indicator of the integrity of the binding (27).

2.4.3 CD of nucleic acids

DNA bases absorb light in the UV range (180 – 300 nm) (39), making spectropolarimetry a valuable technique to look at the DNA structure. This technique for obtaining conformational information has many advantages. For example, it is extremely sensitive which allows for working with concentrations as low as 25 µg/mL. This permits experiments with samples of low solubility or which tend to aggregate in extreme solvent conditions. In addition, through changing the path length of the cell, the DNA concentration can effectively be altered up to three

orders of magnitude (40). Secondly, the technique allows for analysis of not only short oligonucleotides, but long strands of DNA. These samples can be easily titrated with various agents to study their conformational isomerizations. Graphical methods of CD data allow for the determination of isodichroic points that often accompany two-state conformational isomerizations. Iosdichroic points are the points at which two spectra for limiting conformers intersect (41). Finally, CD measurements are quick and relatively inexpensive which provides the ability to perform comparative studies of DNA under varying conditions. As a result, CD has become an attractive technique for certain studies over NMR and X-ray crystallography which require a lot of material and extensive amounts of time and effort.

DNA can adopt various forms such as B, A and Z-form which each have distinctive CD spectra. In B-form DNA the peak intensities are relatively low since the base pairs are perpendicular to the double-helix axis, conferring weak chirality on the molecule. However, as the (A + T) content increases, the conformational variability increases which results in a deepening of the minimum. B-DNA is characterized by a maximum between 260 and 280 nm and a minimum at 245 nm. The exact position and magnitude of these peaks are dependent on the nucleic acid sequence, since the chromophores in each base differ and the conformational properties become altered (40).

While DNA is most commonly observed in the B-form, it can transition to A-form (a constitutive form for RNA) in aqueous ethanol and other solutions (42). The CD spectrum of A-form DNA is characterized by a dominant peak ($7 – 12\ M^{-1}cm^{-1}$) at 260 nm in addition to a minimum at 210 nm. The B-A transition observed for DNA is a fast, two-step process which can be observed by the isodichroic points as well as the S-shaped CD spectra obtained with increasing the concentration of the inducing agent (**Figure 2.3**).

The newest observed form of DNA is the left-handed Z-form. The base pairs in the Z-form double helix have an opposite orientation to what is observed in both A- and B-form DNA. As a result, its CD spectrum is nearly the opposite of B-form with a minimum at 290 nm, maximum at 260 nm, and a very characteristic strong minimum at ~205 nm (43). As with the B-A transition, the B-Z transition can also be monitored by CD spectroscopy (**Figure 2.3**) (41). In contrast, the B-Z transition is a much slower process resulting from a required base pair flip which is kinetically a challenging process.

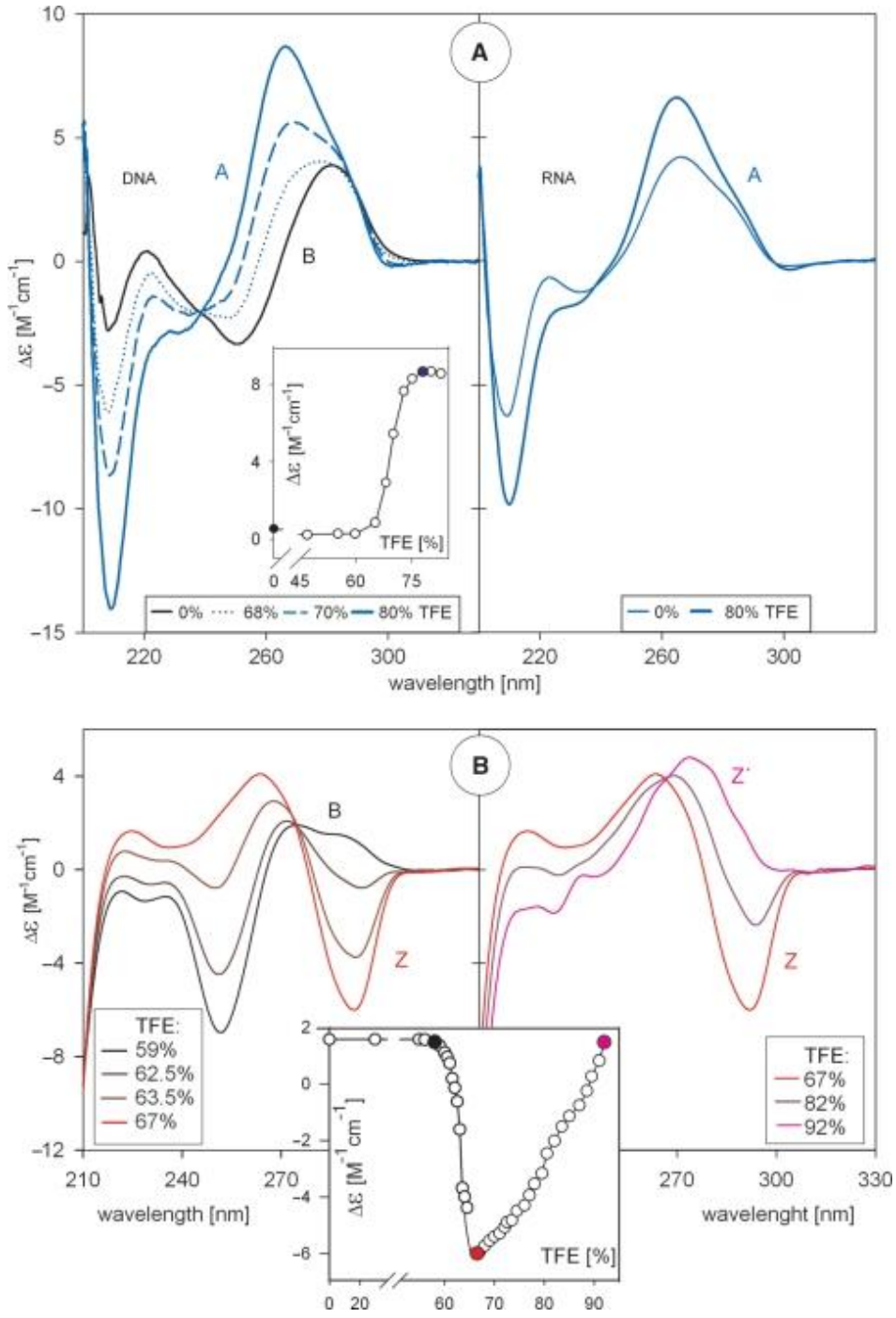

Figure 2.3 Circular dichroism spectra of B-A and B-Z transitions of DNA

A) Left panel: CD spectra reflecting trifluorethanol-induced B–A transition of d(GCGGCGACTGGTGAGTACGC) duplex. Insert: the transition monitored at 266 nm. Right panel: CD spectra of RNA of the same sequence (U instead of T) duplexed with complementary DNA strand.
(B) CD spectra reflecting trifluorethanol-induced B–Z (left panel) and Z–Z′ (right panel) transitions of poly[d(GC)] duplex. Insert: the transitions monitored at 291 nm (40).

2.5 Nuclear magnetic resonance

2.5.1 Theory

NMR is a technique which exploits the magnetic properties of atomic nuclei which have an odd mass number. These atoms (in contrast to those with an even mass number), rotate about a given axis and therefore create a magnetic field. When placed in an external magnetic field (such as in NMR) the magnetic field of the nuclei will orient opposite to that of the external field and precess around the axis of the field vector. A fluctuating decline in the amplitude is observed over time as the nuclear magnetization decreases, referred to as free induction decay (FID) (44). FID is then Fourier transformed from a function of time to that of frequency to generate an NMR spectrum.

A chemical shift (or shielding effect), is observed due to the behaviour of the electrons in the molecule when placed in an external magnetic field. The electrons produce local currents, which generate an alternative field that opposes the external field. As a result, the effective magnetic field is reduced and is dependent on the structural geometry and chemical structure of the molecule. Chemical shifts are usually measured as a shift in Hz from a reference signal which is then converted to a dimensionless unit of parts per million (ppm) (44).

2.5.2 Sample preparation

To prepare a sample for NMR studies, a labelling strategy must be employed. Furthermore, when preparing macromolecular complexes, one must consider whether to label only one partner or differentially label both. For protein/DNA complexes, the most common labelling strategy is to ^{15}N and/or ^{13}C label the protein while leaving the DNA unlabelled,

however it is just as possible to ^{15}N and/or ^{13}C label the DNA while leaving the protein unlabelled. Through implementing one of these labelling strategies, the resonances can be conveniently followed in a titration experiment (45). Proteins can be easily isotopically labelled by growing the protein-expressing cells in a minimal media contain the desired label. On the other hand, DNA labelling is prepared through a chemical or enzymatic method (46).

2.5.3 Secondary structure determination

NMR is another technique which can provide some insight as to the secondary structure of a protein. However, it requires much more material and specific protein labelling to be the method of choice. Thus, it is often used as an initial experiment before proceeding to further experimentation with the same sample. As such, a heteronuclear single quantum coherence spectroscopy (HSQC) spectra is obtained which correlates the backbone amide nitrogen resonances with those of the directly attached protons. For an intrinsically disordered protein (IDP) the peaks in the proton dimension are not widely dispersed and span between 8.0 and 8.5 ppm while in the nitrogen dimension, the peaks are widely spread between 105 and 130 ppm. This difference in dispersion between the two dimension is indicative of disorder in the protein (**Figure 2.4**) (47).

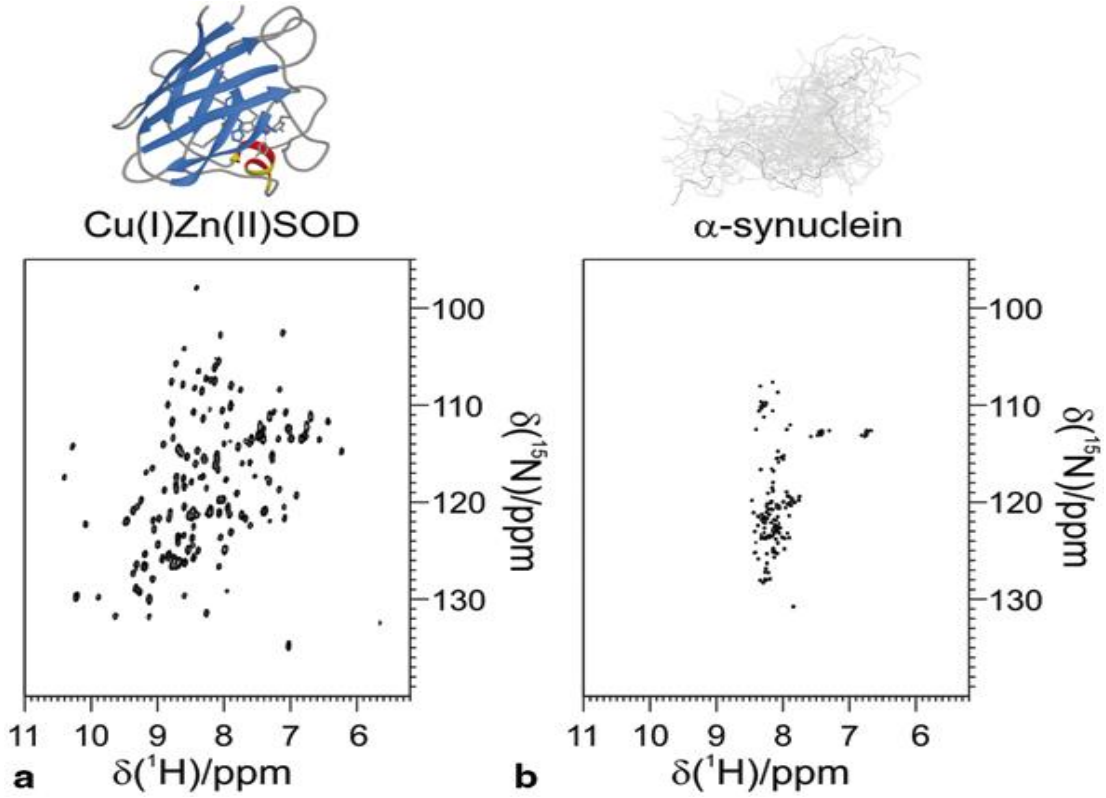

Figure 2.4 Chemical shift dispersion of structured and intrinsically disordered proteins

A 2D ^{1}H-^{15}N correlation spectrum was acquired for a) Cu(I)Zn(II) superoxide dismutase and b) α-synuclein both of similar size. Experiments were acquired on a 700 MHz Bruker AVANCE spectrometer equipped with a CPTXI probe (48).

Determining the amount of secondary structural elements using NMR is not as straightforward as CD since the spectra are not as defined. A group in Brazil developed the PASSNMR program which can extract information from ^{13}C – and ^{15}N – HSQC spectra to predict the amount of α-helix, β-sheet, and non-αβ structure in an unknown protein with 72%, 74%, and 48% reliability, respectively (49). They also claim that information from only the ^{15}N – HSQC is sufficient to obtain an accurate prediction.

2.5.4 Chemical shift mapping

Chemical shift mapping is the easiest NMR method and is based on the detection of the protein backbone ^{15}N and HN chemical shifts in HSQC spectra. From the development of TROSY (Transverse Relaxed Optimized SpectroscopY), it has become possible to record NMR spectra for large complexes with high spectral resolution. To further analyze a protein/DNA complex, the HSQC spectra for the free and bound ^{15}N-labelled must be recorded. From this information, the normalized chemical shift perturbations (CSPs) can be calculated

$$\boldsymbol{CSP} = \sqrt{\Delta\boldsymbol{\delta^2 HN} + \left(\frac{\boldsymbol{\gamma_N}}{\boldsymbol{\gamma_H}}\right) \times \Delta\boldsymbol{\delta^2 N}} \tag{2.20}$$

where ΔδHN and ΔδN are the observed chemical shift variations in parts per million and γ_H and γ_N are the gyromagnetic ratio of each nucleus.

The important CSPs that are observed in a protein/DNA complex are a result of the change in DNA phosphate proximity or the ring currents induced by aromatic bases. As a result, NMR is a method used to rapidly map out the interaction surface on a protein (45). For such methodology to be advantageous, the DNA-binding surface must be in a continuous protein interface and not located in several discontinuous faces.

2.6 Small angle x-ray scattering

2.6.1 Data collection and analysis

SAXS is a low-resolution technique (50 Å – 10 Å) which, similar to X-ray crystallography, places a sample in a collimated beam and detects the scattered X-rays. As the name suggests, the main differences between SAXS and X-ray crystallography is the angle at which these samples are exposed, the nature of the sample (solvated or crystalline), and the

radial symmetry. In SAXS, one exposes the solvated sample to X-rays at low angles without the need to rotate it during data collection, as is necessary for crystalline samples. However, buffer measurements must also be precisely collected and subtracted to accurately analyze the solute data. This makes SAXS particularly attractive for nucleic acids as they have a higher contrast than proteins and can therefore be analyzed at lower concentrations (50). The ability to obtain useful information on a sample is contingent on the contrast between the solute and bulk solvent. For lengths larger than 15 Å, the scattering density for biomolecules is relatively homogeneous. Therefore, any internal structural features can be ignored at resolutions lower than $q = 0.2$ Å^{-1}.

The scattering curve obtained in a SAXS experiment, *I(q)*, is a function of momentum transfer of the photons $q = (4\pi\sin\theta)/\lambda$, where 2θ is the scattering angle, and λ is the wavelength of the incident X-ray beam (51). For a homogeneous, monodisperse sample that lacks long-range interactions in solution, the scattering curve can be derived from the electron distribution of the particle (known as the pair-distance distribution function, *P(r)*):

$$\boldsymbol{I(q) = 4\pi \int_0^{D_{max}} P(r) \frac{\sin(qr)}{qr} dr} \quad (2.21)$$

Where D_{max} is the maximal particle dimension (50).

The lowest resolution obtained in a given experiment is based on the radius of gyration (R_g), which is the square root of the average squared distance of each scatterer from the particle center. Similar to R_h and Stoke's radius (R_s), R_g is shape-dependent and therefore a poor measure of the actual molecular weight of the molecule of interest. A Guinier approximation can describe the solution scattering at low resolution:

$$\boldsymbol{I(q) = I(0) \exp\left[-\frac{q^2 {R_g}^2}{3}\right]} \quad (2.22)$$

Moreover, a Guinier plot of $\log(I(q))$ against q^2 results in a straight line by which R_g and *I(0)* can be determined. As common practice, SAXS experiments should be done over a range of

concentrations in the same buffer, and as such the R_g should not vary. An additional parameter that is gained from a Guinier plot is the *I(0)* which can only be obtained through extrapolation as it is coincident with the direct beam. *I(0)* is defined as the square of the number of electrons in the scatterer and is unaffected by particle shape. This parameter becomes useful upon determining the molecular weight of the sample (52).

Analyzing the region of higher *q* values provides information regarding the molecular shape of a particle. For folded macromolecules, the intensity of the scattering falls off by Porod's law as $I(q) \propto q^{-4}$ (53). This relationship assumes a uniform density of the scatterer and does not hold for high *q* values at which point atomic resolution information begins to contribute significantly. As a result, SAXS is a useful technique to identify unfolded domains. This can be observed directly through a Kratky plot [$q^2I(q)$ as a function of *q*] which can be calculated from the original scattering curve. Folded domains appear roughly as a parabolic curve with the position of the peak providing some information on overall size (but not molecular weight). Conversely, random coil or unstructured proteins lack the characteristic folded peak and instead are linear with respect to *q* in the large *q* region (54).

The *P(r)* function is an autocorrelation function that provides information about the distances between pairs of electrons in the scattering particle. It can be directly calculated from an indirect Fourier transform of the scattering curve. The indirect Fourier transform essentially constructs trial *P(r)* functions that are Fourier transformed and then compares them to the experimental scattering. The GNOM program uses a regularizing multiplier to balance the smoothness of the trial *P(r)* functions to the goodness of fit to the data (55). Theoretically, the *P(r)* function is zero at $r = 0$ and $r \geq D_{max}$, and therefore typically constraints are introduced to set these parameters to zero. In well-behaved, globular samples, these constraints are often

unnecessary. However, in unfolded proteins the *P(r)* function is often not zero at $r = 0$, and non-zero values at $r = D_{max}$ suggest protein aggregation or poor buffer subtraction. D_{max} is a useful parameter for characterizing a sample and heavily relied on for calculation of the *P(r)* function, but accurately determining it can be a challenge. In practice, estimating the D_{max} by inverse Fourier transformation involves choosing multiple D_{max} values and evaluating the resulting *P(r)* function for its fit to the experimental scattering. Due to the fact that the *P(r)* function is small near the D_{max} it has little contribution to the experimental scattering and errors associated with it are hard to identify.

The *P(r)* plot has an advantage over the Guinier plot in that it is able to use all of the collected data to calculate the R_g and *I(0)*. This becomes advantageous in samples which become complicated with small amounts of aggregation. Additionally, a *P(r)* function can give some initial indication of the overall sample shape. For example, spherical-shaped proteins would generate a bell curve which would be easily distinguishable from a rod-shaped protein. A *P(r)* plot analyzed in conjunction with a Kratky plot can provide useful information about a protein without any low-resolution modelling.

2.6.2 Ab-inito modelling

Programs have been developed and refined to be able to produce a 3D reconstruction from a 1D scattering curve. Successful reconstructions generate a low-resolution envelope which is analogous to the averaged reconstruction in electron microscopy (EM). The *ab-initio* envelope used in these programs is comprised of all the independent experimental values extracted from the scattering curve in addition to external information. The term *ab-initio* refers to this lack of pre-defined structure input. In order to reduce the search space, these programs perform on the

assumption that the shape is a continuous object. DAMMIN and DAMMIF are two of multiple programs which represent particles by a large number of densely packed beads inside a designated search volume. The program begins with an arbitrary initial model and assigns each bead to either the particle or solvent through simulated annealing until the interconnected model yields a scattering pattern that fits the experimental data (56, 57). External constraints such as smoothness, connectivity, and particle symmetry can be implemented to further reduce the search space (58). In practice, a comparison of the output of multiple independent modeling runs is undergone and the agreement between the models is characterized by the normalized spatial discrepancy (NSD). For identical models, the NSD is zero.

Other programs such as MONSA (an extended version of DAMMIN) fit simultaneously multiple curves. This provides the ability to deconvolute the contribution of protein and DNA in a complex *ab initio* model (57, 59). Comparing the hydrodynamic parameters of these models with those determined experimentally, as well as the goodness of fit parameters, one can be confident that the model produced accurately represents the solution behaviour of the molecules. SAXS has become a very useful technique which when combined with X-ray crystallography is very powerful.

References

1. Fekete, S., Beck, A., Veuthey, J.-L., and Guillarme, D. (2014) Theory and practice of size exclusion chromatography for the analysis of protein aggregates. *J. Pharm. Biomed. Anal.* **101**, 161–173
2. Hong, P., Koza, S., and Bouvier, E. S. P. (2012) Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J. Liq. Chromatograpy Relat. Technol.* **35**, 2923–2950
3. Vander Heyden, Y., Popovici, S. T., and Schoenmakers, P. J. (2002) Evaluation of size-exclusion chromatography and size-exclusion electrochromatography calibration curves. *J. Chromatogr.* **957**, 127–137
4. Arakawa, T., Ejima, D., Li, T., and Philo, J. S. (2010) The critical role of mobile phase composition in size exclusion chromatography of protein pharmaceuticals. *J. Pharm. Sci.* **99**, 1674–1692
5. Kamberi, M., Chung, P., DeVas, R., Li, L., Li, Z., Ma, X., Fiels, S., and Riley, C. M. (2004) Analysis of non-covalent aggregation of synthetic hPTH (1-34) by size-exclusion chromatography and the importance of suppression of non-specific interactions for a precise quantitation. *J. Chromatogr.* **810**, 151–155
6. Ejima, D., Yumioka, R., Arakawa, T., and Tsumoto, K. (2005) Arginine as an effective additive in gel permeation chromatography. *J. Chromatogr.* **1094**, 49–55
7. Barth, H. G., and Saunders, G. D. (2012) The state of the art and future trends of size-exclusion chromatography packing and columns. *LCGC North Am.* **30**, 544–563
8. Stetefeld, J., Mckenna, S. A., and Patel, T. R. (2016) Dynamic light scattering : a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8**, 409–427

9. Lorber, B., Fischer, F., Bailly, M., Roy, H., and Kern, D. (2012) Protein analysis by dynamic light scattering: Methods and techniques for students. *Biochem. Mol. Biol. Educ.* **40**, 372–382

10. Harding, S. E., and Jumel, K. (1998) Light Scattering. in *Current Protocols in Protein Science* (Coligan, J. E., Dunn, B. M., Ploegh, H. L., Speicher, D. W., and Wingfield, P. T. eds), John Wiley & Sons, New York

11. Harding, S. E. (1999) Protein Hydrodynamics. in *Protein: a comprehensive treatise*, vol 2 (Allen, G. ed), pp. 271–305, JAI Press, Greenwich

12. Pecora, R. (ed.) (1985) *Dynamic Light Scattering - Applications of Photon Correlation Spectroscopy*, Plenum Press, New York

13. Smith, C. A. (1988) Estimation of Sedimentation Coefficients and Frictional Ratios of Globular Proteins. *Biochem. Educ.* **16**, 104–106

14. Malvern (2016) Intensity-Volume-Number. [online] http://www.malvern.com/en/pdf/secure/TN101104IntensityVolumeNumber.pdf (Accessed January 30, 2017)

15. Cole, J. L., Lary, J. W., Moody, T., and Laue, T. M. (2009) Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods Cell Biol.* **84**, 1433–179

16. Correia, J. J., and Stafford, W. F. (2015) Sedimentation Velocity: A Classical Perspective. in *Methods in Enzymology*, pp. 49–81, Elsevier Inc.

17. Fujita, H. (1975) *Foundations of ultracentrifugal analysis*, Wiley, New York

18. Schuck, P. (2013) Analytical Ultracentrifugation as a Tool for Studying Protein Interactions. *Biophys. Rev.*

19. Laue, T. M. (1996) *Choosing Which Optical System of the Optima^{TM} XL-I Analytical*

Ultracentrifuge to Use, Beckman Instruments, Inc.

20. Yphantis, D. (1964) Equilibrium ultracentrifugation of dilute solutions. *Biochemistry*. **3**, 297–317

21. Zhao, H., Brown, P. H., and Schuck, P. (2011) On the distribution of protein refractive index increments. *Biophys. J.* **100**, 2309–2317

22. Salvay, A. G., Santamaria, M., Le Maire, M., and Ebel, C. (2007) Analytical ultracentrifugation sedimentation velocity for the characterization of detergent-solubilized membrane proteins Ca ++-ATPase and ExbB. *J. Biol. Phys.* **33**, 399–419

23. DeRosier, D., Munk, P., and Cox, D. (1972) Automatic Measurement of Interference Photographs for the Ultracentrifuge. *Anal. Biochem.* **50**, 139–153

24. Schuck, P., and Demeler, B. (1999) Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. *Biophys. J.* **76**, 2288–96

25. Dam, J., and Schuck, P. (2004) Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. *Method Enzymol.* **384**, 185–212

26. Brown, P. H., and Schuck, P. (2008) A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. *Comput. Phys. Commun.* **178**, 105–120

27. Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta*. **1751**, 119–139

28. Johnson Jr., W. C. (1996) *Circular Dichroism Instrumentation in Circular Dichroism and The Conformational Analysis of Biomolecules* (Fasman, G. D. ed), Plenum Press, New York

29. Kelly, S. M., and Price, N. C. (2000) The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **1**, 349–84

30. Berova, N., Nakanishi, K., and Woody, R. W. (eds.) (2000) *Circular Dichroism: Principles and Applications*, 2nd Ed., Wiley-VCH, New York

31. Sreerama, N., and Woody, R. W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209**, 32–44

32. Manavalan, P., and Johnson Jr., W. C. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **167**, 76–85

33. Johnson Jr., W. C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins Struct. Funct. Genet.* **35**, 307–312

34. Andrade, M. A., Chacón, P., Merelo, J. J., and Morán, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised neural network. *Protein Eng.* **6**, 383–390

35. Provencher, S. W., and Glöckner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry*. **20**, 33–37

36. Whitmore, L., and Wallace, B. a (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668-73

37. Wallace, B. A., and Janes, R. W. (2001) Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics. *Curr. Opin. Chem. Biol.* **5**, 567–571

38. Krittanai, C., and Johnson Jr., W. C. (1997) Correcting the circular dichroism spectra of

peptides for contributions of absorbing side chains. *Anal. Biochem.* **253**, 57–64

39. Johnson Jr., W. C. (1996) *Determination of the Conformation of Nucleic Acids by Electronic CD*, Plenum Press, New York

40. Kypr, J., Kejnovská, I., Renčiuk, D., and Vorlíčková, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713–1725

41. Kypr, J., and Vorlíčková, M. (1986) Graphical analysis of circular dichroic spectra distinguishes between two-state and gradual alterations in DNA conformation. *Gen. Physiol. Biophys.* **5**, 415–422

42. Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamentetskii, M. D., and Schyolkina, A. K. (1974) The B to A transition of DNA in solution. *J. Mol. Biol.* **87**, 817–833

43. Pohl, F. M., and Jovin, T. M. (1972) Salt-induced cooperative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.* **67**, 375–396

44. Poulsen, F. M. (2002) A brief introduction to NMR spectroscopy of proteins. [online] https://www.cs.duke.edu/brd/Teaching/Bio/asmb/Papers/Intro-reviews/flemming.pdf (Accessed March 21, 2017)

45. Campagne, S., Gervais, V., and Milon, A. (2011) Nuclear magnetic resonance analysis of protein − DNA interactions. *J. R. Soc. Interface*. **8**, 1065–1078

46. Jamin, N., and Toma, F. (2001) NMR studies of protein-DNA interactions. *Prog. Nucl. Magn. Res. Sp.* **38**, 83–114

47. Tompa, P., and Fersht, A. (2009) *Structure and Function of Intrinsically Disordered Proteins*, Champan & Hall

48. Brutscher, B., Felli, I. C., Gil-Caballero, S., Hošek, T., Kümmerle, R., Piai, A., Pierattelli, R., and Sólyom, Z. (2015) NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. in *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*, pp. 49–122, Springer International Publishing Switzerland, Switzerland, 10.1007/978-3-319-20164-1_3

49. Moreau, V. H., Valente, A. P., and Almeida, F. C. L. (2006) Prediction of the amount of secondary structure of proteins using unassigned NMR spectra: a tool for target selection in structural proteomics. *Genet. Mol. Biol.*

50. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. a (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285

51. Boldon, L., Laliberte, F., and Liu, L. (2015) Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Rev.*

52. Kikhney, A. G., and Svergun, D. I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577

53. Porod, G. (1951) X-ray small angle scattering of close packed colloidal systems. *Kolloid Zeitschrift*. **124**, 83–114

54. Kratky, O., and Porod, G. (1949) Roetgenuntersuchung Geloester Fadenmolekuele. *Recl. des Trav. Chim. des Pays-Bas*. **68**, 1106–1122

55. Semenyuk, A., and Svergun, D. I. (1991) GNOM - a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **24**, 537–540

56. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346

57. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886

58. Koch, M. H., Vachette, P., and Svergun, D. I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* **36**, 147–227

59. Svergun, D. I., and Nierhaus, K. H. (2000) A map of protein-rRNA distribution in the 70S Escherichia coli ribosome. *J. Biol. Chem.* **275**, 14432–14439

Chapter 3：Nano-scale assembly of high-mobility group AT-hook 2 protein with DNA replication fork

This chapter is based on the following manuscript submitted to the journal *"Biophysical Journal"*. Full citation of the paper is as follows:

Krahn, N., Meier, M., To, V., McEleney, K., O'Neil, J., Klonisch, T., Patel, T.R., Stetefeld, J. (2017). Nano-scale assembly of high-mobility group AT-hook 2 protein with DNA replication fork. *Biophys. J.* **Accepted for publication.**

This chapter investigates a protein (HMGA2) which has been found at high expression levels in tumour cells. As a result, it has been proposed as a potential broad-range biomarker for diagnosis of numerous types of cancer. Among its many roles, HMGA2 assists in the stabilization of stalled replication forks, such that they can continue through replication, avoiding apoptosis. With this information, I studied the interaction of HMGA2 with replication fork DNA such that I could begin to understand the mechanisms which aid in tumour growth.

Abstract

High mobility group AT-hook 2 (HMGA2) protein is composed of three AT-hook domains that have a higher binding affinity to distorted DNA, such as replication forks (RFs), compared to typical B-form dsDNA. HMGA2 expresses at high levels in embryonic stem cells (ESCs) and cancer cells where it stabilizes RFs. This ultimately results in increased cell proliferation that is necessary for growth and development in ESCs but becomes problematic in adults upon the onset of tumour growth. Although it is known that HMGA2 interacts with RFs, the solution structure of the complex, and the conformational changes in HMGA2 upon

interaction with RF have been largely unexplored. Using biophysical techniques, we sought to unveil the mode of interaction which leads to cell proliferation in ESCs and tumour formation in cancer cells. Circular Dichroism and Nuclear Magnetic Resonance confirmed the disordered state of unbound HMGA2. Dynamic light scattering and sedimentation velocity experiments determined that HMGA2 and RF are monodispersed in solution and form an equimolar complex. Furthermore, small-angle X-ray scattering extended this knowledge to reveal that HMGA2 binds in a side-by-side orientation to RF where 3 AT-hooks act as a "clamp" and wrap around a distorted RF stabilizing it. Cell biology experiments revalued that indeed, absence of HMGA2 reduces cell proliferation. Thus, this work allows an understanding of how HMGA2 interacts with stalled RFs and its biological significance.

3.1 Introduction

The high-mobility group (HMG) family of proteins consist of HMGA (HMG-I/Y), HMGB (HMG-1) and HMGN (HMG-14/17) superfamilies. They play significant roles in a number of biological activities, including chromatin remodelling and DNA repair pathways (1, 2). The HMGA group of proteins have preferential binding to the minor groove of AT-rich regions in B-form DNA as well as many bent or distorted DNA structures (3). HMGA proteins are highly expressed in self-renewing human embryonic stem cells (ESCs) but upon differentiation, the levels of HMGA drops to undetectable levels due to reduction in *Lin28b* transcription (4). This then leads to suppression of HMGA production by let-7 microRNA (5).

In mammals, HMGA is composed of two functional members, HMGA1 (a.k.a. HMGI-Y) and HMGA2 (a.k.a HMGI-C), which are encoded by separate genes (1, 6). HMGA1 has three experimentally observed splice variants (HMGA1a, HMGA1b, and HMGA1c) (7) while only

one major transcript has been observed for HMGA2 (HMGA2a) (8). *In vivo*, HMGA1 and HMGA2 regulate transcription of many of the same genes; however, they also assume distinct roles. For example, HMGA1 is involved in cardiomyocyte cell growth (9) and regulation of glucose and insulin pathways (10), whereas HMGA2 aids in controlling skeletal size (11) and is a critical component for normal spermatogenesis (2). The high expression of HMGA2 in ESCs is necessary for their proper growth and development (11). Unfortunately, if HMGA2 expresses at high levels in adult somatic cells, the effects can be detrimental. This occurs when let-7 microRNA gets degraded in somatic cells, relieving the inhibition of HMGA2 expression. The presence of HMGA2 in the nucleus leads to high cell proliferation and formation of tumours (5). Thus, HMGA2 has become a useful biomarker for various cancers (12–14), for pathological subtyping of testicular germ cell tumours (15) and for predicting poor prognosis in breast cancer (16).

HMGA2a (referred as HMGA2), having a molecular weight of 11.8 kDa, (**Figure 3.1**), contains three independent AT-hook DNA binding motifs (Pro-Arg-Gly-Arg-Pro) which are flanked by positively charged Lys/Arg residues (17). HMGA2 also contains a high number of prolines (~15%) and is highly basic in nature. The AT-hook binding motif is conserved in HMGAs throughout evolution, from bacteria to humans, and is also found in many transcription factor proteins involved in chromatin remodeling complexes (17, 18). AT-hooks are DNA binding motifs that insert into the minor groove of DNA. It is speculated that the proline residues are necessary to maintain the rigidity of this domain, while the arginine residues in the Arg-Gly-Arg motif are interacting with the bases of the DNA (18). These three AT-hooks are responsible for DNA binding in transcriptional regulation (19), base excision repair (20), and the stabilization of stalled replication forks (RFs) (21).

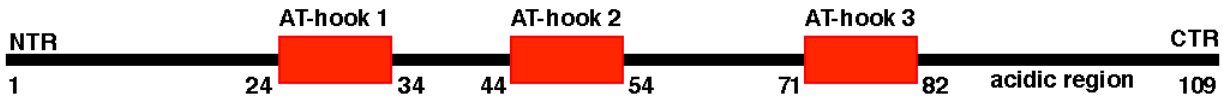

Figure 3.1 Schematic of the domains of human HMGA2 highlighting the AT-hook domains

During replication, a fraction of RFs are arrested as they translocate along the parental DNA. In order for the RF to resume replication, adult cells are equipped with a RF protection complex that recognizes ssDNA and attempts to stabilize the RF (22). The RFs which are unable to be stabilized lead to the collapse of the DNA and eventual cell death. In ESCs and cancer cells, the overexpression of HMGA2 provides additional support to the fork protection complex and is one of the reasons for the observed increase in cell proliferation (21).

In this study, we characterized HMGA2 and a DNA RF individually and in complex using biophysical methods. Our results confirm that human HMGA2 exists as a disordered monomer in solution (in contrast to the dimeric nature of mouse HMGA2 (23)); however, upon interaction with a DNA RF, it forms an equimolar complex by binding in a side-by-side orientation. Our study suggests that it is necessary for HMGA2 to be in a disordered state such that it has the flexibility to stabilize a stalled RF by binding to the necessary AT-rich sites. HMGA2 creates a "clamp" along the RF by taking advantage of the rigidity of the double-stranded region to stabilize the strands being newly synthesized. Furthermore, in-cell studies have confirmed a positive role for HMGA2 in regulating cell proliferation.

3.2 Materials and Methods

3.2.1 Protein expression and purification

A pet28a(+) vector containing *Escherichia coli* optimized codons (24) for the production of human HMGA2 with a thrombin cleavable N-terminal hexa-histidine tag was transformed into

E.coli BL21 cells. Cells were inoculated into 1.6 L Luria Broth (LB) containing 50 µg/mL kanamycin and grown at 37 °C until the cell density reached an OD_{600} of 0.6. Cells were then induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1mM) and protein was expressed for 3 hours before harvesting the cells by centrifugation at 6500 rpm for 20 mins. The cell pellet was frozen at -80 °C until further processing. Purification and removal of the N-terminal hexa-histidine tag was performed as previously described (25).

3.2.2 Formation of RF DNA with and without HMGA2

Polyacrylamide Gel Electrophoresis (PAGE) purified oligonucleotides oSG309 (TTTTTTATAATGCCAACTTAGTATAAAAAAGCTGAACGAGAAACGTAAAA), oSG606 (TTTTACGTTTCTCGTTCAGCTTTTTTATACTAACTTGAGCGAAACGGGAA), and oSG607 (TTCCCGTTTCGCTCAAGGTTGGCATTATAAAAAA) (Alpha DNA, Montreal, QC) were annealed as described previously (21) to form the RF. Assembled RF was separated from free oligonucleotides by means of size exclusion chromatography (SEC) using a HiLoad Superdex 75 column (GE Healthcare, Fairfield, CT) in 10 mM HEPES, pH 7.5, 154 mM NaCl buffer. To purify the DNA:protein complex using SEC, equal volumes of RF (2 µM) and HMGA2 (3 µM) were mixed together and incubated for 30 mins at room temperature. The complex was then concentrated and loaded onto a Superdex 200 column (GE Healthcare, Fairfield, CT) in 10 mM HEPES, pH 7.5, 154 mM NaCl buffer to remove any unbound species.

3.2.3 Coomassie SDS-PAGE

HMGA2 protein before and after thrombin cleavage treatment was mixed 1:1 with reduced Lämmli buffer and heated at 95°C for 5 mins before loaded onto a 16% SDS-PAGE gel

containing 6 M urea. The gel was run at 5 W for 2.5 hrs and then stained with Coomassie Brilliant Blue.

3.2.4 DNA gel shift assays

DNA-protein interactions were observed by incubating a constant amount of RF (200 nM) with various amounts of HMGA2 (0 – 3 μM) in 10 mM HEPES, pH 7.5, 154 mM NaCl for 20 mins. The binding reactions were resolved by electrophoresis on a 10 % Tris-borate Ethylenediaminetetraacetic acid (TBE) native gel. DNA was stained using SYBR Gold fluorescent nucleic acid dye (Invitrogen, Carlsbad, CA) and visualized on a FluorChem Q imaging system (Cell Biosciences, Santa Clara, CA) (25).

3.2.5 Hydrodynamic characterization

The homogeneity of HMGA2, RF and their complex was studied by dynamic light scattering (DLS) and sedimentation velocity (SV) methods in 10 mM HEPES, pH 7.5, 154 mM NaCl buffer. SEC purified samples were subjected to 0.1 μm filtration before equilibrating at 20 ºC for 5 mins for DLS analysis. The hydrodynamic radius (R_h) distributions were obtained using a Nano-S Zetasizer (Malvern, Toronto, ON) at concentrations between 1 – 10 mg/mL as described previously (26–28). SV experiments were performed as previously described (27, 29) using a ProteomeLab^{TM} XL-I analytical ultracentrifuge and an An50Ti 8-cell rotor (Beckman Coulter Inc., Mississauga, ON). Standard 12 mm double-sector cells were filled with 400 μL of sample and 420 μL of buffer. A total of 120 absorbance and interference radial scans were performed at a speed of 45,000 revolutions per minute (163, 296 x g) at every 11.5 minute interval with a concentration range of 0.2 – 2 mg/mL. Data were analysed using the SEDFIT

program to determine the sedimentation coefficients of each macromolecule ($s_{20,b}$) (30, 31). The sedimentation coefficients were then corrected to standard solvent conditions using the buffer viscosity (η_0) and density (ρ_0) calculated from SEDNTERP ($s_{20,w}$) (32). Plotting the $s_{20,w}$ values for multiple concentrations provided the Svedberg constant at infinite dilution ($s^0_{20,w}$) which combined with the R_h allows calculation of the molecular mass (28, 33):

$$M_w = \frac{6\pi\eta_0 R_h N_A s^0{}_{20,w}}{1-\bar{\nu}\rho_0} \quad (3.1)$$

where N_A is Avogadro's number (6.022 x 10^{23}) and $\bar{\nu}$ is the partial specific volume.

3.2.6 Spectropolarimetry

Far UV CD spectra were recorded on a J-810 spectropolarimeter (Jasco Inc., Easton, PA) that was calibrated using 2.583 mM (1S)-(+)-camphor-10-sulfonic acid monohydrate (Alfa Aesar, Ward Hill, MA). The protein and RF spectra were measured at concentrations of 0.25 mg/mL and 0.21 mg/mL, respectively, in 10 mM sodium phosphate, pH 7.5, 154 mM NaF buffer at 20.0 °C. The RF/HMGA2 complex was measured in the same phosphate buffer at 0.17 mg/mL DNA concentration. Data were collected using an integration time of 8 s/nm in a 1 mm cell. The spectra were normalized by the light path length and the number of peptide bonds (in case of protein) or nucleotides (in case of RF or RF/HMGA2 complex) per unit volume as described previously (27, 29).

3.2.7 Chemical shift mapping of amide backbone

HMGA2 was expressed from *E. coli* cells grown in M9 minimal media with $^{15}NH_4Cl$ as the sole nitrogen source to label the nitrogens and purified as described in section 3.2.1. Pure HMGA2 was concentrated to 500 – 700 µM with and without the presence of unlabelled RF in

NMR buffer (10 mM HEPES, pH 7.5, 154 mM NaF, 10 % D_2O, 75 µM 2,2-dimethyl-2-silapentane-5-sulfonate (DSS)). The ^{15}N-HSQC spectra (34) were collected with 1024 x 128 complex points at 25 °C on a Varian INOVA 600 MHz spectrometer. Sixteen scans were performed for each sample with a ^{1}H sweep of 16.0 ppm and a ^{15}N sweep of 32.0 ppm. Data were processed using NMRPipe (35) and visualized using Sparky (36).

3.2.8 Low resolution solution scattering

SAXS data for SEC purified (*i*) HMGA2 were collected at 6, 7, 8 and 10 mg/mL, (*ii*) RF at 2, 3, 4, and 5 mg/mL and (*iii*) complex at 2, 3 and 4 mg/mL in 10 mM HEPES, pH 7.5, 154 mM NaCl buffer using a Rigaku 3-pinhole camera (S-MAX3000) equipped with a Rigaku MicroMax+002 microfocus sealed tube (Cu Kα radiation at 1.54 Å) and a Confocal Max-Flux (CMF) optics system operating at 40 W (Rigaku, USA). Scattering data for all three species were recorded using a 200 mm multiwire 2D detector and processed as previously described (27, 37, 38). Briefly, the buffer scattering data was subtracted from the sample scattering data using the program PRIMUS (39) followed by merging of buffer-subtracted data for all concentrations of each sample and the complex. The data were further processed using the GNOM program (40) to obtain the radius of gyration (R_g) and maximum particle dimension (D_{max}). The *ab initio* models for RF were calculated using the program DAMMIF (41) whereas the models for the complex were done using DAMMIN and MONSA (42). The quality of the models was verified by the goodness of fit parameter (χ value) for each model of RF and complex. The *ab initio* models for RF and the complex were then rotated and averaged using the program DAMAVER (43).

3.2.9 Calculation of hydrodynamic parameters from ab initio models.

The hydrodynamic properties of RF and complex were calculated for each model using the program HYDROPRO (44) as described previously (27, 37, 38). The density (1.005 g/mL) and viscosity (1.025 Poise) of buffer was calculated using SEDNTERP (32). The molecular weight and partial specific volume for HMGA2 were calculated based on amino acid sequence using SEDNTERP (32) whereas NucProt Calculator (45) was used for RF. The partial specific volume of complex was calculated using Equation 1 from Dzananovic *et.al* (38).

3.2.10 Knockdown of HMGA2

The HEK293T cell line was a gift from Dr. Thomas Klonisch. Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum as described previously (46). For siRNA transfections, cells were reverse transfected with Lipofectamine RNAiMAX and 1.5 pmoles siRNA (Life Technologies, Waltham, MA) according to the manufacturer's protocol. Transfection of a non-targeting control siRNA and siRNA targeting HMGA2 were performed simultaneously. Cell viability was monitored by an MTT assay according to a previously described protocol (47). Cell lysis and western blot analysis were performed as described previously (25) using an anti-HMGA2 antibody (Cat # 5269) (Cell Signaling Technology, Danvars, MA) and an anti-Tubulin antibody (T6074) (Sigma-Aldrich, St Louis, MO) for detection.

3.3 Results

3.3.1 HMGA2, RF and the complex can be purified to homogeneity

Codon optimized (24) HMGA2 was recombinantly expressed in *E. coli* and affinity purified by immobilised metal ion affinity chromatography (IMAC) using a hexa-histidine tag.

The tag was subsequently removed by thrombin cleavage. As a secondary purification step, HMGA2 was subjected to size exclusion chromatography (SEC). The presence of a single peak at ~15.9 mL (**Figure 3.2**) confirmed the homogeneity of HMGA2 as observed by Coomassie SDS-PAGE (**Figure 3.3**).

The RF was assembled by annealing three oligonucleotides (oSG309, oSG606, and oSG607) (see Materials and Methods for details) followed by purification using a HiLoad SEC column that yielded two distinctive peaks at ~120 mL and ~130 mL (**Figure 3.2**). The second, more prominent peak, was collected to use for further studies.

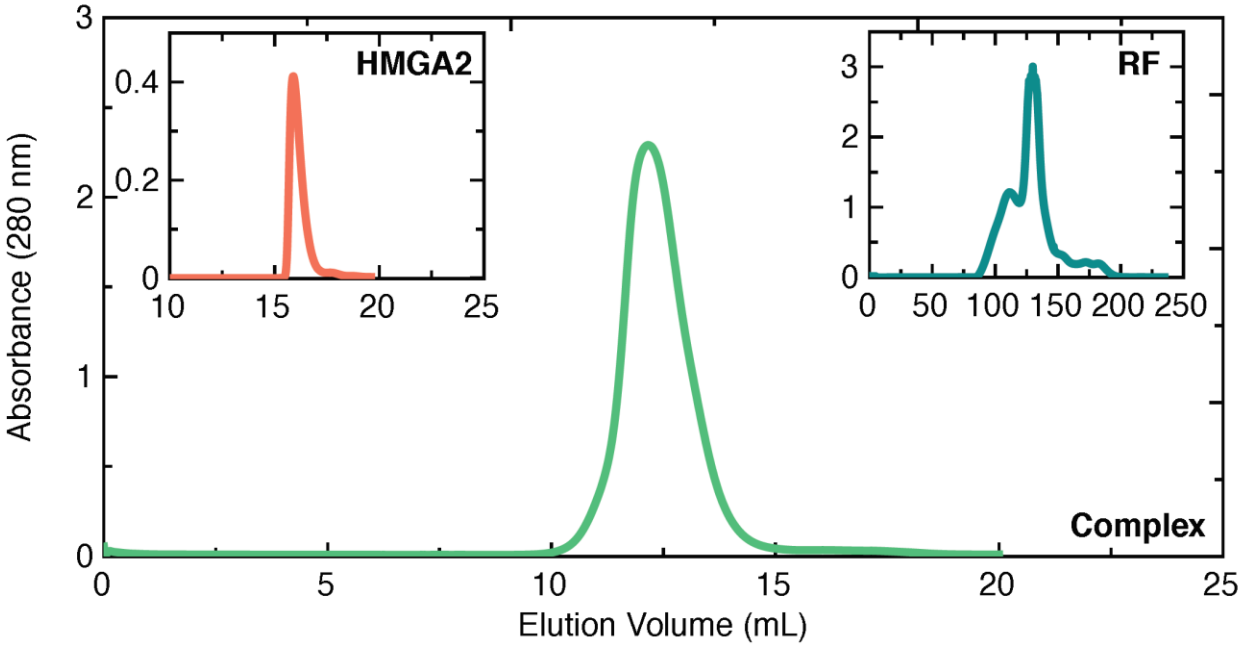

Figure 3.2 Purity of HMGA2, RF and complex

Elution profiles of HMGA2 (coral), RF (teal), and complex (sea green) from the Superdex 75 10/300 GL, HiLoad 16/600 Superdex 75, and Superdex 75 10/300 GL SEC columns, respectively, indicating purity of both individual components as well as the complex.

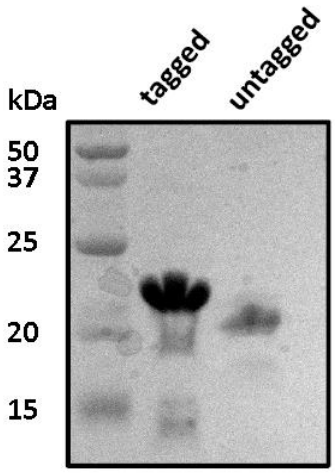

Figure 3.3 Coomassie SDS-PAGE of purified and cleaved HMGA2

Purified HMGA2 before (tagged) and after (untagged) thrombin cleavage was denatured and reduced before run on a 16% SDS-PAGE containing 6 M urea. A single band was observed indicating purity of sample.

Initial studies using electrophoretic mobility shift assays (EMSA) suggested that recombinantly-produced HMGA2 binds to the SEC purified RF (**Figure 3.4**), as we observed a shift in the RF gel position to higher molecular weight with increasing concentration of HMGA2. Based on this, we prepared the complex by incubating RF with a small excess of HMGA2 (to ensure that all DNA was in a bound state) before purifying it by SEC. The complex eluted as a single peak at ~12.5 mL with no unbound HMGA2 (**Figure 3.2**) Purified complex was used for the subsequent biophysical studies.

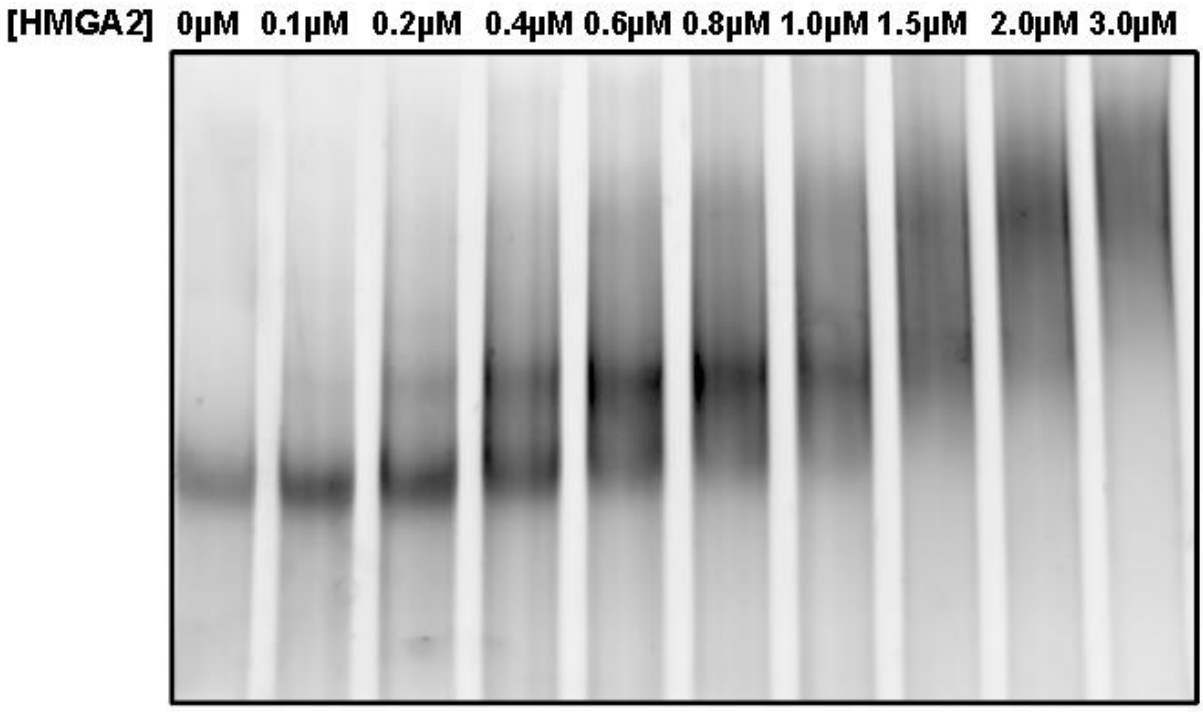

Figure 3.4 Electrophoretic mobility shift assay of RF interaction with HMGA2

A 10% PAGE gel indicates slower migration of RF (200 nM) upon interaction with HMGA2 (0 – 3 μM).

3.3.2 HMGA2 and RF are monomeric in solution and form an equimolar complex

DLS analysis of HMGA2, RF, and complex at multiple concentrations suggest that they are highly monodisperse (**Figure 3.5A**). Determination of hydrodynamic radii (R_h) for each species at multiple concentrations demonstrated the concentration independence of HMGA2 (2.58 ± 0.25 nm), RF (4.12 ± 0.24 nm), and complex (4.65 ± 0.15 nm) (**Figure 3.5B, Table 3.1**). SV analysis using analytical ultracentrifugation (AUC) confirmed a monodisperse sedimentation coefficient distribution of all macromolecules (**Figure 3.5C**). Measurements taken at concentrations between 0.1 mg/mL and 2 mg/mL displayed concentration independence, affirming homogeneity of the components and a tight interaction of HMGA2 to RF in the complex. Sedimentation coefficients corrected to standard solvent conditions (pure water at

20°C, $s_{20,w}$) were extrapolated from individual concentrations to infinite dilution yielding a Svedberg constant ($s^0_{20,w}$) of 1.12 ± 0.12 S (HMGA2), 3.86 ± 0.12 S (RF), and 4.06 ± 0.11 S (complex). The modest increase in R_h and $s^0_{20,w}$ values for RF upon incubation with HMGA2 indicated the formation of a protein/DNA complex. We used the Svedberg equation, that combines the R_h (DLS) and $s^0_{20,w}$ (SV) to calculate the molecular weight (M_w) for each species in solution (28, 33). We found that the M_w of HMGA2 (11.80 kDa) and the RF (43.03 kDa) agrees with the sequence based M_w (11.70 kDa and 41.40 kDa, respectively), implying that both species are monomeric in solution. Furthermore, the obtained M_w of 55.47 kDa for HMGA2-RF complex demonstrated that the molecules interact in a 1:1 stoichiometry (**Table 3.1**).

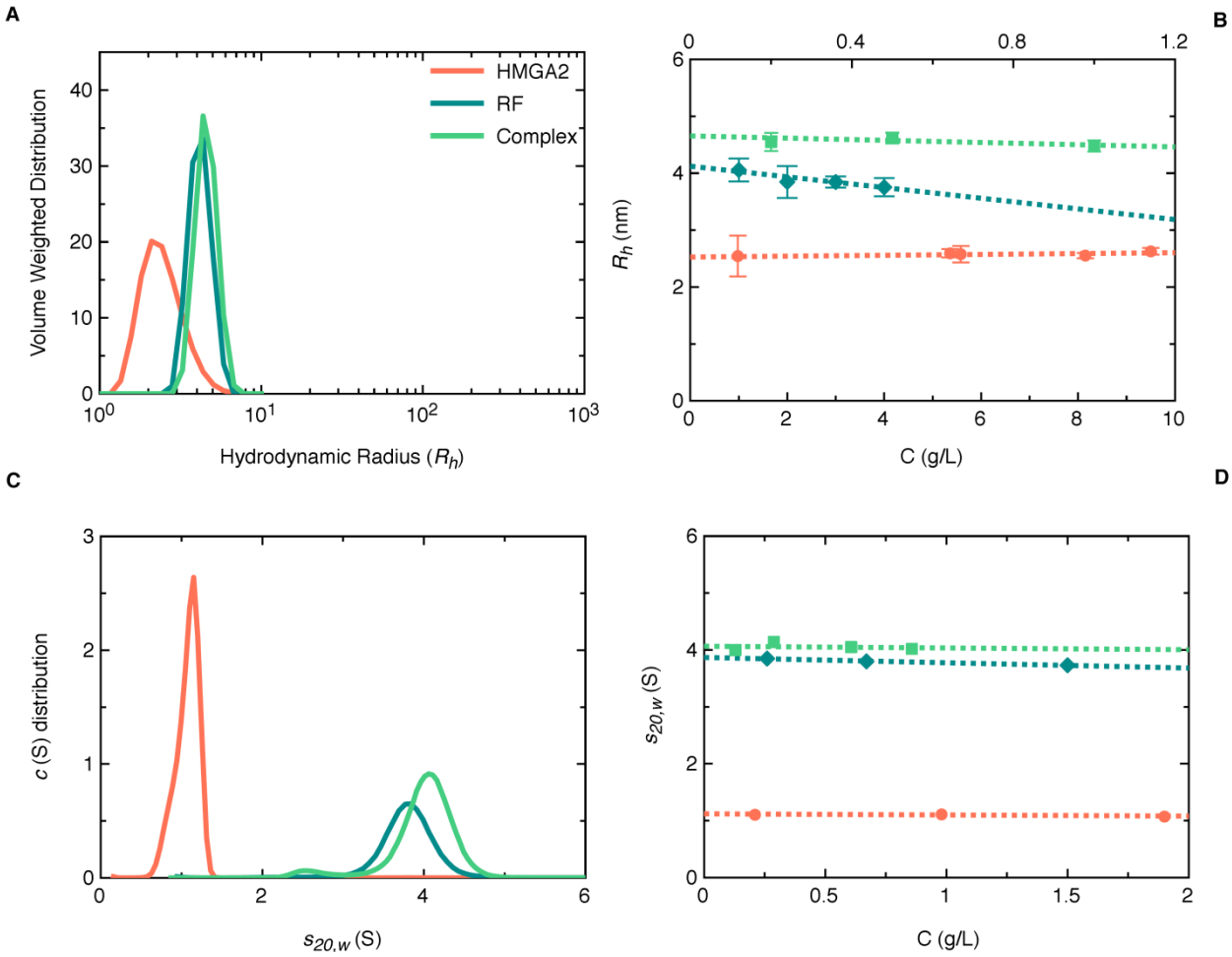

Figure 3.5 Monodisperse hydrodynamic radius and sedimentation coefficient distributions of HMGA2, RF, and complex

(A) Hydrodynamic radius (R_h) distribution and (C) sedimentation coefficient ($s_{20,w}$) distribution of HMGA2 (coral), RF (teal), and complex (sea green) at 1 mg/mL suggest monodisperse samples. (B) The R_h for each sample was plotted at multiple concentrations (1 – 10 mg/mL) with vertical error bars indicative of the standard deviation from five measurements. (D) The Svedberg coefficient in water ($s_{20,w}$) was plotted at multiple concentrations (0.2 – 2 mg/mL) for each sample with vertical error bars indicative of the standard error from multiple scans. Concentration independence was observed for all samples under the experimental conditions.

Table 3.1 Summary of Hydrodynamic Parameters for HMGA2, RF, and complex

Parameters	**HMGA2**	**RF**		**HMGA2 – RF complex**		
		Exp	HYDROPRO	Exp	HYDROPRO	HYDROPRO MONSA
R_h (nm)[a]	2.58 ± 0.25	4.12 ± 0.24	4.06 ± 0.05	4.65 ± 0.15	4.57 ± 0.04	4.37 ± 0.10
R_g (nm)[b]	2.45 ± 0.06	4.54 ± 0.14	4.51 ± 0.12	4.85 ± 0.06	4.88 ± 0.07	4.77 ± 0.01
D_{max} (nm)[b]	8.0	15.4	15.7 ± 0.27	15.0	15.3 ± 0.12	15.0 ± 0.20
$s^0_{20,w}$ (S)[c]	1.12 ± 0.12	3.86 ± 0.12	4.03 ± 0.05	4.06 ± 0.11	4.11 ± 0.05	4.31 ± 0.10
M_w (kDa)[d]	11.70	41.40	-	53.10	-	-
M_w (kDa) [e]	11.80	43.03	-	55.47	-	-
Chi[f]		1.2[h]		1.0[h] 1.2, 1.2, 1[i]		
NSD[g]		0.73 ± 0.08[h]		0.83 ± 0.03[h] 0.87 ± 0.06[i]		

[a] experimentally measured using DLS with errors obtained from linear regression analysis
[b] experimentally measured using GNOM analysis from SAXS data
[c] experimentally measured using AUC
[d] calculated from amino acid/nucleotide sequence
[e] calculated from R_h and $s^0_{20,w}$
[f] goodness of fit parameter suggesting agreement between raw data and data back calculated from *ab initio* model
[g] normalized spatial discrepancy indicating agreement between individual models
[h] data from DAMMIN
[i] data from MONSA

3.3.3 HMGA2 is disordered in solution

Based on the amino acid sequence of HMGA2, bioinformatics tools (PrDOS (48), PSIPRED (49, 50) and RONN (51)) predicted a disordered secondary structure (**Figure 3.6**).

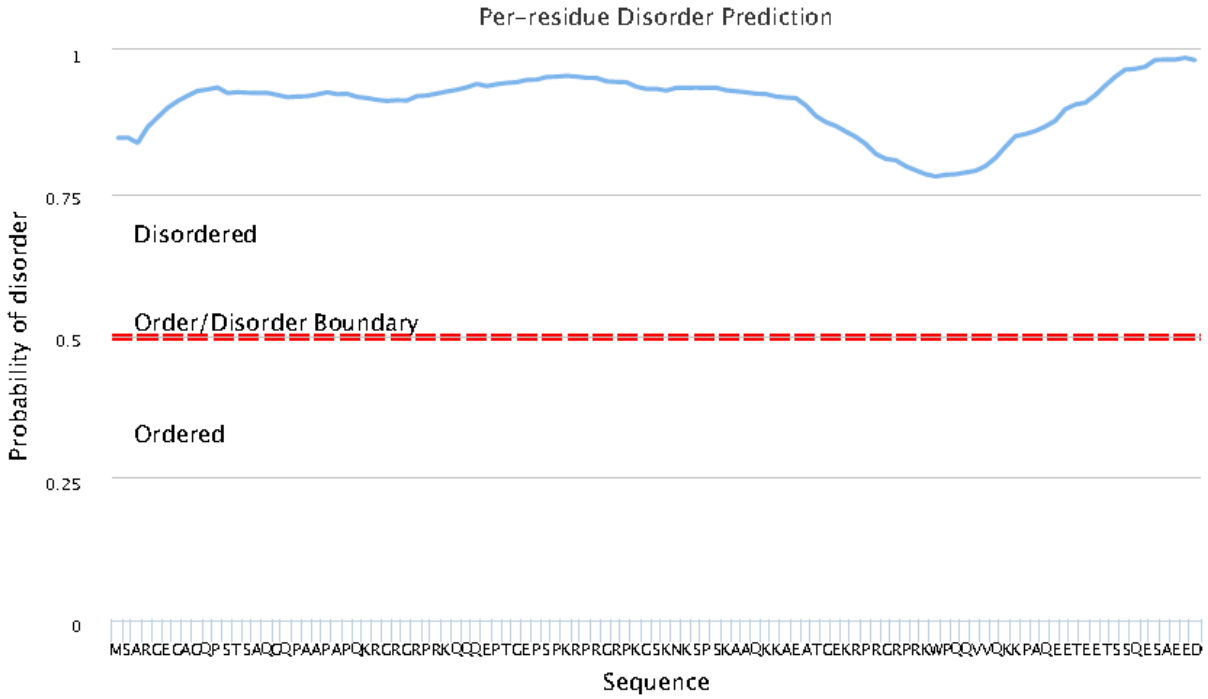

Figure 3.6 HMGA2 sequence prediction

The entire sequence of HMGA2 is predicted to be disordered according to the program RONN (51).

To experimentally validate the disordered nature of HMGA2, we employed three separate techniques; Circular Dichroism (CD), Nuclear Magnetic Resonance (NMR) heteronuclear single quantum coherence (HSQC) spectroscopy, and Small Angle X-ray Scattering (SAXS). Far-UV CD spectropolarimetry of HMGA2 produced a spectrum with a minimum at 199 nm followed by a flat line from 210 – 250 nm (**Figure 3.7**). This spectrum appeared identical to that of a disordered protein (52). To investigate the secondary structure elements by NMR spectroscopy, HMGA2 was labelled with ^{15}N and an HSQC spectrum was acquired. The HSQC spectrum of

HMGA2 (**Figure 3.8**) displayed a very narrow chemical shift range in the proton dimension which is highly characteristic of a disordered protein.

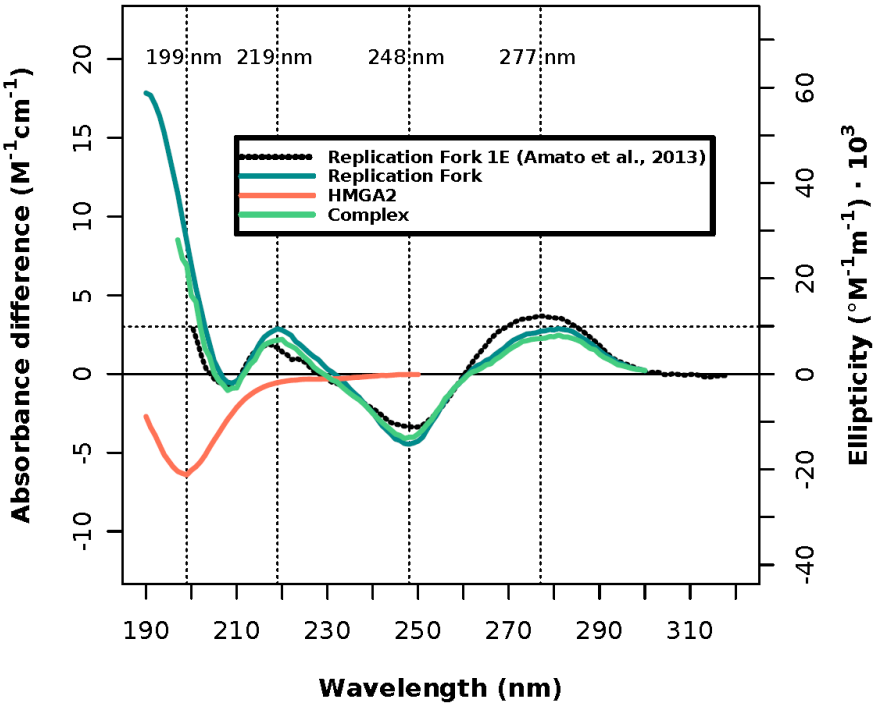

Figure 3.7 Secondary structure analysis of HMGA2, RF, and complex

Far-UV CD spectra of HMGA2, RF, and complex displayed as an average of three measurements indicating disordered HMGA2 and confirmation of RF formation. A reference RF (dotted black) spectrum (53) was overlaid onto the experimental RF (teal) for comparison.

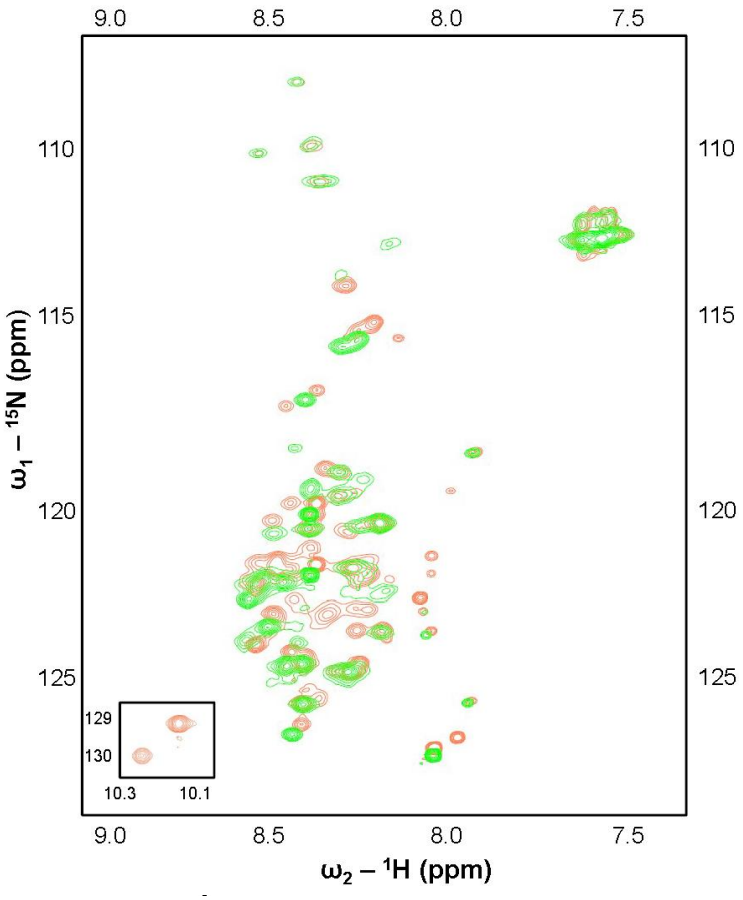

Figure 3.8 Secondary structure analysis of bound and unbound HMGA2

^{15}N HSQC spectral overlay of free HMGA2 (coral) with HMGA2 in complex with RF (green) suggesting HMGA2 retains flexibility upon interaction with RF.

Finally, solution scattering of HMGA2 provided low-resolution information on the protein behaviour at multiple concentrations. Scattering curves for each concentration were buffer subtracted and merged to generate the SAXS plots (**Figure 3.9A-C**). In general, the Kratky analysis of SAXS data for proteins provided qualitative information on the overall shape of the biomolecule of interest (54). The ascending linear curve (coral) observed for HMGA2 in the Porod-Debye Plateau region was indicative of a disordered protein (54) (**Figure 3.9B**) and therefore agreed with results of the CD and NMR analyses.

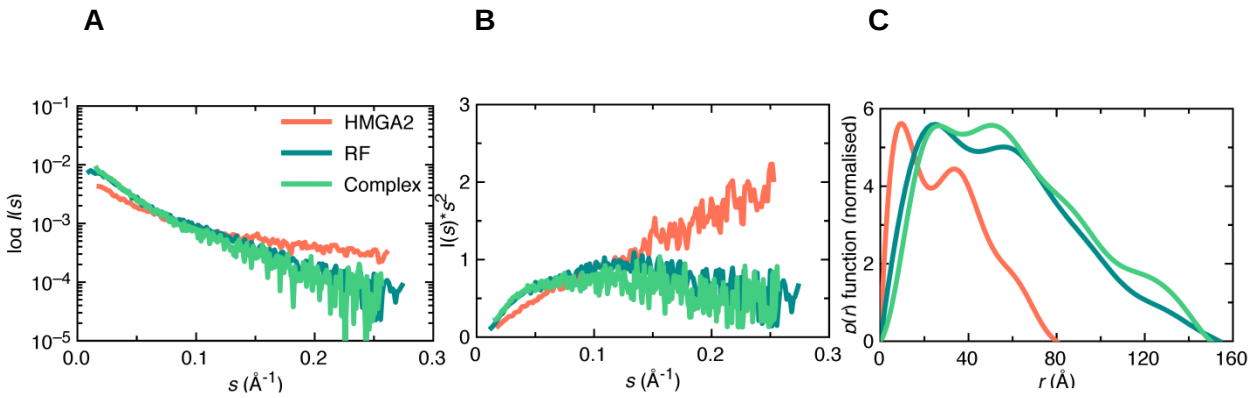

Figure 3.9 SAXS plots of HMGA2, RF, and complex merged from multiple concentrations

(A) Merged SAXS data of HMGA2 (coral), RF (teal), and complex (sea green) obtained from multiple concentrations. The corresponding (B) Kratky plot suggests that HMGA2 (coral) is a disordered protein and upon interaction with RF undergoes a conformational change. (C) *P(r)* function of HMGA2 (coral), RF (teal) and complex (sea green) outlining the multi-domain nature of the molecules and their D_{max}.

3.3.4 CD spectropolarimetry confirms assembly of RF

To validate the assembly of RF by annealing of three DNA oligonucleotides, far-UV CD analysis was performed (**Figure 3.7**). The CD spectrum of the RF had extrema at positions which correlated with that of B-DNA; two distinct maxima at 219 nm and near 280 nm and a minimum at 248 nm (55) (**Figure 3.7**). A properly assembled RF can be distinguished from B-DNA by a depressed peak magnitude at 219 and 248 nm. We have overlaid our spectrum with a spectrum of RF 1E published by Amato *et. al* for comparison (53). Our peak heights correspond to those of their RF, indicating that our RF assembly was successful.

3.3.5 HMGA2 retains flexibility when in complex with RF

CD spectropolarimetry, NMR, and SAXS studies confirmed that HMGA2 is disordered in solution. Furthermore, SEC, DLS and SV experiments have demonstrated that the individual species as well as the complex between them are highly pure. Our next goal was to study changes in secondary structure of HMGA2 upon interaction with RF. We first analysed the

complex using CD spectropolarimetry which is very sensitive to changes in the secondary structure. The CD signal of nucleic acids originates from the glycosidic bond angles. The spectrum below 230 nm contains contributions from DNA and protein, whereas above 230 nm the protein signal is zero. We have therefore ignored the protein contribution when we normalized the spectra, so the spectrum of the complex can be directly compared to the spectrum of the RF in the range below 230 nm. The CD spectra of our RF in the presence and absence of HMGA2 are identical within experimental uncertainty, indicating that the RF maintains its conformation when the protein is bound. Next, we studied the complex through HSQC NMR using ^{15}N-labelled HMGA2 bound to unlabelled RF. The HSQC spectrum of bound HMGA2 (green) is superimposed on the spectrum of unbound HMGA2 (coral) in **Figure 3.8**. A 70-75% decrease in peak intensity was observed upon complex formation. This is evidence that HMGA2 is in a bound state with the addition of the RF resulting in slower tumbling. Interestingly, the chemical shift range in the complex is nearly identical to that observed for the free protein suggesting that HMGA2 is disordered even when bound to the RF. One noticeable difference between the two spectra was the disappearance of the tryptophan (W83) side chain peak in the bound form (bottom left inset to **Figure 3.8**). This infers that W83 is near or part of a binding site in the interaction with RF. Finally, Kratky analysis of SAXS data for the complex displays a curved shape (sea green) that approaches the baseline at high *s*, which is indicative of a macromolecule with only partial flexibility. This is opposed to an ascending linear curve in the Porod-Debye Plateau region for unbound HMGA2 (coral) that suggests a fully disordered protein (54) (**Figure 3.9B**).

3.3.6 HMGA2 interacts with RF in a side-by-side manner

SAXS data for HMGA2, RF, and the complex were collected at multiple concentrations, buffer subtracted, and merged to study their size and shape in solution (**Figure 3.9A and C**). As mentioned, the Kratky analysis of SAXS data for HMGA2 suggests that it is a disordered protein (**Figure 3.9B**), therefore we only studied the radius of gyration (R_g = 2.45 ± 0.06 nm) and the maximum particle dimension (D_{max} = 8 nm) using the pair-distance distribution function ($P(r)$, **Figure 3.9C**), without performing low-resolution shape analysis. The *P(r)* plot for HMGA2 (coral) has three defined peaks suggesting that the protein is composed of three domains. A similar observation can be made for the RF, in which multiple peaks are present in the *P(r)* plot (teal). This implies that the RF is a multi-domain macromolecule with R_g of 4.54 ± 0.14 nm and D_{max} of 15.4 nm (**Figure 3.9C**). The low-resolution structural studies performed based on the SAXS data (**Figure 3.9A and C**) demonstrate that the RF adopts an extended and curved structure in solution (**Figure 3.10A**). The goodness of fit parameter (χ value) of 1.2 indicates an agreement between the experimentally collected SAXS data and the back-calculated scattering curve derived from individual models. Furthermore, the normalized spatial discrepancy (NSD) of 0.73 ± 0.08 indicates that individual models calculated for RF based on the SAXS data agree well with each other (**Table 3.1**). In addition, the X-ray diffraction crystal structure of a shorter RF (14.1 kDa, PDB: 5D9I) (56) fits well with the low-resolution structure of our RF (41.4 kDa) (**Figure 3.10A**).

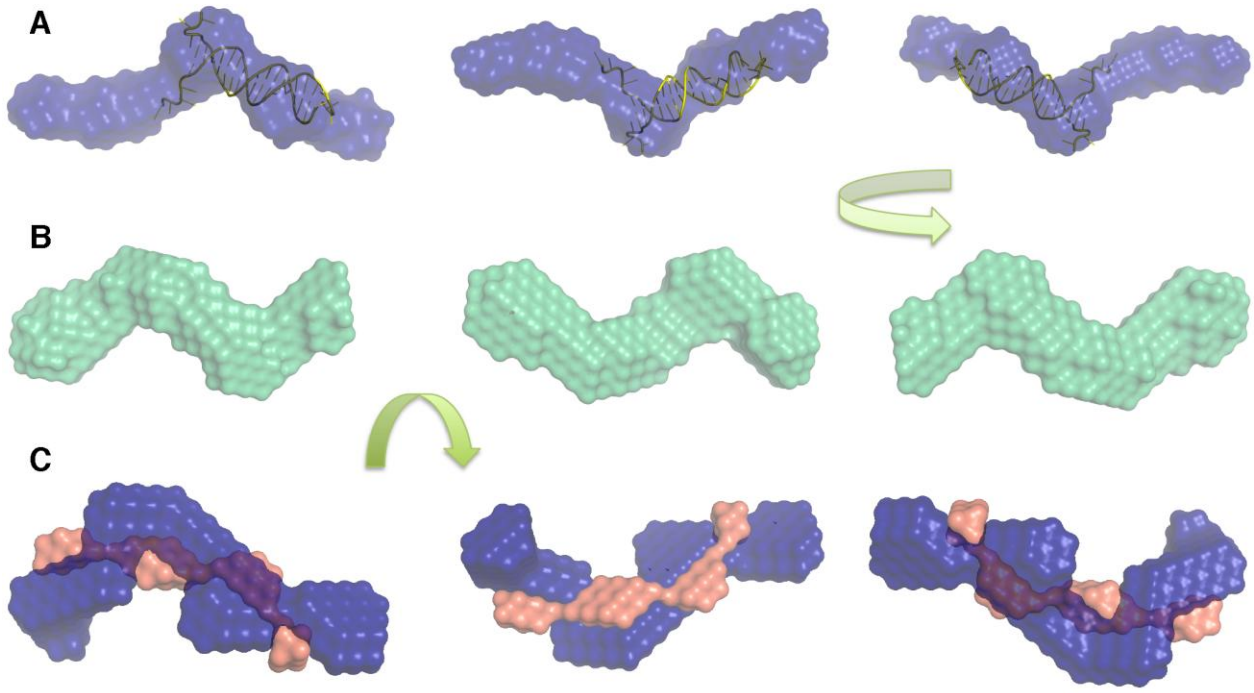

Figure 3.10 Low resolution models of RF and complex suggest RF interacts with HMGA2 in a side-by-side orientation

Low resolution models of (A) RF (extended shape) superimposed with a high resolution RF structure (PDB: 5D9I) (56) and (B) complex determined by DAMMIN. (C) Modelling by MONSA distinguishes HMGA2 (coral) from RF (blue) in the complex and suggests that HMGA2 interacts along the side of the RF.

The *P(r)* function calculated from the SAXS data of the complex differs from that of the RF (**Figure 3.9B**). The merged SAXS data for the complex were processed using the *P(r)* analysis to obtain the R_g of 4.85 ± 0.06 and D_{max} of 15 nm (**Figure 3.9C**). The increase, albeit relatively small, in the R_g for the complex compared to that of RF (4.85 ± 0.06 *vs*. 4.54 ± 0.14 nm) is indicative of complex formation between HMGA2 and RF. Low-resolution shape reconstruction using the DAMMIN program also suggested an extended and curved structure for the complex (**Figure 3.10B**) as with the unbound RF. However, as reflected by the shape of the *P(r)* plot, a distinguishably different pattern of curved domains was observed compared to the low-resolution shape of the RF (**Figure 3.10B**). The χ value of 1.0 and the NSD value of 0.83 ±

0.03 indicated that individual models for the complex agree well with the SAXS data and that the models are in good agreement with each other (**Table 3.1**).

Having established that HMGA2 and RF form a complex, we employed the program MONSA to deconvolute the location and orientation of HMGA2 and RF in the complex. MONSA is a multiphase bead modelling program which is used to fit multiple curves simultaneously (42). Using the MONSA program, we could observe that HMGA2 wraps around the RF and that HMGA2 has three distinguished AT-hook structures (**Figure 3.10C**). The χ value of 1.2, 1.2, and 1 for HMGA2, RF and complex, respectively, and the NSD value of 0.87 ± 0.06 indicate that the individual models agree well with the SAXS data and the models also agree with each other (**Table 3.1**). In summary, the low-resolution structural studies suggest that HMGA2 stretches across the RF to interact in a side-by-side orientation.

3.3.7 HMGA2 plays a role in cell proliferation

To investigate the effect that HMGA2 has on cell proliferation of HEK293T cells and compare to similar studies in cancer cell lines, the expression of HMGA2 was knocked down using siRNA transfection against the mRNA of HMGA2 and confirmed using a western blot analysis (**Figure 3.11A**). At 48 and 72 hrs post transfection, a significant reduction in viability ($p < 0.05$) was observed when compared to the HEK293T cells transfected with a control siRNA. However, after 96 hrs, the viability of both transfected cell lines was found to be within error of the untreated cells (**Figure 3.11B**).

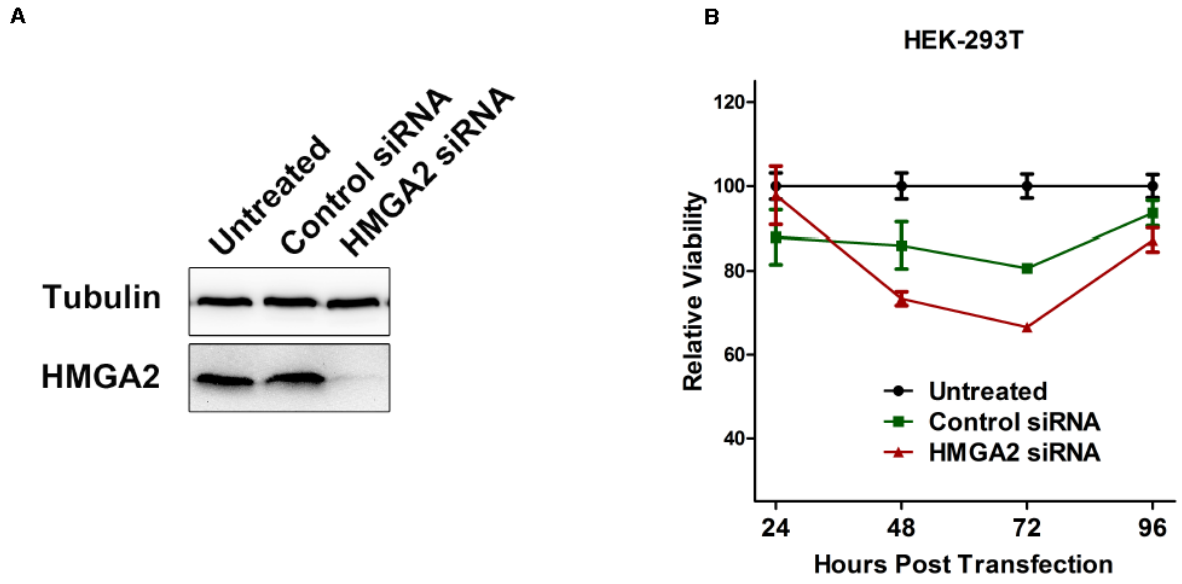

Figure 3.11 Knockdown of HMGA2 suggests role in cellular proliferation

(A) Western blot of HMGA2 knockdown at 48 hrs post transfection using an anti-Tubulin (control) and anti-HMGA2 antibody.
(B) The effect of HMGA2 knockdown on cell viability was monitored for HEK293T cells at 24 hr intervals over 96 hrs. A significant loss in viability was observed upon transfection with the HMGA2 siRNA. Data points are shown as an average of eight measurements with error bars corresponding to the standard deviation.

3.4 Discussion

During the process of DNA replication, various stress inducers can cause the leading strand of a RF to stall the replisome which, unless rescued, can cause its collapse and eventual cell death (57). In ESCs as well as cancer cells, HMGA2 plays a significant role in stabilizing RFs (21). This interaction is essential to maintain a high cell proliferation rate for ESCs, but becomes a major concern when it occurs in somatic cells. In addition to stabilizing RFs, the DNA binding capability of HMGA2 has a significant role in other biological processes such as chromatin remodeling, DNA repair pathways, and regulating target genes (1, 2, 58). While the importance of HMGA proteins in several biological processes is well understood, their solution behaviour and interaction with RFs and DNA in general are largely unexplored. Previous structural studies using X-ray crystallography and NMR spectroscopy have investigated truncated versions of HMGA1 and their interaction with dsDNA. This research concluded that the AT-hook domains on HMGA1 are disordered in the apo-state but adopt a defined structure upon binding to the minor groove (59, 60). Furthermore, solution studies with mouse HMGA2 also determined that, similar to HMGA1, the protein is disordered but presented additional findings that in solution HMGA2 is present in a dimeric state (23). The goal in our research was to further the current binding studies performed with dsDNA by investigating RF interactions with human HMGA2. This interaction has biological significance in growth and development through ESCs (11) but also plays a large role in tumour progression (12–14). Elucidating the binding interaction of HMGA2 with RF will provide insight into the mechanism of RF stabilization for a potential target in cancer diagnosis or treatment.

Our biophysical studies revealed that human HMGA2 is monomeric in solution (**Table 3.1**) which is contrary to that of a previous study with mouse HMGA2 (23). This can be explained by the presence of a cysteine residue in the murine amino acid sequence which forms a homodimer through a disulphide linkage. As for HMGA2 in humans as well as for many other species, there is no cysteine to form these bridges (61) and therefore the protein exists in the observed monomer up to concentrations of 10 mg/mL (**Figure 3.5, Table 3.1**). The absence of this disulphide bridge between AT-hook domains 1 and 2 would allow HMGA2 to have more flexibility in binding to DNA that would otherwise be restricted if in a homodimeric state. It has been previously documented that telomerase activity is higher in mice compared to humans (62) and recently discovered that HMGA2 plays a role in telomere stability (63). As a result, it is possible that in order for HMGA2 to be successful in stabilizing telomeres in an environment with higher telomerase activity, HMGA2 must be present as a dimer. Alternatively, the disulphide bridge could be required in binding of mouse HMGA2 to one of its protein binding partners which is not present in humans. The underlying reason for this difference is unclear, however it does suggest that there is a functional significance that is yet to be revealed.

The homogeneity and monomeric state of HMGA2 and RF was confirmed using DLS and SV studies (**Figure 3.5**) where the experimentally calculated M_w of both species agree with those based on primary sequences. Furthermore, the formation of complex and its purity was evident from the SEC, DLS and SV experiments (**Figure 3.5**). The M_w of 55.47 kDa for the complex agrees with the sequence-determined M_w (53.10 kDa) and unequivocally concludes formation of a 1:1 complex. This is contrary to what was observed for HMGA1 binding to dsDNA where a truncated version of HMGA1 (containing two AT-hook domains) was shown to bind to two dsDNA molecules (59). A decade later, Cui and Leng determined that each AT-hook domain of

HMGA2 recognizes a specific type of 15 bp sequence; first five bp are AT-rich, the middle four or five are GC-rich and the last five or six are AT-rich (64). Since Huth *et. al* used an 18-mer DNA segment in their binding studies (59), it follows that only one of the two AT-hook domains was capable of binding to it, leaving the second one free to bind an additional 18-mer segment of DNA. We designed our RF such that it was large enough to accommodate binding of all three AT-hook domains of HMGA2, and a closer representative to what occurs *in vivo* during replication. As a result, HMGA2 indeed binds to RF in a 1:1 stoichiometric ratio, suggesting that all three AT-hook domains are involved in stabilizing the RF.

The disordered nature of HMGA2 predicted by bioinformatics analysis of the sequence was confirmed by CD, NMR, and SAXS experiments (**Figure 3.7 and Figure** 3.8). HMGA2 is composed of three AT-hook DNA-binding domains situated in the middle of the protein sequence and attached together by linkers (10-20 amino acids in length, **Figure 3.1**). These linkers are flexible in nature and do not adopt any distinct structure. Interestingly, previous NMR studies suggest that in the apo state, the AT-hook domains themselves are also unstructured (59, 64). Furthermore, the remaining ~50% of the protein that could potentially be structured is split up into the N- and C-termini with each containing ~25 amino acids. Structural predictions deem these residues to have a high probability of disorder, which supports the experimental evidence on unbound HMGA2 (**Figure 3.6, Figure** 3.7**, Figure** 3.8**, and Figure** 3.9**B**). It is this disordered and flexible nature of HMGA2 that facilitates its diverse functionality. In addition to the AT-hook domains being used in DNA binding to alter chromatin structure (6) and stabilize stalled and reversed RFs (21), HMGA2 has also been found to have protein binding partners such as telomeric repeat-binding factor 2 (TRF2) which play a role in telomere dysfunction (63). The composition of the linkers between the AT-hook domains of HMGA2 is

critical for its interaction with such protein binding partners. For example, mutations in the AT-hook domains of HMGA2 only reduce its binding, but mutations in the linker regions completely abolishes interaction of HMGA2 with TRF2 (63). This wide range of activities and binding partners requires HMGA2 to be flexible in solution such that it can adapt to the various binding sites present on the same or different molecules (6). Moreover, this enhanced flexibility is inherent in its role in chromatin restructuring and formation of "enhanceosomes" for regulation of gene expression (19).

We have shown the stable formation of a complex between recombinantly produced HMGA2 and a RF formed through annealing DNA. In DLS and SV we observed an increase in the R_h (4.12 nm to 4.65 nm) and $s^0_{20,w}$ (3.86 S to 4.06 S) of the RF, respectively (**Table 3.1**). Through CD, we are able to detect that there is no significant structural changes to the RF upon binding to HMGA2 (**Figure 3.7**). Although NMR (59) and X-ray crystallography (60) studies reveal that the minor groove widens and kinks are introduced due to AT-hook binding, such changes would not be visible by CD. Additionally, since the CD spectrum of the protein is convoluted with that of the DNA (65), we are unable to directly detect any secondary structural changes of the protein in the complex (**Figure 3.7**). However, through NMR we were able to detect the protein in the complex and determine that there was no significant secondary structure changes of HMGA2 upon binding to RF. The resulting HSQC showed a nearly identical chemical shift range to unbound HMGA2 (**Figure 3.8**) with decreased peak intensity. Atomic level data suggests that amino acids in the AT-hook domain adopt a defined structure when bound (59), but the remaining residues of HMGA2 are still largely in a disordered state. Since it is widely known that AT-hooks are involved in binding to DNA (18), it is inherent that W83 is

adjacent to an AT-hook. The disappearance of the tryptophan side chain peak in the complex confirms that AT-hook 2 is involved in RF binding.

SAXS studies provided low-resolution structural information on the interaction of HMGA2 and RF (**Figure 3.9**). This low-resolution structure shows that HMGA2 interacts with the RF along nearly the entire molecule in a side-by-side orientation. It is known that AT-hooks penetrate deep into the minor groove, with flanking residues forming electrostatic and hydrophobic interactions with the floor of the minor groove (59). Therefore, one could suggest that upon deconvolution, each AT-hook of HMGA2 is represented by a lobe as it interacts with the various AT-rich regions of the RF (**Figure 3.9C**). With HMGA2 stretched across the RF it suggests that HMGA2 is not localized at the replicating end of the RF, but rather uses all AT-hook domains to bind to both sides of the replication bubble acting as a "clamp" until replication can continue. Although it has been shown that a single AT-hook can bind to DNA (59), our model supports previous mutational data which demonstrates that all three AT-hook domains need to be present in order to be successful in stabilizing the RF (21).

The important role that HMGA2 has in stabilizing RFs, it would follow that knockdown of HMGA2 in cells would have a significant effect on their cell viability. This was evidenced for HEK293T cells which is consistent with previous studies on prostate cancer PC3 cells (66), nasopharyngeal cancer cells (67), and breast adenocarcinoma (MDA-MB-468) cells (68) in which all observed a decrease in cell viability upon HMGA2 knockdown. In our studies, after 96 hrs it was found that the cell viability of both transfected cell lines recovered as a result of the control samples reaching confluence. As cells reach confluence metabolism and growth rates are altered it affects the results of the MTT assay (47). However, it is also plausible that over time the cells adapt and activate other cellular pathways to compensate for the lack of HMGA2.

Through our extensive biophysical approach to elucidate the interaction between HMGA2 and RF, we have shed light on the mechanism of this biologically relevant process. CD, NMR, and SAXS analysis determined that HMGA2 is inherently disordered which provides it with the flexibility necessary to bind to distorted DNA. Combined DLS and SV data confirm that a single HMGA2 binds to one RF to form a monodisperse complex. Ab-initio modelling of SAXS data concludes that all three AT-hook domains of HMGA2 interact with RF in a side-by-side orientation binding on both sides of the point of replication. It is this "clamped" orientation that indeed allows HMGA2 to stabilize stalled RFs *in vivo* and provides ESCs and cancer cells the added RF protection that is absent in somatic cells. This added protection is what maintains the cellular viability for increased proliferation.

References

1. Reeves, R. (2015) High mobility group (HMG) proteins: Modulators of chromatin structure and DNA repair in mammalian cells. *DNA Repair (Amst).* **36**, 122–136
2. Reeves, R., and Adair, J. E. (2005) Role of high mobility group (HMG) chromatin proteins in DNA repair. *DNA Repair (Amst).* **4**, 926–938
3. Reeves, R., and Nissen, M. S. (1990) The AT-DNA-binding domain of mammalian high mobility group I chromosomal proteins. *J. Biol. Chem.* **265**, 8573–82
4. Natarajan, S., Hombach-Klonisch, S., Dröge, P., and Klonisch, T. (2013) HMGA2 Inhibits Apoptosis through Interaction with ATR-CHK1 Signaling Complex in Human Cancer Cells. *Neoplasia*. **15**, 263–280
5. Copley, M. R., Babovic, S., Benz, C., Knapp, D. J. H. F., Beer, P. a, Kent, D. G., Wohrer, S., Treloar, D. Q., Day, C., Rowe, K., Mader, H., Kuchenbauer, F., Humphries, R. K., and Eaves, C. J. (2013) The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nat. Cell Biol.* **15**, 916–25
6. Pfannkuche, K., Summer, H., Li, O., Hescheler, J., and Dröge, P. (2009) The high mobility group protein HMGA2: a co-regulator of chromatin structure and pluripotency in stem cells? *Stem Cell Rev.* **5**, 224–30
7. Ozturk, N., Singh, I., Mehta, A., Braun, T., and Barreto, G. (2014) HMGA proteins as modulators of chromatin structure during transcriptional activation. *Front. Cell Dev. Biol.* **2**, 1–9
8. Wu, J., and Wei, J.-J. (2013) HMGA2 and high-grade serous ovarian carcinoma. *J. Mol. Med.* **91**, 1155–65
9. Fedele, M., Fidanza, V., Battista, S., Pentimalli, F., Klein-Szanto, A., Visone, R., De

Martino, I., Curcio, A., Morisco, C., Del Vecchio, L., Baldassarre, G., Arra, C., Viglietto, G., Indolfi, C., Croce, C. M., and Fusco, A. (2006) Haploinsufficiency of the Hmga1 gene causes cardiac hypertrophy and myelo-lymphoproliferative disorders in mice. *Cancer Res.* **66**, 2536–43

10. Foti, D., Chiefari, E., Fedele, M., Juliano, R., Brunetti, L., Paonessa, F., Manfioletti, G., Barbetti, F., Brunetti, A., Croce, C., Fusco, A., and Brunetti, A. (2005) Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. *Nat. Med.* **11**, 765–73
11. Papaioannou, G., Inloes, J. B., Nakamura, Y., Paltrinieri, E., and Kobayashi, T. (2013) let-7 and miR-140 microRNAs coordinately regulate skeletal development. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E3291-300
12. Jun, K., Jung, J., Choi, H., Shin, E., and Chin, H. (2015) HMGA1/HMGA2 protein expression and prognostic implications in gastric cancer. *Int. J. Surg.* **24**, 39–44
13. Pallante, P., Sepe, R., Puca, F., and Fusco, A. (2015) High mobility group A proteins as tumor markers. *Front. Med.* **2**, 15
14. Lee, Y. S., and Dutta, A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev.* **21**, 1025–1030
15. Kloth, L., Gottlieb, A., Helmke, B., Wosniok, W., Löning, T., Burchardt, K., Belge, G., Günther, K., and Bullerdiek, J. (2015) HMGA2 expression distinguishes between different types of postpubertal testicular germ cell tumour. *J. Pathol. Clin. Res.* **1**, 239–51
16. Wu, J., Zhang, S., Shan, J., Hu, Z., Liu, X., Chen, L., Ren, X., Yao, L., Sheng, H., Li, L., Ann, D., Wang, J., and Wang, X. (2016) Elevated HMGA2 expression is associated with cancer aggressiveness and predicts poor outcome in breast cancer. *Cancer Lett.* **376**, 284–

92

17. Reeves, R. (2001) Molecular biology of HMGA proteins: hubs of nuclear function. *Gene*. **277**, 63–81

18. Aravind, L., and Landsman, D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.* **26**, 4413–4421

19. Cattaruzzi, G., Altamura, S., Tessari, M. A., Rustighi, A., Giancotti, V., Pucillo, C., and Manfioletti, G. (2007) The second AT-hook of the architectural transcription factor HMGA2 is determinant for nuclear localization and function. *Nucleic Acids Res.* **35**, 1751–1760

20. Summer, H., Li, O., Bao, Q., Zhan, L., Peter, S., Sathiyanathan, P., Henderson, D., Klonisch, T., Goodman, S. D., and Dröge, P. (2009) HMGA2 exhibits dRP/AP site cleavage activity and protects cancer cells from DNA-damage-induced cytotoxicity during chemotherapy. *Nucleic Acids Res.* **37**, 4371–84

21. Yu, H., Lim, H. H., Tjokro, N. O., Sathiyanathan, P., Natarajan, S., Chew, T. W., Klonisch, T., Goodman, S. D., Surana, U., and Dröge, P. (2014) Chaperoning HMGA2 protein protects stalled replication forks in stem and cancer cells. *Cell Rep.* **6**, 684–97

22. Branzei, D., and Foiani, M. (2010) Maintaining genome stability at the replication fork. *Nat. Rev. Mol. Cell Biol.* **11**, 208–19

23. Frost, L., Baez, M. a. M., Harrilal, C., Garabedian, A., Fernandez-Lima, F., and Leng, F. (2015) The Dimerization State of the Mammalian High Mobility Group Protein AT-Hook 2 (HMGA2). *PLoS One*. **10**, 1–18

24. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. **16**, 276–7

25. Booy, E. P., Meier, M., Okun, N., Novakowski, S. K., Xiong, S., Stetefeld, J., and McKenna, S. A. (2012) The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.* **40**, 4110–4124

26. Patel, T. R., Nikodemus, D., Besong, T. M. D., Reuten, R., Meier, M., Harding, S. E., Winzor, D. J., Koch, M., and Stetefeld, J. (2015) Biophysical analysis of a lethal laminin alpha-1 mutation reveals altered self-interaction. *Matrix Biol.* **49**, 1–13

27. Meier, M., Patel, T. R., Booy, E. P., Marushchak, O., Okun, N., Deo, S., Howard, R., McEleney, K., Harding, S. E., Stetefeld, J., and McKenna, S. A. (2013) Binding of G-quadruplexes to the N-terminal recognition domain of the RNA helicase associated with AU-rich Element (RHAU). *J. Biol. Chem.* **288**, 35014–35027

28. Stetefeld, J., Mckenna, S. A., and Patel, T. R. (2016) Dynamic light scattering : a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8**, 409–427

29. Krahn, N., Spearman, M., Meier, M., Dorion-Thibaudeau, J., McDougall, M., Patel, T. R., De Crescenzo, G., Durocher, Y., Stetefeld, J., and Butler, M. (2017) Inhibition of glycosylation on a camelid antibody uniquely affects its FcγRI binding activity. *Eur. J. Pharm. Sci.* **96**, 428–439

30. Schuck, P. (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.* **75**, 1503–1512

31. Dam, J., and Schuck, P. (2004) Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. *Method Enzymol.* **384**, 185–212

32. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided

interpretation of analytical sedimentation data for proteins. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, A. J. R. S. E., and Horton, J. C. eds), pp. 90–125, Royal Society of Chemistry, Cambridge, United Kingdom

33. Patel, T. R., Winzor, D. J., and Scott, D. J. (2016) Analytical ultracentrifugation : A versatile tool for the characterisation of macromolecular complexes in solution. *Methods*. **95**, 55–61

34. Kay, L., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–5

35. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*. **6**, 277–93

36. Goddard, T. D., and Kneller, D. G. SPARKY 3

37. Deo, S., Patel, T. R., Dzananovic, E., Booy, E. P., Zeid, K., McEleney, K., Harding, S. E., and McKenna, S. A. (2014) Activation of 2’ 5’-oligoadenylate synthetase by stem loops at the 5’-end of the West Nile virus genome. *PLoS One*. **9**, 1–11

38. Dzananovic, E., Patel, T. R., Deo, S., McEleney, K., Stetefeld, J., and McKenna, S. A. (2013) Recognition of viral RNA stem-loops by the tandem double-stranded RNA binding domains of PKR. *RNA*. **19**, 333–344

39. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282

40. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform

methods using perceptural criteria. *J. Appl. Crystallogr.* **25**, 495–503

41. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346
42. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886
43. Volkov, V. V., and Svergun, D. I. (2003) Small-Angle Scattering Techniques. *J. Appl. Crystallogr.* **36**, 860–864
44. Ortega, A., Amorós, D., and García de la Torre, J. (2011) Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* **101**, 892–898
45. Voss, N. R., and Gerstein, M. (2005) Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J. Mol. Biol.* **346**, 477–92
46. Booy, E. P., McRae, E. K. S., Howard, R., Deo, S. R., Ariyo, E. O., Dzananovic, E., Meier, M., Stetefeld, J., and McKenna, S. A. (2016) RNA Helicase Associated with AU-rich Element (RHAU/DHX36) Interacts with the 3'-Tail of the Long Non-coding RNA BC200 (BCYRN1). *J. Biol. Chem.* **291**, 5355–72
47. Mosmann, T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods*. **65**, 55–63
48. Ishida, T., and Kinoshita, K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460-4
49. Buchan, D., Minneci, F., Nugent, T., Bryson, K., and Jones, D. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, W340–W348

50. Jones, D. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. **292**, 195–202

51. Yang, Z., Thomson, R., McNeil, P., and Esnouf, R. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*. **21**, 3369–76

52. Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta*. **1751**, 119–139

53. Amato, N. J., Mwai, C. N., Mueser, T. C., and Bryant-Friedrich, A. C. (2013) Thermodynamic and structural analysis of DNA damage architectures related to replication. *J. Nucleic Acids*. **2013**, 867957

54. Rambo, R., and Tainer, J. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*. **95**, 559–71

55. Kypr, J., Kejnovská, I., Renčiuk, D., and Vorlíčková, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713–1725

56. Meinke, G., Phelan, P. J., Fradet-Turcotte, A., Bohm, A., Archambault, J., and Bullock, P. A. (2011) Structure-based analysis of the interaction between the Simian Virus 40 T-Antigen origin binding domain and single-stranded DNA. *J. Virol.* **85**, 818–827

57. Cortez, D. (2015) HHS Public Access. *DNA Repair (Amst).* **32**, 149–157

58. Li, O., Li, J., and Dröge, P. (2007) Architectural factor and proto-oncogene HMGA2 regulates key developmental genes in pluripotent human embryonic stem cells. *FEBS Lett.* **581**, 3533–3537

59. Huth, J. R., Bewley, C. A., Nissen, M. S., Evans, J. N. S., Reeves, R., Gronenborn, A. M., and Clore, G. M. (1997) The solution structure of an HMG-I(Y)-DNA complex defines a

new architectural minor groove binding motif. *Nat. Struct. Biol.* **4**, 657–665

60. Fonfría-Subirós, E., Acosta-Reyes, F., Sapera, N., Pous, J., Subirana, J. A., and Campos, J. L. (2012) Crystal structure of a complex of DNA with one AT-Hook of HMGA1. *PLoS One*. **7**, e37120

61. Altschul, S. F., Madden, Thomas L. Schäffer, Alejandro A. Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402

62. Horikawa, I., Chiang, Y. J., Patterson, T., Feigenbaum, L., Leem, S.-H., Michishita, E., Larionov, V., Hodes, R. J., and Barrett, J. C. (2005) Differential cis-regulation of human versus mouse TERT gene expression in vivo: identification of a human-specific repressive element. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 18437–42

63. Natarajan, S., Begum, F., Gim, J., Wark, L., Davie, J. R., Hombach-klonisch, S., and Klonisch, T. (2016) High Mobility Group A2 protects cancer cells against telomere dysfunction. *Oncotarget*. **7**, 12761–82

64. Cui, T., and Leng, F. (2007) Specific recognition of AT-rich DNA sequences by the mammalian high mobility group protein AT-hook 2: A SELEX study. *Biochemistry*. **46**, 13059–13066

65. Kelly, S. M., and Price, N. C. (2000) The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **1**, 349–84

66. Cai, J., Shen, G., Liu, S., and Meng, Q. (2016) Downregulation of HMGA2 inhibits cellular proliferation and invasion, improves cellular apoptosis in prostate cancer. *Tumor Biol.* **37**, 699–707

67. Xia, Y., Yin, L., Jiang, N., Guo, W.-J., Tian, H., Jiang, X.-S., Wu, J., Chen, M., Wu, J.-Z.,

and He, X. (2015) Downregulating HMGA2 attenuates epithelial-mesenchymal transition-induced invasion and migration in nasopharyngeal cancer cells. *Biochem. Biophys. Res. Commun.* **463**, 357–363

68. Mansoori, B., Mohammadi, A., Goldar, S., Shanehbandi, D., Mohammadnejad, L., Baghbani, E., Kazemi, T., Kachalaki, S., and Baradaran, B. (2016) Silencing of High Mobility Group Isoform I-C (HMGI-C) Enhances Paclitaxel Chemosensitivity in Breast Adenocarcinoma Cells (MDA-MB-468). *Adv. Pharm. Bull.* **6**, 171–177

Chapter 4 : Inhibition of glycosylation on a camelid antibody uniquely affects its FcγRI binding activity

This chapter is based on the following manuscript published in the journal *"European Journal of Pharmaceutical Sciences"*. Full citation of the paper is as follows:

Krahn, N., Spearman, M., Meier, M., Dorion-Thibaudeau, J., McDougall, M., DeCrescenzo,G., Durocher, Y., Stetefeld, J., Butler, M. (2017). Inhibition of glycosylation on a camelid monoclonal antibody uniquely affects its FcγRI binding activity. *Eur. J. Pharm. Sci.* **96**,428-439

In this chapter I compared the N-glycan effect of a unique chimeric camelid antibody, EG2-hFc, to a full-sized IgG1, DP-12. EG2-hFc is targeted to the epidermal growth factor receptor (EGFR) which is highly expressed on tumour cells. Additionally, its small size (~ 80 kDa) makes it a preferred candidate for cancer therapy as it can penetrate further into tumours while still performing the effector functions of a full-sized mAb. Moving forward with EG2-hFc, it was necessary to determine whether changing the N-glycosylation pattern had an effect on the hydrodynamic or Fc binding properties of the mAb. These results are crucial studies in determining the suitability of a mAb for cancer therapy.

Abstract

Glycoengineering of mAbs has become common practice in attempts to generate the ideal mAb candidate for a wide range of therapeutic applications. The effects of these glycan modifications on the binding affinity of IgG mAbs for FcγRIIIa and their cytotoxicity is well known. However, little is understood about the effect that these modifications have on binding to

the high affinity FcγRI receptor. This study analysed the effect of variable N-glycosylation on a human-llama hybrid mAb (EG2-hFc, 80 kDa) binding to FcγRI including a comparison to a full-sized IgG1 (DP-12, 150 kDa). This was achieved by the addition of three glycosylation inhibitors (swainsonine, castanospermine, and kifunensine) independently to Chinese hamster ovary (CHO) cell cultures to generate hybrid and high mannose glycan structures. Biophysical analysis by circular dichroism, dynamic light scattering and analytical ultra-centrifugation confirmed that the solution-behaviour of the mAbs remained constant over multiple concentrations and glycan treatments. However, changes were observed when studying the interaction of FcγRI with variously glycosylated mAbs. Both mAbs were observed to have a decreased binding affinity upon treatment with swainsonine which produced hybrid glycans. Following de-glycosylation the binding affinity for EG2-hFc was only marginally reduced (6-fold) compared to a drastic (118-fold) decrease for DP-12. In summary, our data suggest that the relatively low molecular weight of chimeric EG2-hFc may contribute to its enhanced stability against glycan changes making it a highly suitable mAb candidate for therapeutic applications.

4.1 Introduction

Monoclonal antibodies are used to treat a broad range of diseases such as cancer (1), macular degeneration (2), Crohn's disease (3) and autoimmune diseases (4). However, designing therapeutic mAbs that bind tightly to a specific antigen and alter the function of the target system in a controlled manner remains challenging (1). For effective cancer treatment, it is desirable that the Fc-fragment of an antibody has a high affinity interaction with Fc receptors, primarily the Fcγ receptors (FcγR) (5). This interaction induces secondary responses through signalling

pathways in the cell such as the antibody-dependent cell-mediated cytotoxicity (ADCC) pathway that is primarily activated through FcγRIIIa (CD16a). To achieve strong affinity to receptors, many strategies such as amino acid mutations and N-glycosylation manipulations have been implemented in the past (6, 7). Notably, genetic engineering (8) and siRNA (9) have been introduced to interrupt the well-characterized glycan processing pathway (10, 11) and successfully reduce fucosylation, resulting in an increase in the effector function of mAbs (8, 12).

An alternative method for modifying glycosylation is the addition of glycoprotein processing inhibitors directly into the growth media (13). Castanospermine (Cas) is a glucosidase inhibitor that prevents the removal of terminal glucose sugars (in a three glucose chain) from the high mannose oligosaccharide already attached to the protein in the endoplasmic reticulum (ER). The presence of these glucose residues redirects improperly folded glycoproteins back through the calnexin/calreticulin pathway in the ER for proper processing (14). Kifunensine (Kif) is a mannosidase I inhibitor that blocks the removal of mannose sugars in the ER (15) and Golgi (13), thus preventing further processing of the N-glycan into a complex oligosaccharide. Swainsonine (Swa) is a mannosidase II inhibitor that allows production of hybrid glycans by blocking the removal of mannose from the α6 arm only. This results in the α6 arm containing high mannose structures, but allowing proper processing of the α3 branch to complex sugars, including the addition of fucose (16). Studies have shown that through the use of these inhibitors, glycans can be modified to tailor the function of IgGs without affecting secretion of the mAb (17–20).

The human Fc receptor FcγRI (CD64) has the highest affinity to IgG antibodies and is the only receptor that binds monomeric IgG as well as immune complexes (IC). This is

advantageous in the early immune response when antibody concentrations are low (10). FcγRI is normally found on macrophages, monocytes and $CD34^+$ cells enabling them to mediate ADCC activity (through monocytes), endocytosis and phagocytosis (21, 22). Supporting evidence suggests that the activation of FcγRI can elicit inflammatory cytokines such as tumour necrosis factor (TNF)-α and interleukin (IL)-6, which can induce apoptosis in tumour cells (23). Several studies have measured the effects of glycosylation modifications on the interaction of IgG with FcγRI but the effects of fucosylation and sialyation on binding remain unclear (8, 20, 24, 25). However, in the absence of modifications, the affinity of IgGs to FcγRI can be dramatically decreased, and even abolished due to conformational changes of the C_H2 domain (24, 26).

EG2-hFc is an engineered 80 kDa chimeric antibody designed with a camelid fragment variable (Fv) attached by a hinge to a humanized fragment crystallizable (Fc) region (27). Camelid antibodies (produced in Bactrian camels, dromedaries, and llamas) are composed of heavy chains only, and lack the C_H1 domain that binds to the light chains through disulphide bonds (28). Human Fc receptors recognize the EG2-hFc antibody by the human Fc region while its camelid Fv region (V_{HH}) targets EGFR, a current target of biopharmaceuticals such as Cetuximab (29) and Panitumumab (30). This design is advantageous not only to prevent EGF ligands from binding to EGFR overexpressed in tumour cells, but also to trigger proapoptotic pathways through Fc receptor binding (10). As EG2-hFc is smaller (80 kDa) than a typical IgG (150 kDa) (**Figure 4.1**) it could penetrate smaller areas and achieve a higher therapeutic index than current mAb biopharmaceuticals on the market (31).

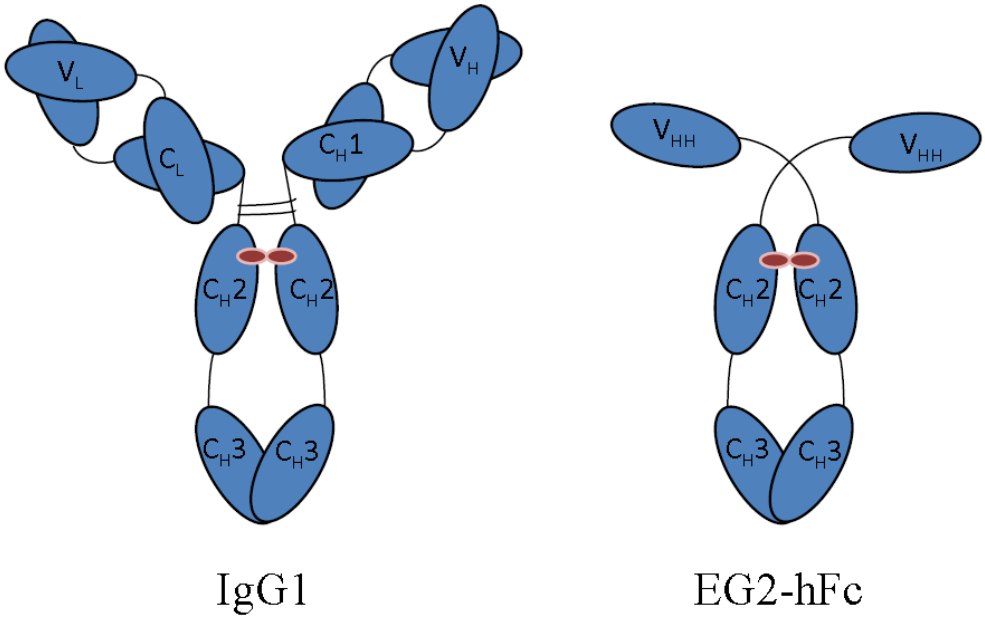

Figure 4.1 Antibody domain schematic of IgG1 compared to EG2-hFc highlighting the position of their N-glycosylation sites on the C_H2 domains (red oval).

In this study, we have investigated the effects of N-glycosylation on structure and function of a chimeric mAb using a variety of methods. The glycoprotein processing inhibitors Kif, Swa, and Cas were added to the growth media of CHO cells expressing EG2-hFc mAb and changes in N-glycosylation patterns were studied using a combination of biochemical and biophysical techniques. First, modifications of the glycans were analyzed via hydrophilic interaction liquid chromatography (HILIC). Changes in physico-chemical properties of the mAb were then determined using CD, DLS and AUC. Finally, FcγRI binding properties were studied with surface plasmon resonance (SPR) biosensors. These results were then compared to a full sized, humanized IgG1 antibody, DP-12, (ATCC® CRL-12445™, clone #1934) produced under similar conditions. The EG2-hFc glycosylation was shown to have high levels of galactosylation with a galactosylation index (GI) of 0.72, whereas DP-12 has a much lower GI (0.30) (32). Addition of the glycoprotein processing inhibitors produced both mAbs with high mannose, high mannose with attached glucose and hybrid glycans, in addition to the native complex sugars.

EG2-hFc and DP-12 binding to FcγRI was found to be decreased with hybrid glycan structures while complete removal of the N-glycan had a much stronger effect on DP-12 binding (118-fold reduction) compared to EG2-hFc (6-fold reduction).

4.2 Materials and Methods

4.2.1 Mammalian cell culture

CHO EG2-hFc is a stably transfected CHO (DUXB) cell line as described previously (33) expressing the antibody EG2-hFc (27) and the CHO cell line DP-12 [clone #1934 from American type culture collection (ATCC)] produces a human IgG1 mAb against IL-8. These cells were grown as described previously (32). Cells were seeded at 2.5 x 10^5 cells/mL in the presence of glycoprotein processing inhibitors Kif (2.2 and 4.4 μM), Cas (0.23 and 0.46 mM), and Swa (8.7 and 17.3 μM) (Toronto Research Chemicals, New York, ON). Total cell counts and cell viability were measured at regular intervals by Trypan dye exclusion (0.5%) using a Cedex Image Analyzer (Innovates AG, Bielefeld, Germany). After four days of growth, cells were centrifuged at 1500 x g for five mins to collect the cultured supernatant that was filtered through a 0.2 μm Steritop filter (EMD Millipore, Etobicoke, ON) before frozen at -20 ^{o}C.

4.2.2 FcγRI production and purification

The CHO cell line expressing a truncated EBNA1 protein (CHO-3E7) (34), was cultured in FreeStyle™ F17 media (Life Technologies, Burlington, ON). *In-cell* biotinylation of FcγRI was produced by transient co-transfection of two plasmids coding for the FcγRI ectodomain followed by a biotin acceptor peptide (BAP) sequence and BirA enzyme. At a cell density of 1.5

x 10^6 cells/mL, 50% (w/w) FcγRI plasmid, 3% (w/w) BirA plasmid, 5% (w/w) green fluorescent protein (GFP) and 27% (w/w) salmon sperm DNA (SSDNA) were transfected using 2.62 mg PEImax (Polysciences, Warrington, PA) as the transfecting agent (35). Culture feed, temperature switch and purification steps were done as described previously (36, 37). Purification of FcγRI was performed through the adaptation of the protocol described by Tom and colleagues (38).

4.2.3 Purification and concentration determination of mAbs

The EG2-hFc and DP-12 mAb concentrations were measured from cell culture supernatant using an ELISA as described previously (32). Monoclonal antibodies were purified from culture media using a HiTrap™ Protein A HP 1 mL column (GE Healthcare, Fairfield, CT). Previously filtered cell culture supernatant was applied to the Protein A column, where it was washed with PBS (pH 7.4) and eluted with 0.1 M glycine-HCl, pH 2.7 (Sigma-Aldrich, St. Louis, MO). The eluted protein was neutralized immediately using 1 M Tris-HCl, pH 9.0 (Thermo Fisher Scientific, Waltham, MA) and concentrated using an Amicon ultra 30 kDa MWCO filter (EMD Millipore, Etobicoke, ON). Non-glycosylated antibody was prepared by immobilising 1 mg onto the Protein A HP SpinTrap™ column (GE Healthcare, Fairfield, CT), and incubated for 48 hr with the addition of 25 U of Peptide-N-Glycosidase F (PNGase F) enzyme (Promega, Madison, WI). The cleaved N-glycan and the PNGase F enzyme were washed through the column with 20 mM phosphate buffer, pH 7.0 before elution and neutralization of the non-glycosylated mAb. Deglycosylation of the mAbs was confirmed by gel electrophoresis using an 8-16% Criterion™ TGX™ gel (Bio-Rad, Mississauga, ON). Glycosylated and non-glycosylated mAbs were prepared in loading buffer (0.0625 M Tris HCl at pH 6.8, 18.75% glycerol, 1.86 mM bromophenol blue, 2% w/v SDS) with and without β-mercaptoethanol as a

reducing agent. Samples were loaded alongside a Precision Plus Protein™ Standard (Bio-Rad, Mississauga, ON) and the gel was run at 200 V before stained with Coomassie Brilliant Blue for protein detection.

4.2.4 N-glycosylation analysis

The N-glycans from purified mAbs were removed using an in-gel method and analyzed by HILIC-HPLC as described previously (32). Glycans were analyzed by HILIC with fluorescent detection using an X-Bridge 3.5 μm amide column (4.6 x 250 mm) (Waters, Milford, MA) heated to 30 °C with a flow rate of 0.86 mL/min. Glycans were eluted using an initial ratio of 20:80 of 50 mM ammonium formate, pH 4.4: acetonitrile, followed by a gradient to 50:50 over 48 minutes. Peaks were calibrated with a 2-amino benzamide (AB) labelled glucose ladder and glycan standards [M5, M9, A2G2S2, F(6)A2G2S2, and A3G3] (Prozyme, Hayward, CA) before compared to a glycan database (www.glycobase.nibrt.ie) for preliminary identification (39). Structures were confirmed as described previously using exoglycosidase enzymatic digestions (32) according to the method of Royle et al (40). Sialidase from recombinant A. *ureafaciens* was used to digest the neuraminic acid from the 2AB-labelled glycans and separate digest arrays had further additions of bovine testis β (1-3,4)-galactosidase, jack bean mannosidase, jack bean β-N-acetylhexosaminidase, and bovine kidney fucosidase (Prozyme, Hayward, USA). Upon confirmation of the glycan structures, the GI was calculated for each treated sample using the equation :

$$GI = (0.5*G1 + G2) / (G0 + G1 + G2)$$

where G0 = non-galactosylated glycans, G1 = mono-galactosylated glycans, and

G2 = di-galactosylated glycans.

4.2.5 Spectropolarimetry

Far UV CD spectra were recorded on a J-810 spectropolarimeter (Jasco Inc., Easton, PA) calibrated using 2.583 mM CSA monohydrate (Alfa Aesar, Ward Hill, MA). Measurements were taken from 190 to 250 nm with a 1.0 mm quartz cuvette and an integration time of 4 s/nm. Additional measurements of EG2-hFc samples were performed from 190 – 200 nm with an integration time of 8 s/nm to reduce some deviation observed with the original 4 s exposure. DP-12 and EG2-hFc mAb samples were prepared at concentrations of 1.5 µM and 2.5 µM, respectively, in 10 mM sodium phosphate, pH 7.5, 154 mM NaF buffer. In all cases, sample and buffer were measured in triplicate, corrected with the calibration factor, averaged and finally baseline subtracted. All displayed spectra were normalized by the number of peptide bonds per unit volume.

4.2.6 Hydrodynamic studies

DLS studies for both mAbs were performed as described earlier at concentrations ranging from (0.07 – 0.7 mg/mL) (41). Melting temperatures were determined at 0.7 mg/mL using DLS by performing five separate measurements every 2 °C. Homogeneity of mAbs was studied using an SV method in a ProteomeLab^{TM} XL-I analytical ultracentrifuge with an An50Ti 8-cell rotor (Beckman Coulter Inc., Mississauga, ON). Standard 12 mm double-sector cells were filled with 400 µL of mAb in the sample channel and 420 µL of 10 mM sodium phosphate, pH 7.5, 154 mM NaF buffer in the buffer channel. For the analysis of EG2-hFc, a total of 90 absorbance and interference radial scans were collected at 20.0 °C every 11.5 mins at 30,000 rpm. DP-12 mAbs

were analyzed with a total of 70 absorbance and interference radial scans measured at a speed of 25,000 rpm. Data analysis was performed using the SEDFIT program (42, 43) to determine the sedimentation coefficients of each mAb sample ($s_{20,b}$). The sedimentation coefficients were then converted to standard solvent conditions ($s_{20,w}$) using the buffer viscosity and density calculated from SEDNTERP (44) and finally extrapolated to infinite dilution ($s^0_{20,w}$).

4.2.7 FcγRI receptor interaction

Receptor binding was studied on a Biacore T100 instrument (GE Healthcare, Fairfield, CT) at a flow rate of 50 µL/min at 25 °C using HBS-P, pH 7.4 (GE Healthcare, Fairfield, CT). The Biotin CAPture kit (GE Healthcare, Fairfield, CT) was used to stably capture biotinylated FcγRI at the surface of the biosensor via biotin/streptavidin interactions, with a flow rate of 30 µL/min for 15 s (0.1 µg/mL, 70 RU) while no receptor was injected over the reference surface. Antibody solutions (30, 50, 100, and 300 nM) and buffer were then injected over captured FcγRI and reference surfaces for 60 s, causing a net increase in signal. This was followed by the final injection of running buffer for an additional 740 s to monitor dissociation of the receptor/mAb complex. Surface regeneration was performed by injection of 6 M guanidine-HCl, 0.25 M NaOH included in the kit. Binding was recorded in real-time for several concentrations of each mAb (30 – 300 nM) run in duplicate. The global analysis of the control-corrected sensorgrams, assuming a 1:1 interaction with the Biacore T100 Evaluation software, allowed the determination of both the association and dissociation kinetic constants for the mAb/receptor interactions (k_a and k_d). The ratio of these kinetic constants provided the apparent thermodynamic dissociation constant for the interaction, K_D (45).

4.2.8 Homology modeling

A homology modeling approach was used to generate 3D structures of glycosylated EG2-hFc. The crystal structure coordinates from Protein Data Bank (PDB) code: 1HZH, a neutralizing human IgG against HIV-1 (46), which has 38% sequence identity and about 54% sequence similarity were used as a template for model building. For sugar attachment studies, the human IgG-Fc glycoforms based on PDB code: 1H3T, 1H3U, 1H3Y (47) were employed. The *DeepView* Tool (48) of the SWISS-MODEL web server was employed to align the sequences in order to generate the template for the model building software *Modeller* (49). After model building, hydrogen atoms were added in *DeepView*. All generated models were evaluated for plausibility by means of the Protein Structure & Model Assessment Tools available on the SWISS-MODEL web server, *i.e.* Procheck (50). The ramifications of introduced point mutations to the models were characterized using conjugated gradient minimization in combination with molecular mechanisms (MM2) and a final molecular dynamics simulation at 1000 k for 50 ps for further optimization.

4.3 Results

4.3.1 Growth and productivity

The effect of the glycoprotein processing inhibitors on cell growth and viability of CHO EG2-hFc was tested at two different concentrations of the inhibitors over a four-day culture period as before (16, 19). This experiment was conducted to establish concentrations such that minimal effects would be observed on viable cell density (VCD) and volumetric mAb productivity while still resulting in significantly modified glycosylation (**Figure 4.2A and**

Figure 4.2B). In this section, the given percentages referring to VCD and volumetric productivity on day four are with respect to the control culture. At the lower inhibitor concentrations, the VCD was decreased to 83%, 74%, and 74% for Cas, Swa, and Kif, respectively. Upon increasing the concentration two-fold, the VCD was further reduced to 67%, 43%, and 58% (Cas, Swa, Kif).

For CHO EG2-hFc, the effect of the inhibitors on the volumetric mAb productivity followed a similar trend as that of growth, with higher concentrations of each inhibitor reducing productivity (**Figure 4.2B**). At the low concentrations of inhibitor, Cas had the greatest effect on productivity (38%), followed by Swa (69%) with Kif having no significant effect ($P = 0.25$). A further decrease in productivity was observed with a two-fold increase in the inhibitor to 31%, 42%, and 70% for Cas, Swa, and Kif, respectively.

The effects of the processing inhibitors on VCD and mAb productivity were also determined for CHO DP-12 on day four (**Figure 4.2C and** Figure 4.2**D**). Due to the strong inhibitory impact of higher inhibitor concentrations on the CHO cell line, only the lower concentrations were analyzed. Similar effects on VCD were observed in CHO DP-12 as with CHO EG2-hFc: Cas reduced the VCD to 82%, Kif to 77% and Swa to 65%. In terms of the volumetric mAb productivity, similar effects were observed for Kif and Swa while Cas was observed to have no inhibitory effect, in contrast to what was observed for EG2-hFc.

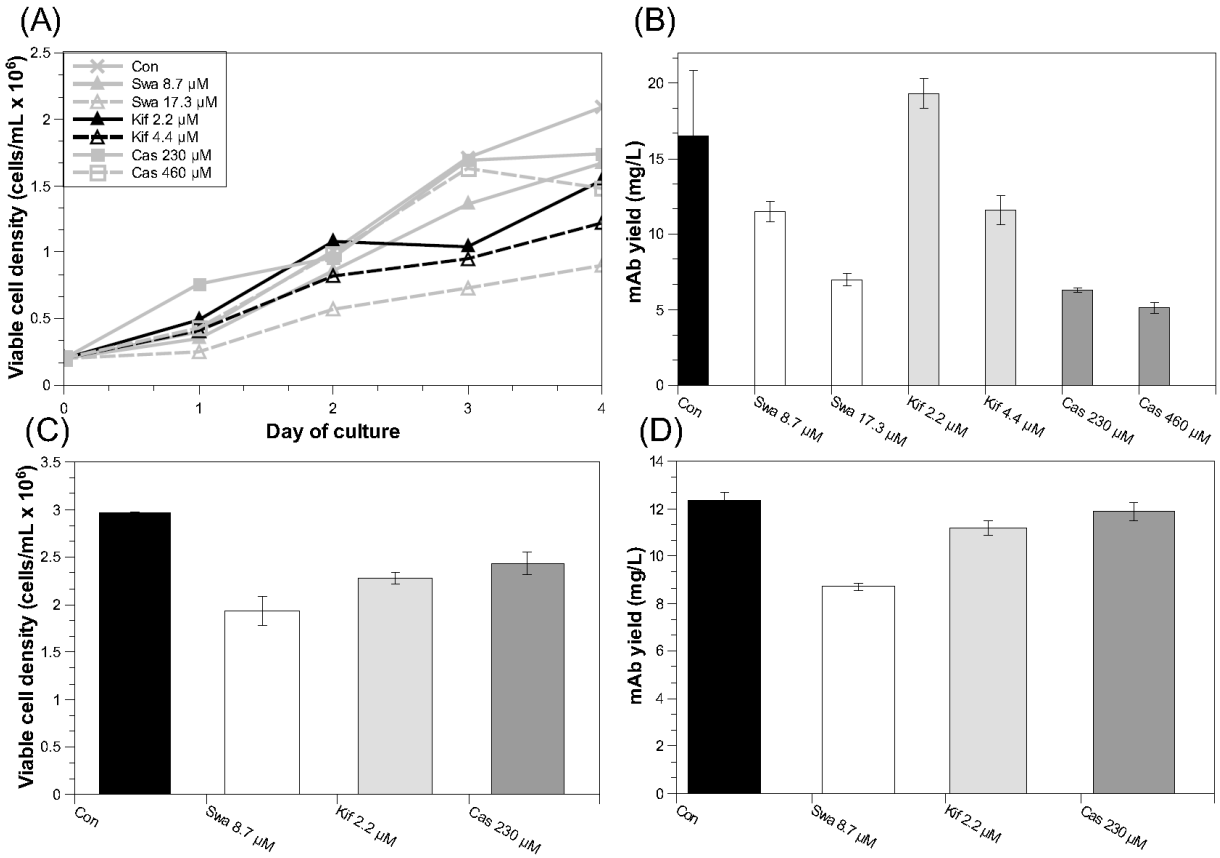

Figure 4.2 Effect of glycoprotein processing inhibitors on the cell growth and mAb yield

The cell growth (A and C) and mAb yield (B and D) of CHO EG2-hFc (A and B) and DP-12 (C and D) in the presence of glycoprotein processing inhibitors. CHO EG2-hFc cells were cultured for four days in the presence of two different concentrations of Swa, Kif or Cas compared to a control culture (Con) with no inhibitor. DP-12 was cultured for four days in the presence of Swa (8.7 µM), Kif (2.2 µM) or Cas (230 µM) compared to the control culture with no inhibitor added. Cultures were performed in duplicate and for each culture cell density and viability measurements were determined twice daily. Cell viabilities remained at >80% throughout each culture. On day 4, media cultures were analyzed for the amounts of EG2-hFc and IgG mAb using ELISA (n = 2, five dilutions each).

4.3.2 *MAb glycan microheterogeneity*

Gel electrophoresis analysis was used to confirm full removal of the glycan from PNGaseF digests of EG2-hFc and DP-12 (**Figure 4.3**) while the remaining glycosylated mAbs were analyzed using HILIC.

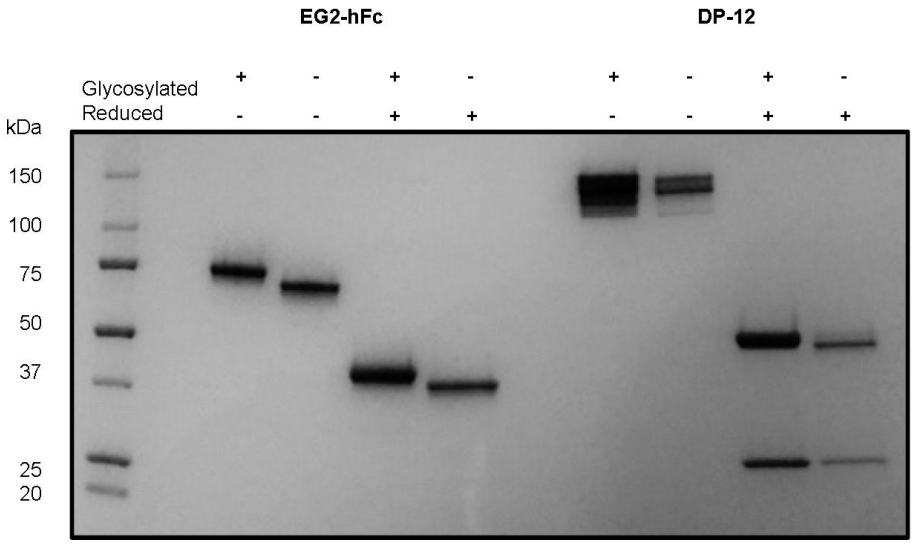

Figure 4.3 Gel electrophoresis analysis of glycosylated and non-glycosylated EG2-hFc and DP-12 mAbs

Reduced and non-reduced samples were prepared in loading buffer (with and without β-mercaptoethanol, respectively) and run at 200 V on an 8-16% Criterion™ TGX™ gel. Coomassie Brilliant Blue staining allowed detection of the mAbs with a ~5 kDa (reduced) and ~10 kDa (non-reduced) decrease in the molecular weight of the non-glycosylated mAbs.

Glycosylation analysis of the chimeric camelid mAb EG2-hFc suggested a high level of galactosylation (GI = 0.72), where the predominant glycan was a fucosylated biantennary structure with two galactose residues [F(6)A2G2] and lesser amounts of fucosylated biantennary glycans with one [F(6)A2G1] or no galactose residues [F(6)A2G0] (**Figure 4.4A**). Smaller amounts of fucosylated biantennary glycans with one or two sialic acids [F(6)A2G2S1 and F(6)A2G2S2] were also detected, as well as non-fucosylated glycans [A2G0, A2G1, A2G2, A4G3, A4G3Sn]. The glycosylation profile of the conventional IgG1 produced in CHO DP-12 resembled typical IgG1 glycosylation with much lower galactosylation; a high amount of F(6)A2G0 and lower amounts of F(6)A2G1 and F(6)A2G2 (**Figure 4.4B**).

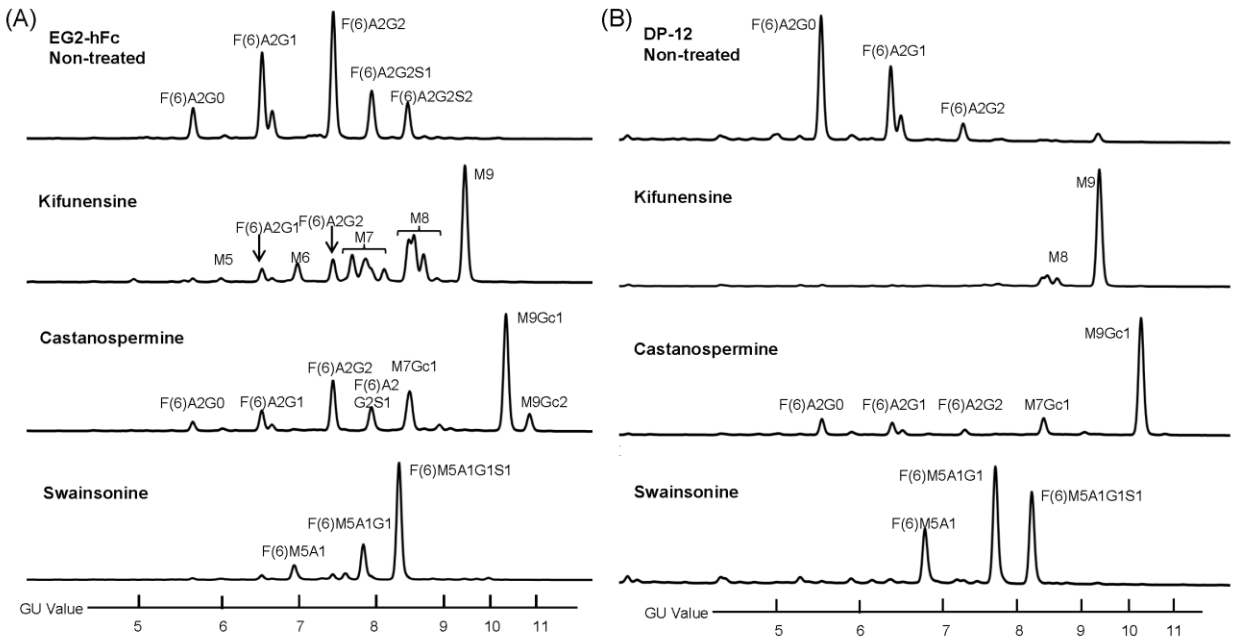

Figure 4.4 HILIC-HPLC glycan analysis of mAb glycans

CHO EG2-hFc (A) and CHO DP-12 (B) cell lines were cultured in the presence of the inhibitors Kif (2.2 μM), Cas (230 μM) and Swa (8.7 μM). Glycans were isolated from Protein A purified mAb, labelled with fluorescent 2-AB and separated using HILIC analysis on a Waters 3.5 μm X-Bridge column which had been standardized using a dextran ladder. Glucose unit (GU) values were calculated for the eluted peaks based on the dextran ladder profile. Structures were tentatively assigned based on the GU values, and then confirmed using exoglycosidase digestion.

Addition of the glycoprotein processing inhibitors to the CHO cultures over four days produced unique mAb glycan profiles compared to the control cultures (**Figure 4.4**). Kif-treated mAbs acquired primarily high mannose structures on both EG2-hFc and DP-12 with overall similar values (85 ± 1% and 84 ± 14%, respectively,**Table 4.1**). In the case of DP-12, these high mannose structures consisted predominantly of a nine-mannose glycan (Man9) and lesser amounts of Man8 and Man7. Significant amounts of Man8, Man7 and Man6 were detected in addition to the Man9 glycans for EG2-hFc (**Figure 4.4**). The treatment of EG2-hFc with Cas, generated predominantly high mannose glycans with the addition of glucose [Man9Glc1, Man9Glc2, and smaller amounts of Man7Glc1]. In the Cas-treated cultures, the ratio of

Man9Glc1 to Man7Glc1 was higher in DP-12 mAb (7.5) than EG2-hFc mAb (2.4). Also present were significant amounts of the typical fucosylated complex glycans in both EG2-hFc (22 ± 16%) and in DP-12 (29 ± 15%). Finally, Swa treatment created mAbs containing hybrid glycans (84 - 87%). EG2-hFc was observed to have much higher amounts of the sialylated hybrid glycan F(6)M5A1G1S1 than DP-12, which instead had higher levels of the non-sialylated [F(6)M5A1G1] and non-galactosylated [F(6)M5A1] hybrid glycan. This result was consistent with the control cultures in which DP-12 mAbs contained a lower amount of galactosylation and hence sialylation.

Table 4.1 Summary of the composition of N-glycans attached to N297 for EG2-hFc and DP-12 samples treated with glycoprotein processing inhibitors

EG2-hFc	**% Fucosylation**	**% Hybrid**	**% High Mannose**	**Galactosylation Index (GI)**
Control (No inhibitor)*	88 ± 3	ND	ND	0.72 ± 0.05
Swainsonine **	78 ± 9	87 ± 10	ND	0.44 ± 0.02
Kifunensine **	8 ± 4	ND	85 ± 12	0.07 ± 0.02
Castanospermine **	23 ± 16	ND	67 ± 12	0.23 ± 0.12
DP-12				
Control (No inhibitor)***	82 ± 10	ND	ND	0.30 ± 0.03
Swainsonine ***	87 ± 8	84 ± 34	ND	0.37 ± 0.01
Kifunensine ***	8 ± 8	ND	84 ± 14	0.04 ± 0.04
Castanospermine ***	29 ± 15	ND	56 ± 22	0.12 ± 0.06

The mAbs were treated with glycoprotein processing inhibitors Swa (8.5 μM), Kif (2.2 μM), and Cas (230 μM). Values are means from five different cultures representing three separate experiments periods (*), three different cultures representing three separate experiment periods (**) or two different cultures representing two separate experiment (***) ± standard deviation. ND = not detected.
Galactosylation Index (GI) = 0.5(G1) + G2/(G0 + G1 + G2), where G0 = non-galactosylated glycans, G1 =mono-galactosylated glycans, G2 = di-galactosylated glycans.

Swainsonine treated mAbs were digested with exoglycosidases for confirmation of the peak assignment. Sialidase and galactosidase digests of both mAbs produced a single peak that corresponded to F(6)M5A1. Further processing of the DP-12 sample with hexosaminidase

digested the peak to F(6)M5. On the other hand, further processing with mannosidase or fucosidase to the EG2-hFc sample digested the peak to F(6)A1 or M5A1, respectively (**Figure 4.5**).

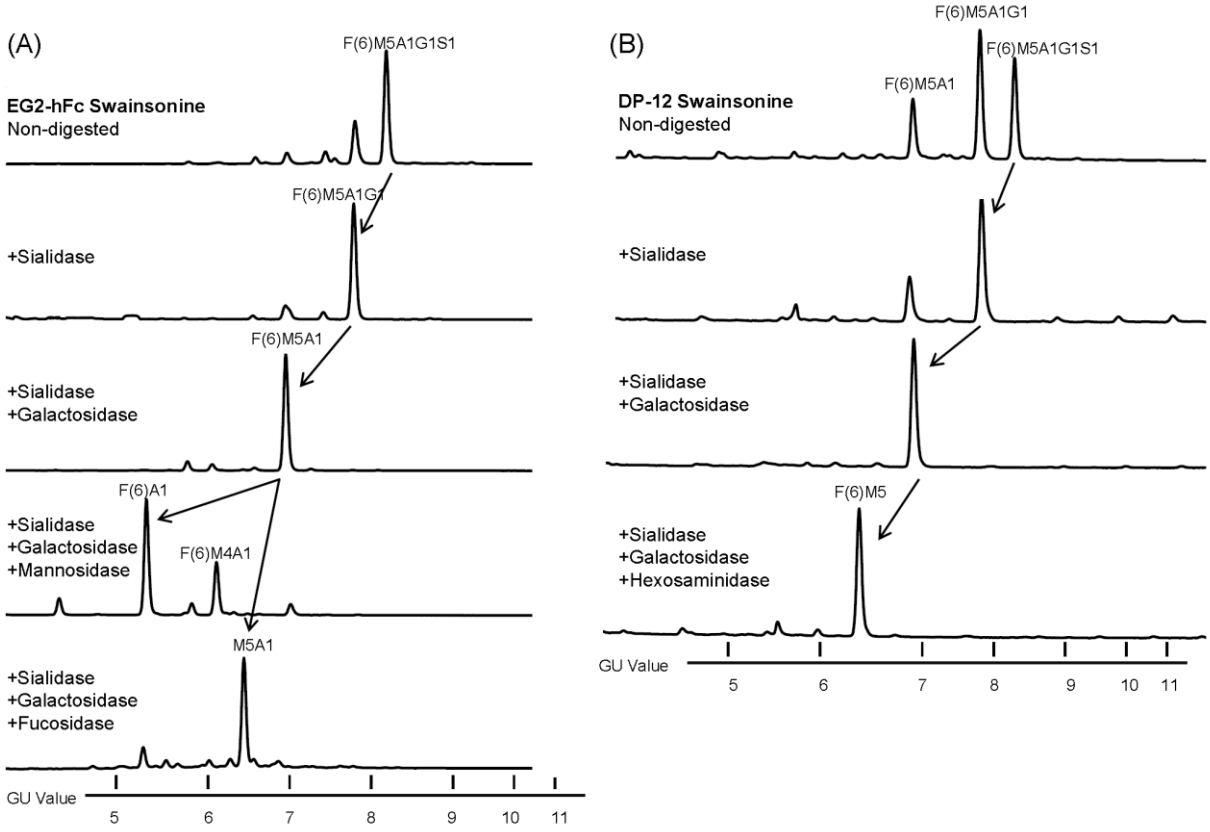

Figure 4.5 Glycan profiles of exoglycosidase digests for swainsonine treated mAbs

In addition to glycan structural assignment using standard GU values from HILIC analysis in GlycoBase, arrays of exoglycosidase digests were used to confirm the assigned structures of Swa treated EG2-hFc (A) and DP-12 (B) mAbs. Sialidase from recombinant A. *ureafaciens* was used digest the neuraminic acid from the 2AB-labelled glycans and separate digest arrays had further additions of bovine testis β (1-3,4)-galactosidase, jack bean mannosidase, jack bean β-N-acetylhexosaminidase, and bovine kidney fucosidase. Following the digests the glycans were re-analyzed by HILIC to determine the digested structures based on GU values and shifts in GU values.

4.3.3 Secondary structure analysis

The secondary structure elements present in the mAbs were determined using DichroWeb, a program which allows online analysis of protein CD data through the use of five

algorithms and seven different reference databases (51). **Figure 4.6** presents the molar ellipticity and absorbance difference of the mAb at each wavelength from 190 nm – 250 nm. From this data it was calculated that both the EG2-hFc (**Figure 4.6A**) and DP-12 (**Figure 4.6B**) mAbs contained 48% β-sheet structure and 48% random coil or unstructured protein. The peak observed at 200 nm for both EG2-hFc and DP-12 mAbs and the minimum observed at 215 nm for DP-12 is characteristic of β-sheet structure. Upon treatment with the glycoprotein processing inhibitors, the CD spectra of each mAb was identical to that of the control, indicating that the secondary structures of EG2-hFc or DP-12 were not significantly affected. This result was expected as earlier studies had shown that removal of the glycan on a full IgG1 mAb did not affect the secondary or even tertiary structure (52).

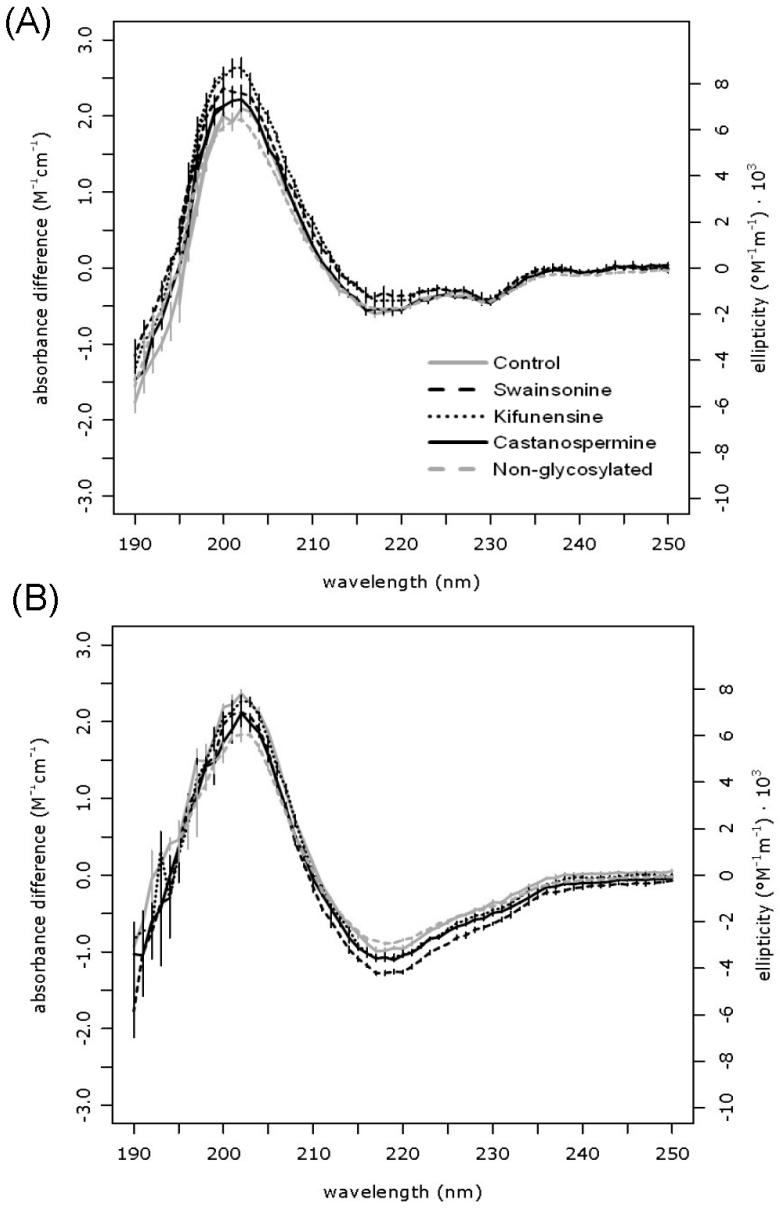

Figure 4.6 Effect of glycoprotein processing inhibitors on the secondary structure of mAbs determined by CD

Far UV spectra of EG2-hFc (A) and DP-12 (B) mAbs were obtained on a J-810 spectropolarimeter, CSA-calibrated, baseline corrected, and displayed in terms of molar ellipticity as a function of the wavelength (n = 3). An integration time of 4 s/nm was measured for all samples from 190-250 nm with additional measurements taken for EG2-hFc from 190-200 nm (8 s/nm integration).

4.3.4 Homogeneity of mAbs

Post-translational modifications, especially glycosylation can influence many properties of macromolecules including aggregation and pharmacokinetics. The impact of altered glycosylation profiles on diffusion and sedimentation properties that provide information on aggregation of the mAbs was investigated using DLS and AUC, respectively. **Figure 4.7** displays the hydrodynamic radius R_h distribution at ~ 0.7 mg/mL for treated mAbs indicating that irrespective of the presence of sugars, the solution is free of aggregates. DLS analysis suggests that, depending on the glycan attachment, the R_h values obtained for EG2-hFC and DP-12 were in the range of 4.13 - 4.68 nm and 5.18 - 5.35 nm, respectively (**Table 4.2**).

Table 4.2 Summary of the hydrodynamic and physical properties determined for EG2-hFc and DP-12 samples treated with glycoprotein processing inhibitors

mAb	Treatment	M_w calculated (kDa)[a]	$\bar{v}$ (cm^3/g)[a]	T_m (°C)[b]	R_h (nm)[b]	$s^0_{20,w}$ (S)[c]	M_w experimental (kDa)[d]
EG2-hFc	NG	79.66	0.72638	64	4.41 ± 0.16	4.72 ± 0.10	89.6
	Con	83.53	0.72260	66	4.13 ± 0.12	4.55 ± 0.04	79.6
	Cas	84.15	0.72078	66	4.35 ± 0.11	4.58 ± 0.02	83.7
	Kif	83.79	0.72121	66	4.68 ± 0.15	4.58 ± 0.02	90.4
	Swa	83.48	0.72237	66	4.26 ± 0.13	4.58 ± 0.02	82.6
DP-12	NG	147.29	0.72714	68	5.20 ± 0.20	6.21 ± 0.03	139
	Con	150.47	0.72560	70	5.18 ± 0.33	6.41 ± 0.02	142
	Cas	151.78	0.72401	68	5.30 ± 0.18	6.52 ± 0.02	147
	Kif	151.42	0.72426	70	5.35 ± 0.15	6.50 ± 0.02	148
	Swa	150.78	0.72503	70	5.28 ± 0.28	6.50 ± 0.01	147

[a] Calculated based off of the amino acid sequence and the most prominent N-glycan present.
[b] Determined from DLS data at ~ 0.7 mg/mL for each antibody
[c] The Svedberg coefficient extrapolated to infinite dilution and to standard solvent conditions
[d] M_w calculated from the Svedberg equation using experimentally determined R_h from DLS and $s^0_{20,w}$ from AUC

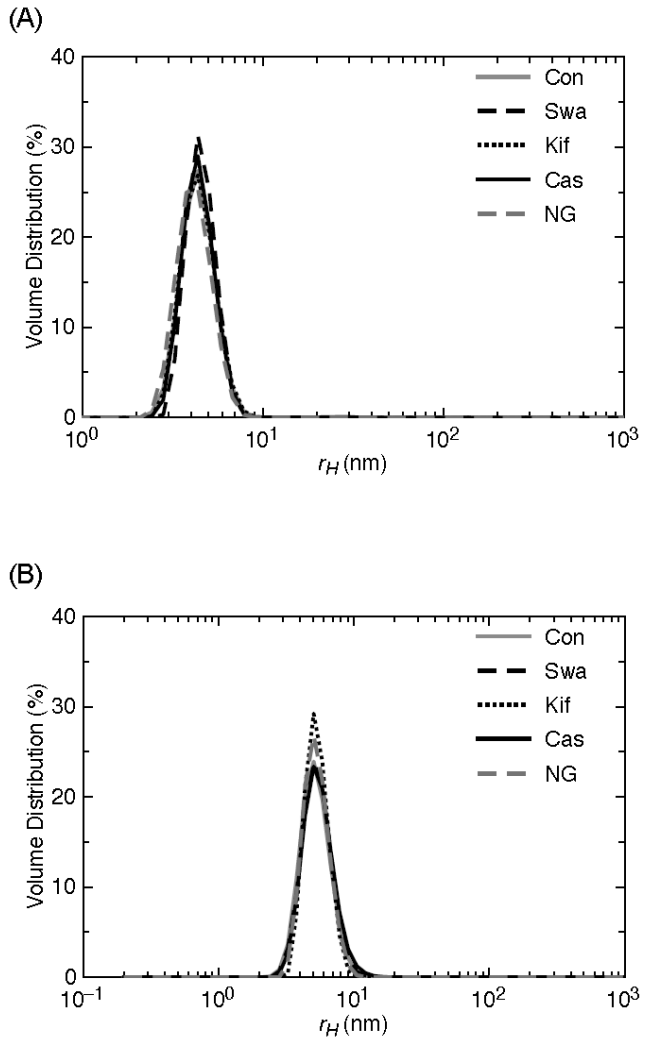

Figure 4.7 Monodispersed hydrodynamic radius (R_h) distribution by volume of variously glycosylated mAbs

Variously glycosylated EG2-hFc (A) and DP-12 (B) mAbs show a monodispersed sample with no light scattering observed at larger R_h to imply aggregation is present. Data was obtained using a Zetasizer Nano S machine and shown as an average of five measurements at 0.7 mg/mL.

The sedimentation coefficient parameter, which is dependent on the molecular weight and the overall shape of each treated mAb, was obtained at multiple concentrations to determine homogeneity of those antibodies (**Figure 4.8**). Both EG2-hFc and DP-12 mAbs containing complex glycans (Con) were found to be in a monomeric state with a Svedberg coefficient extrapolated to infinite dilution ($s^0_{20,w}$) of 4.72 ± 0.10 S and 6.21 ± 0.03 S, respectively (**Table**

4.2). Additionally, the mAbs with high mannose (with or without glucose), hybrid glycan structures, or even the absence of the attached glycan showed no significant change in the Svedberg constant. For example, the sedimentation coefficient values for EG2-hFC were in the range of 4.55 - 4.72 S whereas those for DP-12 were in the range of 6.21 - 6.52 S. Most importantly, these samples were found to be monodisperse and aggregation free.

The R_h and sedimentation coefficient values were used to calculate the M_w of both mAbs prepared with and without the presence of inhibitors using the previously mentioned Equation 1 (53). An average of 85.2 kDa and 145 kDa was calculated for all treated EG2-hFc and DP-12 mAbs, respectively, which is consistent with monomeric antibody. A reasonable agreement between the sequence based molecular weight and the experimental molecular weight calculated from R_h and $s^0_{20,w}$ was obtained as presented in **Table 4.2**.

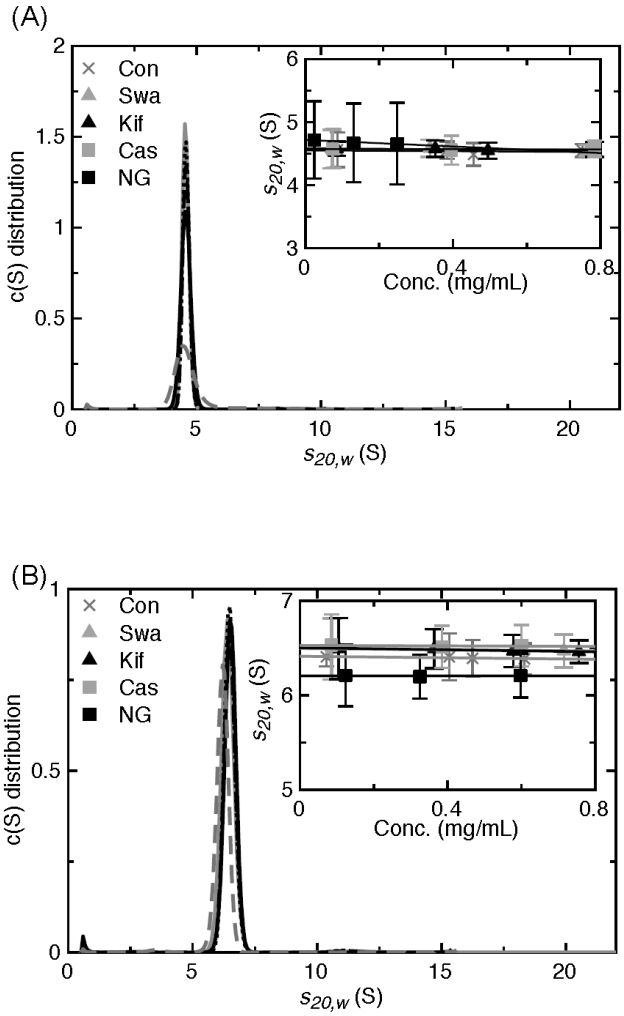

Figure 4.8 Monodisperse sedimentation coefficient distribution in water of variously glycosylated mAbs

Treated EG2-hFc (A) and DP-12 (B) mAbs show a monodispersed sample. Data is shown as an average of 70-90 scans at 0.4 mg/mL. Analysis of the data was performed using the SEDFIT program to obtain a distribution of the sedimentation coefficients. The Svedberg coefficient ($s_{20,w}$) of EG2-hFc (inset to Panel A) and DP-12 mAbs (inset to Panel B) were constant over a range of concentrations (0.07 – 0.7 mg/mL) and irrespective of the N-glycan attached. Extrapolation to zero concentration provides the Svedberg coefficient at infinite dilution ($s^0_{20,w}$) displayed in Table 2. Error bars are from the standard error associated with the multiple scans.

4.3.5 *Stability of mAbs*

DLS was also employed to determine the melting temperature of each treated mAb. This was designated as the point at which the R_h drastically increased upon increasing the temperature (**Table 4.2**). Both mAbs were observed to have a stable melting temperature at around 66 °C and 70 °C for EG2-hFc and DP-12, respectively, regardless of the presence or absence of a glycan structure. Notably, removal of the glycan caused visible aggregation at an antibody concentration higher than 1 mg/mL and therefore studies were performed at concentrations below this threshold. The lack of glycan on mAbs and their propensity to aggregate has been observed previously with non-glycosylated mAbs being the most unstable out of all forms of glycosylated IgG (54, 55).

4.3.6 *FcγRI binding affinity*

Glycosylation, specifically fucosylation at Asn297, affects the interaction of the Fc region with Fc receptors on natural killer (NK) cells, lymphocytes and B-cells (8, 12). The affinity of EG2-hFc and DP-12 mAbs to FcγRI, which can induce ADCC, endocytosis and phagocytosis, was determined by SPR. The recorded sensorgrams are shown in **Figure 4.9** which corresponds to injections of mAbs over immobilised FcγRI. SPR analysis suggested that without treatment, EG2-hFc had a lower affinity (K_D = 4.53 ± 0.04 nM) than DP-12 (K_D = 1.44 ± 0.03 nM) (**Table 4.3**). The affinity was found to be dependent on the glycosylation attached at Asn297 for both mAbs; however, the effect of glycosylation on each was different. The affinity of EG2-hFc treated with Cas (high mannose structures with glucose) was unchanged, however the removal of the glucose residues (Kif) caused a decrease in the affinity (from 4.53 to 8.09 nM). This was not observed with DP-12, which remained unchanged in both treatments.

Treatment of Swa, resulting in the formation of hybrid glycan structures both fucosylated and sialylated, caused a decrease in affinity of both mAbs by approximately two-fold. A novel observation was that the non-glycosyated EG2-hFc had only a 6-fold decrease in affinity in comparison to DP-12 where, though binding was still observed, the affinity decreased ~118-fold.

Table 4.3 Effect of N-glycan on affinity (K_D) of EG2-hFc and DP-12 to FcγRI determined using SPR

Treatment	EG2-hFc (nM)	DP-12 (nM)
Control (no inhibitor)	4.53 ± 0.04	1.44 ± 0.03
Swainsonine	10.82 ± 0.83	2.30 ± 0.02
Castanospermine	5.47 ± 0.17	0.82 ± 0.02
Kifunensine	8.09 ± 0.06	0.83 ± 0.02
Non-glycosylated	28.95 ± 0.07	170 ± 41

Data sets corresponding to various injections of mAbs at different concentrations (30 to 300 nM) in duplicate (independent dilution series) were double-referenced and globally analyzed with a simple binding model. Reported apparent K_D values were calculated from the ratio of the kinetic constants ± standard error.

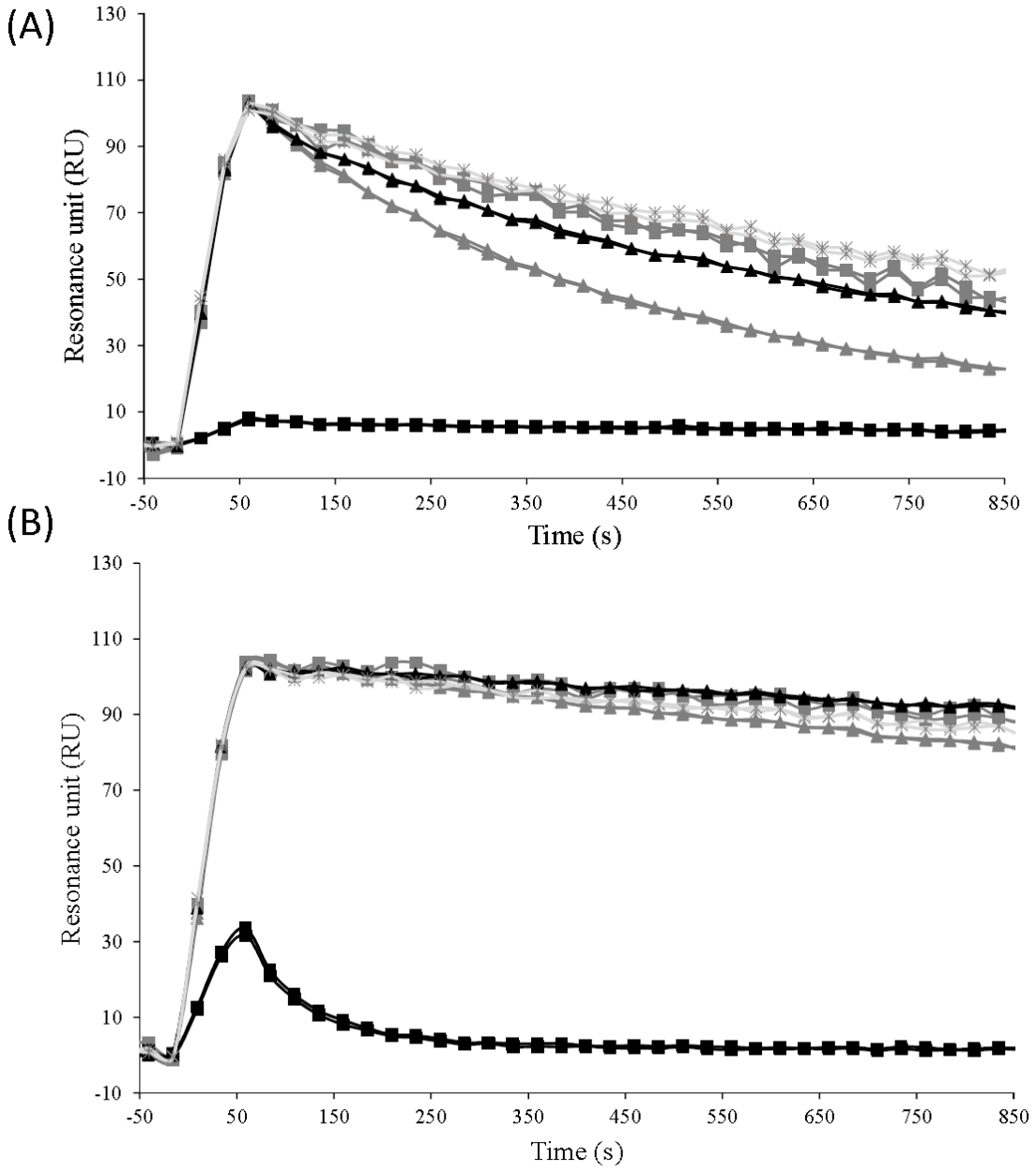

Figure 4.9 Effect of glycoprotein processing inhibitors on the binding of mAbs to FcγRI

The results shown are sensorgrams from duplicate injections where EG2-hFc (A) and DP-12 (B) mAb injections (300 nM) were performed from 0-50 s over a sensor chip on which FcγRI has been tethered. Buffer was injected for another 740 s to dissociate the antibody from its receptor, prior to surface regeneration. Control (light grey stars), Cas (grey squares), Swa (grey triangles), Kif (black triangles), non-glycosylated (black squares).

4.4 Discussion

In this study the effects of modifiying the Fc glycosylation of a heavy chain chimeric camelid antibody, EG2-hFc, were compared to that of a full sized human IgG1, DP-12. The glycosylation at Asn297 in the Fc region of IgG as a requirement for the activation of secondary effector functions is well documented (56–58). The removal of these oligosaccharides results in loss of ADCC activity due to conformational changes in the FcγR binding region. Reduction in fucosylation of the intact glycan has been found to increase binding to the FcγRIIIa receptor on NK cells and therefore increases ADCC activity (8, 12). However, there is less consistent evidence to identify the effects of glycosylation modifications on the affinity of IgG for FcγRI. Because of the similarity in structure of the extracellular domains of these receptors (59), it is likely that the impact of mAb glycoforms on binding to FcγRI would be similar to that previously determined for FcγRIIIa. A set of glycoprotein processing inhibitors were used to modify the Fc glycan of two mAbs to compare the effects that N-glycosylation has on their physical properties as well as their binding affinity to FcγRI.

Initially, two concentrations of inhibitors were used, but due to negative effects on VCD and mAb productivity of the higher concentration, the lower concentration was used for the remainder of the study. The effect of the glycoprotein processing inhibitors on mAb yield showed Cas to have a much greater inhibitory effect for EG2-hFc than DP-12, which may suggest that CHO EG2-hFc cells are more sensitive to the blockage of removing the terminal glucose residues on the immature glycan. Disruption of the normal processing and retention of the glucose residues may re-direct the mAb through the calnexin/calreticulin pathway responsible for proper folding of the molecule (14), instead of allowing further processing in the

Golgi and therefore secretion of the intact mAb. This much-reduced mAb yield in Cas-treated EG2-hFc may also be a result of the higher content of glucose-containing high mannose residues in EG2-hFc mAbs than in DP-12 (66% compared to 56%).

Typically, IgG1 mAbs have a low level of galactosylation, as with DP-12 (GI = 0.30), however EG2-hFc displayed a high level of galactosylation (GI = 0.72), with low levels of sialylation (32). This GI difference may reflect greater accessibility of the glycan for the addition of galactose and terminal sialic acid, or possible differences in the galactosyltransferase and sialyltransferase activities in the two cell lines. The amino acid sequence of the Fc region of EG2-hFc is similar to that of DP-12 suggesting that the hinge region and the attached V_{HH} of EG2-hFc may affect the folding conformation of the Fc region of the mAb. It has been reported that the protein backbone of the antibody directs the glycosylation, limiting the processing to biantennary complex glycans with little sialic acid (60), and mutation of certain residues can lead to an increase in galactosylation and sialylation (25). EG2-hFc has two mutations in the Fc region, D270G and Y278H, which may potentially affect interactions with the Asn297 glycan, thus causing the observed increase in galactosylation and sialylation. Mutations F241A, F243A, V264A, D265A, and R301A in the Fc region of IgG3 and F243A of IgG1 cause an increase in galactosylation and sialylation with up to 85% of glycan structures containing two terminal galactose sugars and 73% of these structures containing one or two sialic acid residues (25, 61).

N-glycosylation analysis of the mAbs expressed in the presence of the processing inhibitors revealed that the effects on glycosylation for both EG2-hFc and DP-12 are similar. Kif increased the amount of high mannose glycans in both mAbs, Cas resulted in high mannose with attached glucose glycans and Swa allowed for fucosylation of hybrid structures with and without sialylation. A high-resolution structure of an IgG Fc region suggested that with Swa-treatment

the hybrid glycan structure remains “trapped” in an artificial conformation that allows increased accessibility of the galactosyltransferase and sialyltransferase to add a galactose and sialic acid, respectively, onto the N-glycan (62).

The secondary structure of the variously glycosylated mAbs were also observed to be unchanged. It has shown that the particular position of the α2,6 sialic acid of the Fc glycan can cause shifts in the CD spectra, allowing a more closed C_H2 conformation as the sialic acid of the α1,3 arm interacts with the protein core of the C_H2 domain, representing an anti-inflammatory state (63). Since CHO cells do not have the α2,6 sialyltransferase enzyme, EG2-hFc α2,3 sialylation is only present in both the control mAb and the hybrid structure. The differences in CD spectra observed between EG2-hFc and DP-12 mAbs at wavelengths 215 – 225 nm (**Figure 4.6**) reflects the presence of light chains linked through disulphide bonds to the C_H1 domain in DP-12, which are absent in EG2-hFc (**Figure 4.10**). The CD spectra for variously treated and untreated mAbs presented no significant differences suggesting that removal of sugars do not cause detectable conformational changes in both mAbs. If alteration of the N-glycan only caused a local conformational change, the resulting change in structure would not be able to be resolved by CD (64).

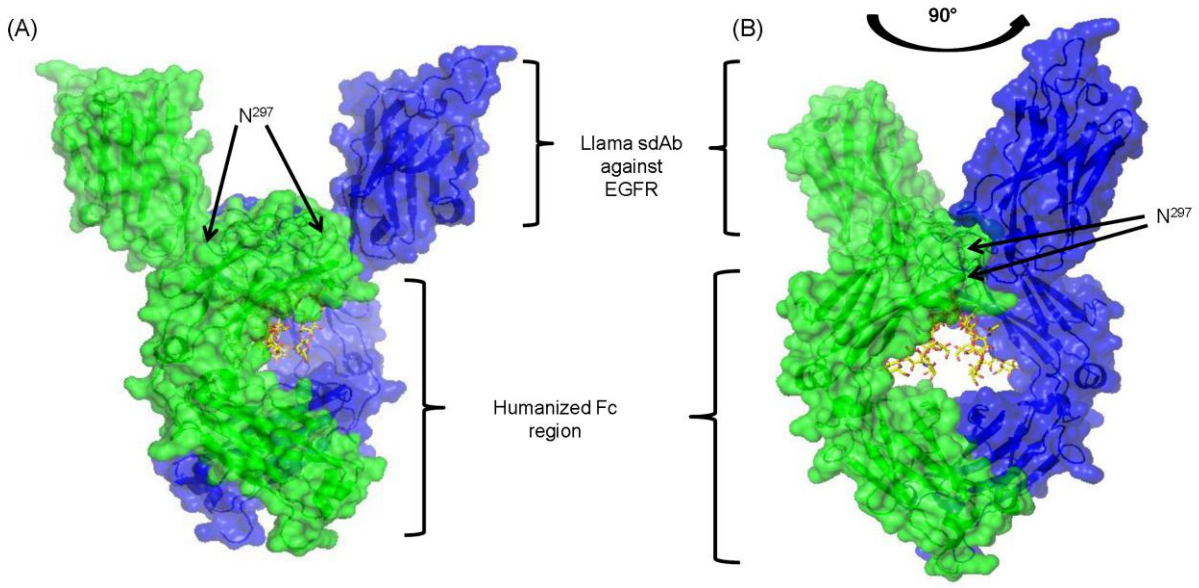

Figure 4.10 Structural interpretation of N-glycosylated EG2-hFc using homology modeling

(A) The overall model of the two EG2-hFc chains (shown in green and blue) bound together by disulphide bonds to form the well-known antibody dimer. The N-terminal EGFR binding region is found at the top of the figure which is then attached to the humanized Fc-fragment at the bottom. The position of Asn297 is pointed out as the attachment point for the N-glycosylation which is displayed as stick-and-ball models. (B) The same structural model is also rotated 90° to expose the position of the N-glycosylation in the Fc "pocket".

To avoid any potential side effects of therapeutic antibodies, it is desirable to achieve an aggregation free preparation. DLS analysis on the effects of controlled glycan modifications on the mAbs showed no changes in terms of homogeneity. Comparing the unmodified mAbs, it was observed that EG2-hFc had a smaller R_h value (20% reduction) compared to full IgG1. This can be explained by its smaller size (80 kDa compared to 150 kDa), due to the absence of the C_H1 domains (**Figure 4.10**) that suggest the presence of the terminal sugars on the N-glycan can affect the conformation of the Fc pocket (47, 65). Similarly, AUC results also indicated that both preparations of mAbs with the presence and absence of inhibitors are highly pure. The AUC data collected at multiple concentrations for all preparations suggest concentration independent behaviour of mAbs in solution (inset to **Figure 4.8A** and **B**).

The binding affinity of EG2-hFc and DP-12 for the FcγRI receptor was affected by the N-glycosylation at Asn297 (**Figure 4.9**). Kif treatment (producing M9 glycans) decreased the affinity of EG2-hFc to FcγRI by 79% (**Table 4.3**), but upon the addition of one glucose sugar to the end of the M9 glycan, the affinity was restored to that of the control mAb. The EG2-hFc data showed a greater effect of Man9 glycosylation on binding affinity compared with a previous study that found IgG1 containing Man8/9 glycans to have minimal effect on FcγRI binding affinity (20). On the other hand, the presence of hybrid glycan structures decreased the binding affinity of both the mAbs (140% decrease for EG2-hFc and 60% decrease for DP-12 (**Figure 4.9, Table 4.3**). The decrease in affinity can be explained by the increase in sialic acid that is observed in the Swa-treated mAbs compared to the untreated samples. As EG2-hFc has a greater proportion of sialylation than DP-12, it has a larger reduction in binding affinity to FcγRI. Bowden *et al*. proposed that the artificial conformation of the hybrid glycan in the Fc region may alter the conformation of the molecule to reduce the binding to the FcγRI (62). Asialo and asialo-agalacto IgG produced in lectin mutants showed very similar K_D values for FcγRI binding compared to the control IgG, but an IgG with only Man5 glycans had a 7-fold decrease in affinity (57). Thus, it is not necessarily the loss of galactose and sialic acid that influences receptor binding, but the presence of a truncated high mannose arm could also affect the binding. The lower GI (**Table 4.1**) for all the inhibitor-treated EG2-hFc mAbs may account for the reduced binding affinity that is observed for the FcγRI receptor.

X-ray crystallography studies have shown that the addition of galactose onto the N-glycans causes the C_H2-C_H2 interface to form an “open” conformation. This increases the accessibility of the binding site to FcγRs, which changes the effector function of mAbs (47, 66). However, structural studies performed using NMR suggests that the glycans in the Fc region are

not as constrained as previously deduced from X-ray crystallography. The glycan can be found to exist in two different states, with the more exposed glycan termini allowing for increased sialylation. *In vitro* sialylation of the IgG is not observed to be inhibited or have altered specificity due to the presence of the Fc polypeptide, and instead is able to successfully add α2-6 linked sialic acids on all the glycans (67).

Removal of the N-glycan resulted in a 118-fold decrease in binding affinity for DP-12 to FcγRI, but only a 6-fold decrease for EG2-hFc (**Table 4.3**). This suggests that the conformation of the chimeric heavy chain EG2-hFc mAb is less sensitive to non-glycosylation than the full sized IgG1. Non-glycosylation of IgG mAbs is known to significantly reduce binding to FcγRI and even abolish the binding to other Fcγ receptors as the glycan structure is necessary to maintain the conformation of the FcγR binding sites (56, 58, 68). Complete removal of IgG glycans induces conformational changes in the C_H2 region (closing the Fc pocket) which has been measured by differential scanning calorimetry (69), hydrogen/deuterium exchange mass spectrometry (68), and X-ray crystallography (70). These conformational changes make it difficult, if not impossible for Fc receptors to bind. However, this is in contrast to more recent work with small angle X-ray scattering (SAXS) which provided low-resolution shape information that found non-glycosylation allows a more open C_H2 orientation (71). The authors suggest that the non-glycosylated form may be able to adopt more than one conformation which would again make binding of the Fc receptor difficult. Both explanations support the data observed for the removal of the N-glycan on EG2-hFc and DP-12 with the EG2-hFc mAb capable of stabilizing its Fc region. Since the Fc region is 97.8% identical between the two mAbs, it is possible that the hinge region between the Fc and Fv region or even the Fv region

itself also plays a role in the stability of the C_H2-C_H2 distance (47) affecting the interaction to Fc receptors.

Specific glycan structures may exhibit a more finely-tuned control over Fc receptor binding than just maintaining the Fc region in a specific conformation. It has been proven that changes to the glycan structure, such as the presence of high mannose glycans can have up to a 79% decrease on FcγRI binding by EG2-hFc without measureable changes in the secondary structure, hydrodynamic radius, or sedimentation coefficient. On the other hand, that similar glycan pattern was found not to affect the binding affinity of DP-12. This suggests that the hinge and variable region may influence how the Fc region is affected by glycan changes. The variable region attached to the Fc region is known to cause an increase in the melting temperature (63, 72), and therefore the presence of this highly thermostable V_{HH} domains could create a stabilizing effect on the Fc region. Melting temperatures display that even at 80 kDa, EG2-hFc remains stable and is not affected with changes in the glycosylation (**Table 4.2**). As binding of the Fc receptors occurs near the hinge region (24), the length of the hinge and subsequently the position of the variable region could also be a factor affecting the accessibility of the Fc receptor to the Fc region. The hinge region of EG2-hFc positions the variable domain much closer to the Fc region than what would be observed in a typical IgG1, which could cause this enhanced stabilization.

Our biochemical and biophysical analysis of EG2-hFc and DP-12 mAbs treated with glycoprotein processing inhibitors suggest that the secondary structure of the mAbs remains unchanged and that all the preparations are monodispersed and aggregation free. However, the changes in N-glycosylation pattern affect the binding properties to FcγRI as observed by SPR. The results for DP-12 are in agreement with previously published data for full IgG mAbs. Our

data suggest that the design of EG2-hFc imparts a stabilizing effect on the Fc region resulting in unique binding properties. This stabilizing effect is hypothesized to be a combination of the hinge region length and the properties of the V_{HH} domain attached to the Fc. Therefore, we strongly believe that the smaller sized (80 kDa) EG2-hFc is an attractive candidate for biopharmaceutical applications and that the enhanced stabilization it imparts onto this mAb due to glycan changes is also advantageous.

References

1. Scott, A. M., Wolchok, J. D., and Old, L. J. (2012) Antibody therapy of cancer. *Nat. Rev. Cancer*. **12**, 278–87
2. Rosenfeld, P. J., Brown, D. M., Heier, J. S., Boyer, D. S., Kaiser, P. K., Chung, C. Y., and Kim, R. Y. (2006) Ranibizumab for Neovascular Age-Related Macular Degeneration. *N. Engl. J. Med.* **355**, 1419–1431
3. Van Dullemen, H. M., Van Deventer, S. J. H., Hommes, D. W., Bijl, H. A., Jansen, J., Tytgat, G. N. J., and Woody, J. (1995) Factor Chimeric Monoclonal Antibody (cA2). *Gastroenterology*. **109**, 129–135
4. Bruno, V., Battaglia, G., and Nicoletti, F. (2011) The advent of monoclonal antibodies in the treatment of chronic autoimmune diseases. *Neurol. Sci.* **31**, 283–288
5. Lux, A., Yu, X., Scanlan, C. N., and Nimmerjahn, F. (2013) Impact of immune complex size and glycosylation on IgG binding to human FcγRs. *J. Immunol.* **190**, 4315–23
6. Romain, G., Senyukov, V., Rey-Villamizar, N., Merouane, A., Kelton, W., Liadi, I., Mahendra, A., Charab, W., Georgiou, G., Roysam, B., Lee, D. A., and Varadarajan, N. (2014) Antibody Fc engineering improves frequency and promotes kinetic boosting of serial killing mediated by NK cells. *Blood*. **124**, 3241–3249
7. Abès, R., and Teillaud, J. L. (2010) Impact of glycosylation on effector functions of therapeutic IgG. *Pharmaceuticals*. **3**, 146–157
8. Shields, R. L., Lai, J., Keck, R., O'Connell, L. Y., Hong, K., Meng, Y. G., Weikert, S. H. a, and Presta, L. G. (2002) Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* **277**, 26733–40

9. Mori, K., Kuni-Kamochi, R., Yamane-Ohnuki, N., Wakitani, M., Yamano, K., Imai, H., Kanda, Y., Niwa, R., Iida, S., Uchida, K., Shitara, K., and Satoh, M. (2004) Engineering Chinese hamster ovary cells to maximize effector function of produced antibodies using FUT8 siRNA. *Biotechnol. Bioeng.* **88**, 901–8

10. Nimmerjahn, F., and Ravetch, J. V (2007) Antibodies, Fc receptors and cancer. *Curr. Opin. Immunol.* **19**, 239–45

11. Berger, M., Kaup, M., and Blanchard, V. (2011) Protein Glycosylation and Its Impact on Biotechnology. *Adv. Biochem. Eng. Biotechnol.* **127**, 165–185

12. Chung, S., Quarmby, V., Gao, X., Ying, Y., Lin, L., Reed, C., Fong, C., Lau, W., Qiu, Z. J., Shen, A., Vanderlaan, M., and Song, A. (2012) Quantitative evaluation of fucose reducing effects in a humanized antibody on Fcγ receptor binding an antibody-dependent cell-mediated cytotoxicity activities. *MAbs*. **4**, 326–340

13. Elbein, A. D., Tropea, J. E., Mitchell, M., and Kaushal, G. P. (1990) Kifunensine , a Potent Inhibitor of the Glycoprotein Processing Mannosidase I. *J. Biol. Chem.* **265**, 15599–15605

14. Van Leeuwen, J. E. M., and Kearse, K. P. (1997) Reglucosylation of N-Linked Glycans Is Critical for Calnexin Assembly with T Cell Receptor (TCR) Proteins but Not TCR Proteins. *J. Biol. Chem.* **272**, 4179–4186

15. Weng, S., and Spiro, R. G. (1996) Endoplasmic reticulum kifunensine-resistant alpha-mannosidase is enzymatically and immunologically related to the cytosolic alpha-mannosidase. *Arch. Biochem. Biophys.* **325**, 113–23

16. Crispin, M., Harvey, D. J., Chang, V. T., Yu, C., Aricescu, a. R., Jones, E. Y., Davis, S. J., Dwek, R. a., and Rudd, P. M. (2006) Inhibition of hybrid- and complex-type

glycosylation reveals the presence of the GlcNAc transferase I-independent fucosylation pathway. *Glycobiology*. **16**, 748–756

17. Rothman, R. J., Perussia, B., Herlyn, D., and Warren, L. (1989) Antibody-dependent cytotoxicity mediated by natural killer cells is enhanced by castanospermine-induced alterations of IgG glycosylation. *Mol. Immunol.* **26**, 1113–23

18. Kanda, Y., Yamada, T., Mori, K., Okazaki, A., Inoue, M., Kitajima-Miyama, K., Kuni-Kamochi, R., Nakano, R., Yano, K., Kakita, S., Shitara, K., and Satoh, M. (2006) Comparison of biological activity among nonfucosylated therapeutic IgG1 antibodies with three different N-linked Fc oligosaccharides: The high-mannose, hybrid, and complex types. *Glycobiology*. **17**, 104–118

19. van Berkel, P. H. C., Gerritsen, J., van Voskuilen, E., Perdok, G., Vink, T., van de Winkel, J. G. J., and Parren, P. W. H. I. (2010) Rapid production of recombinant human IgG With improved ADCC effector function in a transient expression system. *Biotechnol. Bioeng.* **105**, 350–357

20. Yu, M., Brown, D., Reed, C., Chung, S., Lutman, J., Stefanich, E., Wong, A., Stephan, J. P., and Bayer, R. (2012) Production, characterization, and pharmacokinetic properties of antibodies with N-linked mannose-5 glycans. *MAbs*. **4**, 475–487

21. Olovnikova, N. I., Grigorieva, O. V, and Petrov, A. V (2012) Effector Properties and Glycosylation Patterns of Recombinant Human Anti-D-IgG1 Antibodies Produced by Human PER.C6® Cells. *Immunol. Microbiol.* **154**, 245–249

22. Li, X., Ptacek, T. S., Brown, E. E., and Edberg, J. C. (2009) Fcγ Receptors: Structure, Function and Role as Genetic Risk Factors in SLE. *Genes Immun*. **10**, 380–389

23. Bevaart, L., Goldstein, J., Vitale, L., Russoniello, C., Treml, J., Zhang, J., Graziano, R. F.,

Leusen, J. H. W., van de Winkel, J. G. J., and Keler, T. (2006) Direct targeting of genetically modified tumour cells to Fc gammaRI triggers potent tumour cytotoxicity. *Br. J. Haematol.* **132**, 317–25

24. Lu, J., Ellsworth, J. L., Hamacher, N., Oak, S. W., and Sun, P. D. (2011) Crystal structure of Fcγ receptor I and its implication in high affinity γ-immunoglobulin binding. *J. Biol. Chem.* **286**, 40608–13

25. Lund, J., Takahashi, N., Pound, J. D., Goodall, M., and Jefferis, R. (1996) Multiple Interactions of IgG with Its Core Oligosaccharide Can Modulate Recognition by Complement and Human Fcy Receptor I and Influence the Synthesis of Its Oligosaccharide Chains. *J. Immunol.* **157**, 4963–4969

26. Dashivets, T., Thomann, M., Rueger, P., Knaupp, A., Buchner, J., and Schlothauer, T. (2015) Multi-Angle Effector Function Analysis of Human Monoclonal IgG Glycovariants. *PLoS One*. **10**, e0143520

27. Zhang, J., Liu, X., Bell, A., To, R., Baral, T. N., Azizi, A., Li, J., Cass, B., and Durocher, Y. (2009) Transient expression and purification of chimeric heavy chain antibodies. *Protein Expr. Purif.* **65**, 77–82

28. Harmsen, M. M., and De Haard, H. J. (2007) Properties, production, and applications of camelid single-domain antibody fragments. *Appl. Microbiol. Biotechnol.* **77**, 13–22

29. Baselga, J. (2001) The EGFR as a target for anticancer therapy--focus on cetuximab. *Eur. J. Cancer*. **37 Suppl 4**, S16-22

30. Addeo, R., Caraglia, M., Cerbone, D., Frega, N., Cimmino, G., Abbruzzese, A., and Del Prete, S. (2010) Panitumumab: a new frontier of target therapy for the treatment of metastatic colorectal cancer. *Expert Rev. Anticancer Ther.* **10**, 499–505

31. Bell, A., Wang, Z. J., Arbabi-Ghahroudi, M., Chang, T. a, Durocher, Y., Trojahn, U., Baardsnes, J., Jaramillo, M. L., Li, S., Baral, T. N., O'Connor-McCourt, M., Mackenzie, R., and Zhang, J. (2010) Differential tumor-targeting abilities of three single-domain antibody formats. *Cancer Lett.* **289**, 81–90

32. Liu, B., Spearman, M., Doering, J., Lattová, E., Perreault, H., and Butler, M. (2014) The availability of glucose to CHO cells affects the intracellular lipid-linked oligosaccharide distribution, site occupancy and the N-glycosylation profile of a monoclonal antibody. *J. Biotechnol.* **170**, 17–27

33. Agrawal, V., Slivac, I., Perret, S., Bisson, L., St-laurent, G., Murad, Y., Zhang, J., and Durocher, Y. (2012) Stable expression of chimeric heavy chain antibodies in CHO cells. *Methods Mol. Biol.* **911**, 287–303

34. Raymond, C., Robotham, A., Kelly, J., Lattova, E., Perreault, H., and Durocher, Y. (2012) Production of Highly Sialylated Monoclonal Antibodies. in *Glycosylation*, 10.5772/51301

35. Dorion-Thibaudeau, J., St-Laurent, G., Raymond, C., De Crescenzo, G., and Durocher, Y. (2016) Biotinylation of the Fcγ receptor ectodomains by mammalian cell co-transfection: application to the development of a surface plasmon resonance-based assay. *J. Mol. Recognit.* **29**, 60–69

36. Sunley, K., Tharmalingam, T., and Butler, M. (2008) CHO cells adapted to hypothermic growth produce high yields of recombinant beta-interferon. *Biotechnol. Prog.* **24**, 898–906

37. Furukawa, K., and Ohsuye, K. (1998) Effect of culture temperature on a recombinant CHO cell line producing a C-terminal α-amidating enzyme. *Cytotechnology*. **26**, 153–64

38. Tom, R., Bisson, L., and Durocher, Y. (2008) Purification of His-Tagged Proteins Using Fractogel-Cobalt. *Cold Spring Harb. Protoc.* 10.1101/pdb.prot4980

39. Guile, G. R., Rudd, P. M., Wing, D. R., Prime, S. B., and Dwek, R. a (1996) A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. *Anal. Biochem.* **240**, 210–26

40. Royle, L., Radcliffe, C. M., Dwek, R. A., and Rudd, P. M. (2006) Detailed Structural Analysis of N -Glycans Released From Glycoproteins in SDS-PAGE Gel Bands Using HPLC Combined With Exoglycosidase Array Digestions. in *Methods in Molecular Biology*, 347th Ed. (Borckhausen, I. (Humana P. I. . ed), pp. 125–143, Humana Press, Totowa, NJ, **347**, 125–143

41. Patel, T. R., Nikodemus, D., Besong, T. M. D., Reuten, R., Meier, M., Harding, S. E., Winzor, D. J., Koch, M., and Stetefeld, J. (2015) Biophysical analysis of a lethal laminin alpha-1 mutation reveals altered self-interaction. *Matrix Biol.* **49**, 1–13

42. Dam, J., and Schuck, P. (2004) Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. *Method Enzymol.* **384**, 185–212

43. Schuck, P. (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.* **75**, 1503–1512

44. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, A. J. R. S. E., and Horton, J. C. eds), pp. 90–125, Royal Society of Chemistry, Cambridge, United Kingdom

45. Dorion-Thibaudeau, J., Raymond, C., Lattová, E., Perreault, H., Durocher, Y., and De Crescenzo, G. (2014) Towards the development of a surface plasmon resonance assay to evaluate the glycosylation pattern of monoclonal antibodies using the extracellular

domains of CD16a and CD64. *J. Immunol. Methods*. **408**, 24–34

46. Saphire, E. O., Parren, P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R., and Wilson, I. A. (2001) Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science*. **293**, 1155–9

47. Krapp, S., Mimura, Y., Jefferis, R., Huber, R., and Sondermann, P. (2003) Structural Analysis of Human IgG-Fc Glycoforms Reveals a Correlation Between Glycosylation and Structural Integrity. *J. Mol. Biol.* **325**, 979–989

48. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis*. **18**, 2714–2723

49. Šali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815

50. Laskowski, R. A., Chistyakov, V. V., and Thornton, J. M. (2004) PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* **33**, D266–D268

51. Whitmore, L., and Wallace, B. a (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668-73

52. Zheng, K., Bantog, C., and Bayer, R. (2011) The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs*. **3**, 568–76

53. Deo, S., Patel, T. R., Dzananovic, E., Booy, E. P., Zeid, K., McEleney, K., Harding, S. E., and McKenna, S. A. (2014) Activation of 2’ 5’-oligoadenylate synthetase by stem loops at the 5’-end of the West Nile virus genome. *PLoS One*. **9**, 1–11

54. Kayser, V., Chennamsetty, N., Voynov, V., Forrer, K., Helk, B., and Trout, B. L. (2011) Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. *Biotechnol. J.* **6**, 38–44

55. Latypov, R. F., Hogan, S., Lau, H., Gadgil, H., and Liu, D. (2012) Elucidation of acid-induced unfolding and aggregation of human immunoglobulin IgG1 and IgG2 Fc. *J. Biol. Chem.* **287**, 1381–1396

56. Mimura, Y., Sondermann, P., Ghirlando, R., Lund, J., Young, S. P., Goodall, M., and Jefferis, R. (2001) Role of oligosaccharide residues of IgG1-Fc in Fc gamma RIIb binding. *J. Biol. Chem.* **276**, 45539–45547

57. Wright, A., and Morrison, S. L. (1998) Effect of C2-associated carbohydrate structure on Ig effector function: studies with chimeric mouse-human IgG1 antibodies in glycosylation mutants of Chinese hamster ovary cells. *J. Immunol.* **160**, 3393–3402

58. Yamaguchi, Y., Nishimura, M., Nagano, M., Yagi, H., Sasakawa, H., Uchida, K., Shitara, K., and Kato, K. (2006) Glycoform-dependent conformational alteration of the Fc region of human immunoglobulin G1 as revealed by NMR spectroscopy. *Biochim. Biophys. Acta - Gen. Subj.* **1760**, 693–700

59. Lu, J., Chu, J., Zou, Z., Hamacher, N. B., Rixon, M. W., and Sun, P. D. (2015) Structure of FcγRI in complex with Fc reveals the importance of glycan recognition for high-affinity IgG binding. *Proc. Natl. Acad. Sci.* **112**, 833–838

60. Barb, A. W., and Prestegard, J. H. (2011) NMR analysis demonstrates immunoglobulin G N-glycans are accessible and dynamic. *Nat. Chem. Biol.* **7**, 147–153

61. Raymond, C., Robotham, A., Spearman, M., Butler, M., Kelly, J., and Durocher, Y. (2015) Production of α2,6-sialylated IgG1 in CHO cells. *MAbs*. **7**, 571–583

62. Bowden, T. A., Baruah, K., Coles, C. H., Harvey, D. J., Yu, X., Song, B.-D., Stuart, D. I., Aricescu, a R., Scanlan, C. N., Jones, E. Y., and Crispin, M. (2012) Chemical and structural analysis of an antibody folding intermediate trapped during glycan biosynthesis. *J. Am. Chem. Soc.* **134**, 17554–63

63. Sondermann, P., Pincetic, A., Maamary, J., Lammens, K., and Ravetch, J. V (2013) General mechanism for modulating immunoglobulin effector function. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9868–72

64. Zheng, K., Yarmarkovich, M., Bantog, C., Bayer, R., and Patapoff, T. W. (2014) Influence of glycosylation pattern on the molecular properties of monoclonal antibodies. *MAbs*. **63**, 649–658

65. Crispin, M., Bowden, T. A., Coles, C. H., Harlos, K., Aricescu, A. R., Harvey, D. J., Stuart, D. I., and Jones, E. Y. (2009) Carbohydrate and domain architecture of an immature antibody glycoform exhibiting enhanced effector functions. *J. Mol. Biol.* **387**, 1061–1066

66. Ha, S., Ou, Y., Vlasak, J., Li, Y., Wang, S., Vo, K., Du, Y., Mach, A., Fang, Y., and Zhang, N. (2011) Isolation and characterization of IgG1 with asymmetrical Fc glycosylation. *Glycobiology*. **21**, 1087–1096

67. Barb, A. W., Meng, L., Gao, Z., Johnson, R. W., Moremen, K., and Prestegard, J. H. (2012) NMR Characterization of Immunoglobulin G Fc Glycan Motion on Enzymatic Sialylation. *Biochemistry*. **51**, 4618–4626

68. Houde, D., Peng, Y., Berkowitz, S. A., and Engen, J. R. (2010) Post-translational Modifications Differentially Affect IgG1 Conformation and Receptor Binding. *Mol. Cell. Proteomics*. **9**, 1716–1728

69. Hristodorov, D., Fischer, R., Joerissen, H., Müller-Tiemann, B., Apeler, H., and Linden, L. (2012) Generation and Comparative Characterization of Glycosylated and Aglycosylated Human IgG1 Antibodies. *Mol. Biotechnol.* **53**, 326–35

70. Feige, M. J., Nath, S., Catharino, S. R., Weinfurtner, D., Steinbacher, S., and Buchner, J. (2009) Structure of the Murine Unglycosylated IgG1 Fc Fragment. *J. Mol. Biol.* **391**, 599–608

71. Borrok, M. J., Jung, S. T., Kang, T. H., Monzingo, A. F., and Georgiou, G. (2012) Revisiting the Role of Glycosylation in the Structure of Human IgG Fc. *ACS Chem. Biol.* **7**, 1596–1602

72. Liu, C., Vailhe, C., Samuel, N., and Min, Q. (2009) A simple and reproducible approach to characterize protein stability using rheology. *Int. J. Pharm.* **374**, 1–4

Chapter 5: Conclusion

5.1 Thesis summary

With over one third of the Canadian population expected to be diagnosed with cancer, advancement in the diagnosis and treatment of the various types of cancer becomes pertinent to our society (1). This thesis explored two unique proteins: a transcriptional activating factor, HMGA2, which has been proposed as a potential broad range biomarker, and a chimeric hcAb, EG2-hFc, a potential biotherapeutic. Each protein was investigated using biophysical techniques including CD, DLS, and SV-AUC in addition to other biochemical or biotechnical experiments. From this, I was able to obtain significant results which furthered the understanding of the role of these proteins in cancer.

HMGA2 consists of an acidic C-terminal tail with three AT-hook DNA-binding domains (2). According to the amino acid sequence, this protein was predicted to be intrinsically disordered which was further confirmed experimentally through CD, NMR and SAXS. Moreover, DLS and SV-AUC analysis determined that HMGA2 behaves as a monomer in solution (3). This result distinguished itself from mouse HMGA2 which is found in solution as a dimer through a disulphide linkage (4, 5). I then investigated the interaction that HMGA2 has with a 41.40 kDa RF. The size of the RF was specifically chosen to contain a sufficient number of residues by which the three AT-hook domains could bind. Contrary to previous studies which investigated smaller DNA segments or truncated versions of HMGA (6, 7), I was able to conclude from HSQC NMR and SAXS studies that all three AT-hook domains in HMGA2 bound to the RF in a 1:1 ratio (3). This supported previous results which suggested that all AT-hooks in HMGA2 are necessary for stabilizing RFs (8). Additionally, from the low-resolution models I obtained, I proposed that HMGA2 harnesses the stability of the double-stranded region

of the RF to stabilize the replication bubble and newly synthesized strands (3). This additional stabilization prevents the cells from entering apoptosis and instead allows them to continue cell proliferation. It is this continual cell growth in adult cells which forms cancerous tumours and the driving force of this study.

As mentioned previously, the high amounts of HMGA2 present in various cancers has labelled it is a potential biomarker in the diagnosis and prognosis of cancer. Much research on HMGA2 has been focused on determining its accuracy in diagnosing cancer in patients and detecting its presence in yet another type of cancer. These overwhelming results makes HMGA2 stand out from other biomarkers used in clinical settings. However, few studies have actually begun to investigate the molecular mechanism by which HMGA2 supports the formation of tumours. The results displayed in this thesis are the first implications as to the binding mechanism of HMGA2 with RF DNA. With further studies, one could then begin to truly understand how the replication process is stabilized and determine its suitability as a target for cancer therapy.

EG2-hFc was designed as a potential biotherapeutic that is smaller in size but with equivalent or increased cytotoxicity to the current biopharmaceuticals on the market (9, 10). To achieve high tumour killing capabilities, a glycoengineering approach was implemented. Castanospermine (11), kifunensine (12), and swainsonine (13) are glycoprotein processing inhibitors which were added to the growth media of CHO EG2-hFc cells to alter the N-glycosylation pattern produced on the final mAb product. I analyzed the effect of these changes to the hydrodynamic properties, secondary structure, and Fc binding affinities using the methods of DLS and SV-AUC, CD, and SPR, respectively. Furthermore, I investigated the effect on complete removal of the N-glycan and compared results to that of a full IgG1 mAb (DP-12). I

found that although the hydrodynamic properties and secondary structure of the mAbs did not significantly change with alteration of the N-glycosylation, binding affinities to FcγRI were affected (14). In both mAbs, hybrid glycans were found to reduce their binding affinity to FcγRI by approximately two-fold, with the presence of sialylated glycans further reducing binding. Interestingly, upon removal of the glycan at Asn297, EG2-hFc had only a marginal decrease in binding affinity when compared to previous studies of a full IgG1 (14–16). I proposed that this was due to the position of the variable region with respect to the Fc pocket containing the N-glycans. The shorter hinge positions the V_{HH} domain closer to the Fc region which could impart additional stability to the mAb when sugars are absent. Aside from the small size of EG2-hFc, this newly discovered capability of the mAb to bind FcγRI with sufficient affinity without N-glycans present makes it attractive for biopharmaceuticals.

Biotechnology has altered their focus to designing smaller versions of mAbs for therapy for many reasons; cheaper production, deeper tumour penetration, and increased avidity. However, few have been successful enough to make it through all phases of clinical trials. Variation in N-glycosylation is a culprit for causing adverse effects (17) and preventing a mAb candidate from making it to the biopharmaceutical market. As such, discovering in this thesis that non-glycosylated EG2-hFc still has significant binding to FcγRI (and most probably its other Fc receptors) demonstrates its uniqueness when compared to full IgG1 mAbs. This characteristic in addition to its small size is what marks this mAb as the new mAb biopharmaceutical which would have potential to minimize the adverse effects and enhance tumour penetration.

5.2 Future perspectives

In moving forward, there is still much to uncover about HMGA2 and EG2-hFc and their role in cancer diagnosis, treatment and chemoresistance. Although HMGA2 has a significant role in cancer, it is definitely under-studied at the protein level. My thorough analysis brought about some controversy when compared to previous studies, but also proposed some additional questions that should be investigated. On the other hand, the effect of N-glycans on mAb binding affinity to FcγRIIIa is deeply studied. Analyzing the glycan changes on a novel mAb and its interaction with FcγRI provided striking results which raised inquiries about how it compares to other studied mAbs. With this I have discussed the newly raised questions and presented suggestions about how one would accomplish this.

Through biophysical techniques I was able to determine the stoichiometry of HMGA2 and RF binding as well as produce a low-resolution model to describe the interaction. Following that X-ray crystallography should be employed to provide a high-resolution structure of the interaction. This can then be superimposed with the low-resolution model to have an accurate representation of the mechanism of interaction in solution. However, since HMGA2 is still found to be largely disordered even when bound to the RF, crystal formation might not be plausible. An alternate method to achieve a high-resolution structure would be to further the research using NMR. Although time consuming, HMGA2 has been shown to express in significant quantities under minimal growth conditions and might prove to be the more successful route. Obtaining this high-resolution information will provide information on the interaction sites between these two molecules leading to the mechanism by which HMGA2 stabilizes the RF. From this, we can begin to design molecules which block this interaction and presumably reduce tumour growth.

However, HMGA2 is also known to have protein binding partners. The question might then become whether HMGA2 works alone in the stabilization of stalled RFs *in vivo,* or whether it recruits some additional proteins to form a stabilizing complex. If this would be the case, then the protein-protein interactions could also be a therapeutic target. Additionally, if this protein were to be necessary for stabilization of RFs and promoting cell growth, then using both as biomarkers would enhance the accuracy of diagnosing cancer patients.

After analyzing the effect that the N-glycan presence and composition had on EG2-hFc binding to FcγRI, it would follow that this study be repeated with FcγRIIIa. However, the mAb sequence which was used for the binding studies with FcγRI had been previously mutated such that it would not bind to FcγRIIIa (14). As a result, these two point mutations would have to be rectified to investigate the glycan effect on binding to FcγRIIIa, a low-affinity receptor found on natural killer (NK) cells. NK cells are typical targets in ADCC assays where the cytotoxicity of a mAb is investigated (18). Therefore, instead of just exploring the effect of N-glycosylation on binding affinity, one could also look at the cytotoxic effect through ADCC. Furthermore, without altering the current EG2-hFc construct, the ADCC analysis could also be performed with neutrophils which are more abundant with FcγRI. Achieving this data will provide conclusive evidence as to the N-glycosylation composition that is necessary for high cytotoxicity.

With this information, we would then need to engineer a homogeneous population of EG2-hFc with the appropriate N-glycosylation and test its capability at penetrating and reducing tumour growth using *in vivo* studies. If this glycoengineered EG2-hFc has equivalent cytotoxicity and increased penetration compared to current biopharmaceuticals on the market, the research can develop into upscaling mAb production in preparation for clinical trials. With a

smaller mAb, therapeutic treatments should become more efficient and thorough then what is currently being used.

References

1. *Canadian Cancer Statistics 2016* (2016) Toronto, ON
2. Summer, H., Li, O., Bao, Q., Zhan, L., Peter, S., Sathiyanathan, P., Henderson, D., Klonisch, T., Goodman, S. D., and Dröge, P. (2009) HMGA2 exhibits dRP/AP site cleavage activity and protects cancer cells from DNA-damage-induced cytotoxicity during chemotherapy. *Nucleic Acids Res.* **37**, 4371–84
3. Krahn, N., Meier, M., To, V., McEleney, K., O'Neil, J., Klonisch, T., Patel, T. R., and Stetefeld, J. (2017) Nano-scale assembly of high-mobility group AT-hook 2 protein with DNA replication fork. *Biophys. J.***(Pending Revisions)**
4. Altschul, S. F., Madden, Thomas L. Schäffer, Alejandro A. Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
5. Frost, L., Baez, M. a. M., Harrilal, C., Garabedian, A., Fernandez-Lima, F., and Leng, F. (2015) The dimerization state of the mammalian high mobility group protein AT-Hook 2 (HMGA2). *PLoS One*. **10**, 1–18
6. Huth, J. R., Bewley, C. A., Nissen, M. S., Evans, J. N. S., Reeves, R., Gronenborn, A. M., and Clore, G. M. (1997) The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* **4**, 657–665
7. Fonfría-Subirós, E., Acosta-Reyes, F., Sapera, N., Pous, J., Subirana, J. A., and Campos, J. L. (2012) Crystal structure of a complex of DNA with one AT-Hook of HMGA1. *PLoS One*. **7**, e37120
8. Yu, H., Lim, H. H., Tjokro, N. O., Sathiyanathan, P., Natarajan, S., Chew, T. W., Klonisch, T., Goodman, S. D., Surana, U., and Dröge, P. (2014) Chaperoning HMGA2

protein protects stalled replication forks in stem and cancer cells. *Cell Rep.* **6**, 684–97

9. Bell, A., Wang, Z. J., Arbabi-Ghahroudi, M., Chang, T. a, Durocher, Y., Trojahn, U., Baardsnes, J., Jaramillo, M. L., Li, S., Baral, T. N., O'Connor-McCourt, M., Mackenzie, R., and Zhang, J. (2010) Differential tumor-targeting abilities of three single-domain antibody formats. *Cancer Lett.* **289**, 81–90
10. Butler, M. (2009) Strategies for the production of single-type glycoform monoclonal antibodies
11. Hashim, O. H., and Cushley, W. (1987) Role of processing of N-linked oligosaccharides in control of immunoglobulin secretion from rat hybridomas. *Mol. Immunol.* **24**, 1087–1096
12. Elbein, A. D., Tropea, J. E., Mitchell, M., and Kaushal, G. P. (1990) Kifunensine , a Potent Inhibitor of the Glycoprotein Processing Mannosidase I. *J. Biol. Chem.* **265**, 15599–15605
13. Hori, H., Kaushal, G. P., and Elbein, A. D. (1985) Fucosylation of Membrane Porteins in Soybean Cultured Cells. *Am. Soc. Plant Biol.* **77**, 687–694
14. Krahn, N., Spearman, M., Meier, M., Dorion-Thibaudeau, J., McDougall, M., Patel, T. R., De Crescenzo, G., Durocher, Y., Stetefeld, J., and Butler, M. (2017) Inhibition of glycosylation on a camelid antibody uniquely affects its FcγRI binding activity. *Eur. J. Pharm. Sci.* **96**, 428–439
15. Yamaguchi, Y., Nishimura, M., Nagano, M., Yagi, H., Sasakawa, H., Uchida, K., Shitara, K., and Kato, K. (2006) Glycoform-dependent conformational alteration of the Fc region of human immunoglobulin G1 as revealed by NMR spectroscopy. *Biochim. Biophys. Acta - Gen. Subj.* **1760**, 693–700

16. Houde, D., Peng, Y., Berkowitz, S. A., and Engen, J. R. (2010) Post-translational Modifications Differentially Affect IgG1 Conformation and Receptor Binding. *Mol. Cell. Proteomics*. **9**, 1716–1728

17. Solá, R. J., and Griebenow, K. (2011) Glycosylation of Therapeutic Proteins: An Effective Strategy to Optimiza Efficacy. *BioDrugs*. **24**, 9–21

18. Jimenez, I., Jedrzejewski, P. M., Exley, K., Sou, S. N., Kyriakopoulos, S., Polizzi, K. M., and Kontoravdi, C. (2012) Application of Quality by Design Paradigm to the Manufacture of Protein Therapeutics. in *InTech*, pp. 353–396, Creative Commons Attribution License, 10.5772/50261