



**Biophysical studies of dynamic CD4 changes  
implicated in HIV-1 infection**

---

Jennifer Anne CHANNELL

*A thesis submitted in fulfillment of the requirements  
for the degree of Doctor of Philosophy*

Keele University

March 2019

## PGR STUDENT: ACADEMIC HONESTY DECLARATION FORM

### To be completed by the student:

Section A: Student Details			
Name of student	Jennifer Channell	Student No.	14020321
Research Home:	ISTM <input type="checkbox"/> PCHS <input type="checkbox"/> IACS <input type="checkbox"/> HUMSS <input type="checkbox"/> NATSCI <input checked="" type="checkbox"/>		
Lead supervisor:	Prof. V. Trevor Forsyth		
Degree for which thesis being submitted:	Doctor of Philosophy	Submission Date: <i>Must comply with Regulation 2D</i>	29/06/2018
Title of Thesis:	Biophysical studies of dynamic CD4 changes implicated in HIV-1 infection		
This thesis contains confidential information and is subject to the protocol set down for the submission and examination of such a thesis: <i>(if yes please complete the relevant section of the Thesis Submission Form)</i>			Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>

Section B: Thesis Details	
<b>I certify that:</b>	
The thesis being submitted for examination is my own account of my own research	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
My research has been conducted ethically. Where relevant a letter from the approving body confirming that ethical approval has been given has been bound into the thesis as an Annex	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
The data and results presented are the genuine data and results actually obtained by me during the conduct of the research	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
Where any collaboration has taken place with one or more other researchers, I have included within an 'Acknowledgements' section in the thesis a clear statement of their contributions, in line with the relevant guidance document <i>(see note*)</i>	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
The greater portion of the work described in the thesis has been undertaken subsequent to my registration for the higher degree for which I am submitting for examination	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
Where part of the work described in the thesis has previously been incorporated in another thesis submitted by me for a higher degree (if any), this has been identified and acknowledged in the thesis	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
The thesis submitted is within the required word limit as specified in the Regulations	Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>
Total words in submitted thesis (including text and footnotes, but excluding references and appendices)	71711

Student Signature:		Date:	28/06/2018
--------------------	---	-------	------------

*“Pour prier comme il nous plaît, si nous sommes religieux. Pour écrire ce qu’il nous plaît, si nous sommes poètes.”*

Antoine de Saint-Exupéry



# Abstract

Cluster of differentiation 4 (CD4) is a receptor protein found on the surface of cells of the immune system and is the primary receptor of the human immuno-deficiency virus-1 (HIV-1) envelope glycoprotein 120 (gp120). The extracellular portion of CD4 is comprised of four immunoglobulin-like domains of which domains 1, 2 and 4 contain disulphide bonds. The disulphide bond in domain 2 is an allosteric disulphide bond which can alter protein function through redox shuffling. Reduction of this bond is essential for gp120 binding and subsequent HIV-1 host cell entry. However, reduced monomeric CD4 does not have any known physiological function in CD4's normal immune response. CD4 reduced in domain 2 is a functionally distinct redox isoform of CD4 implicated in HIV-1 infection and is therefore, a target for an anti-HIV-1 vaccine.

The aim of this work was to obtain biophysical and small-angle X-ray and neutron scattering (SAXS and SANS) data which would explain how reduction of the second domain disulphide bond affects CD4 structure and therefore its ability to bind to gp120. A novel cell-free protein expression (CFPE) protocol was developed to produce recombinant deuterium labelled and un-labelled wild-type two domain CD4 (2dCD4-WT). 2dCD4-WT produced using CFPE is demonstrated to be functional, correctly folded and suitable for SAXS and SANS by a series of biochemical and biophysical techniques. Ablation of the second domain disulphide bond is shown to cause relaxation of the domain so that 2dCD4-WT reduced in domain 2 has a smaller hydrodynamic volume than its fully oxidised and fully reduced counterparts. For the first time, deuterium labelled 2dCD4-WT in the gp120-bound and -unbound state is described by SANS contrast variation analysis. Finally, size-exclusion chromatography coupled to SAXS data on gp120-bound 2dCD4-WT is presented and validates the use of CFPE for the production of recombinant protein for small-angle scattering studies.



# Acknowledgements

There are a lot of people to thank in helping this thesis come to fruition!

Firstly, I would like to thank my supervisors Prof. Trevor Forsyth, Prof. Ed Mitchell and Dr. Michael Haertlein for their help and support, not just during my thesis but over the last six years. It is hard to believe I first stepped foot on the EPN campus as an undergraduate and now I am leaving with a PhD. You have provided me with such a wonderful opportunity and helped me believe in myself.

I would also like to thank Prof. Maria Papathanasopoulos at the HIV Pathogenesis Research Unit for welcoming me into her lab in South Africa. Thank you to everyone at the HPRU for providing me with support during the course of this thesis work. I have thoroughly enjoyed working with you all and I believe it has been a fruitful collaboration. I would like to extend a special thanks to Dr. Nichole Cerutti for welcoming me into her home during the first three months of my PhD. I would also like to thank Dr. Mark Killick for teaching me about mammalian cell culture, Dr. Gavin Owen for collaborating with me on the disulphide bond ablation work and Nancy Tumba for her help with the pseudo-viral neutralisation assay.

Thank you also to Drs. Bruno Tillier and Sandra Cortes from Synthelis, France for your help and collaboration. Vinesh, I think I owe you about 500 pints of beer for all the cell-free reactions you set-up for me! Without you I would never have got any meaningful results so thank you!

I would like to acknowledge Dr. Anne Martel for all her help with the SANS experiments on D22 at the ILL. I would also like to acknowledge Drs. Bart Van der Laer, Mark Tully and Martha Brennich for your assistance during the SAXS experiments on BM29 at the ESRF.

This work used the platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). I would like to extend my thanks to; Drs. Adrien Favier and Alicia Vallet from the NMR platform; Dr. Luca Signor from the

mass spectrometry platform; Dr. Caroline Mas from the biophysical platform and Dr. Jean-Pierre Andrieu from the amino acids determination platform, for assistance with and access to these platforms.

It goes without saying that the Life Sciences Group at the ILL have been a constant source of help and support throughout this journey that is a PhD. Merci Val d'avoir partagé 209 avec moi c'était un plaisir! Merci Juliette pour ta bonne humeur et tes méthodes rigoureuses. Merci surtout Martine pour tout ton aide pendant ma thèse. Tu as toujours été quelqu'un de calme pendant les périodes de stress. Merci Benjamin d'avoir partagé cette aventure avec moi je te souhaite le meilleur pour le Master ! Thanks to Dr. Ashley Jordan for your comments regarding this manuscript, helping me with the SAXS experiment when I broke my arm and for being my Keele point of reference! Thank you Lindsay for all the work you did for me when I broke my arm. Thank you Sarah for all the cake!

J'ai une spéciale dedicace pour tous mes copains d'ARTURIA. Nous avons passé des moments vraiment inoubliables ensemble... Votre créativité m'inspirera pour toujours. Je vous aime.

Merci à Léo tu m'as montré ta France : ses coins incroyables, sa cuisine délicieuse, sa riche culture, son peuple adorable et ses chiens de merde. Je t'aime fort !

Merci à tous mes colocataires pendant ces années. Je remercie surtout les colocataires du Château de La Tronche de m'avoir accueilli pendant les derniers mois de la période le plus intense : la rédaction.

Thank you to my Grenoble family "Regimented Fun" you are so important to me and were always there for me when my family couldn't be. You inspire me daily and I will miss you so much.

I'd also like to say a big thanks to all my friends at home who always welcomed me back with open arms. I have missed you over the past years. You always made it feel like I'd never been away and that, for me, is so important.

Finally, I would like to thank my family and especially my parents for their continued love and support. Thanks to my Father who supervised PhD students but never got to do one himself.



# Contents

<b>Abstract</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Human Immuno-deficiency Virus (HIV)	1
1.1.1 Discovery of HIV	2
1.1.2 Genome and Structure	3
1.1.3 Replicative Cycle	5
1.2 The primary host-cell/viral interaction	8
1.2.1 The viral glycoprotein 120 (gp120)	8
gp120 function in HIV-1 infection	8
Atomic resolution structure of gp120	8
1.2.2 The cluster of differentiation 4 (CD4) receptor protein	10
Biological function of CD4 and its role in HIV-1 infection	10
The atomic resolution structure of CD4	10
1.2.3 The CD4-gp120 interaction	14
1.2.4 gp120 and CD4 as a target for vaccines	15
1.3 Allosteric disulphide bonds and redox biology	17
1.3.1 Disulphide bond classification	17
1.3.2 Proteins containing allosteric disulphide bonds	18
1.4 The importance of redox biology in HIV entry	20
1.4.1 The role of CD4 redox biology in HIV entry	20
1.4.2 The role of gp120 redox biology in HIV entry	22
1.5 Rationale of the PhD thesis	24
1.6 A guide to the CD4 acronyms used in this thesis	27

<b>2</b>	<b>Experimental Methods</b>	<b>31</b>
2.1	Molecular Biology . . . . .	31
2.1.1	Plasmid DNA preparation . . . . .	31
	Plasmid DNA purification . . . . .	32
2.1.2	Protein expression . . . . .	34
	Bacterial protein expression . . . . .	34
	Cell-free protein expression . . . . .	35
	Mammalian protein expression . . . . .	36
2.2	Biochemical characterisation . . . . .	37
2.2.1	Protein purification . . . . .	37
	Affinity chromatography . . . . .	37
	Size-exclusion chromatography . . . . .	39
2.2.2	Concentration determination of proteins and nucleic acids . . . . .	40
2.2.3	Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)	41
2.2.4	Western blotting . . . . .	42
2.2.5	Enzyme-linked immunosorbant assay (ELISA) . . . . .	43
2.2.6	Pseudo-viral neutralisation assay . . . . .	43
2.2.7	N-terminal protein sequencing . . . . .	44
2.3	Biophysical characterisation . . . . .	46
2.3.1	Mass spectrometry . . . . .	46
	Liquid chromatography/electrospray ionisation time-of-flight mass spec-	
	trometry . . . . .	47
	Matrix assisted laser desorption/ionisation time-of-flight mass spectrom-	
	etry . . . . .	47
2.3.2	SEC-MALLS-RI . . . . .	48
2.3.3	1D Nuclear Magnetic Resonance . . . . .	49
2.4	Small angle scattering (SAS) . . . . .	51
2.4.1	Scattering theory . . . . .	52
	A monodisperse, ideal solution . . . . .	54
2.4.2	Small-angle X-ray scattering (SAXS) . . . . .	55
	The BM29 instrument at the ESRF . . . . .	55
2.4.3	Small-angle neutron scattering (SANS) . . . . .	56

	The D22 instrument configuration at the ILL . . . . .	57
	Contrast variation . . . . .	57
	Deuterium labelling of proteins . . . . .	60
2.4.4	Data analysis . . . . .	61
	Guinier analysis . . . . .	61
	Pair-distance distribution function (P(r)) . . . . .	62
	Kratky analysis . . . . .	62
<b>3</b>	<b>Concerning the expression and purification of 2dCD4</b>	<b>65</b>
3.1	Abstract . . . . .	65
3.2	Introduction . . . . .	66
3.3	Materials and methods . . . . .	68
3.3.1	Plasmid DNA preparation . . . . .	68
	Plasmid DNA mini prep . . . . .	69
	Plasmid DNA giga prep . . . . .	70
3.3.2	2dCD4 Expression in <i>Escherichia coli</i> . . . . .	71
	<i>E. coli</i> transformation . . . . .	71
	Expression of 2dCD4 in <i>E. coli</i> . . . . .	71
	Purification of 2dCD4 expressed in <i>E. coli</i> . . . . .	71
	2dCD4-WT expressed in <i>E. coli</i> refolding . . . . .	72
	Size-exclusion chromatography purification of refolded 2dCD4-WT . . . . .	72
3.3.3	Deuteration of 2dCD4 in <i>E. coli</i> . . . . .	73
	Adaptation of <i>E. coli</i> to deuterated growth conditions . . . . .	73
	Fed-batch fermentation to produce a high cell density culture . . . . .	73
3.3.4	2dCD4 Expression in <i>Brevibacillus choshinensis</i> . . . . .	74
	Immobilised metal affinity chromatography (IMAC) purification of 2dCD4- WT expressed in <i>B. choshinensis</i> . . . . .	74
	Size Exclusion Chromatography . . . . .	75
3.3.5	Deuteration of 2dCD4 in <i>B. choshinensis</i> . . . . .	75
	Expression tests of 2dCD4-WT in <i>B. choshinensis</i> in deuterated media . . . . .	75
	Purification tests of deuterated 2dCD4-WT . . . . .	76
3.3.6	Cell-free protein expression of 2dCD4 . . . . .	77

	Plasmid DNA preparation . . . . .	77
	Expression of 2dCD4-WT using CFPE . . . . .	77
	Purification of 2dCD4-WT produced using CFPE . . . . .	78
	Deuteration of 2dCD4-WT using CFPE . . . . .	78
3.3.7	Expression and purification of the 2dCD4 domain 2 disulphide bond knockout variant (2dCD4-D2A) . . . . .	78
	2dCD4-D2A expression tests . . . . .	79
	2dCD4-D2A purification test . . . . .	79
3.4	Results . . . . .	79
3.4.1	Expression of 2dCD4-WT in <i>E. coli</i> . . . . .	79
	Purification of 2dCD4-WT from inclusion bodies . . . . .	80
	Optimising 2dCD4-WT expression in <i>E. coli</i> . . . . .	82
3.4.2	Deuteration of 2dCD4-WT using <i>E. coli</i> . . . . .	83
	Deuteration of 2dCD4-WT in <i>E. coli</i> . . . . .	83
3.4.3	Expression and secretion of 2dCD4-WT from <i>B. choshinensis</i> . . . . .	85
	Purification of protiated 2dCD4-WT from <i>B. choshinensis</i> secretion . . . . .	85
	Deuterated 2dCD4-WT expression tests in <i>B. choshinensis</i> . . . . .	89
	Deuterated 2dCD4-WT expression tests in <i>B. choshinensis</i> using D-Silantes	91
	Purification tests of 2dCD4-WT expressed in D-Silantes from <i>B. choshinensis</i>	92
3.4.4	Cell-free protein expression and purification of 2dCD4-WT . . . . .	94
	Expression of protiated 2dCD4-WT . . . . .	94
	Expression of deuterated 2dCD4-WT . . . . .	97
	Scaling-up and purification of cell-free protein expression of protiated and deuterated 2dCD4-WT . . . . .	98
3.4.5	Expression of 2dCD4-D2A by CFPE . . . . .	102
	2dCD4-D2A purification . . . . .	103
	Stabilising 2dCD4-D2A produced by CFPE . . . . .	106
3.5	Discussion . . . . .	108
3.5.1	Poor 2dCD4-WT yield from refolded <i>E. coli</i> expressed inclusion bodies . .	108
3.5.2	Expression and secretion of 2dCD4-WT by <i>B. choshinensis</i> . . . . .	109
3.5.3	Production of protiated and deuterated 2dCD4-WT using cell-free pro- tein expression . . . . .	110

3.5.4	2dCD4-D2A variant production using CFPE . . . . .	111
3.5.5	Future work . . . . .	112
3.6	Conclusions . . . . .	112
<b>4</b>	<b>Biophysical characterisation of cell-free expressed 2dCD4</b>	<b>115</b>
4.1	Abstract . . . . .	115
4.2	Introduction . . . . .	116
4.3	Materials and methods . . . . .	118
4.3.1	2dCD4 expression and purification . . . . .	118
4.3.2	Gp120 expression and purification . . . . .	118
4.3.3	Functional assays . . . . .	118
	Enzyme-linked immunosorbant assay . . . . .	118
	Pseudo-viral neutralisation assay . . . . .	118
4.3.4	1D nuclear magnetic resonance . . . . .	120
4.3.5	LC-ESI-TOF Mass-spectrometry . . . . .	120
4.3.6	N-terminal protein sequencing . . . . .	121
4.3.7	SEC-MALLS-RI . . . . .	121
4.4	Results . . . . .	122
4.4.1	Enzyme-linked immunosorbant assay (ELISA) . . . . .	122
4.4.2	Pseudo-viral neutralisation assay . . . . .	123
4.4.3	One dimensional proton nuclear magnetic resonance (1D NMR) . . . . .	126
4.4.4	Mass spectrometry analysis of 2dCD4-WT . . . . .	127
	Percentage deuteration of 2dCD4-WT produced using CFPE . . . . .	129
4.4.5	N-terminal protein sequencing of 2dCD4-WT . . . . .	130
4.4.6	Size-exclusion chromatography coupled to multiple-angle laser light scattering and refractive index (SEC-MALLS-RI) . . . . .	130
4.5	Discussion . . . . .	133
4.5.1	2dCD4-WT produced using cell-free synthesis is functional and correctly folded . . . . .	133
4.5.2	Truncation products are produced using cell-free synthesis . . . . .	135
4.5.3	Future work . . . . .	137
4.6	Conclusions . . . . .	138

<b>5</b>	<b>The effect of disulphide bond ablation on 2dCD4-WT structure</b>	<b>139</b>
5.1	Abstract . . . . .	139
5.2	Introduction . . . . .	140
5.2.1	Previous studies of 2dCD4-WT by SAXS . . . . .	144
5.2.2	Experiments to test the effect of reduction on 2dCD4-WT structure . . . . .	144
5.3	Materials and Methods . . . . .	146
5.3.1	Recombinant protein expression and purification . . . . .	146
5.3.2	DTT titration . . . . .	146
5.3.3	SEC Purification of 2dCD4-WT redoximers . . . . .	146
	2dCD4-WT SEC Purification . . . . .	146
	Superdex 75 16/600 calibration . . . . .	146
	SDS-PAGE analysis of fractions . . . . .	147
5.3.4	SAXS analysis of the effect of DTT on the size of 2dCD4-WT . . . . .	147
	Concentration series . . . . .	147
	DTT titration . . . . .	147
	Redox SEC-SAXS . . . . .	148
	Treatment of the data . . . . .	148
	SEC-MALLS-RI . . . . .	149
5.4	Results . . . . .	149
5.4.1	DTT titration causes a shift in <i>E. coli</i> expressed and refolded 2dCD4-WT redoximer distribution . . . . .	149
5.4.2	Partially reduced 2dCD4-WT has a smaller hydrodynamic volume than oxidised 2dCD4-WT . . . . .	150
5.4.3	Choosing a 2dCD4-WT concentration . . . . .	152
5.4.4	Increasing DTT concentration causes a loss of 2dCD4-WT structure . . . . .	153
5.4.5	2dCD4-R2 is larger than 2dCD4-R1 . . . . .	157
5.5	Discussion . . . . .	164
5.5.1	Relating the data to the disulphide bonds in 2dCD4-WT . . . . .	166
5.5.2	2dCD4-WT previously measured by SAXS . . . . .	168
5.5.3	Future work . . . . .	169
5.6	Conclusion . . . . .	170

<b>6</b>	<b>Characterisation of gp120 and small-angle scattering analysis of the complex</b>	<b>171</b>
6.1	Abstract . . . . .	171
6.2	Introduction . . . . .	172
6.3	Part I: Biophysical characterisation of gp120 . . . . .	175
6.3.1	Materials and methods . . . . .	175
	Stably transfected HEK293 cell culture . . . . .	175
	gp120 expression . . . . .	175
	gp120 purification . . . . .	176
	Matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) . . . . .	176
	Size-exclusion chromatography coupled to multi-angle laser light scattering and refractive index (SEC-MALLS-RI) . . . . .	177
	gp120 monomer/dimer separation . . . . .	177
6.3.2	Results . . . . .	177
	gp120 purification . . . . .	177
	Determination of percentage glycosylation of gp120 by MALDI-TOF-MS .	178
	Estimation of the composition of the gp120 glycans . . . . .	181
	SEC-MALLS-RI shows a monomer/dimer distribution . . . . .	181
	gp120 dimers can be separated by treatment with DTT and SEC . . . . .	182
6.4	Part II: SANS analysis of d-2dCD4-WT, h-gp120 and the complex . . . . .	185
6.4.1	Match-out point determination of protiated gp120 . . . . .	185
	Sample preparation . . . . .	186
	Experimental contrast match-point determination of h-gp120 . . . . .	186
6.4.2	Determination of the contrast-match point of deuterated 2dCD4-WT . . .	187
	Sample preparation . . . . .	188
	Experimental contrast-match point determination of deuterated 2dCD4-WT . . . . .	189
6.4.3	Contrast variation analysis of the h-gp120/d-2dCD4-WT complex . . . .	189
	Sample preparation . . . . .	190
	Data treatment . . . . .	190
	Analysis of the size parameters from the SANS data . . . . .	191
	Molecular mass analysis from the SANS data . . . . .	192

	Analysis of the graphical representation of the d-2dCD4 bound and un-bound SANS data . . . . .	193
	Analysis of the graphical representation of the gp120 monomer and dimer SANS data . . . . .	196
	Comparison of the SANS and X-ray crystal structure data of 2dCD4-WT and the complex . . . . .	198
	<i>Ab initio</i> modelling of gp120-bound and -unbound 2dCD4-WT . . . . .	199
6.5	Part III: Small-angle X-ray scattering analysis of 2dCD4-WT, gp120 and the complex . . . . .	205
6.5.1	SAXS analysis of samples previously measured by SANS . . . . .	205
	Sample preparation . . . . .	205
	Sample measurement . . . . .	205
	Analysis of the size parameters of the SANS samples determined by SAXS	206
	Analysis of the graphical representation of the SAXS data of d-2dCD4-WT	206
	Analysis of the graphical representation of the SAXS data of the d-2dCD4-WT/h-gp120 complex . . . . .	208
	Analysis of the graphical representation of the SAXS data of the gp120 monomer and dimer . . . . .	209
6.5.2	Size-exclusion chromatography coupled to small angle X-ray scattering analysis of the 2dCD4-WT/gp120 complex . . . . .	212
	Sample preparation and measurement . . . . .	212
	Selecting the buffer and sample frames . . . . .	213
	Size parameter analysis of the complex determined from SEC-SAXS data	214
	Analysis of the graphical representation of the SEC-SAXS data of the complex . . . . .	215
6.6	Discussion . . . . .	216
6.6.1	Analysis of gp120 . . . . .	217
	gp120 is expressed as a monomer and a non-physiological dimer . . . . .	217
	gp120 glycosylation complicates analysis of SAS data . . . . .	217
6.6.2	gp120-bound and un-bound 2dCD4-WT analysis . . . . .	218
	2dCD4-WT is more stable when in complex with gp120 . . . . .	218



d-2dCD4-WT may exist in a more compact form when bound to gp120 as shown by SANS . . . . .	219
SAS data does not match that of the X-ray crystal structure . . . . .	220
The relationship between the structure of 2dCD4-WT, its dynamics and gp120 binding . . . . .	222
6.6.3 Using SAS to study 2dCD4-WT, gp120 and the complex . . . . .	224
SEC-SAS is a more suitable method for carrying out SAS measurements on 2dCD4-WT . . . . .	224
SAS data quality . . . . .	225
6.6.4 The use of <i>B. choshinensis</i> and CFPE as recombinant protein expression methods . . . . .	227
6.6.5 Future work . . . . .	227
6.7 Conclusions . . . . .	230
<b>7 Conclusions and future work</b>	<b>231</b>
7.1 Conclusions . . . . .	231
7.2 Future work . . . . .	234
<b>Bibliography</b>	<b>237</b>
<b>A Plasmids and cDNA sequences</b>	<b>251</b>
A.1 2dCD4 clones . . . . .	251
A.2 gp120 clones . . . . .	254
<b>B Buffer recipes</b>	<b>257</b>
B.1 Bacterial growth media . . . . .	257
B.1.1 Liquid growth media . . . . .	257
B.1.2 Enfors minimal medium . . . . .	257
B.1.3 Solid growth media . . . . .	258
B.2 Antibiotic stock solutions . . . . .	258
B.3 Purification buffers . . . . .	259
B.3.1 <i>E. coli</i> purification buffers . . . . .	259
B.3.2 <i>B. choshinensis</i> purification buffers . . . . .	259
B.3.3 CFPE purification buffers . . . . .	259

B.4	SDS-PAGE . . . . .	260
B.4.1	Coomassie blue staining . . . . .	261
B.4.2	Western blotting . . . . .	261
<b>C</b>	<b>Mass spectrometry analysis of 2dCD4-WT analogues</b>	<b>263</b>
C.1	Mass-spectra of 2dCD4-WT analogues . . . . .	263
C.2	Percentage deuteration determination of 2dCD4-WT . . . . .	271
<b>D</b>	<b>List of activities</b>	<b>273</b>
	Publications . . . . .	273
	Conferences with oral contribution . . . . .	273
	Awards . . . . .	273
	Courses and other conferences . . . . .	274
	Student days and user meetings . . . . .	274

# List of Figures

1.1	Map of the global distribution of HIV infection . . . . .	2
1.2	HIV-1 genome . . . . .	3
1.3	The HIV-1 structure . . . . .	4
1.4	HIV-1 replicative cycle . . . . .	7
1.5	The unglycosylated gp120 core . . . . .	9
1.6	The ectodomans of the CD4 receptor . . . . .	13
1.7	The gp120/2dCD4 binding site . . . . .	14
1.8	Broadly neutralising antibodies bound to conserved epitopes on the viral Env spike . . . . .	16
1.9	Schematic representation of a disulphide bond . . . . .	18
1.10	Schematic of PDI's association with CD4 and gp120 during viral entry . . . . .	24
2.1	Plasmid purification protocols . . . . .	33
2.2	Nickel affinity chromatography . . . . .	38
2.3	Size-exclusion chromatography . . . . .	40
2.4	Pseudo-viral neutralisation assay . . . . .	45
2.5	N-terminal sequencing schematic . . . . .	46
2.6	Sinapinic acid . . . . .	48
2.7	1D HET-SOFAST NMR quantitative scale . . . . .	51
2.8	The BioSAXS beamline BM29 at the ESRF . . . . .	56
2.9	The D22 diffractometer at the ILL . . . . .	57
2.10	Scattering-length density and match-out point of biological macromolecules . . . . .	59
2.11	Typical contrast variation experiment of a protein complex . . . . .	59
2.12	Typical Kratky plots for biomacromolecules of different compactness in solution . . . . .	63
3.1	Nickel IMAC purification of 2dCD4-WT from <i>E. coli</i> inclusion bodies . . . . .	80

3.2	Size-exclusion chromatogram and SDS-PAGE analysis of 2dCD4-WT refolded from <i>E. coli</i> expressed inclusion bodies . . . . .	82
3.3	Western blot analysis of 2dCD4-WT deuteration in <i>E. coli</i> . . . . .	84
3.4	Vivaflow@cassette versus ammonium sulphate precipitation concentration of expression medium . . . . .	86
3.5	SEC purification of 2dCD4-WT produced in <i>B. choshinensis</i> . . . . .	88
3.6	<i>B. choshinensis</i> growth curve in different media with differing levels of deuteration	89
3.7	Deuterated 2dCD4-WT expression tests in <i>B. choshinensis</i> . . . . .	90
3.8	2dCD4-WT epxression test in <i>B. choshinensis</i> cultured in Silantes media . . . . .	92
3.9	Analytical purification test of 2dCD4-WT expressed in <i>B. choshinensis</i> cultured at different levels of deuteration . . . . .	93
3.10	Western blot depicting N- and C-terminally TEV-His-tagged 2dCD4-WT expression at 20°C and 30°C . . . . .	95
3.11	Anti-His Western Blot showing the expression of N- and C-terminally TEV-His-tagged 2dCD4-WT under reducing and non-reducing conditions . . . . .	96
3.12	Anti-His Western blot of the expression of C-terminally TEV-His-tagged 2dCD4-WT in different reaction volumes, filtered using different filter sizes. . . . .	97
3.13	Anti-His Western Blot to determine the concentration of Mg <sup>2+</sup> ions and deuterated amino acids required for deuterated 2dCD4-WT expression. . . . .	98
3.14	Nickel IMAC purification of 50 mL CFPE reaction of h-2dCD4-WT . . . . .	99
3.15	Nickel IMAC purification of 50mL CFPE reaction of d-2dCD4-WT . . . . .	100
3.16	SEC purification of d-2dCD4-WT . . . . .	101
3.17	SEC purification of d-2dCD4-WT . . . . .	102
3.18	50mL 2dCD4-D2A CFPE nickel IMAC . . . . .	104
3.19	2dCD4-D2A SEC purification . . . . .	105
3.20	2dCD4-D2A expression test with diffrent temperatures, with and without agitation	106
3.21	2dCD4-D2A expression test with different detergent concentrations . . . . .	107
3.22	2dCD4-D2A test purification . . . . .	107
4.1	2dCD4-WT ELISA . . . . .	122
4.2	Pseudo-viral neutralisation assay of CD4 variants . . . . .	125
4.3	Standard 1D NMR experiment for CFPE produced 2dCD4-WT . . . . .	126

4.4	1D HETSOFAST NMR experiment for CFPE produced 2dCD4-WT . . . . .	127
4.5	Normalised SEC-MALLS-RI results of d-2dCD4-WT before and after freezing . .	131
4.6	SEC-MALLS-RI d-2dCD4-WT peak overlay . . . . .	132
5.1	SDS-PAGE analysis of 2dCD4-WT redox isomers . . . . .	141
5.2	SDS-PAGE analysis of gp120 binding to 2dCD4-WT . . . . .	142
5.3	SDS-PAGE of a 2dCD4-WT titration . . . . .	150
5.4	Separation of 2dCD4-WT redox isomers by SEC . . . . .	151
5.5	1D scattering curve of a 2dCD4-WT concentration series . . . . .	153
5.6	1D scattering curve of a 2dCD4-WT DTT titration . . . . .	154
5.7	Kratky plot of a 2dCD4-WT DTT titration . . . . .	156
5.8	DTT titration of 2dCD4-WT producing using CFPE . . . . .	157
5.9	2dCD4-WT with and without 50 mM DTT SEC-SAXS spectra comparison . . . .	158
5.10	$R_g$ and intensity versus frame number plot for 2dCD4-WT with and without 50 mM DTT . . . . .	158
5.11	1D SEC-SAXS scattering curves of 2dCD4-WT with and without 50 mM DTT . .	159
5.12	Pair-distance distribution function for 2dCD4-WT with and without 50 mM DTT	160
5.13	Kratky plot of 2dCD4-WT with and without 50mM DTT . . . . .	161
5.14	SEC-MALLS-RI data for 2dCD4-WT with and without 50 mM DTT . . . . .	163
5.15	Hydrodynamic volume size order of 2dCD4 redox isomers . . . . .	167
6.1	SDS-PAGE analysis of gp120 . . . . .	178
6.2	MALDI-TOF mass spectrometry analysis of gp120 . . . . .	180
6.3	SEC-MALLS-RI of gp120 <sub>ZA</sub> CAP45 . . . . .	182
6.4	SEC purification of the gp120 <sub>ZA</sub> CAP45 monomer and dimer . . . . .	183
6.5	Separation of gp120 CAP45 monomers and dimers . . . . .	184
6.6	gp120 <sub>ZA</sub> CAP45 contrast match-out point plot . . . . .	187
6.7	Deuterated 2dCD4-WT contrast match-point plot . . . . .	189
6.8	Log-log, Guinier plot, P(r) function and Kratky plots from SANS data of d- 2dCD4-WT and gp120 . . . . .	195
6.9	Log-log, Guinier plot, P(r) function and Kratky plots from SANS data of the gp120 monomer and dimer . . . . .	197
6.10	CRYSON analysis of the SANS data for d-2dCD4-WT and the complex . . . . .	199

6.11	<i>ab initio</i> model of d-2dCD4-WT in 0% D <sub>2</sub> O . . . . .	202
6.12	<i>ab initio</i> model of d-2dCD4-WT in 45.2% D <sub>2</sub> O . . . . .	203
6.13	<i>ab initio</i> model of d-2dCD4-WT in complex with gp120 . . . . .	204
6.14	Log-log, Guinier plot, P(r) function and Kratky plots from SAXS data of d-2dCD4-WT . . . . .	208
6.15	Log-log, Guinier plot, P(r) function and Kratky plots from SAXS data of the d-2dCD4-WT/h-gp120 complex . . . . .	209
6.16	Log-log, Guinier plot, P(r) function and Kratky plots from SAXS data of the gp120 monomer and dimer . . . . .	211
6.17	Buffer and sample frame selection from SEC-SAXS of the 2dCD4-WT/gp120 complex . . . . .	213
6.18	Log-log, Guinier plot, P(r) function and Kratky plots from SECSAXS data of the 2dCD4-WT/gp120 complex . . . . .	216
6.19	2Fo-Fc and Fo-Fc electron density maps of the disulphide bonds in domains 1 and 2 of 2dCD4-WT bound to gp120 . . . . .	221
6.20	Domain 2 and 4 C- $\alpha$ distances along strands $\beta$ C and $\beta$ F . . . . .	223
A.1	2d-CD4-WT and -D2A sequence alignment . . . . .	251
A.2	pNCMO2 vector map . . . . .	252
A.3	pET15b vector map . . . . .	252
A.4	pETm11 vector map . . . . .	253
A.5	piVEX2.3d vector map . . . . .	253
A.6	gp120 BAL and ZA CAP45 sequence alignment . . . . .	254
A.7	pcDNA 3.1(-)_A012 vector map . . . . .	255
A.8	pCIneo vector map . . . . .	255
C.1	Mass spectrum of <i>E. coli</i> expressed h-2dCD4-WT in 0 mM DTT . . . . .	264
C.2	Mass spectrum of <i>E. coli</i> expressed h-2dCD4-WT in 50 mM DTT . . . . .	265
C.3	Mass spectrum of <i>B. choshinensis</i> expressed h-2dCD4-WT in 0 mM DTT . . . . .	266
C.4	Mass spectrum of <i>B. choshinensis</i> expressed h-2dCD4-WT in 50 mM DTT . . . . .	267
C.5	Mass spectrum of CFPE produced h-2dCD4-WT in 0 mM DTT . . . . .	268
C.6	Mass spectrum of CFPE produced h-2dCD4-WT in 50 mM DTT . . . . .	269
C.7	Mass spectrum of CFPE produced d-2dCD4-WT in 0 mM DTT . . . . .	270

C.8	Mass spectrum of CFPE produced d-2dCD4-WT in 50 mM DTT . . . . .	271
C.9	Percentage deuteration determination of CFPE produced 2dCD4-WT . . . . .	272





# List of Tables

1.1	Conserved viral Env epitopes and their broadly neutralising antibodies . . . . .	15
1.2	Visual guide to the 2dCD4-WT acronyms . . . . .	27
1.3	Visual guide to the 2dCD4 variant proteins . . . . .	27
1.4	Visual guide to the 2dCD4 redox isomer acronyms . . . . .	28
1.5	Visual guide to the 2dCD4 variant proteins . . . . .	29
2.1	X-ray and neutron scattering lengths . . . . .	58
4.1	ELISA control values . . . . .	123
4.2	IC <sub>50</sub> values for 2dCD4-WT neutralisation of tier 1 and 2 HIV . . . . .	124
4.3	Expected vs experimental molecular masses of 2dCD4-WT analogues . . . . .	128
5.1	Schematic showing 2dCD4-WT redox isomers and disulphide bond knockout variants . . . . .	143
5.2	Published SAXS parameters for 2dCD4-WT . . . . .	144
5.3	Table summarising hydrodynamic volume analysis of 2dCD4-WT . . . . .	152
5.4	2dCD4-WT size parameters as a function of DTT concentration as determined by SAXS . . . . .	154
5.5	Size parameters of 2dCD4-WT treated with 0 mM and 50 mM DTT . . . . .	159
6.1	Published SAXS parameters for 2dCD4-WT, gp120 and the complex . . . . .	173
6.2	Estimation of the glycan composition of gp120 . . . . .	181
6.3	Experiment configuration for SANS experiments on D22 . . . . .	185
6.4	SANS parameters for d-2dCD4-WT, h-gp120 and their respective complex . . . . .	192
6.5	Dimensions of the damaver models of d-2dCD4-WT . . . . .	201
6.6	Experiment configuration for SAXS experiments on BM29 . . . . .	206
6.7	Size parameters for the SANS samples determined by SAXS . . . . .	207

6.8	SAXS parameters of 2dCD4-WT from different expression systems in complex with gp120 . . . . .	214
6.9	Domain 2 and 4 C- $\alpha$ distances along strands $\beta$ C and $\beta$ F . . . . .	224
B.1	Ingredients for 12% Tris-Tricine SDS-PAGE . . . . .	260

# List of Abbreviations

<b>1D</b>	<b>1 Dimensional</b>
<b>1dCD4-D1</b>	<b>One domain CD4 Domain 1</b>
<b>1dCD4-D2</b>	<b>One domain CD4 Domain 2</b>
<b>2dCD4-D1A</b>	<b>Two domain CD4 Domain 1 Cysteine to Alanine variant</b>
<b>2dCD4-D2A</b>	<b>Two domain CD4 Domain 2 Cysteine to Alanine variant</b>
<b>2dCD4-Ox</b>	<b>Two domain CD4 Oxidised</b>
<b>2dCD4-R1</b>	<b>Two domain CD4 Reduced in 1 domain</b>
<b>2dCD4-R2</b>	<b>Two domain CD4 Reduced in 2 domains</b>
<b>2dCD4-WT</b>	<b>Two domain CD4 Wild Type</b>
<b>2SY</b>	<b>2 times concentrated SoyTone Yeast extract expression medium</b>
<b>2YT</b>	<b>2 times concentrated Yeast extract pepTone expression medium</b>
<b>4dCD4</b>	<b>Four domain CD4 (Domains 1-4)</b>
<b>17b</b>	<b>anti-CD4bs (gp120) antibody</b>
<b>AIDS</b>	<b>Acquired ImmunoDeficiency Syndrome</b>
<b>au</b>	<b>arbitrary units</b>
<b>Au</b>	<b>Absorbance units</b>
<b>ALP</b>	<b>ALkaline Phosphatase</b>
<b>AMP</b>	<b>Adenoside MonoPhosphate</b>
<b>APC</b>	<b>Antigen Presenting Cell</b>
<b>ATP</b>	<b>Adenoside TriPhosphate</b>
$b_i$	<b>scattering length of N atoms</b>
$b_N$	<b>neutron scattering length</b>
$b_X$	<b>X-ray scattering length</b>
<b>BM29</b>	<b>Bending Magnet 29 (SAXS beamline at the ESRF)</b>
<b>bNAb</b>	<b>broadly Neutralising Antibodies</b>
<b>B<sub>0</sub></b>	<b>magnetic field</b>
<b>C</b>	<b>Concentration (g L<sup>-1</sup>)</b>
<b>CA</b>	<b>CApsid protein</b>
<b>CCR5</b>	<b>Chemokine Receptor type 5</b>
<b>CD3</b>	<b>Cluster of Differentiation 3</b>
<b>CD4</b>	<b>Cluster of Differentiation 4</b>
<b>CD4+</b>	<b>Cluster of Differentiation 4 positive expression of CD4 on cells</b>
<b>CD8</b>	<b>Cluster of Differentiation 8</b>
<b>CD19</b>	<b>Cluster of Differentiation 19</b>
<b>CD4bs</b>	<b>CD4 binding site</b>
<b>cDNA</b>	<b>complementary DNA</b>
<b>CDR3</b>	<b>Complementarity Determining Region 3</b>
<b>CFPE</b>	<b>Cell Free Protein Expression</b>
<b>CFPS</b>	<b>Cell Free Protein Synthesis</b>
<b>CMP</b>	<b>Contrast Match Point</b>
<b>CMV</b>	<b>CytoMegaloVirus</b>
<b>Co</b>	<b>Cobalt</b>

CO <sub>2</sub>	Carbon diOxide
CT	Cytosolic Tail
C-terminal	Carboxy <b>terminal</b>
CV	Column Volume
CXCR4	Chemokine Receptor type 4
D1-4	Domains 1-4
D22	Diffractionmeter 22 (SANS instrument at the ILL)
D <sub>2</sub> O	heavy water
Da	Dalton
D <sub>max</sub>	<b>maximum Dimension</b>
DDM	n-Dodecyl β-D-maltoside
DEAE	DiEthylAminoEthyl
DNA	DeoxyyRibonucleicAcid
D-Lab	Deuteration <b>Laboratory</b>
DLS	Dynamic Light Scattering
DMEM	Dulbecco's Modified Eagle Medium
dsDNA	double stranded DNA
DTT	DiThioThreitol
ECL	Electro ChemiLuminescent
EDTA	EthyleneDiamineTetraAcetic acid
ELISA	Enzyme-Linked ImmunoSorbant Assay
ESRCT	Endosomal Sorting Complexes Required For Transport
ESRF	European Synchrotron Radiation Facility
EM	Electron Microscopy
Env	<b>Envelope</b> glycoprotein complex
FBS	Foetal Bovine Serum
FPLC	Fast Protein Liquid Chromatography
G-418	Geneticin antibiotic
g	gram
GFC	Gel Filtration Chromatography
GPC	Gel Permeation Chromatography
gp120	glycoprotein 120
gp41	glycoprotein 41
GPCR	G-Protein Coupled Receptor
GSH	Glutathione reduced
GSSG	Glutathione oxidized
GuHCl	Guadinium Chloride
H <sub>2</sub> O	light water
<i>h</i>	Planck's constant
h	hour
HAART	Highly Active Anti-Retroviral Treatment
HCl	HydroChloric acid
HCN	Hydrogen Cyanide
HEK293FS	Human Embryonic Kidney 293 FreeStyle
HET-SOFAST	<b>Heterogeneity- Selective Optimised- Flip- Angle Short- Transient</b>
HIV	Human Immunodeficiency Virus
HPRU	HIV Pathogenesis Research Unit
HPLC	High Perfomance Liquid Chromatography
HR	Heptad Repeat
HRP	Horse Radsh Peroxidase

I(0)	forward scattering Intensity at 0 angle
<i>I</i>	nuclear spin
IAM	IodoAcetaMide
IC50	50% maximal Inhibitory Concentration
Ig	Immunoglobulin
ILL	Institute Laue Langevin
IMAC	Immobilised Metal Affinity Chromatography
IN	INtegrase
IPTG	IsoPropylThionylGalactoside
I(q)	scattering intensity
ITAM	Immunoreceptor Tyrosine Activation Motif
<i>k</i>	incident wavevector
<i>kb</i>	kilo base
$K_D$	Dissociation constant
kDa	kilo Daltons
<i>k'</i>	incident wavevector
$K^*$	constant describing vertically polarised incident light
L	Litre
<i>l</i>	path length
LB	Luria Broth
Lck	Lymphocyte-specific protein tyrosine Kinase
LC-ESI-TOF-MS	Liquid Chromatography Electro Spray Ionisation Time-of-Flight Mass Spectrometry
LC-MALDI-TOF-MS	Liquid Chromatography Matrix Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry
LDS	Lithium Dodecyl Sulphate
MA	Matrix protein
MALLS	Multi Angle Laser Light Scattering
MA	MAtrix protein
mg	milligram
Mg	Magnesium
MHCII	Major Histocompatibility Complex II
M	Molar
MetAP	Methionine AminoPeptidase
MM	Molecular Mass
<i>m</i>	metre
mM	milli Molar
mm	milli metre
MMP	Methyl- $\alpha$ -D-MannoPyranoside
$MM_{H_2O}$	Molecular Mass determined in $H_2O$ buffer
$^1HMM$	protiated molecular mass
MPER	Membrane Proximal Extended Region
mRNA	messenger RNA
<i>M<sub>w</sub></i>	Molecular weight
MWCO	Molecular Weight Cut Off
<i>m/z</i>	mass to charge ratio
$N_A$	Avogadro's Number
NaCl	Sodium Chloride
NC	NucleoCapsid
ng	nano gram

$b_N$	neutron scattering length
$(\text{NH}_4)_2\text{SO}_4$	ammonium sulphate
Ni	Nickel
$\text{NiSO}_4$	Nickel Sulphate
NIH	National Institute of Health
NiNTA	Nickel NitriloTriaceticAcid
NMR	Nuclear Magnetic Resonance
N-MX	Neutron Macromolecular crystallography
NOE	Nuclear Overhauser Effect
$N_{\text{non-exH}}$	Number of <b>non-exchangeable</b> Hydrogens
N-terminal	Amino ( <b>NH<sub>2</sub></b> ) <b>terminal</b>
$\text{OD}_{600}$	Optical Density at <b>600 nm</b>
PAGE	PolyAcrylamide Gel Electrophoresis
PBS	Phosphate <b>B</b> uffered Saline
PBS-T	Phosphate <b>B</b> uffered Saline with Tween-20
PCR	Polymerase Chain Reaction
PDI	Protein Disulphide Isomerase
PDB	Protein Data <b>B</b> ank
PEG	PolyEthyleneGlycol
pH	potential Hydrogen
PIC	Pre- Integration Complex
PITC	Phenyl IsoThioCyanate
P(q)	form factor
PR	<b>PR</b> otease
P(r)	Pair-distance distribution function
PSB	Partnership for <b>S</b> tructural <b>B</b> iology
P-TEFb	Positive Transcription Elongation Factor <b>b</b>
PTH	Phenyl Thiohydantonin
q	scattering vector <b>q</b>
$R_g$	Radius of <b>gyration</b>
R5	<b>CCR5</b>
RI	Refractive Index
RNA	RiboNucleic Acid
rpm	revolutions <b>p</b> er <b>m</b> inute
RT	ReverseTranscriptase <b>H</b>
$R_\theta$	Rayleigh ratio
S75 ag	Superdex <b>75</b> analytical <b>g</b> rade
S75 pg	Superdex <b>75</b> <b>p</b> reperative <b>g</b> rade
S200 ag	Superdex <b>200</b> analytical <b>g</b> rade
S200 pg	Superdex <b>200</b> <b>p</b> reperative <b>g</b> rade
SAS	Small Angle Scattering
SANS	Small Angle Neutron Scattering
SAXS	Small Angle X-ray Scattering
SDS	Sodium Dodecyl Sulphate
SEC	Size Exclusion Chromatography
SEC-MALLS-RI	Size Exclusion Chromatography coupled to <b>M</b> ulti <b>A</b> ngle <b>L</b> aser <b>L</b> ight Scattering and <b>R</b> efractive <b>I</b> ndex
SH2	Src Homology <b>2</b>
SLD	Scattering Length Density
SN	SuperNatant
SOSIP	Soluble Env glycoprotein trimer with stabilising modifications

<b>SPC 1 &amp; 2</b>	<b>Spacer 1 &amp; 2</b> peptides
<b>S(q)</b>	structure factor
<b>Src</b>	Proto-oncogene tyrosine-protein kinase
<b>ssDNA</b>	single stranded <b>DNA</b>
<b>TCID<sub>50</sub></b>	<b>Tissue Culture Infectious Dose 50</b>
<b>TCR</b>	<b>T-Cell Receptor</b>
<b>TF</b>	<b>Tissue Factor</b>
<b>TFA</b>	<b>TriFluoroacetic Acid</b>
<b>TM</b>	<b>Trans-Membrane</b>
<b>T<sub>m</sub></b>	<b>melting Temperature</b>
<b>TMB</b>	<b>3,3',5,5'-tetramethylbenzidine</b>
<b>TOF</b>	<b>Time-of-Flight</b>
<b>tRNA</b>	<b>transfer RNA</b>
<b>Trx</b>	<b>Tioredoxin</b>
<b>UV</b>	<b>UltraViolet</b>
<b>UV<sub>260 nm</sub></b>	<b>UV absorbance at 260 nm</b>
<b>UV<sub>280 nm</sub></b>	<b>UV absorbance at 280 nm</b>
<b>V</b>	<b>Volume or Volts</b>
<b>V<sub>H</sub></b>	<b>Hydrodynamic Volume</b>
<b>V<sub>ref</sub></b>	<b>reference nucleus</b>
<b>VSV-G</b>	<b>Vesicular Stomatitis Virus-G</b>
<b>V<sub>x</sub></b>	<b>nucleus of interest</b>
<b>X4</b>	<b>CXCR4</b>
<b>X-MX</b>	<b>X-ray Macromolecular crystallography</b>
<b>WHO</b>	<b>World Health Organisation</b>
<b>Å</b>	<b>Ångström</b>
<b>γ</b>	<b>gyromagnetic ratio</b>
<b>γNOE</b>	<b>contrast</b>
<b>Δρ<sub>M</sub></b>	<b>chemical dispersion or shift</b>
<b>δ</b>	<b>molar extinction coefficient</b>
<b>ε</b>	<b>incident wavelength of light</b>
<b>λ<sub>O</sub></b>	<b>viscosity</b>
<b>η</b>	<b>scattering length density distribution</b>
<b>ρ(r)</b>	<b>SLD</b>
<b>ρ<sub>s</sub></b>	<b>nuclear magnetic moment</b>
<b>μ</b>	<b>micro Litre</b>
<b>μL</b>	<b>micro Molar</b>
<b>μM</b>	
<b>Amino acids</b>	<b>conventional one letter and three letter codes</b>
<b>Nucleobases</b>	<b>conventional one letter code</b>





To my Father who always wanted better for me...



## Chapter 1

# Introduction

### 1.1 The Human Immuno-deficiency Virus (HIV)

According to the World Health Organisation, HIV is one of the world's leading infectious killers, with approximately 37 million people living with the HIV infection in 2016. HIV is particularly prevalent in low and middle income communities with more than 70% of the global total of HIV cases arising in sub-Saharan Africa [1]. Figure 1.1 presents an overview of the global distribution of HIV cases for adults aged 15-49, from which it is evident that Africa is the most affected WHO area with 4% of the population testing positive for HIV. The advent of highly active anti-retroviral therapy (HAART) has meant that an HIV diagnosis is no longer a death sentence and has become a 'liveable' chronic disease, but only to those who can afford the ten thousand dollar price tag per year. With more than 50% of those infected with HIV unaware of their HIV status and approximately seven thousand individuals newly infected every day [1], it is imperative that a prophylactic vaccination be developed, preventing the spread of HIV.

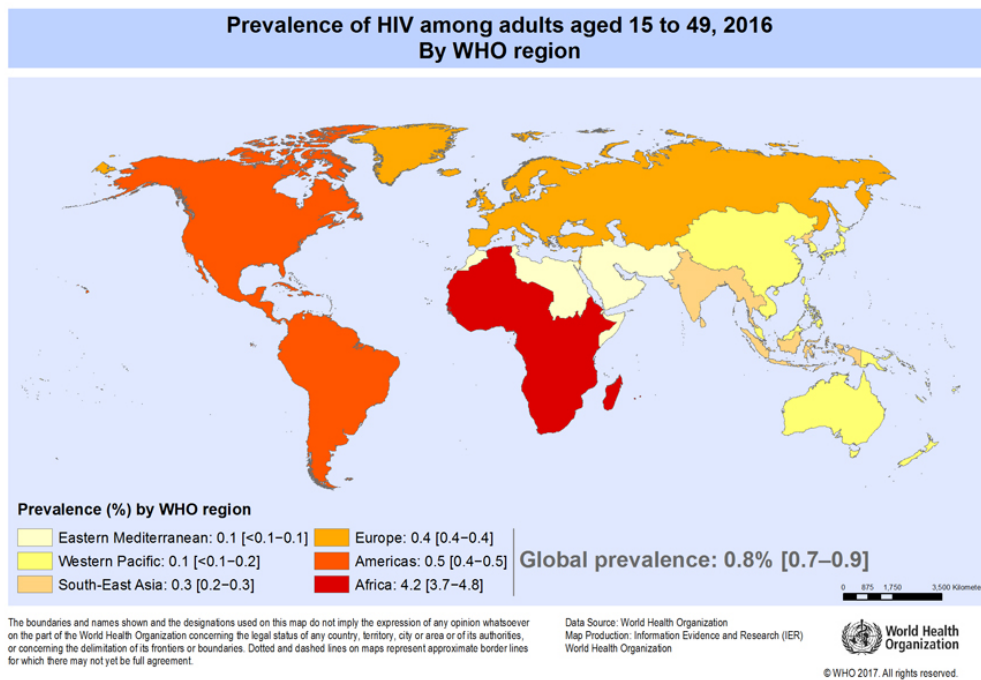


FIGURE 1.1: Map showing the global distribution of HIV among adults aged 15-49 in 2016 by World Health Organization region. Africa is evidently the continent most heavily affected by HIV with approximately 4.2% adults aged 15-49 testing positive for the virus. From: [1]

### 1.1.1 Discovery of HIV

HIV is a lentivirus of the retroviridae family. It was first identified as the infectious agent resulting in the onset of acquired immuno-deficiency syndrome (AIDS) in 1983, when a reverse transcriptase containing virus was isolated from the lymph node of a patient suffering with persistent lymphadenopathy syndrome [2]. Later, three prototype viruses isolated from AIDS patients were found to have identical properties: those of a lentivirus. This AIDS-causing lentivirus was finally named the human immunodeficiency virus. Shortly after the discovery and classification of HIV, a separate subtype was isolated in Western Africa, resulting in the designation of the HIV-1 and HIV-2 subtypes known today [2].

Progression of the disease from HIV to AIDS is characterised by a CD4+ cell count of less than 200 per mm<sup>3</sup> of blood plasma, when the CD4+ cell count (cells on which CD4 is expressed) is less than 14% of the total white blood cell population, or when an individual presents at least one AIDS defining secondary illness including a myriad of invasive cancers and opportunistic infections [3].

### 1.1.2 Genome and Structure

The HIV-1 genome consists of 2 copies of single stranded RNA molecules [4], each 9.8 kb in length and by means of differential RNA splicing, the HIV-1 genome encodes 9 different gene products. The 9.2 kb unspliced genome gives rise to the Gag and Pol polyproteins, a singly spliced 4.5 kb mRNA transcript encodes the Env, Vif, Vpr and Vpu proteins and finally the multiply spliced 2 kb genome codes for the Tat, Rev and Nef proteins. Figure 1.2 shows the different reading frames giving rise to the different genes and gene products.

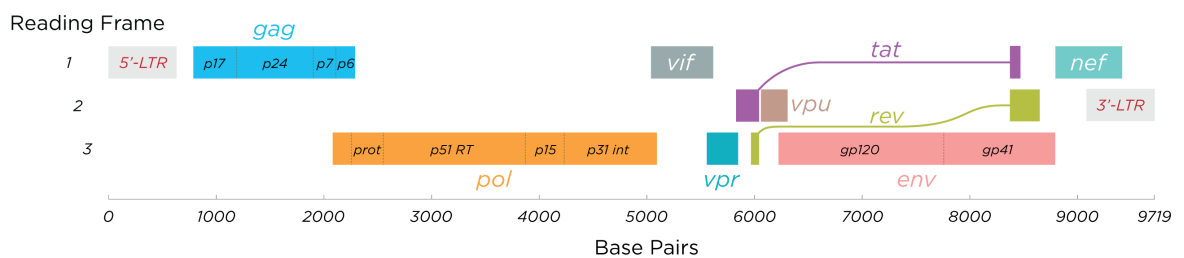


FIGURE 1.2: Schematic showing the 3 reading frames of the HIV-1 genome and the resultant gene products after splicing. From: [5]

The Gag, Pol and Env polyproteins undergo cleavage to produce their respective functional products. The Gag polyprotein is cleaved by viral protease to produce the matrix (MA), capsid (CA) and nucleocapsid (NC) structural proteins of the virus as well as the smaller spacer 1 & 2 (SPC1 & 2) peptides and the P6 protein [6, 7]. The Pol polyprotein is processed to produce the viral enzymes: reverse transcriptase (RT), RNase H, integrase (IN) and HIV protease (PR)[8]. The gp160 Env precursor protein is cleaved by furin proteins [9] from the host cell to produce the gp41 and gp120 monomers which associate into a trimer of heterodimers to form the Env spike (which is the viral attachment point to the human host cells).

The diagram depicted in figure 1.3 shows the configuration of the mature HIV-1 particle [10]. HIV-1 is comprised of a host-cell derived lipid membrane on which the only surface-exposed viral protein is the Env-glycoprotein complex (Env spike). Under the surface of this plasma membrane is the matrix formed by the MA protein. Within the viral matrix is the capsid which forms a fullerene-like cone shape due to the pentamer of CA hexamers at one end and the heptamer of CA hexamers at the opposite end [11]. The viral genome can be found within the capsid and is closely associated with RT and NC [2].

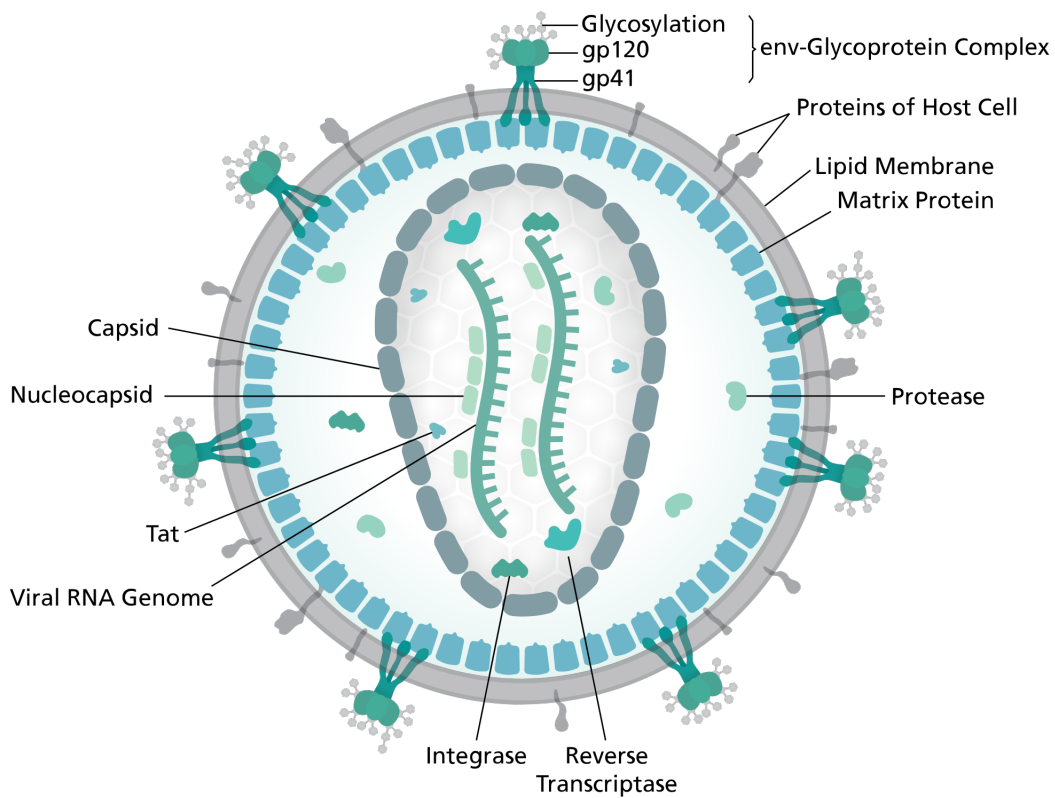


FIGURE 1.3: A section through HIV-1 showing the structure and organisation of the mature viral particle. The Env-glycoprotein complex (Env), consisting of gp41 and gp120 is the only viral protein complex exposed on the surface of the viral particle and is heavily glycosylated. From: [10]

### 1.1.3 Replicative Cycle

Figure 1.4 outlines the major steps of the HIV-1 infection cycle. After entry of the HIV-1 particle into the host organism, the virus must bind to susceptible cell types.

1. Attachment: The infection cycle begins by initial attachment of the viral Env spike to CD4+ cells, key interactions are described in detail in section 1.2.3. During this attachment, gp120 binds to domain 1 (D1) of the CD4 receptor. This initial binding event triggers a conformational change within the Env spike [12], causing exposure of the CD4 induced conformation and of the secondary receptor binding site which binds to the chemokine receptor on the host-cell [13, 14]. The secondary co-receptor depends on the tropism of the virus and the CD4+ cell type. X4 tropic viruses will bind to the CXCR4 G-protein coupled receptor (GPCR) and R5 tropic viruses bind to the CCR5 GPCR [15–17].
2. Fusion: Upon co-receptor binding, further structural realignment within the envelope causes exposure of gp41 and formation of the pre-fusion complex in which the fusion peptide located at the tip of gp41 is inserted into the host membrane [18]. The trimerised carboxy- and amino-terminal heptad repeat sequences of gp41 wrap around each other to form a six-helix hairpin structure which bring the host and viral membranes within proximity to allow viral/host-cell membrane fusion [19, 20].
3. Uncoating: Upon membrane fusion the fullerene-like viral capsid cone, containing the viral RNA, RT and IN, as well as other accessory proteins, is released into the cell where it is uncoated and the contents are released [6].
4. Reverse transcription: Viral RNA is reverse transcribed into proviral cDNA by viral reverse transcriptase [21].
5. Nuclear import: The viral cDNA and MA, and IN proteins then assemble into the pre-integration complex (PIC) prior to nuclear import through the nuclear pore complex, directed by Vpr [22, 23].
6. Integration: Within the nucleus the proviral DNA is irreversibly incorporated into the host genome by the viral integrase enzyme which forms a tetrameric structure. Viral IN is helped by host regulatory proteins [22, 23].

7. Transcription: Viral Tat interacts with positive transcription elongation factor (P-TEFb) to hijack host cell RNA polymerase II elongation control machinery and promote elongation of viral mRNA [24, 25].
8. RNA export: The spliced viral mRNA transcripts are then exported for translation with the help of the viral Rev protein. Rev binds to the Rev response element RNA to form a ribonucleoprotein able to bind the Crm1 host export factor [26].
9. Translation: The viral mRNA is translated into protein using host ribosomes, tRNAs and regulatory proteins into viral precursor proteins: Gag-Pol and Pol and the various viral accessory proteins [27].
10. Assembly: The translated viral proteins are then packaged into new virus particles with the help of the host ESCRT machinery [28]. Env is inserted into the membrane and Gag and Gag-Pol bind the inner side of the membrane [29].
11. Budding: Viral particle budding off from the host-membrane is facilitated by interactions between Gag and vacuolar protein sorting proteins, which are subunits of the ESCRT machinery. As the viral particle forms it acquires host-cell lipid membrane [28, 30].
12. Release: The budding viral particle pinches off from the host membrane to become an individual, immature virion with the help of Vpu which interrupts tetherin membrane tethering [28].
13. Maturation: The viral protease enzyme cleaves the viral precursor proteins to form the matrix, capsid and nucleocapsid proteins as well as the protease, integrase and reverse transcription enzymes, during the process of viral maturation in preparation for the next cycle of infection [31].



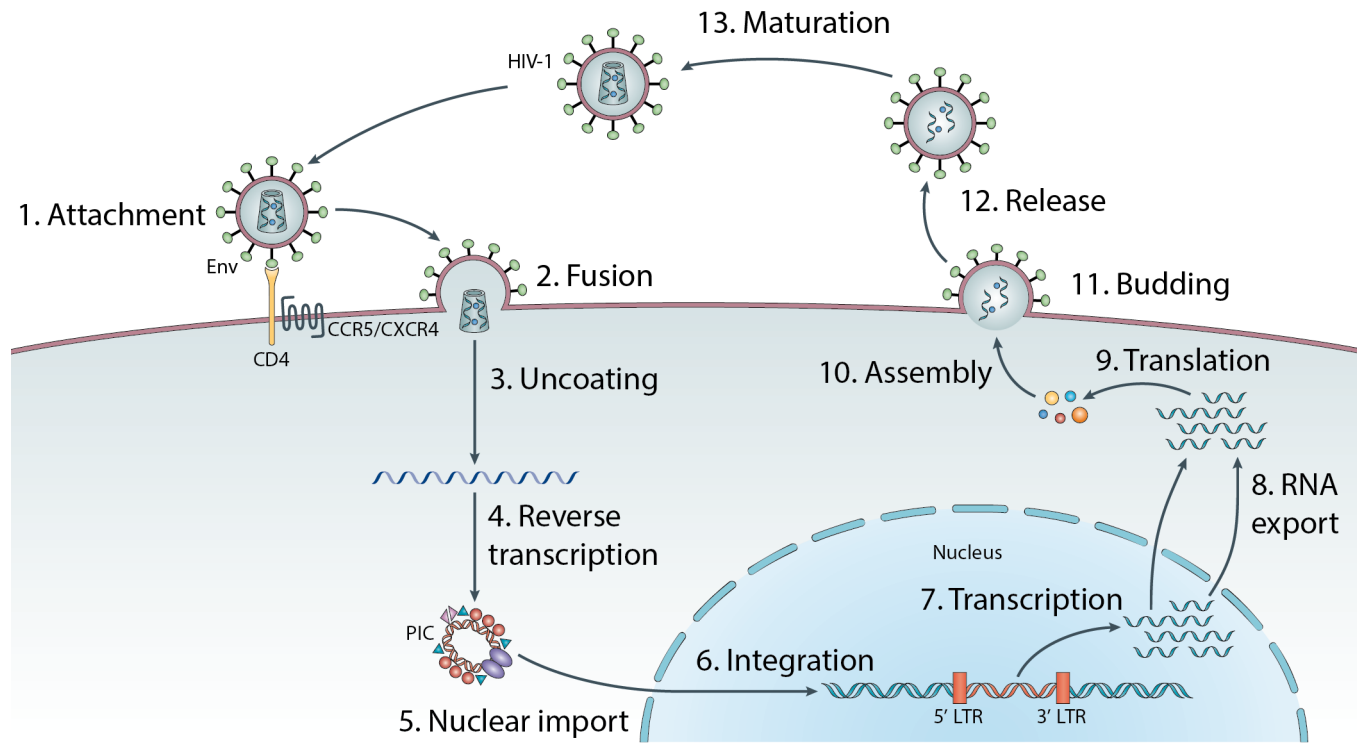


FIGURE 1.4: Schematic depicting the HIV-1 replicative cycle. (1) Attachment: viral gp120 binds CD4 and the CXCR4 or CCR5 co-receptor. (2) Fusion: viral and host cell membranes fuse. (3) Uncoating: the viral capsid is uncoated releasing regulatory proteins and the viral genome into the host cell. (4) Reverse transcription: viral RNA is reverse transcribed into cDNA by viral reverse transcriptase. (5) Nuclear import: the viral genome is imported into the host cell nucleus. (6) Integration: the viral genome is incorporated into the host cell genome. (7) Transcription: the viral genome is transcribed using host cell machinery. (8) RNA export: mRNA coding for viral proteins is exported from the host cell nucleus. (9) Translation: host cell machinery is used to transcribe viral mRNA into viral precursor proteins. (10) Assembly: the translated viral proteins are assembled into a new viral particle. (11) Budding: the new viral particle buds off from the host cell membrane using host ESCRT machinery. (12) Release: the budding viral particle is released from the host membrane. (13) Maturation: the immature viral particle becomes mature after viral protease cleaves the viral precursor proteins.

Adapted from [32]

## 1.2 The primary host-cell/viral interaction

### 1.2.1 The viral glycoprotein 120 (gp120)

#### gp120 function in HIV-1 infection

Glycoprotein 120 (gp120) is one of the two proteins along with gp41 to form the viral Env-glycoprotein complex, located on the surface of the viral particle [33]. The function of gp120 is to identify the correct cell type for infection followed by exposure of the cell-membrane fusion machinery. gp120 binds to domain 1 of the CD4 receptor of human immune cells [34] which induces structural realignments allowing the secondary receptor, either CXCR4 or CCR5, to bind to the secondary receptor binding site of the CD4 induced gp120 conformation [14, 35]. Further structural rearrangements, after secondary receptor binding, result in the unveiling of gp41, whose function is to bring the viral and host-cell membranes into close proximity for membrane fusion. The gp41 fusion peptide anchors into the host-cell membrane as a type of harpoon, triggering reorganisation of the trimerised N- and C-terminal heptad repeat motifs [18]. The HR motifs wind around each other to ratchet the viral and host-cell membranes into proximity for subsequent membrane fusion and HIV entry [19, 20] (fig. 2).

#### Atomic resolution structure of gp120

Kwong *et al.* 1998 were the first to describe the structure of the gp120 core by X-ray crystallography [34]. The core of gp120 is formed of an inner and outer domain with respect to Env, which are linked by a four beta-stranded bridging sheet (fig. 1.5). The inner and outer domains lie approximately parallel to one another with the bridging sheet connecting the two domains at their termini distal from the viral membrane. gp120 presents five constant regions (C1-5) and five variable regions (V1-5) of which the latter are important for viral immune evasion [36]. gp120 has many N-linked and a variable number of O-linked glycosylation sites which are glycosylated with complex glycans to make up approximately 50% of its mass [37] and are important for immune evasion [38]. gp120 also has 9 conserved disulphide bonds: C57-C74, C119-C205, C126-C196, C131-C157, C218-C247, C228-C239, C296-C331, C385-C418 and C378-C445 [37, 39]. Their significance is discussed in section 1.4.2.

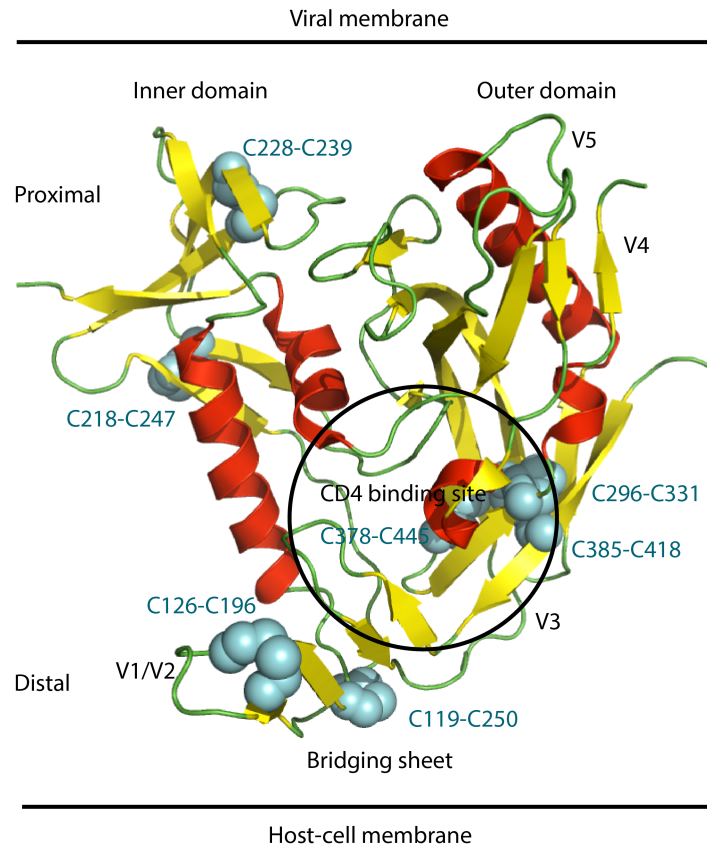


FIGURE 1.5: The gp120 core is comprised of an inner and an outer domain, connected via a bridging sheet of beta-strands at the distal end from the viral membrane. The inner and outer domains are formed by a complex fold involving beta-strands (yellow) and alpha-helices (red). Flexible loops are shown in green. gp120 has 5 variable loops (V1-5) and 9 disulphide bonds of which 7 can be seen from this PDB entry (1G9M) and are indicated in cyan. CD4 binds the CD4 binding site in the crevice formed by the connection of the bridging sheet with the inner and outer domains (black circle). Adapted from: [34]

With the advent of high-resolution cryo-electron microscopy (cryo-EM) and the development of stable trimeric complexes [40] there has been an explosion of trimeric Env structures being submitted to the protein data bank (PDB). This has allowed the determination of unliganded [41], CD4-induced pre-fusion intermediate [42–45], and fusion [46] conformations, which have helped in understanding the HIV-1 entry mechanism. They have shown that unliganded Env adopts a closed position, whereas receptor bound Env adopts an open conformation whereby the gp120 protomers are displaced out and away from the gp41 trimer situated beneath, allowing fusion peptide insertion and subsequent membrane fusion [47]. These high-resolution structures are also important for structure-based immunogen design strategies for raising broadly neutralising antibodies in the race to develop a prophylactic anti-HIV vaccine 1.2.4.

## 1.2.2 The cluster of differentiation 4 (CD4) receptor protein

### Biological function of CD4 and its role in HIV-1 infection

The CD4 receptor protein plays a role in both the innate and adaptive immune responses. However, it is also the primary receptor for the human immunodeficiency virus, binding the gp120 component of the viral Env trimer [48]. CD4 is a 55 kDa type 1 integral membrane glycoprotein expressed on the surface of immune cells such as monocytes, monocyte-derived macrophages and dendritic cells [2, 49]. However, it is found primarily on T-lymphocytes where its principle function is to bind cooperatively the major histocompatibility complex class 2 (MHCII) expressed on antigen presenting cells (APCs) in combination with the T-cell receptor (TCR) [50]. This interaction results in the activation of the intracellular p56<sup>Lck</sup> src-family tyrosine kinase [51] associated with the cytoplasmic tail of CD4.

The pseudo-dimer model proposed by Li *et al.* 2004 suggests that CD4 brings together endogenous and agonist MHCII complexes which allows activation of many TCRs [51]. This concomitant signal activation contributes towards the high sensitivity of T-cells to antigen as determined by the correlation between p56<sup>Lck</sup> recruitment to CD4 and T-cell sensitivity [51]. Lck phosphorylates the immunoreceptor tyrosine activation motifs (ITAMS) on the cytoplasmic portion of CD3 (another receptor) [52]. SH2 domain tyrosine kinases are recruited to the phosphorylated ITAMs to effect downstream signalling events which result in activation of certain transcription factors [53]. This signalling cascade triggers activation of T-cells [53, 54].

MHCII and gp120 have been shown to bind functionally distinct regions of CD4 [55]. CD4 dimerisation has been shown to trigger activation of T-cells [56, 57] and is necessary for binding to MHCII [58] which has been shown to be inhibitory to gp120 binding [59, 60]. Whilst the precise mechanism for CD4 dimerisation is unclear [61–63], it is clear that gp120 and MHCII bind functionally and structurally unique isoforms of CD4.

### The atomic resolution structure of CD4

The human CD4 monomer is comprised of 4 ecto-domains (D1-D4, residues 1-371), a trans-membrane domain (TM, residues 372-395) and a cytoplasmic tail (CT, residues 396-433) [64–

[67]. The extracellular portion of CD4 is composed of 4 concatenated immunoglobulin-like domains of which each domain, except for D3, contains a single disulphide bond (fig. 1.6) [64, 67, 68]. Ig domains are commonly found in cell-surface receptors involved in the immune system [69] and are defined by a 7-9  $\beta$ -strand configuration forming a beta sandwich. The conserved core consists of 4 beta-strands B, C, E and F with between 3 and 5 additional strands: A, C', C'', D and G [64, 70]. The strands arrange themselves into two opposing beta-sheets to form a beta-sandwich.

Three key crystallographic studies in the early 90's allowed the determination of 2 two-domain CD4 high-resolution structures: human 2dCD4 domains 1 and 2 [66, 67] and rat 2dCD4 domains 3 and 4 [68] providing insights into the structure of the protein and its interactions with viral gp120 and MHCII. Garrett *et al.* 1993 [71] built on these structures to solve the structure of four-domain CD4 which indicates that there may be a flexible hinge between domains 2 and 3 required for function [72].

In domain 1 of CD4, the first beta sheet is composed of strands B, E and D and the second beta-sheet contains the strands A, G, F, C, C' and C'', which folds across the first in an antiparallel fashion. Domain 1 has the hallmarks of a variable immunoglobulin domain with a conserved Arg residue at position 1 of the D-strand which forms a salt bridge to an Asp residue at the EF junction and the FG corner, analogous to the CDR3 region of an Ig variable domain [64, 67]. There is also a conserved Trp28 residue in sheet C which is packed against the Cys16 residue positioned in strand B that forms a disulphide bond with Cys84 in strand F [66, 67]. Domain 1 also is suggested to have a flexible region at the C'C'' beta-turn [67].

Domain 2 is truncated to around 75 residues, with respect to the 100 of domain 1. Domain 2 therefore, has a different topology and displays characteristics of a constant Ig domain [64, 68]. The first beta sheet is comprised of strands A, B and E with the second containing strands C', C, F and G. Strand A of domain 2 is a continuation of strand G from domain 1 suggesting a certain rigidity of the two-domained structure, with possible flexibility in the highly-conserved AB and EF loops [66, 68]. The conserved Trp residue is lost in strand C of domain 2 and replaced by a Cys residue to form a disulphide bond with the Cys residue in strand F (C130-C159) which is a non-canonical Ig domain disulphide bond configuration [66, 67].

The formation of a disulphide bond between C130 and C159 in D2 is rare because this positioning is unfavourable for the typical geometry of a disulphide bond [67, 73]. The disulphide bond in domain 1 is positioned between the two beta-sheets whereas the disulphide bond in domain 2 is positioned between beta-strands of the same sheet which causes puckering of the sheet [74]. Hence the geometry of this disulphide bond follows an unusual configuration classed as -RHStaple [73], which is known as an allosteric disulphide bond and is described further in section 1.3.

Similar anti-parallel beta-sandwich structures to domains 1 and 2 are observed in domains 3 and 4. Domain 3 has 9 beta-strands, similar to those of domain 1 and can be classed as a variable Ig domain [64, 65, 68]. However, domain 3 lacks the disulphide bond usually found between strands B and F [68]. Domain 4, like domain 2 has 7-beta strands and is classed as a constant Ig domain [64, 65, 68]. Although unlike domain 2, domain 4 has a canonical Ig disulphide bridge between cysteine residues in strands B and F, such as that found in domain 1. Domain 4 therefore has a more compact structure whereas domain 3 has a more open structure [68]. The G to A strand continuation for domains 1 and 2 is conserved in domains 3 and 4 suggesting a similar rigidity of the domains [68], linked by a proposed flexible hinge [54, 72].

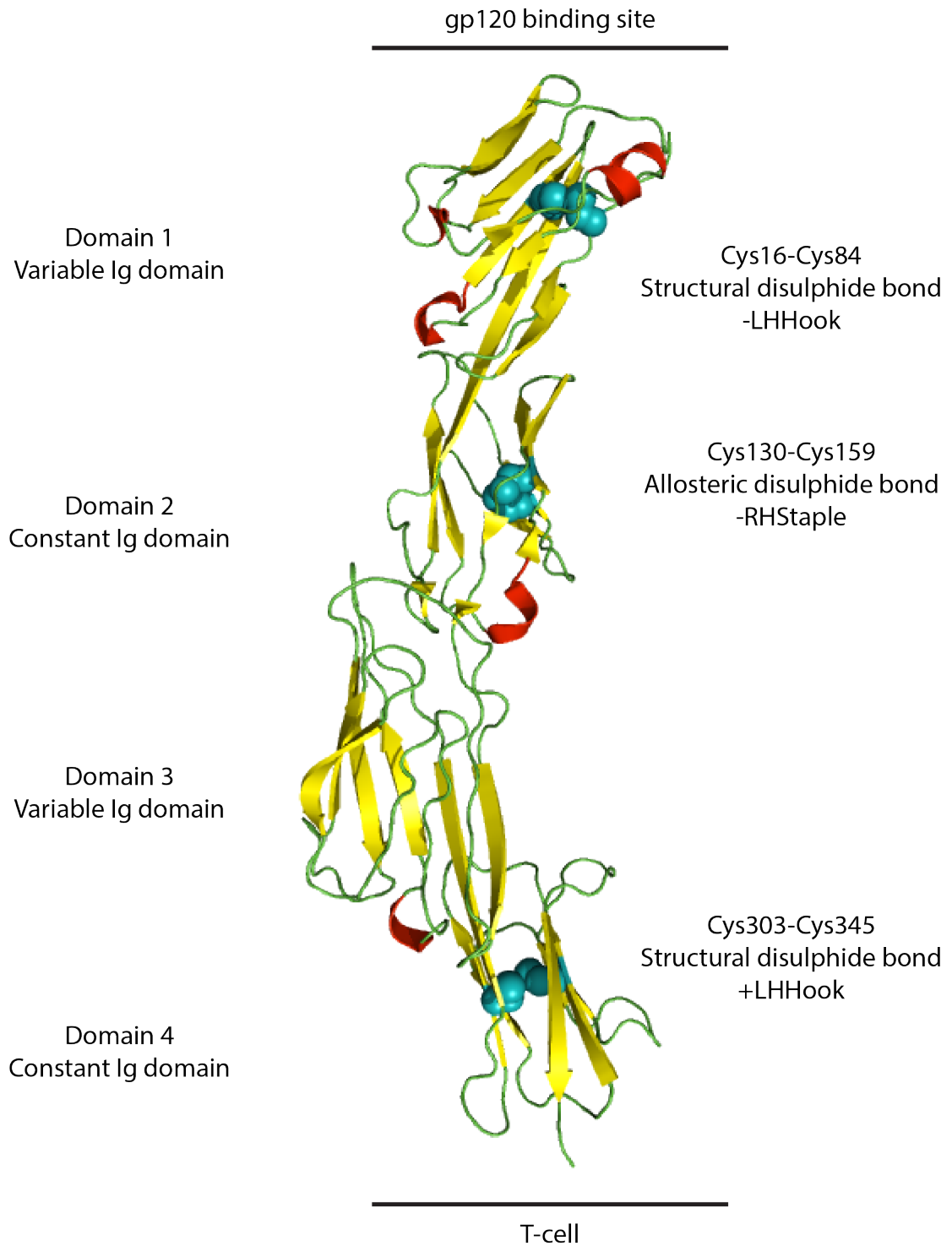


FIGURE 1.6: Beta-strands are shown in yellow, alpha-helices are shown in red, flexible loops are shown in green and the disulphide bonds are shown as spheres in cyan. This figure was made using PDB entry 1WIO [54]. The extra-cellular receptor portion of CD4 is formed from 4 concatenated immunoglobulin-like domains. Each domain is of a beta-sandwich structure formed by two anti-parallel beta-sheets. From this orientation it can be seen that domains 1 and 4 contain disulphide bonds which bridge the two sheets between strands B and F. Whereas the disulphide bond in domain 2 is unusual in that it is formed between neighbouring beta-strands of the same sheet. Domain 3 does not have a disulphide bond and so is the least compact of the 4 domains.

### 1.2.3 The CD4-gp120 interaction

Kwong *et al.* 1998 were the first to structurally describe the CD4-gp120 interaction [34]. CD4 binds gp120 in the crevice formed by the junction of the bridging sheet with the inner and outer domains. The most significant interactions involve CD4 domain 1 residues Phe43 and Arg59 (fig. 1.7) which form several, essential contacts with gp120 residues Asp368, Glu370 and Trp427. Phe43 also interacts with Ile371, Asn425, Met426 and Gly473 and Arg59 also interacts with Val430. There is an overall general complementarity in the electrostatic potential at the interface between the two proteins, however, the positive region on CD4 is slightly displaced with respect to the centre of the negative region on gp120 [34].

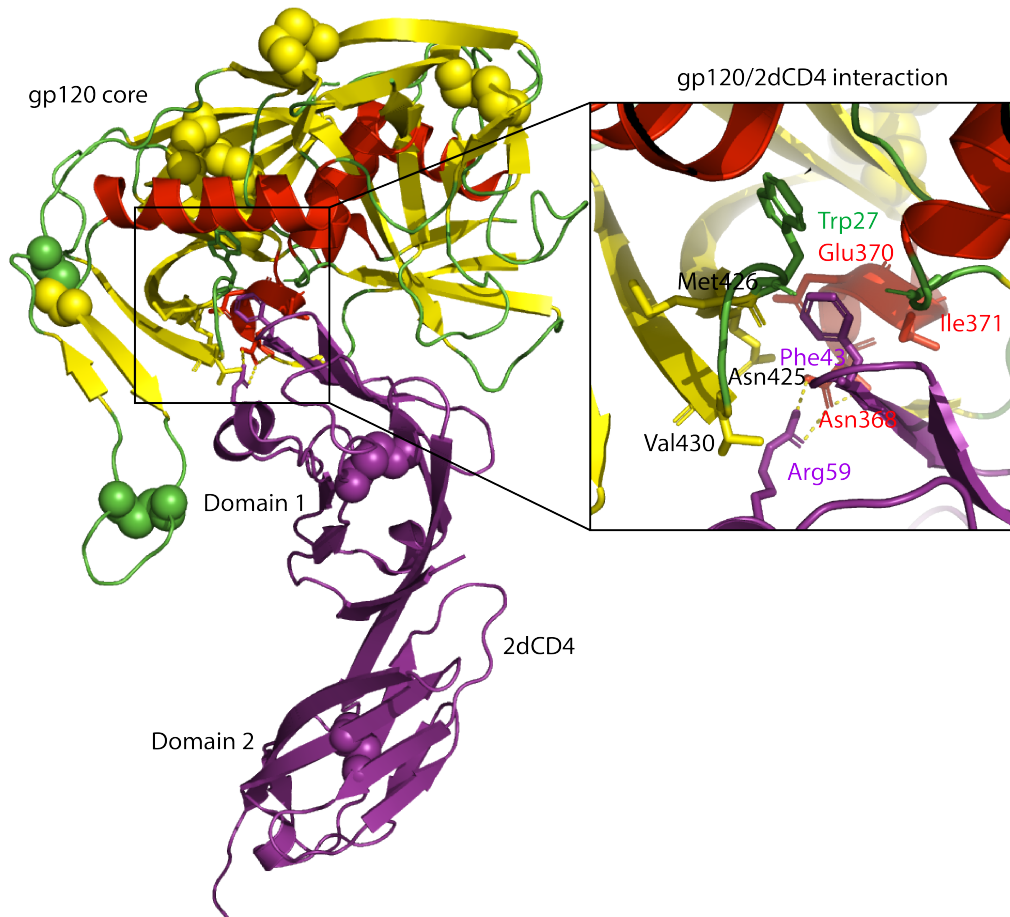


FIGURE 1.7: Domain 1 of CD4 interacts with the CD4 binding site on gp120 (black box). The residues forming the key interactions are shown as sticks. Phe43 and Arg59 from CD4 interact with Glu370 and Trp427 from gp120. Arg59 also forms interactions with Val430 and electrostatic interactions with Asp368 (shown in yellow dashed lines). Phe43 also interacts with Ile371, Asn425, Met426 and Gly473 (not shown). CD4 is shown in purple and gp120 beta sheets are shown in yellow, alpha helices in red and loops in green. Disulphide bonds are shown as space-filled spheres. The diagram was made using PDB entry 1GCI.



### 1.2.4 gp120 and CD4 as a target for vaccines

Anti-HIV vaccine design is far from trivial. The virus displays huge genetic variation and is able to evade the immune system by using a highly error prone reverse transcriptase enzyme which creates vast sequence diversity by recombinant and point mutations over the course of infection within one individual [36, 75]. In addition, the virus employs myriad other strategies including: conformation masking [76], steric occlusion [77], carbohydrate shielding and shifting [38, 78], transient epitope exposure, [79], expression of non-functional Env [80, 81] and general low density of functional Env [12, 82]. So while the immune system can generate antibodies against these changes, their evolution is slower than that of the virus allowing viral immune escape.

Most attempts to develop an anti-HIV vaccine are directed towards the Env complex on the surface of the virus. gp120 and gp41 are the only viral encoded proteins expressed on the surface of the virion so there is not much choice in terms of which protein the immune system can target, in virus directed approaches. Conserved, functional epitopes on the Env complex are the best candidates for immunogen development [75, 83–85]. Broadly neutralising antibodies (bNAbs) to conserved antigens have been isolated from certain HIV-infected individuals called ‘Elite neutralisers’ [85, 86] and have been characterised and mapped to such epitopes (fig 1.8 and table 1.1).

Epitope	Antibody	References
CD4 binding-site	b12	[87]
(CD4bs)	VCR01	[88, 89]
V1/V2	PG9	[90]
	PGT145	[91]
Glycan V3	2G12	[92]
	PGT128	[91]
Membrane proximal extended region	2F5	[93]
(MPER)	10E8	[94]

TABLE 1.1: The main conserved epitopes on the viral Env complex are the CD4 binding-site, the V1/V2 region, a cluster of conserved glycans at the V3 loop and the membrane proximal extended region at the base of gp41. Adapted from: [84]

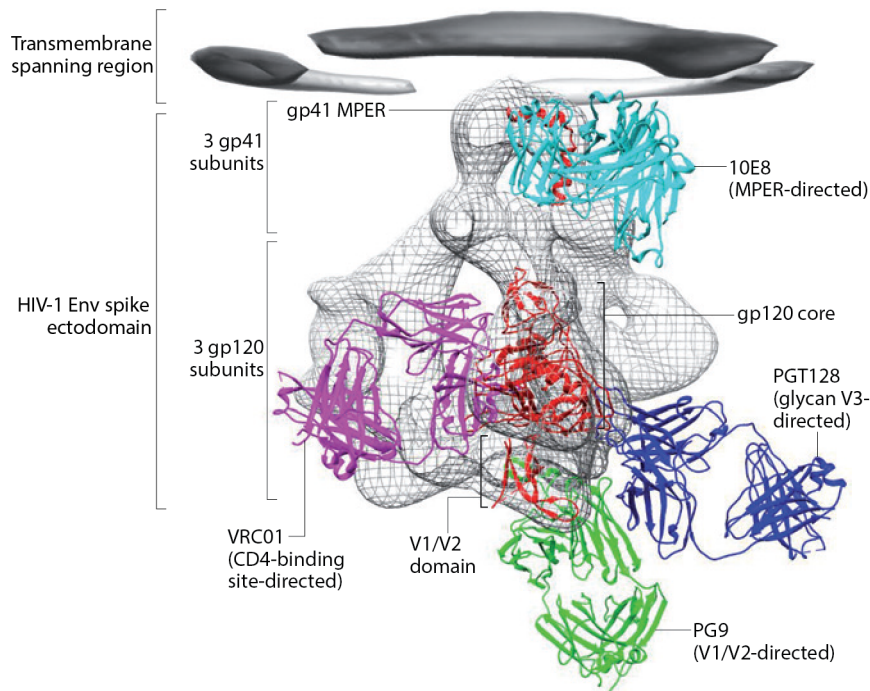


FIGURE 1.8: Conserved, functional epitopes are the target areas on the Env spike that broadly neutralising antibodies can be raised against in the development of an anti-HIV vaccine. These conserved regions comprise the CD4 binding-site which binds VCR01 (pink), the V1/V2 region which binds PG9 (green), a cluster of glycans at V3 which binds PGT128 (dark blue) and the membrane proximal extended region (MPER) at the base of gp41 which binds 10E8 (cyan). The gp120 core is shown in red. From: [84].

The main strategy being exploited for HIV-1 vaccine design is the development of immunogens based on the structure and interaction of bNAbs circulating in the HIV-1 infected population. The immunogens are designed to mimic the Env complex which are introduced into the body to allow B-cells to raise antibodies with a broadly neutralising effect [75, 84, 85]. To this end, soluble Env trimers, called "SOSIP trimers", are the latest and best native Env mimic for eliciting broadly neutralising antibodies against the virus [43, 95].

There are also attempts to target CD4, rather than the Env complex. Since CD4 does not possess the complex immune evasion techniques that Env does, targeting CD4 would be much simpler in this respect. The difficulty in targeting CD4 lies, rather, in the fact that it is a host-cell protein which serves a critical function in the immune response. With this approach it is essential that a neutralising antibody either: (1) prevents gp120 binding by occluding the gp120 binding site on domain 1, whilst still allowing interactions with the MHCII or (2) fixes gp120 to CD4 in the

CD4 bound conformation, without allowing membrane fusion. For example, ibalizumab is a humanised monoclonal antibody of murine origin [96] which potently blocks HIV-1 infection. Ibalizumab binds to an epitope on domain 2 of CD4 in the BC loop at the interface of domains 1 and 2, opposite the MHCII binding epitope [97, 98]. Therefore, it has been shown to allow normal immune function of CD4 [98]. Surprisingly, ibalizumab permits gp120 binding to CD4 [96] and it is therefore believed to exert its anti-viral effect by preventing structural realignment within gp120, after the CD4 binding event, to inhibit gp41 function [98, 99].

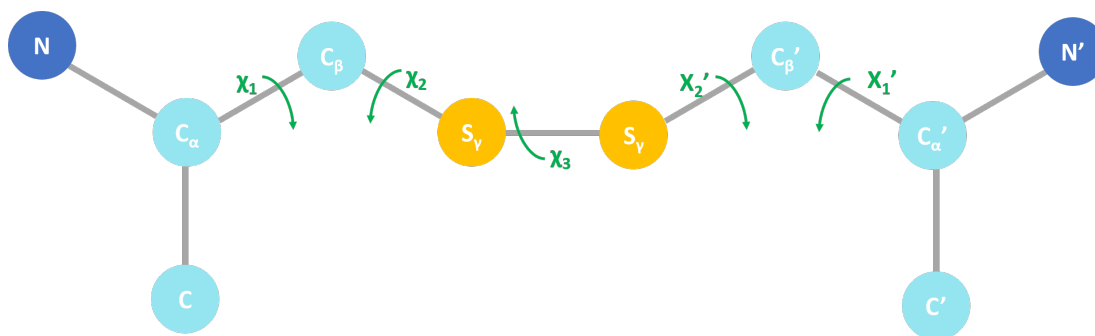
### 1.3 Allosteric disulphide bonds and redox biology

Disulphide bonds are typically considered to serve structural or catalytic functions. Structural disulphide bonds are generally associated with protein folding by providing a protective and stabilising effect and are therefore believed to be inert. Catalytic disulphide bonds are located at the active site of oxidoreductase enzymes catalysing redox shuffling events. Secreted proteins function in the extracellular milieu which is a harsh environment abundant in oxidants and proteolytic enzymes [100]. Disulphide bonds can act as protection against these conditions due to their robust covalent nature. However, another class of disulphide bonds referred to as the allosteric disulphides, such as that found in the second domain of CD4, are believed to impact on protein function by shuffling of their redox state [73, 100]. Before they were officially classified as "allosteric" disulphide bonds [73] they were first referred to as "cross-strand disulphides" [74].

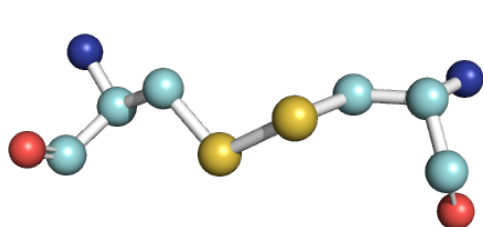
#### 1.3.1 Disulphide bond classification

The geometry of a disulphide bond is first classed by the sign of the  $\chi_2$ ,  $\chi_3$  and  $\chi_2'$  angles (fig. 1.9) into one of three basic categories: spiral, hook or staple and secondly by whether they are left or right handed based on the sign of the  $\chi_3$  angle. Schmidt *et al.* 2006 analysed the geometry and dihedral strains of some 6 874 unique disulphide bonds in over 2 700 X-ray crystal structures and found that of the 20 types of disulphide bond identified from the sign of the five chi angles that make up the bond, the allosteric disulphide bonds were all contained in the -RHStaple configuration [73]. The -RHStaple has subsequently been considered a hallmark of the allosteric disulphide bonds. According to the classification by Schmidt *et al.* 2006 [73],

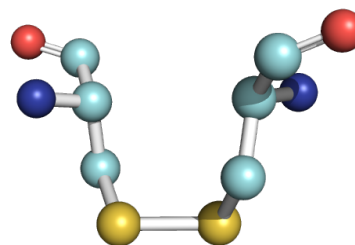
the signs for the chi angles of the -RHStaple class of disulphide bonds are therefore + - + - - or - - + - +. The first domains of CD4 contain a typical structural disulphide bond of -LHHook geometry and domain 2 contains a typical allosteric disulphide bond of -RHStaple geometry (fig. 1.9).



(A) Five  $\chi$  angles of a disulphide bond



(B) Domain 1 disulphide bond:  
-LHHook



(C) Domain 2 disulphide bond:  
-RHStaple

FIGURE 1.9: Schematic of a disulphide bond. (A) The five chi angles that define the classification of a disulphide bond [73]. (B) The disulphide bond in domain 1 is a -LHHook which is a common structural disulphide bond geometry. (C) The disulphide bond in domain 2 is a -RHStaple which is a hallmark allosteric disulphide bond configuration. Grey = bond, yellow = sulphur atom, dark blue = nitrogen atom, light blue = carbon atom. The position of the chi angles are shown in green. The disulphide bonds were taken from PDB entry 3CD4. The classification of the disulphide bond in domain 1 was determined using the online disulphide bond analysis tool (<http://149.171.101.136/python/disulfideanalysis/search.html>) [101].

### 1.3.2 Proteins containing allosteric disulphide bonds

It is becoming increasingly accepted that protein function can be modified by redox shuffling of their allosteric disulphide bonds [102–104]. Many of the proteins which undergo such redox shuffling events have been localised to the cell surface [105, 106] and are involved in immune or homeostatic functions [107–109]. The allosteric disulphide bonds in these proteins may act as protein activation switches [102, 110].

Lawrence *et al.* 1996 found that HIV-infected individuals presented with more surface thiols on CD4+ and CD19+ lymphocytes than CD8+ lymphocytes [111]. They suggest that HIV may preferentially activate CD4+ over CD8+ T-cells which is in keeping with the fact that HIV infects CD4+ cells and necessitates a reduced monomeric isoform of CD4 in order to enter host cells [112, 113]. They also showed that there are important correlations between surface thiol concentrations and cell activation: on activation the concentration of cell-surface thiols increases, suggesting that redox biology is indeed important in activation of these lymphocytes. Finally, they demonstrated that there are surface thiol differences between lymphoid subsets which may explain why different subsets respond differently to diseases that cause oxidative stress such as HIV-1 infection.

Donoghue *et al.* 2000 used GSAO-B, (biotinylated 4-[N-(S-glutathionylacetyl)amino]arsenoxide) which is a trivalent arsenical forming a high affinity ring structure with proximal dithiols, to detect closely spaced thiols within cell-surface proteins by streptavidin-peroxidase detection of the biotin moiety [105]. They focused on the closely spaced thiols as these thiols were more likely to be susceptible to redox shuffling events and thus were most likely implicated in the protein's function. They found that both Thioredoxin (Trx) and protein disulphide isomerase (PDI) incorporated the GSAO-B suggesting that both Trx and PDI contain closely spaced thiols. They also found GSAO-B incorporation in 10 proteins on the surface of endothelial cells and 12 proteins on the surface of fibroblast cells and of these proteins, one of them was found to be PDI. PDI is a known regulator of oxidoreduction at the cell surface and the other proteins found containing proximal dithiols may be other regulators of oxidoreduction or, alternatively, regulated by PDI. Ultimately their research shows that closely spaced thiols, which are likely to impact the function of the protein in which they reside, can be found on different cell types, corresponding to their difference in function.

A number of proteins involved in blood coagulation have been found to possess allosteric disulphide bonds which appear to act as activation switches. Tissue factor (TF) which is involved in the blood coagulation cascade contains two fibronectin type III domains in the extracellular portion, of which the C-terminal domain Cys186-Cys209 disulphide bond is of the -RHStaple configuration and has been shown to be oxidised in active TF but reduced in inactive TF. In addition, thiol alkylating agents have been shown to block TF activity which suggests that this disulphide bond is essential for TF activity and may act as an activation switch [103,

108]. Von Willebrand factor (VWF), which mediates the adhesion of platelets to exposed sub-endothelium, contains a -RHStaple disulphide bond which may be found in lateral association of VWF molecules to form Weibel-Palade bodies [114]. Other proteins found to be controlled by allosteric disulphide bond switches include:  $\beta$ II-Tryptase [107], gp120 [115] and CD4 [63, 112, 113].

## 1.4 The importance of redox biology in HIV entry

The interplay between redox biology and HIV entry has been apparent for over two decades [112, 113, 116, 117]. Allosteric disulphide bonds have been found in both the CD4 receptor on the surface of T-cells [106] and the gp120 component of the Env complex on the surface of the virion [74]. It is believed that CD4 binds gp120 in a monomeric reduced form [63, 113] which is functionally distinct from dimeric, active CD4 [56, 57] and monomeric, oxidised, inactive CD4 [63].

When gp120 binds CD4, there are conformational changes in both proteins. CD4 is believed to bend with a hinge-like action at the interface between domains 2 and 3 [118], bringing the viral and host-cell membranes into proximity. gp120 exposes its CD4 induced epitopes which bind to the secondary GPCR receptor: CXCR4 or CCR5, after which further structural alignments ensue which are believed in part to be as a result of the cleavage of allosteric disulphide bonds within gp120 [119]. This sequence of events results in viral/host-cell membrane fusion and release of the viral genome and regulatory proteins into the host-cell. The specifics of the redox events happening within CD4 (1.4.1) and gp120 (1.4.2) are discussed below.

### 1.4.1 The role of CD4 redox biology in HIV entry

Before the second domain disulphide bond of CD4 was classified as an allosteric disulphide bond [73], it was recognised that disulphide exchange in domain 2 of CD4 is important for HIV-1 entry into host cells [112]. Wouters *et al.* 2009, Matthias *et al.* 2003 and Matthias *et al.* 2003b, describe the unusual geometry of the bond [74, 100, 112]. Unlike most Ig domain disulphide bonds, which link  $\beta$ -strands in opposing sheets, the D2 disulphide bond links  $\beta$ -strands within the same sheet [120] causing a puckering of the sheet [74]. The D2 disulphide

is a -RHStaple disulphide bond which has an unusually short  $C_{\alpha}$ - $C_{\alpha}$  distance of 3.92 Å. The torsion energy is 4.74 kcal/mol which is high compared to 2.28 and 1.71 kcal/mol for the D1 and D4 disulphide bonds, respectively. Finally, from the enthalpy calculations of the three CD4 disulphide bonds, the D2 disulphide contributes least to the stability of the protein [112]. The unusual geometry of the D2 disulphide bond resulting in its high dihedral strain means that it is metastable and susceptible to reduction by oxidoreductases.

Since CD4 is the primary receptor for HIV-1 it was important to determine whether binding of Env is sensitive to CD4 redox biology. Susceptibility of HIV-1 entry into CD4+ cells depending on the redox state of CD4 was evaluated by measuring viral gag DNA accumulation, and results confirmed that reduction of monomeric CD4 was necessary for HIV-1 entry [112]. In a later study, ablation of the D2 disulphide bond by mutation of the domain 2 cysteine pairs to alanine (C130A and C159A) was shown to increase HIV-1 entry and Env-mediated cell-cell fusion, independent of the chemokine receptor tropism of the virus [113]. However, gp120 cannot bind monomeric, fully oxidised CD4 but can bind a monomeric and partially reduced form of CD4 in which either D1 or D2 is reduced [63]. Since D1 contains a stable, structural disulphide bond and D2 contains a metastable, allosteric disulphide bond, CD4 reduced in domain 1 but oxidised in domain 2 is unlikely to exist physiologically. In addition, the evidence from Matthias *et al.* 2002 and 2010 points towards gp120 binding to CD4 reduced in domain 2 [112, 113]. Therefore, the initial attachment of Env to CD4+ cells is determined by a functionally unique CD4 'redoximer' (redox isomer) in which the domain 2 disulphide is reduced.

Investigation into the physiological oxidoreductant which regulates CD4 redox biology supports the idea that the CD4 receptor is a functionally redox active protein. Both the protein disulphide isomerase (PDI) and thioredoxin (Trx) proteins are secreted oxidoreductants which bind to the cell surface to reduce disulphide bonds [63, 112]. Originally, PDI was believed to be the oxidoreductase acting on CD4. However Trx, which is secreted by T-cells and binds to their cell surface and localises to CD4 containing lipid rafts [121] has been shown to reduce two-domain CD4 comprised of domains 1 and 2 (2dCD4) robustly after activation by thioredoxin reductase (TR) and may be involved in CD4 dimerisation [112]. T-cell and macrophage activation has also been correlated with increased reduction of CD4 and secretion of Trx [112].

CD4 is found on active T-cells in a homodimeric form [56, 57] and whilst the precise mechanism for dimerisation is unclear, there is evidence to suggest that the dimerisation mechanism could

be through domain-swapped disulphide bond formation between the second domain cysteine pairs [58, 62]. Bourgeois *et al.* 2006 showed that inhibition of CD4 dimerisation resulted in increased HIV-1 entry [60] which reinforces the findings by Matthias *et al.* 2010 which states that gp120 requires monomeric CD4 for binding [113]. In line with this, gp120 has also been shown to have an inhibitory effect on Trx mediated dimerisation of 2dCD4 [63].

Lynch *et al.* 2006 carried out analysis of the full length (CT, TM & D1-D4) and extra-cellular portions of CD4 (D1-D4) in which they determined the effect of reduction and disulphide ablation by cysteine to alanine mutation on gel mobility [122]. They found that over time there was a shift in band intensity from 55 kDa to 59 kDa and finally to a 110 kDa CD4 dimer in the full length CD4 when exposed to a CD4+ cell stimulator. In addition they found that full length CD4-WT and CD4-D2A (domain 2 disulphide bond knockout by Cys-Ala substitution) resolved to the same molecular weight of 55 kDa whereas D1A and D4A (domain 1 and 4 disulphide bond knockout by Cys-Ala mutation, respectively) show a decrease in gel mobility, resolving to an approximate molecular weight of 57 kDa. They attribute the lack of difference in gel mobility as a function of D2 disulphide bond ablation to a less extensive Ig fold alteration compared to D1A and D4A.

However, it is important to note that Lynch *et al.* 2006 [122] have not assessed the initial status of the CD4 disulphide bonds in the WT protein, before reduction with DTT. Therefore, it is presumed that the D2 disulphide is oxidised and thus the cysteines are in the same redox state as those of D1 and D4. Another interpretation of this same data is that the D2 disulphide in the WT full length CD4 is in fact reduced, hence there is no difference in gel mobility between the WT and D2A proteins due to their being of the same conformation. This could be considered a reasonable assumption taking into account the metastability of the domain 2 disulphide bond described within this thesis.

### 1.4.2 The role of gp120 redox biology in HIV entry

gp120 has 9 disulphide bonds [34] of which 3 have been identified as allosteric [74]. By creating gp120 Cys-Ala mutants Anken *et al.* 2008 [39] found that of the allosteric disulphides in gp120; the Cys126-Cys196 disulphide bond was dispensable for folding; the Cys296-Cys331 disulphide was indispensable for folding and the Cys385-Cys418 disulphide mutant was able



to provide some function despite folding deficiency. While the labile disulphide bonds of gp120 have been shown to be sensitive to both chemical and enzymatic reduction [106], PDI is the proposed physiological oxidoreductase acting on gp120 [123]. This idea is reinforced by the finding that inhibition of lymphocyte surface-associated PDI prevents HIV infectivity [115, 117].

Ryser *et al.* 1994 first showed that PDI activity was necessary for the spread of HIV in cell cultures, as inhibition of PDI by anti-PDI antibodies, bacitracin and DTNB appeared to disrupt HIV infectivity [116]. Fenouillet *et al.* 2001 developed this work to show that PDI activity is necessary for viral/host-cell membrane fusion and subsequent HIV entry, post-CD4 binding [123]. In addition, they showed that surface-associated PDI is localised to CD4-enriched membrane regions and that only the surface-associated PDI and not soluble PDI had an effect on syncytium formation.

Gallina *et al.* 2002 went on to localise the post-CD4 binding action of PDI to the cleavage of disulphide bonds in gp120, which resulted in the opening of a disulphide loop and conformational changes within the viral protein [117]. They also confirmed the association of PDI with CD4 on the surface of the target cells which implies that whilst the virus may be able to attach to CD4 it requires the presence of PDI for membrane fusion and host cell entry. Finally, they noticed that three of gp120's disulphide bonds are located within the vicinity of the CD4 molecule, and therefore may be the disulphide bonds cleaved by PDI. Figure 1.10 suggests how PDI may be associated with CD4 and gp120 leading to membrane fusion and HIV-1 entry.

Barbouche *et al.* 2003 [115] quantified the thiol content of gp120, pre- and post- CD4 and CXCR4 binding (from 0.5-1 mol of thiol/mol gp120 to 4 mol of thiol/mol gp120 pre- and post-binding, respectively). They also found that gp120 with 2 mol of thiol/mol gp120 was able to bind CD4 but not CXCR4 and that PDI inhibition did not affect gp120 binding to CD4 or CXCR4. Therefore, they suggest that two of gp120's disulphides are reduced after CXCR4 binding which results in fusion competence of the Env complex. Their results indicate that these two disulphide bonds may be two of the three disulphide bonds located between the V1/V2 and V4 loops. Interestingly, two of these three disulphide bonds are allosteric disulphide bonds (C126-C196 and C385-C418) [115, 124].

Thioredoxin-1 [125, 126] and glutaredoxin-1 [124] have also been shown to reduce gp120 *in vitro*. Since PDI inhibitors do not fully prevent HIV infectivity [123], it is possible that several

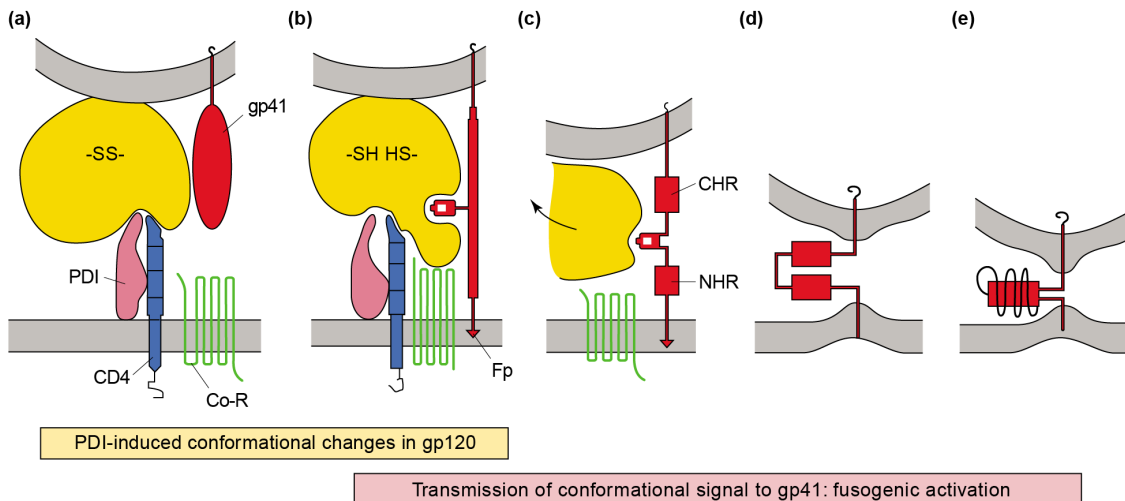


FIGURE 1.10: (a) T-cell surface associated PDI localised to CD4 and may interact with domains 3 and 4 of CD4 so that when CD4 binds gp120, (b) structural realignment of the gp120 protein might include reorganisation of its disulphide bond network. Redox shuffling within gp120 may expose the co-receptor binding site and prime gp41 for fusion by forming the fusion peptide (Fp). (c) gp41 forms N-terminal heptade repeat (NHR) and C-terminal heptade repeat (CHR) helices that (d) fold parallel to one another so that (e) the CHR can wind around the NHR to draw the membranes proximal for fusion. From: [119]

thioredoxin superfamily oxidoreductases are able to catalyse gp120 disulphide bond reduction *in vivo*. Of particular note is the fact that the V3 allosteric disulphide bond (C296-C331) has been shown to be cleaved by Trx-1 [125]. This has been proposed to regulate secondary receptor binding.

It is clear from these studies that reduction of gp120 and CD4 disulphide bonds by one or more oxidoreductases is essential for viral/host-cell membrane fusion and HIV entry. However, whilst there is strong evidence to suggest that gp120's allosteric disulphide bonds are involved in conformational changes resulting in fusion competence, the precise role and reorganisation of the disulphide bond network within gp120 has yet to be elucidated. Whereas for CD4 the questions is why does gp120 necessitate a monomeric reduced form of CD4, since it only forms interactions with domain 1? What is clear is that the redox biology involved in gp120 reduction, HIV entry and the regulation of these process by oxidoreductases is interesting and complex.

## 1.5 Rationale of the PhD thesis

The overall aim of this PhD thesis was to use biochemical, biophysical and structural techniques to characterise the impact of 2dCD4 redox state on its structure and therefore its ability

to bind gp120.

It is becoming increasingly accepted that the redox biology of CD4 plays an essential role in HIV-1 pathogenesis. While it has been shown that HIV-1 requires a partially reduced, monomeric CD4 redox isoform for gp120 binding and subsequent HIV-1 entry into host cells, the structural implications of the redox state of CD4 remain a mystery.

The minimal domain of CD4 required to bind gp120 is domain 1, whilst domain 2 contains the allosteric disulphide bond of interest. Therefore, the idea was to use gp120 in combination with wild-type two domain CD4 (2dCD4-WT) and 2dCD4 variants in which either their domain 1 or 2 disulphide bonds have been knocked out (D1A and D2A, respectively) by cysteine to alanine mutation, in this thesis. The disulphide knockout variant proteins were to be used as redox isomer mimics and were to be compared with 2dCD4-WT. This would allow characterisation the structural changes of the 2dCD4 domains as a function of their redox state and thus the ability of 2dCD4 to bind gp120 using small-angle X-ray and neutron scattering (SAXS and SANS).

SANS was used in contrast variation studies, whereby the 2dCD4 component was match-out labelled with deuterium and gp120 was measured in its unlabelled, protiated state. By using contrast variation (section 2.4.3) it was possible to study the scattering from the complex in 0% D<sub>2</sub>O and importantly the scattering of the individual components of the complex as a function of their match-out point. At the match-out point of gp120 it was possible to study the scattering from the deuterated 2dCD4-WT (d-2dCD4). Unfortunately sample concentrations and volumes did not permit the measurement of the scattering of gp120 in the complex at the match-out point of d-2dCD4. The contrast variation studies were compared to the scattering of the proteins in isolation. It was expected that conformational changes would be observed between the d-2dCD4-WT in isolation and when in complex with gp120 which would be comparable to the structure of 2dCD4-D2A. Unfortunately, sufficient yield of the 2dCD4-D2A mutant was not achieved and therefore this variant protein was not studied by small-angle scattering (SAS).

The results obtained by SANS were followed by SAXS experiments conducted on the same samples in an attempt to corroborate the data obtained through contrast variation studies. In

addition, SEC-SAXS was employed to study the complex in order to ensure an ideal, monodisperse sample with no aggregation. SAXS was also used to study the effect of DTT concentration on 2dCD4-WT thanks to the high-throughput capabilities of BM29 at the ESRF. It was expected that conformational changes or modifications in the folding of 2dCD4-WT as a function of the reduction of the disulphide bonds would be observed. This allowed better understanding of how the disulphide bonds of 2dCD4 affect its structure and therefore function.

In order to achieve these goals, the following was carried out:

1. Optimisation of the expression and purification of hydrogenated and deuterated 2dCD4-WT and variant proteins (chapter 3).
2. Functional, biochemical and biophysical characterisation of homogenous solutions of recombinant hydrogenated and deuterated 2dCD4-WT and variant proteins (chapter 4).
3. Biophysical and SAXS characterisation of the structure of 2dCD4-WT as a function of disulphide bond integrity (chapter 5).
4. Biophysical characterisation of gp120 (chapter 6).
5. Match-out point determination of deuterated 2dCD4 and hydrogenated gp120 and subsequent SANS contrast variation studies of the d-2dCD4/h-gp120 complex (chapter 6).
6. SANS/SAXS studies of the proteins in isolation (chapter 6).
7. SEC-SAXS studies of the 2dCD4/gp120 complex (chapter 6).

Upon reduction of the D2 disulphide it was predicted that there would be significant and previously uncharacterised structural realignment within CD4 to permit gp120 binding. The partially reduced, monomeric CD4 which binds gp120 appears to be functionally distinct from fully oxidised 2dCD4 and dimeric CD4 which binds MHCII and thus, may exist in a unique conformation whose function serves only in binding gp120 of the HIV Env spike. Therefore, the various conformations associated with the different redox states of CD4 present novel targets for anti-HIV pharmaceuticals.

## 1.6 A guide to the CD4 acronyms used in this thesis

Diagram	Acronym	Description
	2dCD4-WT	Wild type two domain CD4 (domains 1 &2)
	d-2dCD4-WT	deuterated 2dCD4-WT
	h-2dCD4-WT	hydrogenated (protiated) 2dCD4-WT
	E-2dCD4-WT	2dCD4-WT expressed in <i>E. coli</i>
	B-2dCD4-WT	2dCD4-WT expressed in <i>B. choshinensis</i>
	CFPE-2dCD4-WT	2dCD4-WT produced using cell-free protein expression

TABLE 1.2: 2dCD4-WT is composed of 2dCD4-O<sub>x</sub>, 2dCD4-R1 and 2dCD4-R2. In this thesis 2dCD4-WT was expressed using: *E. coli* (E-2dC4-WT), *B. choshinensis* (B-2dCD4-WT) and cell-free protein expression (CFPE-2dCD4-WT). 2dCD4-WT was also used in its hydrogenated and deuterated forms (h-2dCD4-WT and d-2dCD4-WT, respectively). Domain 1 is shown in green and domain 2 is shown in blue.

Diagram	Acronym	Description
	1dCD4-D1	One domain CD4 domain 1
	1dCD4-D2	One domain CD4 domain 2

TABLE 1.3: 2dCD4-D2A has had its cysteine residues at positions 130 and 159 replaced by alanine residues and mimics 2dCD4-R1 which is reduced in domain 2 and oxidised in domain 1. 2dCD4-D1A has had its domain 1 cysteine residues at positions 16 and 84 substituted for alanine residues and mimics 2dCD4 which is reduced in domain 1 and oxidised in domain 2. 2dCD4-CΔA has had both its domain 1 and domain 2 cysteine residues substituted for alanine residues and mimics 2dCD4-R2 which is reduced in both domains 1 and 2.

Diagram	Acronym	Description
	2dCD4-Ox	Two domain CD4 (domains 1 &2) oxidised in both
	2dCD4-R1	Two domain CD4 (domains 1 &2) reduced in domain 2 equivalent to 2dCD4-D2A
	2dCD4-R2	Two domain CD4 (domains 1 &2) reduced in both equivalent to 2dCD4-D1A

TABLE 1.4: 2dCD4-WT exists in three different redox forms: fully oxidised (2dCD4-Ox), reduced in domain 2 (2dCD4-R1) and fully reduced (2dCD4-R2). The disulphide bonds can be knocked out by substituting the Cys residues for Ala residues to produce a: domain 2 disulphide bond knockout (2dCD4-D2A), a domain 1 and 2 disulphide bond knockout (2dCD4-CΔA) and a domain 1 disulphide bond knockout (2dCD4-D1A). The table shows the 2dCD4-WT redox isoforms and a schematic to help visualise the redox state of the cysteines. Domain 1 is shown in green and domain 2 is shown in blue.

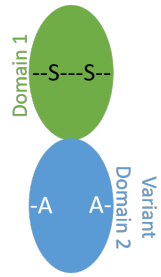
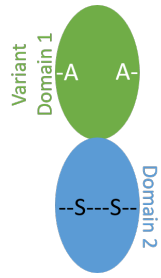
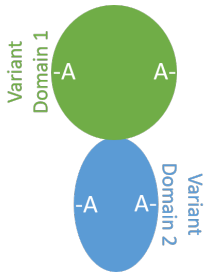
Diagram	Acronym	Description
	2dCD4-D2A	Two domain CD4 (domains 1 &2) domain 2 cys replaced by ala equivalent to 2dCD4-R1
	2dCD4-D1A	Two domain CD4 (domains 1 &2) domain 1 cys replaced by ala
	2dCD4-CΔA	Two domain CD4 (domains 1 &2) reduced in both equivalent to 2dCD4-R2

TABLE 1.5: 2dCD4-D2A has had its cysteine residues at positions 130 and 159 replaced by alanine residues and mimics 2dCD4-R1 which is reduced in domain 2 and oxidised in domain 1. 2dCD4-D1A has had its domain 1 cysteine residues at positions 16 and 84 substituted for alanine residues and mimics 2dCD4 which is reduced in domain 1 and oxidised in domain 2. 2dCD4-CΔA has had both its domain 1 and domain 2 cysteine residues substituted for alanine residues and mimics 2dCD4-R2 which is reduced in both domains 1 and 2.





## Chapter 2

# Experimental Methods

The theory behind the experimental techniques employed in this thesis is outlined in this chapter. The specific protocols used for each technique are detailed in the results chapters: [3](#), [4](#), [5](#) and [6](#).

## 2.1 Molecular Biology

### 2.1.1 Plasmid DNA preparation

Plasmids are circular molecules of supercoiled DNA which can be isolated from prokaryotes. Plasmid DNA is extremely useful for the purpose of protein expression as genes of proteins of interest can be ligated into purpose-designed plasmid DNA, called expression vectors which are typically less than 10 kb in size. Expression vectors have specific features that allow single-celled expression systems, such as bacterial or mammalian cells, to over express recombinant proteins [[127](#)].

Typical features of expression vectors include:

- Antibiotic resistance gene - allows for the selective growth of only cells containing the expression plasmid of interest e.g. ampicillin resistance gene which produces the beta-lactamase protein localised to the periplasmic space that inactivates ampicillin.
- Restriction enzyme sites - allows for the easy cutting of the plasmid DNA with restriction endonucleases e.g. Xho1 and ligation of the gene of interest by DNA ligases into the expression plasmid.

- Affinity purification tag - permits affinity purification of the protein of interest e.g. the polyhistidine-tag which allows immobilised metal affinity purification.
- Origin of replication - recognised by DNA polymerase, DNA replication begins here, allows replication of the plasmid.
- Promoter - recognised by RNA polymerase, transcription of the gene should start here. Can be induced or repressed by regulatory chemicals allowing control of gene expression and in prokaryotes these are often based on operons e.g. the *lac* promoter is induced by addition of isopropylthiogalactoside (IPTG) and the *trp* promoter is repressed by Trp. Expression vectors for expression in mammalian cells use different promoters to plasmids for prokaryotic expression hosts e.g. the cytomegalovirus (CMV) promoter.
- Terminator - recognised by RNA polymerase, transcription of the gene stops here.

### Plasmid DNA purification

Expression vector plasmid DNA, containing the gene or coding DNA for a protein of interest, can be replicated and purified after transformation of highly-competent *E. coli* strains such as Top10 (Invitrogen) and used for recombinant protein expression. This is done by first transforming the expression vector into *E. coli* (see below) and then culturing the transformed cells in varying volumes of culture media which will allow recovery of specific quantities of plasmid DNA.

In order to recover the plasmid DNA the cells are first collected by centrifugation, resuspended and then lysed using the alkaline lysis method. The plasmid DNA is then recovered using spin-column based nucleic acid purification. The purified plasmid DNA can then be resuspended in nuclease-free water and stored at -20°C. Figure 2.1 outlines the steps for mini- and giga-prep DNA purification.

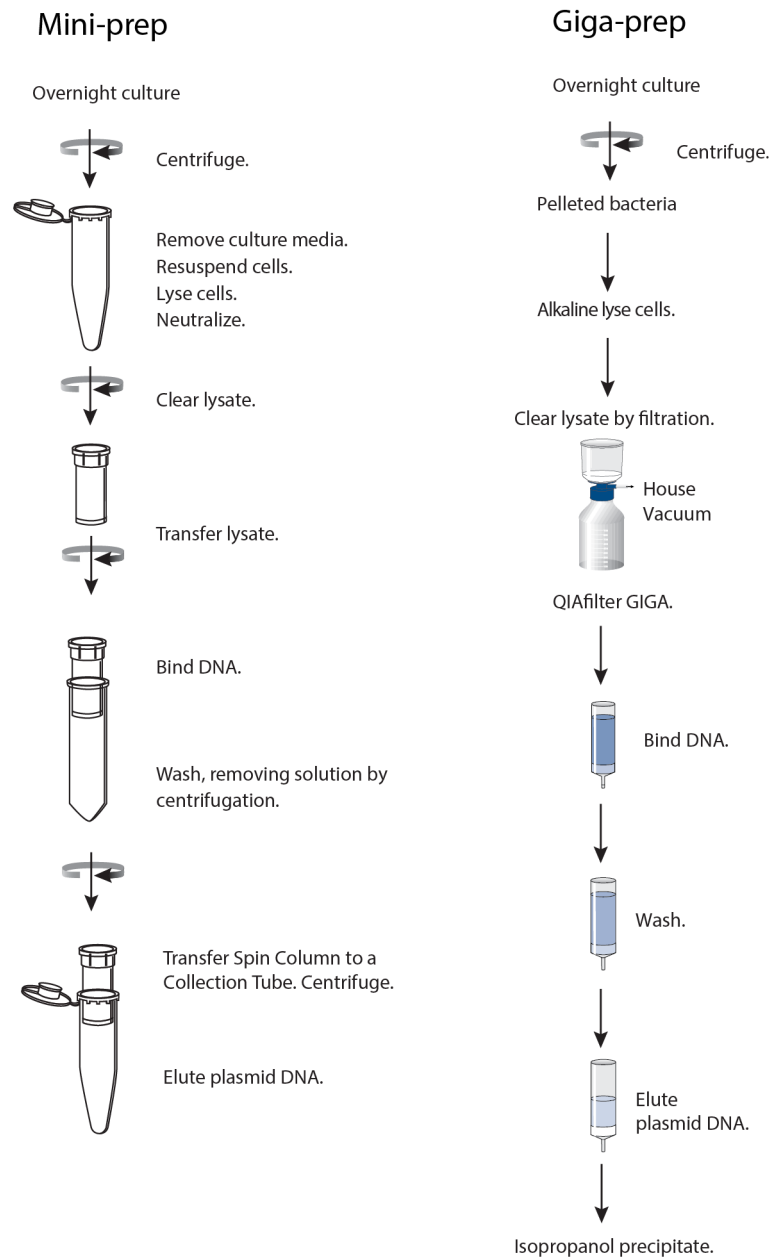


FIGURE 2.1: Schematic outlining the steps for mini-prep purification and giga-prep purification of plasmid DNA. Mini-preps yield approximately 100  $\mu\text{g}$  pure plasmid DNA whereas giga-preps yield around 5 mg plasmid DNA. Adapted from: [128, 129]

## 2.1.2 Protein expression

### Bacterial protein expression

Bacterial expression systems are generally the go-to expression hosts for recombinant protein expression as they are usually well characterised and do not need expensive growth media, producing high yields of recombinant protein. In this thesis work two different approaches towards bacterial protein expression were used: the first was a traditional *Escherichia coli* expression host and the second a novel gram positive expression host called *Brevibacillus choshinensis*.

Expression vectors carrying the genes of interest are transformed into the expression host and grown overnight at a specific temperature (usually around 37°C) on agar/growth medium plates supplemented with an antibiotic specific to the antibiotic resistance marker of the expression plasmid. The growth medium provides nutrients for the bacteria and the agar provides a solid surface for the bacteria to grow on. The presence of the antibiotic ensures that only those bacteria which contain the expression vector after transformation are able to grow.

**Transformation** Transformation is the process by which foreign DNA is introduced into a bacterium. In a laboratory setting, the bacteria are treated chemically or physically in order to become 'competent' allowing the relative ease of DNA uptake. Incubating the cells in ice-cold calcium chloride is a commonly used method for inducing chemical competence, increasing the permeability of the cell wall and membrane to plasmid DNA. The cells are then stimulated to take up the DNA, typically by a short and gentle heat-shock at 42°C for 30 seconds. The Tris-PEG transformation method [130, 131] is another technique which can be used for the transformation of gram positive bacteria and was used in this thesis for the transformation of *B. choshinensis*.

A colony is selected from the agar plate and used to inoculate a small amount of liquid growth medium (usually 5-20 mL) supplemented with the antibiotic corresponding to the antibiotic resistance marker of the expression plasmid. This culture is grown under specific conditions in terms of period of time, temperature and agitation. This small volume of culture is called a starter culture or pre-culture which will be used to inoculate the larger expression culture (usually 500 mL -1 L) which is again inoculated with the appropriate antibiotic. The large expression culture is again cultured at a precise temperature and agitation for a specific length

of time. These parameters need to be determined by experimentation in order to obtain the optimal conditions for high-yields of recombinant protein expression. If the protein coding gene of interest is under the regulation of an inducible promoter then the expression culture should be grown to an optical density at 600 nm ( $OD_{600}$ ) of between 0.6 and 0.8 at which point the bacteria will be in the exponential growth phase. The culture is then grown using the pre-determined precise growth conditions.

### **Cell-free protein expression**

Cell-free protein expression (CFPE), cell-free protein synthesis (CFPS) or *in vitro* protein synthesis utilises the protein synthesis machinery of cells in a test-tube reaction to produce recombinant proteins. For *in vitro* protein synthesis derived from bacterial cells such as *E. coli*, a culture of *E. coli* cells is grown in 2YT media at 37°C and the growth curve is carefully followed by measuring the  $OD_{600}$  regularly. When the cells reach the end of the exponential growth phase, just before they reach the stationary phase, they are harvested by centrifugation because the maximum amount of active ribosomes per cell are present at this stage of the growth curve. The pelleted cells are lysed and the ribosomes (the protein translation machinery) are collected as well as other necessary enzymes for transcription and translation. This 'lysate' forms the main part of the cell-free reaction.

The cell-free reaction is then set-up by the addition of tRNA, plasmid DNA containing the gene or cDNA of interest, amino-acids, nucleotides, salts, cofactors and various other reagents which allow the experimentalist to customise the protein synthesis. Another important element of the CFPE reaction is the addition of an energy regenerating system which allows generation of ATP as translation of mRNA transcripts into a polypeptide chain is an energetic process. The reaction is left for approximately 12 hours at a defined temperature to be optimised by the user, after which the synthesised protein can be purified directly from the reaction mixture.

This technique therefore, has advantages over cellular protein expression because the protein synthesis step can be carried out in a much more controlled environment and importantly uncouples protein expression from cell growth. In cellular protein expression parameters such as temperature and growth media can be controlled but ultimately the experimentalist relies heavily on the ability of the cells to carry out recombinant protein expression without having

access to the protein synthesis step itself. Cell-free synthesis allows the experimentalist to control additional parameters which directly influence the protein synthesis reaction such as the addition of deuterium labelled amino acids for neutron studies or triple-labelled amino acids for high-resolution NMR studies. In addition, *in vitro* protein synthesis is much quicker than classical cellular expression since the reaction can be set-up within an hour and the expression step takes a matter of hours. Whereas with cellular expression the cells must be cultured fresh each time which can take several days or even months when using mammalian expression systems.

### **Mammalian protein expression**

Glycosylated proteins (such as gp120) cannot be expressed in a native form in bacterial expression systems which lack the appropriate environment and enzymes for disulphide bond formation and post-translational modifications. Therefore, immortal mammalian cell lines such as chinese hamster ovary or human embryonic kidney are employed for expression of such complex proteins.

Mammalian cells can be grown in adherent or suspension cultures using complex medium supplemented with Gln and an antibiotic such as Geneticin® (G-418). Once confluent, adherent cells can be grown in ridged roller bottles which allow the cells to adhere to the large surface area and media to pass over the cells as the roller bottles are rotated in a bottle roller. Alternatively, adherent cells can be grown in hyperflasks which have a large surface area due to the many plastic layers inside the flask. Suspension cells can be grown in flasks with gentle agitation. Mammalian cells divide every 24 hours or so compared to 20-30 minutes for bacterial cells so culture growth is much slower. Protein expression is constitutive, meaning that it is continuous, and recombinant proteins are usually tagged with an cell export signal so that the proteins are secreted into the extra-cellular milieu. This means that protein can be collected over the course of about a month by collecting the growth media every few days and exchanging it with fresh growth media to allow the cells to continue to grow and express the recombinant protein.

## 2.2 Biochemical characterisation

### 2.2.1 Protein purification

Prior to protein purification there are some steps that may be required depending on how the protein was expressed. For example if the protein was expressed inside the cell, as is common with bacterial expression, then the cells need to be lysed. This can be done by treating the cells with lysozyme (if they have a bacterial cell wall) followed by mechanical methods such as: sonication, freeze-thaw, cell-disruption or homogenisation. If the proteins have been expressed into insoluble inclusion bodies then the inclusion bodies will need to be solubilised by treatment with a chaotrope such as 6 M GuHCl or 8 M Urea. His-tagged proteins requiring solubilisation from inclusion bodies can be purified in their denatured form because the function of the his-tag is not dependent on the tertiary structure of the protein.

Proteins which are secreted into the extra-cellular milieu can be purified directly from the expression medium. If the volumes are cumbersome, Viva Flow cassettes (Sartorius) can be used to concentrate the protein into a smaller volume of expression medium. If the extra-cellular milieu contains a large number of contaminating proteins then ammonium sulphate precipitation can be used to precipitate the protein out of solution. This is done by adding ammonium sulphate which increases the ionic strength of the solution. Proteins have different solubility at high ionic strengths and so by using this method the number of contaminants can be reduced by "salting out" the protein of interest at a given concentration of ammonium sulphate. The protein salt can then be resolubilised into a smaller volume of buffer to facilitate purification.

### Affinity chromatography

Affinity chromatography is a common purification technique based on the specific interaction of a part of the molecule of interest with a molecule immobilised in a polymer matrix. While antibodies specific to the protein of interest can be used to such effect, usually the easiest way to carry out affinity purification is by introducing a coding sequence for an affinity purification tag into the expression vector. The tag is translated with the protein sequence appearing at either the amino or carboxy terminal depending on its location within the expression vector. One of the most commonly used affinity tags is the polyhistidine-tag (referred to as his-tag

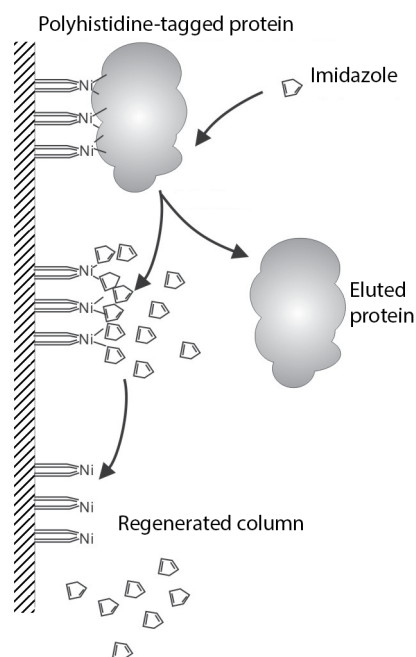


FIGURE 2.2: The polyhistidine-tagged protein binds the  $\text{Ni}^{2+}$  ions. Imidazole has a higher affinity than His for the  $\text{Ni}^{2+}$  ions and the protein is eluted. Imidazole must be washed off the column in order for it to be re-used. Adapted from: [132]

hereinafter), consisting at least six concatenated His moieties, which has micromolar affinity for  $\text{Ni}^{2+}$  and  $\text{Co}^{2+}$  ions, but a lower affinity for these divalent metal ions than imidazole.

The his-tagged protein in its impure crude extract is passed through a column packed with  $\text{Ni}^{2+}$  ions immobilised in a chelator functionalised agarose or Sepharose matrix. The His moieties form ionic bonds with the  $\text{Ni}^{2+}$  ions so that the protein of interest becomes bound to the matrix of the column, whilst the purities flow through the column with little interaction with the column. Since imidazole at high concentrations has a higher affinity for the immobilised metal ions than His it disrupts the binding of the protein to the immobilised metal matrix to recover the purified protein afterwards (fig 2.2). This technique is called immobilised metal affinity chromatography (IMAC) and was used to purify all CD4 analogues as all were his-tagged.

Gp120 was purified using a different form of affinity chromatography called carbohydrate binding affinity chromatography. In this case the gp120 construct was not his-tagged because the glycan shield of gp120 has an affinity for the carbohydrate binding proteins called lectins. The affinity chromatography matrix consists of lectin from *Galanthis nivalis* immobilised on agarose beads. The gp120 sugars bind the lectin proteins as they pass through the column and



gp120 can be eluted from the column using the sugar methyl- $\alpha$ -D- mannopyranoside which competitively binds the lectin proteins with higher affinity than the gp120 glycans.

### Size-exclusion chromatography

Size-exclusion chromatography (SEC) separates biological macromolecules such as proteins based on their hydrodynamic volume, which is related to their molecular mass. Gel-filtration chromatography (GFC) is a SEC technique which separates biological macromolecules in aqueous solutions (as opposed to gel-permeation chromatography (GPC) which uses organic eluents) [133]. GFC and SEC are used interchangeably hereinafter as GPC was not used during this thesis work.

SEC is a fractionation method which allows: (i) purification of samples and (ii) characterisation of the sample size. Samples are fractionated by their passage through a solid matrix column which is comprised of particles with pores which the sample can penetrate, depending on their size. The samples are "flowed" through the column by a mobile phase (buffer) which passes around and through the particles of the matrix. Larger samples will not be able to enter the pores and so have less mobile phase available to them, therefore they elute earlier from the column (fig. 2.3). Smaller samples will be able to penetrate the particles of the matrix and therefore elute later as they have more of the mobile phase to travel [134].

Elution from the column is tracked using in-line spectroscopy. Modern purification systems have several UV detectors which can follow multiple UV traces such as the UV absorbance of protein at 280nm, UV absorbance of nucleic acids at 260nm as well as the conductivity of salt. Fraction collectors connected to the end of the column allow the separated samples to be collected into discrete fractions for further analysis such as SDS-PAGE to identify the contents of each peak and check the purity. The ability of a SEC column to fractionate samples of different sizes is defined by the pore size of the polymer matrix and the size of the column. Smaller pore sizes will be able to resolve smaller samples whereas larger pore sizes will fractionate larger samples. SEC is said to separate macromolecules based on their size but since macromolecules of the same molecular mass can have different shapes, leading to a difference in their ability to access the pores, the separation is actually based on their hydrodynamic volume ( $V_H$ ).  $V_H$  can be calculated from Einstein's viscosity law:

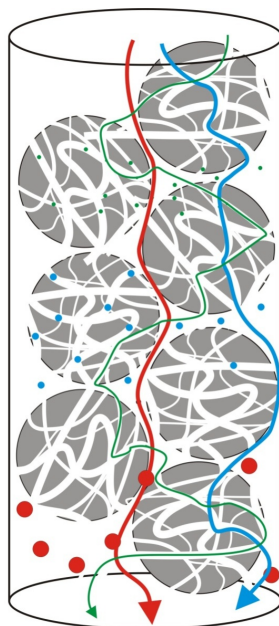


FIGURE 2.3: Size-exclusion chromatography separates proteins based on their size and form and therefore their ability to enter particles of a polymer matrix. The red molecule is large and so cannot enter the particles, therefore it passes straight through the column and elutes first. The blue molecule can enter through some of the pores of the particles of the matrix, slowing down its passage through the column with respect to the red particle and elutes second. The green molecule is the able to enter all the particles of the matrix and therefore has the slowest passage through the column of the three molecules and elutes last. From: [132]

$$[\eta] = 2.5 \frac{NV_H}{MM} \quad (2.1)$$

Where  $MM$  is the molecular mass,  $N$  is the number of particles and  $[\eta]$  is the intrinsic viscosity of the solute.

### 2.2.2 Concentration determination of proteins and nucleic acids

The concentration of proteins can be determined by using a spectrophotometer to measure the UV absorbance at 280 nm ( $Abs_{280nm}$ ) of the aromatic residues Trp and Tyr and disulphide bonds between Cys residues. The  $Abs_{280nm}$  value can be converted into a concentration in  $\text{mol L}^{-1}$  by use of the Beer-Lambert law [135]:

$$A = \epsilon cl \quad (2.2)$$

Where  $l$  is the path length of the light through the sample in cm and  $\epsilon$  is the molar attenuation coefficient in  $\text{M}^{-1} \text{cm}^{-1}$ . The molar attenuation coefficient is dependent on the Trp, Tyr and

disulphide bond (s-s) content [136] of the protein:

$$\epsilon_{280} = (5500 \times n_{Trp}) + (1490 \times n_{Tyr}) + (125 \times n_{s-s}) \quad (2.3)$$

For double stranded DNA (dsDNA) such as plasmid DNA, the UV absorbance is measured at 260 nm ( $Abs_{260nm}$ ) of the purines and pyrimidines. For dsDNA the average attenuation coefficient at 260 nm is  $0.020 \mu \text{ mL}^{-1} \text{ cm}^{-1}$  so that an  $Abs_{260nm}$  of 1 is equivalent to  $50 \mu\text{g/mL}$  dsDNA.

### 2.2.3 Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

In order to quickly and qualitatively analyse a protein sample, sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is the go-to technique. Gel electrophoresis involves separation of macromolecules based on their size (and to some extent their shape) through a polymer matrix driven by an electrical field [137, 138]. For the separation of nucleic acids, agarose is used as the polymer matrix but for proteins, polyacrylamide is used. The percentage of polyacrylamide used in making the gel determines the pore size formed when the gel is set and therefore the resolving power of the gel, similar to the concept of pore size in SEC. Higher percentages of polyacrylamide correspond to smaller pore sizes whereas lower percentages of polyacrylamide correspond to larger pores.

SDS-PAGE is a non-native technique which means that the protein is denatured to form a random-coil. The protein is denatured by boiling for approximately 5 minutes in a buffer containing bromophenol blue (which is used to track the migration through the gel) and SDS which has a two-fold action: (i) protein denaturant (ii) surfactant covering the protein in an anionic charge. The amount of SDS covering the protein and therefore the charge of the SDS-covered-protein roughly corresponds to its molecular weight [139]. In addition to SDS, the sample buffer can also contain DTT which will reduce any disulphide bonds present in the protein of interest as disulphide bonds can affect the migration behaviour of proteins through the gel. SDS-PAGE can therefore be carried out under reducing and non-reducing conditions.

The gel is comprised of two parts: (i) the neutral pH stacking gel which concentrates the proteins before they pass into (ii) the basic pH separating gel. The two gel layers are designed to have different pore sizes and ionic strengths. Prior to running the SDS-PAGE experiment, the

gel is secured into the negative electrode chamber which is inserted into a buffer tank. Cathode buffer is poured into the centre of the negative electrode chamber and anode buffer is poured around the outside into the buffer tank. Samples are loaded into isolated wells of the gel and one well is usually reserved for the molecular weight marker of reference proteins which resolve at specific molecular weights. An electric current is then applied to the gel within the region of 100-200 V. The proteins then migrate through the gel as a function of their mass to charge ratio. After the electrophoresis step the proteins can be visualised on the gel by staining with Coomassie Brilliant Blue which is a blue coloured dye which interacts electrostatically with the amino and carboxyl groups of the protein [140].

(Examples of SDS-PAGE stained with Coomassie blue can be seen extensively in Chapter 3.)

#### 2.2.4 Western blotting

While SDS-PAGE staining with Coomassie blue can be used to visualise protein presence non-specifically, western blotting can be used as an alternative to Coomassie Blue staining to specifically visualise the protein of interest. Western blotting is an immunohistochemistry technique which uses antibodies to identify the protein [141]. Directly after an SDS-PAGE is run the proteins are transferred onto a nitrocellulose membrane by electroblotting. During this process the proteins are pulled from the polyacrylamide gel onto a nitrocellulose membrane. The membrane is then "blocked" using a 5% milk solution to block any subsequent non-specific antibody binding. After blocking an antibody specific to the protein of interest is used. Since most recombinant proteins are tagged with a purification tag, the membrane is often incubated with an antibody specific to the purification tag, such as an anti-his antibody in the case of 2dCD4-WT. If the protein is not tagged then an antibody specific to the protein (e.g. a gp120-antibody in the case of gp120), is necessary.

After incubation with the primary antibody, the membrane is washed to remove excess, unbound primary antibody. The membrane is then incubated with a secondary antibody which has a reporter enzyme such as alkaline phosphatase (ALP) conjugated to it to enhance the signal. When the substrate is added, the ALP produces a chromogenic product indicating the presence of the protein of interest. For higher sensitivity electrochemiluminescent (ECL) substrates can be used. Since the primary antibody is specific to the protein of interest, western

blotting is useful for looking for the presence of the protein of interest in whole cell extracts where expression is weak, since the signal of the protein of interest can be lost in amongst the bands of the other proteins when analysed by SDS-PAGE with Coomassie blue staining. In addition, western blotting has a much higher sensitivity than SDS-PAGE with Coomassie staining as the signal is amplified by the reaction between the substrate and reporter enzyme which can be allowed to reach saturation.

(Examples of western blots using an anti-His primary antibody can be seen in Chapter 3)

### 2.2.5 Enzyme-linked immunosorbant assay (ELISA)

Enzyme-linked immunosorbant assay is a technique which allows the detection (and to some extent quantification) of macromolecules of interest using antibodies conjugated to an enzyme which acts as a reporter label [142–144]. There are different types of ELISA such as the competitive, sandwich or direct ELISA. For this thesis work the indirect ELISA was used.

The basic principle of the indirect ELISA is that the molecule of interest is immobilised on a 96-well plate followed by binding of a primary antibody, specific to the molecule of interest. An enzyme-conjugated secondary antibody, specific to the primary antibody, is subsequently bound, followed by the enzyme's substrate. The reaction between the enzyme and substrate is stopped by the addition of HCl to prevent saturation. The product of the enzymatic reaction is coloured and in modern ELISAs ECL is used so that the absorbance at 450 nm can be measured quantified using a micro-plate reader. This ELISA method is described as indirect because the detection event is of the primary antibody by the secondary antibody, rather than direct detection of the analyte.

### 2.2.6 Pseudo-viral neutralisation assay

The pseudo-viral neutralisation assay is a technique that has been developed to show, typically, the efficacy of antibodies against different tier virions (tier 1 to tier 3) [145]. It functions by incubating the desired protein with the virus and then addition of TZM-bl cells which express CD4, CXCR4 and CCR5 on their surface [146] and contain a luciferase (firefly) reporter gene which is under the regulation of the HIV tat promoter. The protein, or antibody, may then bind to the virion leading to subsequent inhibition of entry of the virion into the TZM-bl cells.

In the case where inhibition is **not** successful, and entry of the virion into the TZM-bl cells can occur, this causes transcription of the HIV *tat* controlled luciferase gene during a normal round of HIV infection, which results in transcription of the luciferase enzyme which catalyses the reaction between Luciferin, oxygen and ATP to produce Oxyluciferin, carbon dioxide, pyrophosphate, AMP and light. Therefore, the greater the inhibition the lower the luminescence. Figure 2.4 outlines how the experiment works.

The pseudo-viral neutralisation assay is therefore an ideal technique to test the functionality of the 2dCD4-WT produced using the different expression systems studied: *E. coli*, *B. choshinensis* and cell-free protein expression. The approach is based on the fact that the purified, recombinant 2dCD4 variants should bind to Env on the virus, preventing the virus from binding to the CD4 on the surface of the TZM-bl cells and therefore inhibiting viral entry into these cells and transcription of the luciferase gene [146].

### 2.2.7 N-terminal protein sequencing

N-terminal sequencing is based on a technique called Edman degradation developed in 1950 [147]. The amino-terminal amino acid residue is labelled with phenyl isothiocyanate (PITC, fig. 2.5) under alkaline conditions to form a cyclical phenyl thiocarbonyl derivative. The phenyl thiocarbonyl is then cleaved under acidic conditions at the peptide bond adjacent to the carbonyl group to form a thiazolinone derivative. The thiazolinone derivative is extracted into an organic solvent and acid treated, forming a phenyl thiohydantoin (PTH) amino acid derivative which can be detected by high performance liquid chromatography (HPLC, fig. 2.5)[148]. Each PTH amino acid derivative has specific HPLC behaviour allowing identification of the amino acid. This cycle is then repeated to identify the following residue in the polypeptide sequence. For the purpose of this thesis, N-terminal sequencing determined the first 6 amino acid residues to confirm the presence of the desired protein by mass spectrometry.

The amino-acid derivatisation was originally performed manually and detection of the liberated amino acid was done using thin layer chromatography [147]. Edman *et al.* 1967 [149] developed the first protein 'sequenator' which automated the process thereby decreasing the sample volume requirements whilst increasing the accuracy. Modern N-terminal sequencing

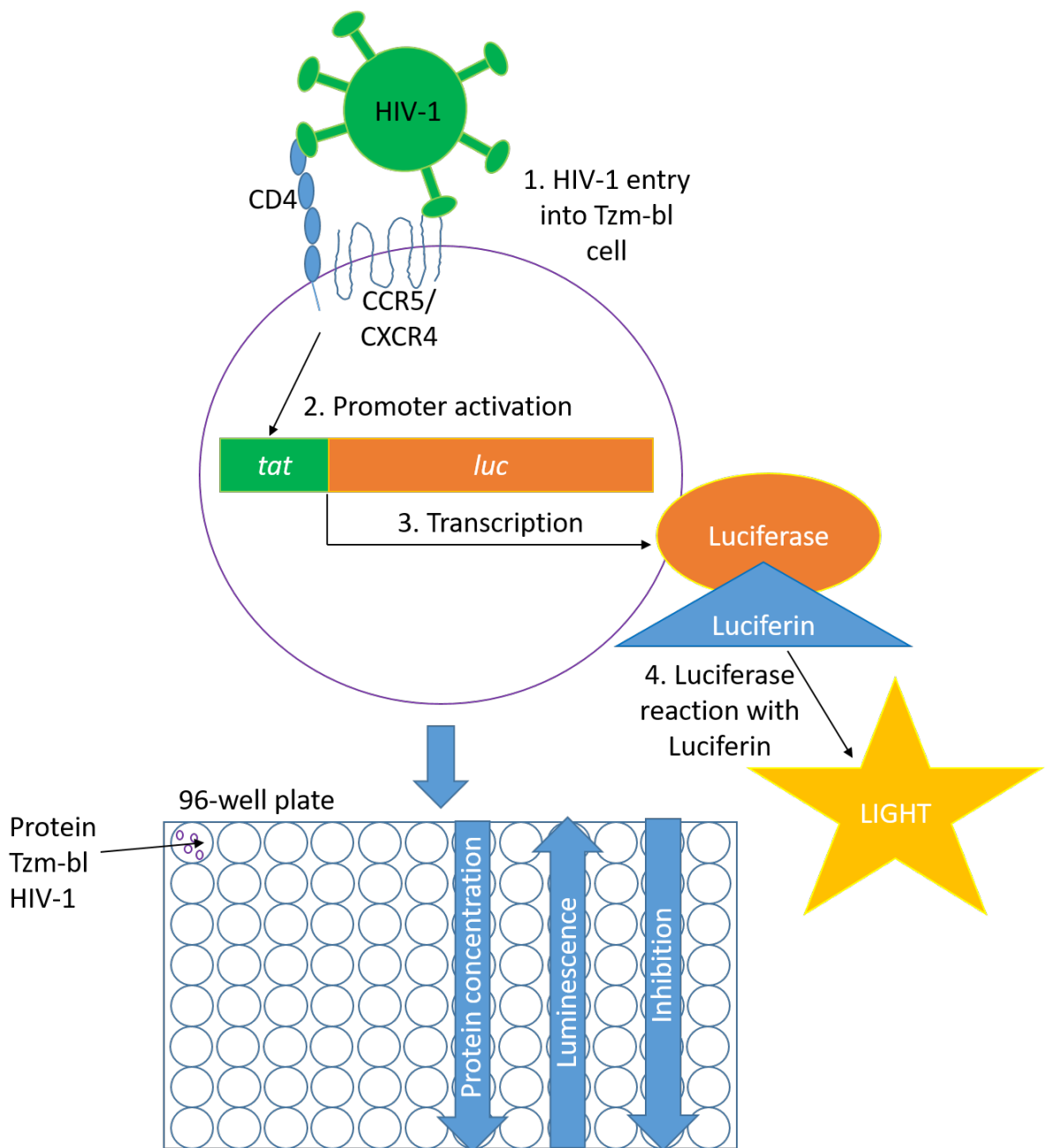


FIGURE 2.4: Pseudo-viral neutralisation assay measures the decrease in HIV-1 *tat*-regulated transcription of the *luc* reporter gene. 1) HIV-1 binds to CD4 and CCR5 or CXCR4 2) triggering activation of the *tat* promoter which triggers 3) transcription of the *luc* gene. 4) The resulting luciferase enzyme reacts with luciferin to produce oxyluciferin and light. Proteins can be added to the reaction to inhibit HIV-1 binding thereby inhibiting *luc* transcription which means this assay can be used to test neutralisation of HIV-1 using different inhibitory proteins (usually antibodies). With increased protein (inhibitor) concentration, there is decreased luminescence showing increased inhibition.

is done using highly sensitive, sophisticated protein sequencers [148] such as the 492 Applied Biosystems gas-phase sequencer with an in-line 140C HPLC unit used in chapter 5.

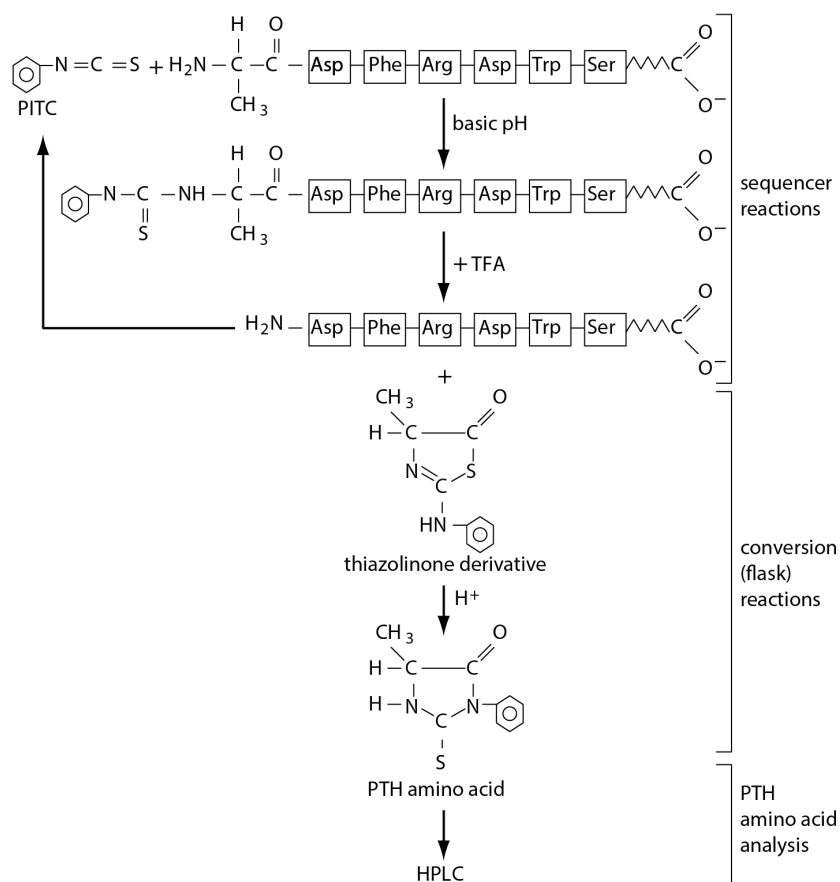


FIGURE 2.5: Schematic detailing the steps for N-terminal sequencing using Edman degradation. Adapted from: [148]

## 2.3 Biophysical characterisation

### 2.3.1 Mass spectrometry

Mass spectrometry is a useful biophysical technique for the accurate determination of the mass of a sample, whether the experimental mass is the same as the expected mass or not. It is therefore used as a quality control step prior to further analysis. Applied to the field of neutron scattering, mass spectrometry is used to estimate the percentage incorporation of deuterium in a recombinant protein produced by *in vivo* or cell-free deuteration techniques.

Biophysical mass spectrometry functions by producing intact gas-phase ions of biological macromolecules of specific mass-to-charge ( $m/z$ ) ratios. The ions are accelerated to specific velocities



by applying an electric current which drives the ions into a mass analyser. In this work time-of-flight (TOF) mass analysers were used. TOF mass analysers function on the principle that ions of different  $m/z$  are accelerated with the same energy as they are produced by the ionisation source but have a different velocity. Therefore, ions with smaller  $m/z$  ratios will arrive at the detector more quickly than ions with higher  $m/z$  ratios [150].

Two different types ionisation technique were used in this work: liquid chromatography electrospray ionisation time-of-flight mass spectrometry (LC/ESI-TOF-MS) and matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF). These are considered 'soft' ionisation methods as the energies used to induce the ionisation event are not high enough to incur fragmentation of the sample.

#### **Liquid chromatography/electrospray ionisation time-of-flight mass spectrometry**

LC/ESI-TOF-MS is a mass spectrometry technique that is useful for studying proteins or peptides with an  $m/z$  ratio of between 30-32 000. This technique was therefore useful for studying 2dCD4-WT. Prior to ionisation, the samples are diluted in acid denaturing conditions and then trapped and desalted on a reverse phase-column [151]. Ionisation can then occur by applying a voltage to the dilute sample in solution as it is ejected from a capillary with a fine gold-plated tip, causing the dispersion of a fine spray of charged droplets [150, 152]. The voltage used for the LC/ESI-TOF-MS experiments in this thesis was 4 kV in the positive ion acquisition mode.

#### **Matrix assisted laser desorption/ionisation time-of-flight mass spectrometry**

MALDI-TOF-MS can be used to study proteins with an  $m/z$  of up to 500 000 and was therefore used for studying gp120. In MALDI-TOF-MS, a laser irradiates a mixture of the sample with a chromophore matrix which expands rapidly into the gas phase. The matrix is designed to absorb the laser radiation and in order to do so, is typically comprised of aromatic compounds with carboxylic acid functional groups. The matrix used in the MALDI-TOF-MS experiments in this thesis was sinapinic acid (fig. 2.6) dissolved in 50/50/0.1 water/acetonitrile/trifluoroacetic

acid. The aromatic groups absorb the radiation from the laser which leads to subsequent desorption of both the matrix and biological macromolecule into the gas phase. There is an ionisation event whereby protons are transferred from excited matrix molecules to the sample and vice versa in both the solid phase and gas phase via collision [150, 153].

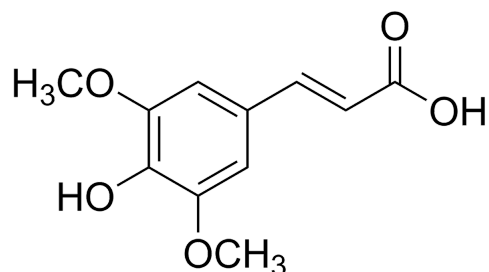


FIGURE 2.6: The matrix used in MALDI-TOF-MS experiments in this thesis work was comprised of sinapinic acid which is an aromatic compound able to absorb laser radiation.

### 2.3.2 SEC-MALLS-RI

Size-exclusion chromatography coupled to multi-angle laser light scattering and refractive index (SEC-MALLS-RI) is a powerful technique, providing useful information regarding the absolute molecular weight, oligomerisation state, concentration, polydispersity and homogeneity of the sample. When coupled to a chromatographic separation technique such as SEC, MALLS allows determination of the presence of different molecular weight species in a sample since SEC permits separation of different sized particles in solution and thus determination of the homogeneity of the sample [154].

The refractive index can inform about the polydispersity of each particle as indicated by the mass, whereby a fluctuation in mass indicates polydispersity. The RI component also allows accurate determination of the concentration by measuring the change in refractive index with respect to the concentration of the protein. The oligomerisation state of a sample can also be determined since the absolute mass is given [150, 155]. SEC-MALLS-RI is, therefore, a useful technique to control the quality of the samples prior to any small angle scattering experiment.

Particles of different size in solution are initially separated by SEC and the UV trace, measured by a UV detector, allows the user to follow the presence of protein (or nucleic acid) as it is eluted from the column [154]. Hereafter, the solution is illuminated by a laser light source as it passes through a flow cell surrounded by a series of light scattering detectors positioned

at different angles which measure the scattered intensity of light produced by the particles. The light scattering can inform about non-protein contaminants which would not be observed by the UV trace during SEC. Finally, the refractive index is measured and recorded by the refractive index detector [13, 155]. The sample can also be recovered by elution into a fraction collector, as is standard with SEC.

The absolute molecular mass of the particles can be derived from a MALLS experiment using the Zimm [156] equation:

$$\frac{(K^*C)}{R(\theta)} = \frac{1}{M_w P(\theta)} + 2A_2C \quad (2.4)$$

Where  $R(\theta)$  is the Rayleigh ratio,  $M_w$  is the weight-averaged molecular mass,  $A_2$  is the second virial coefficient which is measure of the interaction between the solute and solvent,  $C$  is the concentration of the solvent and  $K^*$  is a constant which defines vertically polarised incident light:

$$K^* = \frac{4P^2 n_0^2 (dn/dC)^2}{\lambda_0^4 N_A} \quad (2.5)$$

Where  $N_A$  is Avogadro's number,  $n_0$  is the RI of the solvent,  $\lambda_0$  is the vacuum wavelength of the incident light and  $dn/dC$  is the RI increment [155].

### 2.3.3 1D Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a technique which operates within the radio frequency end of the electromagnetic spectrum. NMR is based on the principle that certain nuclei have a spin which gives them both electrical properties and an angular magnetic momentum [150, 157, 158]. A nucleus of spin  $I$  has  $2I + 1$  energy levels ( $E$ ) which are equally spaced:

$$\Delta E = \frac{\mu B_0}{I} \quad (2.6)$$

When an external magnetic field ( $B_0$ ) is applied at any angle relative to the moment, the nuclear magnetic moment [157, 158] ( $\mu$ ) of the nuclei precesses about the direction of the applied magnetic field with a frequency defined by Larmor:

$$V_0 = \frac{\gamma B_0}{2\pi} \quad (2.7)$$

Where  $\gamma$  is the gyromagnetic ratio. The nuclear magnetic moment ( $\mu$ ) can be defined by:

$$\mu = \frac{\gamma h I}{2\pi} \quad (2.8)$$

Where  $h$  is Planck's constant.

In NMR the chemical dispersion or shift [159] ( $\delta$ ) is measured as the difference in resonance frequencies between that of the nucleus of interest ( $V_x$ ) and that of a reference nucleus ( $V_{ref}$ ):

$$\delta = \frac{V_x - V_{ref}}{V_0} \quad (2.9)$$

The chemical shift of a nucleus is dependent on the surrounding electron density. The  $\delta$  values are recorded on an NMR spectrum in parts per million (ppm) and they define different nuclear environments. Therefore, for a one-dimensional  $^1\text{H}$  NMR experiment, the chemical shift peaks define different proton environments [160]. Most proton resonance peaks within a protein lie within the region  $\delta 1$  to  $\delta 13$  (1 to 13 chemical shift in ppm), with resonance peaks for methyl group protons at around  $\delta 1$  and the amide group protons between  $\delta 7$  and  $\delta 11$  [150]. 1D proton NMR can thus be applied to the study of the extent of folding of a protein.

In an extended, flexible polypeptide chain, the amino acid side groups are exposed to the same solvent environment. This means that repeating amino acid residues throughout the polypeptide chain have a nearly identical  $^1\text{H}\delta$  such that the resonance peaks in the NMR spectrum of a randomly coiled polypeptide, correspond roughly to the sum of their amino acid resonance peaks. However, in a globular, folded, protein the protons experience a range of different environments. Some protons are found in solvent exposed regions but others are shielded from the solvent and surrounded by other fragments of the polypeptide. In the case of the amide protons, which are bound to nitrogen and are therefore labile, there is assumed to be complete exchange with solvent deuterons in an unfolded polypeptide chain, whereas in a globular, folded protein many of the amide groups are shielded from the solvent. Thus, in a 1D proton NMR experiment, it is useful to look for the presence of amide and methyl proton resonances outside of those experienced by a randomly coiled polypeptide, to obtain a qualitative idea of how folded the protein is [150].

The compactness of the protein can also be detected in a quantitative manner using a technique called 1D Heterogeneity-Selective Optimised-Flip-Angle Short-Transient (HET-SOFAST) NMR which measures the difference in the nuclear Overhauser [161] effect  $\lambda_{NOE}$  between a  $^1\text{H}$  reference and a  $^1\text{H}$  saturated spectra.  $\lambda_{NOE}$  reports on the local  $^1\text{H}$  spin network, measuring the proton density and mobility at the amide proton sites. Since the protons of a well-folded globular structure are surrounded by other peptide fragments and generally shielded from the solvent this allows efficient proton spin diffusion. On the other hand, for an unfolded polypeptide chain the proton density is low therefore the proton spin diffusion is poor [162]. Therefore, low values of  $\lambda_{NOE}$  define a compact structure whereas higher values of  $\lambda_{NOE}$  correspond to flexibility and the degree to which the protein is unfolded. Figure 2.7 shows the 1D HET-SOFAST NMR scale of polypeptide classifications.

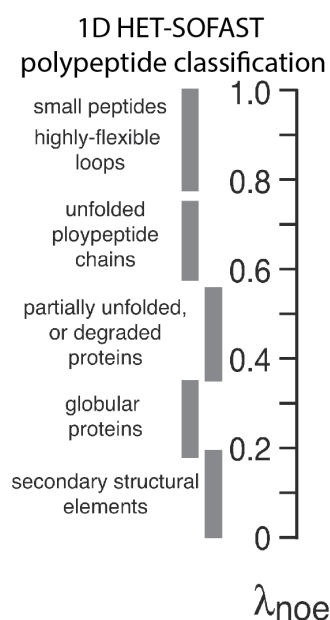


FIGURE 2.7: 1D HET-SOFAST NMR gives a quantitative  $\lambda_{NOE}$  score on a scale from 0 to 1.0 which describes the extent of folding of a protein. Adapted from: [162]

## 2.4 Small angle scattering (SAS)

Small-angle scattering (SAS) is a powerful technique allowing determination of the shape, size and interaction of particles in solution. Applied to biological systems SAS can be useful for: high-resolution structure validation, determination of the low-resolution envelope structure, the stoichiometry of a multimer or complex, conformational changes upon ligand binding. For

this thesis work, the interest lies within the determination of conformational changes upon complex formation.

### 2.4.1 Scattering theory

Whilst X-rays and neutrons are two different types of radiation which interact very differently with their samples (explained later), the underlying principle and fundamental mathematics behind small-angle scattering are common to both techniques. In both instances a monochromatic incident beam ( $\mathbf{k}$ ), interacts with the atoms of the sample with a wavevector of modulus:

$$k = |\mathbf{k}| = 2\pi/\lambda \quad (2.10)$$

To produce spherical scattered waves of which the scattered beam ( $\mathbf{k}'$ ) has a wavevector with a modulus of:

$$k' = |\mathbf{k}'| = k \quad (2.11)$$

Thus, the scattering is elastic because there is no change in the energy, that is to say the wavelength of the incident beam is equal to that of the scattered beam. Elastic scattering arises due to the interference of scattered waves produced by the point scatters (atoms of the sample), which have correlated positions in space ( $\mathbf{r}$ ), within the coherent scattering volume. In elastic scattering only the direction of the wavevector changes, the magnitude stays the same [163]. A wave has momentum:

$$p = \hbar k \quad (2.12)$$

The momentum transfer can also be given by:

$$q = k' - k \quad (2.13)$$

The scattering is isotropic because the particles are randomly oriented in solution with respect to the incident beam. In a monodisperse ideal solution, rotational averaging of the scattered intensity ( $I(q)$ ) provides information on the size and shape of the particles suspended in solution:

$$I(q) = \langle \mathbf{I}(\mathbf{q}) \rangle_{\Omega} \quad (2.14)$$

Where  $\Omega$  is the solid angle in reciprocal space. SAS measurements are made in reciprocal space and are described by the scattering vector,  $q$ . As the scattering is isotropic, because the particles are randomly oriented in a random spacial distribution, the scattered intensity is dependent on the modulus of the momentum transfer:

$$q = |\mathbf{q}| = 4\pi/\lambda \sin\theta \quad (2.15)$$

Where the scattering angle is  $2\theta$  at wavelength  $\lambda$  ( $\lambda=2\pi/k$ ).

The reciprocal-space coordinates ( $q$ ) are related to real-space coordinates ( $r$ ) by Fourier transform. In small-angle scattering inhomogeneities in the scattering-length/electron density determined by the chemical composition of the scatterer are measured and the scattering length density depends on the nuclei and the isotopes that comprise the scatterer. The scattering arises from the contrast between the scattering-length/electron density distribution of the particle ( $\rho(r)$ ) and the homogeneous solvent of constant scattering-length/electron density ( $\rho_s$ ). Since small-angle scattering cannot resolve interatomic distances (atomic length scales are not accessible using this technique) it is not possible to separate the contributions of individual atoms from the total scattering. All the distances within a particle are therefore measured by SAS as a distribution of their average scattering length density  $\rho(r)$ . This is equal to the sum of the total scattering lengths of the atoms per unit volume of the object or particle of interest [163].

The difference in the scattering length density of the particle compared to that of the solvent of equivalent volume, yields the excess scattering length density:

$$\Delta\rho(r) = \rho(r) - \rho_s \quad (2.16)$$

This is related to the scattering amplitude of the particles by a Fourier transform:

$$A(q) = \mathfrak{F}[\rho(r)] = \int_V \Delta\rho(r)e^{iqr} dr \quad (2.17)$$

Where  $V$  is the particle volume and  $A(q)$  is the scattering amplitude which cannot be measured directly. It is possible, however, to measure the intensity ( $I(q)$ ) which is related to the amplitude by the following equation:

$$I(q) = A(q)(A * (q)) \quad (2.18)$$

Which corresponds to the scattered energy. The scattering intensity can also be described by:

$$I(q) = P(q)S(q) \quad (2.19)$$

Where  $P(q)$  is the form factor which corresponds to the scattering amplitude and  $S(q)$  is the structure factor which is the intensity of the scattered radiation.  $S(q)$  describes the scattering that arises from the interaction of the particles in solution, which becomes problematic for determining  $P(q)$ . Therefore, dilution series and extrapolation to zero concentration is used to get around this. In ideal, monodisperse solutions  $S(q) = 1$  so that  $P(q)$  can be determined and therefore the structure of the scatterer can be calculated [164].

### **A monodisperse, ideal solution**

For small-angle scattering it is important for the solution to be as ideal and monodisperse as possible. Ideality refers to the lack of interactions between the particles in suspension, whether they be attractive or repulsive. In other words the particles must be distributed with random position and orientation within the solution with no spatial correlations that could contribute to the scattering intensity. In an ideal solution there is a linear dependence on the concentration. In theory the solution should be infinitely dilute to negate interparticle interactions. Experimentally, it is useful to use a dilution series of the molecule of interest from relatively concentrated to relatively dilute (10 mg/mL to 1 mg/mL). The data can then be extrapolated to zero concentration [163].

Monodispersity refers to there being only one unique particle within the solution. That is to say that the particle must be pure with a defined shape and size. Notably for biological samples, the solution can be simultaneously chemically monodisperse but polydisperse in shape. A common issue with protein samples is aggregation in which large aggregates of the same protein form so that the shape and size of the sample becomes polydisperse whilst maintaining a monodisperse chemical formula [163].

In practice, protein aggregates causing polydispersity can be seen at low- $q$  values: aggregation manifests as an inflection of the scattering data at the lowest- $q$  values measured. This arises from the fact that the scattering is related to the square of the excess scattering length ( $\Delta\rho$ ), therefore, larger particles skew the scattering contribution disproportionately so that few large



aggregates dominate. Prior to carrying out a SAS experiment it is therefore necessary to assess the monodispersity of the sample using separate biophysical measurements by: dynamic light scattering (DLS), multi-angle light laser scattering (MALLS) or analytical ultra-centrifugation. Samples of 95% monodispersity or higher are considered suitable for SAS measurements [163]. MALLS was the method of choice used in this thesis work.

#### **2.4.2 Small-angle X-ray scattering (SAXS)**

In SAXS, the X-ray beam is scattered by the electron cloud surrounding the atom. The scattering is proportional to the atomic number ( $Z$ ) of the atom and SAXS is therefore concerned with the electron density. The X-rays used in SAXS have relatively high energies (10 keV) with wavelengths in the range of 0.1-0.15 nm and are typically produced by synchrotron radiation. All of the SAXS experiments described within this thesis work were carried out on BM29 at the ESRF. This beamline is specially adapted for the measurement of biological samples and its configuration is outlined below.

##### **The BM29 instrument at the ESRF**

BM29, as the name suggests, is situated on bending-magnet 29 of the ESRF. It is a small-angle X-ray scattering beamline operating at a wavelength range between 0.82 and 1.77 Å. BM29 is dedicated to the measurement of biological macromolecules thanks to its automated, high-throughput set-up and high flux at the sample. In addition, specialised optics have increased the  $q$ -range and reduced the parasitic scattering and measuring times, compared to the former BioSAXS beamline: ID14-3 [165]. The diagram in figure 2.8 outlines the set-up of BM29.

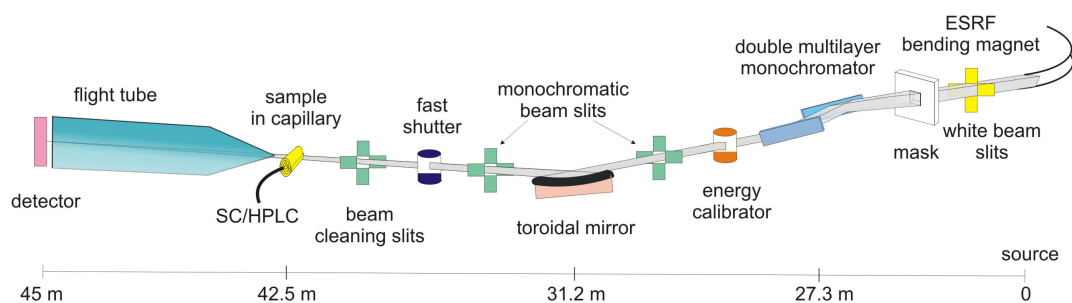


FIGURE 2.8: Diagram depicting the layout of the BioSAXS beamline BM29 at the ESRF, Grenoble, France. From: <https://www.ill.eu/users/instruments/instruments-list/d22/characteristics/> [165]

Samples can be loaded using a state-of-the-art automated and temperature-controlled sample changer, which has the capacity for a 96-well plate, PCR-tube, Eppendorf tube or Falcon tube set-up, giving a maximum capacity of 288 separate aliquots for measurement [166, 167]. In addition the sample changer allows the sample to be flowed through the capillary to reduce the effects of radiation damage [166]. Alternatively, the samples can be loaded via the on-line F/HPLC instrument which allows a purification step prior to SAXS measurement [168, 169]. Samples can also be loaded manually if necessary for specific experiments. The samples are measured in a quartz glass capillary of either 1.8 or 1 mm in diameter. The capillary is washed rigorously between samples as part of the automatic sample changer's protocol.

### 2.4.3 Small-angle neutron scattering (SANS)

Neutrons are scattered by the nucleus of the atom and the scattering is therefore, not proportional to the atomic number of the atom. Therefore SANS is concerned with the scattering-length density and contrast. SANS requires thermal neutrons with wavelengths in the range of 0.10-1.0 nm. Neutrons for SANS experiments can be produced at spallation sources or, in the case of this thesis work, fission sources. The SANS experiments carried out in thesis were conducted on the D22 diffractometer instrument at the ILL.

### The D22 instrument configuration at the ILL

D22 is a large-scale structure diffractometer which specialises in weakly scattering samples, such as dilute proteins in solution, due to its relatively high flux at the sample of  $1.2 \times 10^8$  neutrons  $\text{cm}^2 \text{s}^{-1}$  (for  $\frac{\Delta\lambda}{\lambda} = 10\%$ ) with a wavelength range of 0.45 to 4 nm. The high flux is delivered by using a horizontal cold source, a high-tech velocity selector with a short rotor length (25 cm) and high transmission and a large beam cross-section (55 X 40 mm). The collimation length which defines the virtual source-to-sample distance is achieved by a rotating-drum collimation system providing 8 collimation lengths between 1.4 m and 17.6 m [163].

The sample is measured in a quartz cuvette with a path-length of 1-2 mm, which is contained within a 22 position, thermostat-controlled sample changer. The  $^3\text{He}$  detector has an active area of  $1 \text{ m}^2$  and can be moved horizontally at 50 cm intervals to achieve detector distances between 1.1 and 17.6 m, inside an evacuated tube of 20 m in length and 2.5 m in diameter. This allows D22 to access a large dynamic  $q$  range of  $4 \times 10^{-4}$  to  $0.44 \text{ \AA}^{-1}$  (or  $0.85 \text{ \AA}^{-1}$  with detector offset)[170]. Figure 2.9 shows the configuration of the D22 instrument.

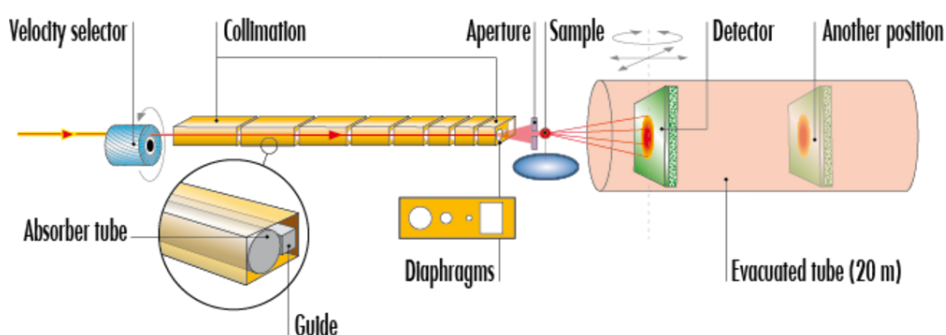


FIGURE 2.9: Diagram depicting the layout of the SANS instrument D22 at the ILL, Grenoble, France. From: [170]

### Contrast variation

Contrast variation is a powerful aspect of SANS which allows the isolation of the scattering of a component in a complex such as of DNA in a protein-DNA complex. As such the exploitation of contrast variation adds an additional level of information which can be obtained from a SANS experiment compared to what is attainable using SAXS. Contrast, as it appears to neutrons, arises from the fact that each atom has a unique coherent and incoherent scattering length.

Atom	H	D	C	N	O	P	S
Atomic mass	1	2	12	14	16	30	32
Z electrons	1	1	6	7	8	15	16
$b_X$ ( $10^{-12}$ cm)	0.282	0.282	1.69	1.97	2.16	3.23	4.51
$b_N$ ( $10^{-12}$ cm)	-0.374	0.667	0.665	0.940	0.580	0.510	0.280

TABLE 2.1: The atomic mass, number of electrons and X-ray and coherent neutron scattering lengths of the atoms most commonly found in biological macromolecules. From: [163]

Coherent scattering contributes to the SANS signal providing information on the structure. However incoherent scattering can be considered as "noise" for the purpose of small-angle neutron scattering studies and mainly arises from  $^1\text{H}$  [163]. Conversely in dynamical studies the incoherent neutron scattering is the signal that informs on the movement of the atoms. Since each atom has a unique coherent and incoherent scattering length, the scattering length density of a molecule, is unique to its molecular composition within its molecular volume  $V_M$  and is represented thus:

$$SLD = \frac{\sum_{i=1}^N b_i}{V_M} \quad (2.20)$$

Where  $b_i$  corresponds to the scattering length of  $N$  atoms.

Table 2.1 presents a list of the coherent neutron scattering lengths for the typical atoms which are found in biological systems compared to their X-ray scattering lengths. Most interesting is the difference between the coherent and incoherent scattering lengths of the two hydrogen isotopes: protium and deuterium. Whilst protium has a negative coherent scattering length and large incoherent scattering length, deuterium on the other hand has a positive coherent scattering length and much smaller incoherent scattering length. In an experimental setting this means that neutron scattering by deuterium contributes greater to the signal with less noise whereas protium contributes strongly to the background. For biological systems this is a useful phenomenon for the purpose of contrast variation studies whereby the experimentalist can play with deuterium/protium contrasts for match-out studies [163].

Natural contrast exists between different biological macromolecules as indicated in figure 2.11. At a given  $\text{H}_2\text{O}/\text{D}_2\text{O}$  ratio in the solvent the scattering length density of the macromolecule is equal to that of the solvent and the macromolecule is rendered invisible to neutrons which is called the match-out point. For hydrogenated proteins approximately 40% of the solvent's hydrogens must be deuterium. The contrast can be increased by deuterium labelling the protein

so that 75% of the protein's non-exchangeable hydrogens are deuterated so that the match-out point is roughly 100% deuterated solvent (fig. 2.10) [171].

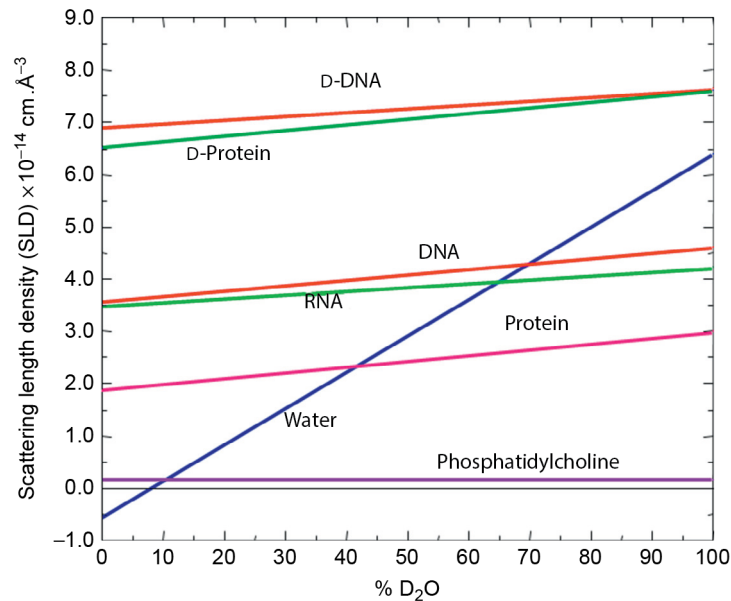


FIGURE 2.10: Different biological macromolecules have different scattering-length densities (SLD) due to the difference in their chemical composition. The difference in SLD corresponds to a difference in the match-out point - the percentage of  $D_2O$  at which the scattering of the solvent is the same as that of the molecule. From: [171]

Deuterium labelling can also be exploited for isolating the scattering of molecules within a complex. For example the solvent H/D ratio can be adjusted to look at the isolated scattering from a protonated protein in complex with a deuterated protein [171]. Figure 2.11 outlines a typical contrast variation SANS study.

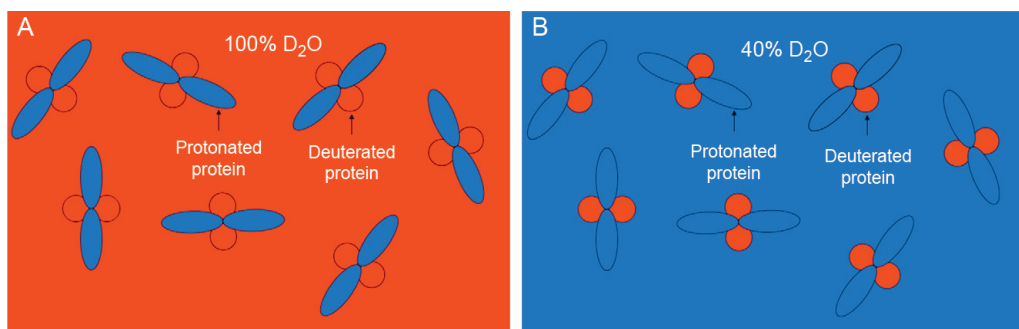


FIGURE 2.11: In 100%  $D_2O$  (A) the SLD of the deuterated protein is the same as that of the solvent so only the protonated protein can be seen. In 40%  $D_2O$  (B) the SLD of the protonated protein is the same as that of the solvent so that the deuterated protein is visible. From: [171]

## Deuterium labelling of proteins

In order to be able to fully exploit contrast variation during SANS studies, as mentioned previously, deuterium labelling is necessary. In order to obtain high levels of deuterium labelling, whether that be match-out labelled (75% deuterated) or perdeuterated (100% deuterated), *in vivo* deuteration is often the method of choice. The Deuteration Laboratory (D-Lab) based at the Institute Laue Langevin is a facility which specialises in the deuteration of biological macromolecules using a variety of expression systems for neutron and NMR studies [171, 172]. All deuteration work carried out in this thesis was done with the D-Lab at the ILL.

For recombinant deuterated protein production the most commonly used expression systems for *in vivo* deuteration are *Escherichia coli* and *Pichia pastoris* which can be adapted to culture in deuterated media and with a deuterated carbon source (in the case of perdeuteration). Since deuterated media is expensive the most cost-effective way of expressing the protein using these two expression systems is by fed-batch fermentation using a minimal medium. Using a fermenter allows the user to control the temperature, agitation, dissolved oxygen, pH and feed to reach high cell-densities before induction. Under certain circumstances, some proteins will not express well in these conditions in which case flask culture can be used. Silantes media (Silantes) is a form of rich medium which exists in protiated and deuterated forms and can be used for deuteration in flasks with a limitation to the  $OD_{600nm}$  which can be achieved, due to the composition of the medium. However, this is a more expensive method since the same high cell-densities cannot be achieved as with fed-batch fermentation [171].

Prior to expression in a flask or fermenter the single-cellular expression host must first be adapted to culture in deuterated conditions. This is done by a gradual process of culturing overnight a small volume of deuterated media (9 mL) with a small volume of the protiated culture (1 mL), followed by several passages of the deuterated culture into fresh deuterated media. Over the course of the passaging process (at least 1 week) the chosen expression host adapts to the deuterated media allowing higher cell-densities to be obtained. Expression tests should also be carried out to ensure that the recombinant protein is still expressed in deuterated conditions. After the final passage, a starter culture of approximately 200 mL is used to inoculate the fermenter. Appendix B.4.2 lists the ingredients for the minimal media used during fed-batch fermentation.

### 2.4.4 Data analysis

#### Guinier analysis

Guinier analysis of a 1-dimensional (1D) scattering curve is carried out by plotting  $\ln I(q)$  vs  $q^2$  and is focused on the low- $q$  range of the plot. Guinier's law [173] is given by:

$$I(q) = I(0)e^{\left(\frac{-q^2 R_g^2}{3}\right)} \quad (2.21)$$

For  $qR_g \leq 1.3$ . From the Guinier analysis the forward scattering intensity at 0 angle ( $I(0)$ ) and the radius of gyration ( $R_g$ ) can be determined. The  $I(0)$  is a useful parameter because it allows determination of the MM of the particle in  $\text{g mol}^{-1}$  using:

$$MM = \frac{I(0) \times N_A}{C(\Delta\rho\bar{v})^2} \quad (2.22)$$

Where  $N_A$  is Avogadro's number,  $C$  is the concentration in  $\text{g cm}^3$ ,  $\Delta\rho$  is the contrast in  $\text{cm}^{-2}$  and  $\bar{v}$  is the partial specific volume of the protein. For proteins  $\bar{v} = 0.743 \text{ g cm}^{-3}$  [174].

The  $I(0)$  must be corrected for the scattering of the cuvette or capillary, the transmission of the sample, normalised by the concentration of the particle in solution and finally placed on an absolute scale. For X-rays this is done by dividing the normalised  $I(0)$  by the measured  $I(0)$  for water and multiplying by the absolute  $I(0)$  of water ( $1.632 \times 10^{-2} \text{ cm}^{-1}$ ). For neutrons the data can be scaled to absolute intensity by using a direct measurement of the flux.

The  $R_g$ , which is defined as the mean square distance from the centre of mass, is another useful parameter describing how the particle tumbles in solution and is, essentially, the measure of a particle's compactness [175].

The  $R_g$  is dictated by the slope of the data at low- $q$  of a Guinier plot ( $\ln I(q)$  vs  $q^2$ ) and should be a straight line. It can also be determined from the pair-distance distribution function. In both instances care must be taken in analysing the  $R_g$  as aggregation manifests strongly at low- $q$  in the Guinier plot as deviation from a straight line and as a long tail at high- $r$  (longer distances) in the  $P(r)$  plot.

### Pair-distance distribution function (P(r))

The pair-distance distribution function (PDDF or  $P(r)$ ) is a real-space representation of the shape and size of the particle between the limits  $0 \leq r \leq \infty$ . Since this interval is not accessible by the scattering curve, the PDDF is calculated by an indirect Fourier transform [176]. Ultimately the PDDF represents probability distribution,  $P(r)$ , of distances joining volume elements pairs within the particle,  $r$ , weighted by the product of the scattering density contrast between the limits 0 and the maximum dimension,  $D_{max}$ :

$$I(q) = 4\pi \int_0^{D_{max}} P(r) \frac{\sin(qr)}{qr} dr \quad (2.23)$$

### Kratky analysis

By plotting the data as  $I(q) \times q^2$  vs  $q$  (Kratky plot, [177]) the shape of the curve alludes to the extent of folding of the particle in solution. The relevant scattering features at high- $q$  are amplified using this plot and it can be used for determination of the compactness of the sample. A globular protein in solution will display a scattering intensity decay of  $q^{-4}$  which corresponds to a bell-shaped Kratky plot. For an unfolded protein the scattering intensity decay is around  $q^{-2}$  [178] which manifests as a plateau at high- $q$  in the Kratky plot. Figure 2.12 depicts the Kratky plot of particles of different compactness in solution, from which it can be seen that unfolded particles display a Kratky plot which never tends to  $I(q) = 0$ .



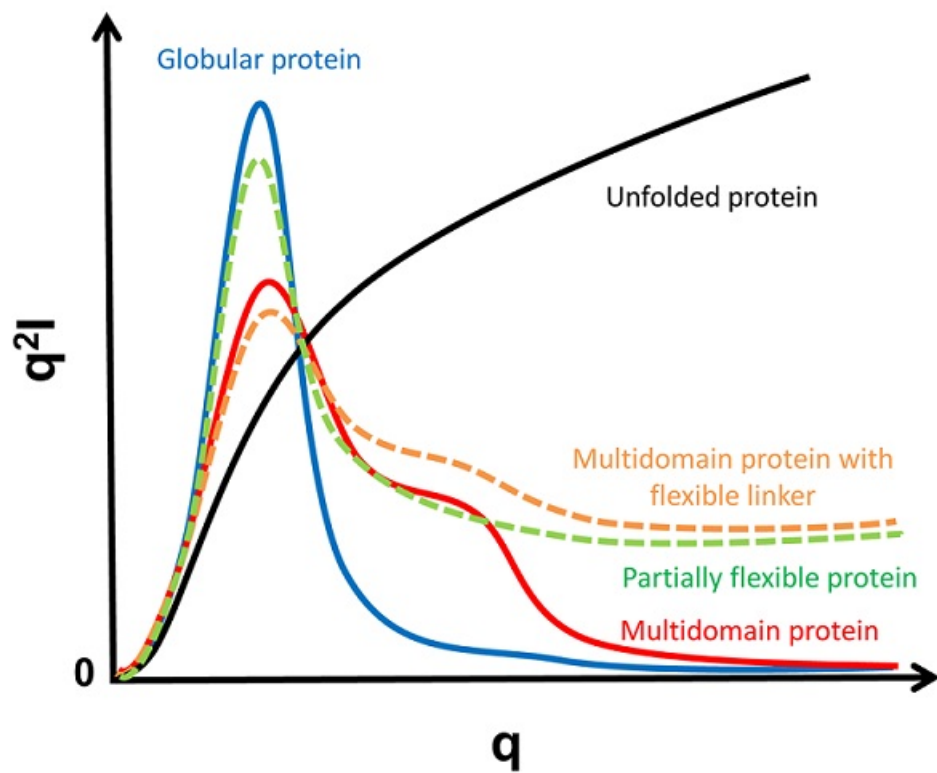


FIGURE 2.12: The Kratky plots for biological macromolecules exhibiting different compactness is displayed. A globular particle exhibits a bell-shaped curve, whereas the curve for an unfolded or flexible particle never converges to  $I(q) = 0$ . From: <https://www-ssrl.slac.stanford.edu/saxs/analysis/assessment.htm>

[179]



## Chapter 3

# Concerning the expression and purification of 2dCD4

### 3.1 Abstract

An important element of this thesis work involved the development of an expression system for wild-type two domain CD4 (2dCD4-WT) which could produce sufficient quantities of pure, soluble protein in both protiated and deuterated forms for structural studies. This task proved to be non-trivial and the main bottle neck in allowing further or deeper analysis to be carried out.

The initial expression system used was the classical *Escherichia coli* expression host. Due to the disulphide bond content of 2dCD4-WT, *E. coli* expresses 2dCD4-WT into insoluble inclusion bodies. In order to obtain soluble, native protein, a denaturing purification followed by a refolding protocol was used. However, much of the protein was lost during the refolding step and the resultant protein was generally not very stable.

The second approach used was a novel bacterial expression system called *Brevibacillus choshinensis* which is a gram-positive bacterium that secretes many proteins into the extra-cellular milieu. This secretion process allows the formation of disulphide bonds due to the oxidising environment outside of the cell. Because the protein is secreted into the growth medium there was a large volume of protein-containing solution to work with. Size-exclusion chromatography also showed that the protein had a tendency to aggregate, so the yields were low and the stability poor.

The final expression system used was set-up in collaboration with the company Synthelis, France, which developed a cell-free protein expression (CFPE) system to produce difficult-to-express proteins. CFPE removes the constraints of *in vivo* protein production and decouples protein synthesis from cell culture. Using their proprietary technology it was possible to produce both hydrogenated and deuterated 2dCD4-WT in soluble, stable forms with which SAXS and SANS experiments could be carried out. Finally, the 2dCD4 domain 2 disulphide bond knockout variant was expressed using CFPE but while the expression was good, there were severe issues with stability and aggregation.

This chapter describes these developments to obtain pure, soluble and stable 2dCD4-WT which was later biochemically and biophysically characterised in chapter 4 and used in SAXS and SANS experiments in chapters 5 and 6.

## 3.2 Introduction

Conventionally, *Escherichia coli* is the traditional expression system, particularly when isotopic labelling is required due to its fast replication rate, adaptability to isotopically labelled media and relatively low cost [171]. However, for proteins containing disulphide bonds, such as CD4, production in *E. coli* results in expression of the protein as insoluble inclusion bodies. Recovery of soluble CD4 from inclusion bodies necessitates a denaturing purification protocol followed by a laborious refolding protocol, often resulting in significant losses of protein and low yield [63, 180].

*Brevibacillus choshinensis* is a laboratory strain of the bacterium *Bacillus brevis* which has advantages over *E. coli* in that proteins can be secreted into the extracellular milieu [181]. This allows the correct formation of disulphide bonds [182, 183], thereby bypassing the refolding steps required during *E. coli* expression. However, while *Brevibacillus choshinensis* grows well in <sup>15</sup>N labelled media [184], there is currently no literature which describes *B. choshinensis* growth in deuterated complex media to produce the yields of deuterated protein required for neutron structural studies.

The cell-free protein expression (CFPE) system is an attractive expression system for the production of high quality protiated and deuterated protein for structural analyses by neutron

scattering (such as small-angle scattering or reflectivity studies). CFPE has notable advantages over other expression systems such as bacterial, yeast and mammalian systems as it liberates the experimentalist from the constraints of cellular expression and decouples protein expression from cell growth. As such, the chemical composition of the cell-free reaction can be tailored to meet the requirements of the experimentalist for the protein of interest.

In contrast to bacterial cell-based methods, cell-free protein expression (CFPE) allows production of good yields of soluble protein in relatively small volumes [185, 186]. In addition, the openness of CPFE means that the reaction composition can be easily adapted to the needs of the experimentalist by the simple addition of specific reagents [187, 188]. For example the reaction can be modified to promote the correct formation of disulphide bonds [189]. In the case of deuteration for NMR or SANS, for example, unlabelled amino acids can be substituted by deuterated ones [190, 191].

CD4, as the primary receptor of viral gp120, has been heavily studied in both its native and recombinant forms [34, 98, 121]. Most structural studies involving CD4 use recombinant CD4 produced using mammalian expression systems; the initial crystal structure study of two-domain CD4 obtained in 1990 was carried out using recombinant CD4 from Chinese hamster ovary cells [67]. While mammalian expression systems are excellent for the production of high yields of soluble, unlabelled protein useful for SAXS or crystallography experiments, they are not well adapted to small angle neutron studies or NMR experiments requiring isotopic labelling due to the requirement for complex growth media and their sensitivity to high deuterium content in the growth media [171].

2dCD4-WT needed to be produced in milligram quantities in both protiated and deuterium labelled forms in order to carry out subsequent analyses in this thesis which would address the question of why 2dCD4-WT reduced in its second domain preferentially binds to gp120. Deuterium labelling of 2dCD4-WT was essential for SANS studies in order that the gp120/2dCD4-WT complex be investigated using contrast variation SANS experiments. 2dCD4 in which its second domain cysteine residues have been replaced by alanine residues (2dCD4-D2A) mimics the partially reduced form of 2dCD4-WT and was therefore also required in milligram quantities in order that biophysical and structural analyses be carried out on this system.

Therefore, this chapter describes the optimisation of expression of deuterium labelled and unlabelled 2dCD4-WT. Expression was developed from the classical *E. coli* expression system, through the novel gram positive *B. choshinensis* expression system to eventually arrive at an *in vitro* cell-free protein synthesis protocol which produced sufficient yields of 2dCD4-WT for subsequent biochemical, biophysical and small-angle scattering studies. Expression of the 2dCD4-D2A was also explored but the protein appeared unstable, prone to aggregation and resulting yields were not sufficient for further analysis.

### 3.3 Materials and methods

Details of the recombinant protein sequences and the expression vectors can be found in appendix A.2. The composition of the media, buffers and general protocols for SDS-PAGE with Coomassie blue and anti-his western blotting can be found in appendix B.4.2.

#### 3.3.1 Plasmid DNA preparation

The same protocol was used for all plasmid DNA preparation. An aliquot of *E. coli* Top10 chemically competent cells (Invitrogen) was incubated with 1 ng of plasmid DNA on ice for 30 minutes. The cells were heat-shocked at 42°C for 30 seconds and placed back on ice for 2 minutes. 250 µL pre-warmed SOC (super optimal broth) media was added to the cells which were incubated at 37°C, 180 rpm for 1 hour. The cells were centrifuged for 10 minutes at 3 000 rpm and resuspended in 100 µL of the same media. The cells were plated on LB (Luria broth) agar plates supplemented with 1 X antibiotic corresponding to the plasmid's antibiotic resistance gene (kanamycin for pETm11 and pMK-T and ampicillin for pET15b, pNCMO2 and piVEX 2.3d). The plates were incubated overnight at 37°C.

The pETm11 and pET15b constructs were provided by the HIV Pathogenesis Research Unit, South Africa. The pMK-T construct was produced by Invitrogen. The pNCM02 construct was produced by Miss Sarah Waldie using the TakaRa *Brevibacillus* Expression System II kit. The piVEX2.3d constructs were produced by ProteoGenix, France and supplied by Synthelis, France.

### Plasmid DNA mini prep

The Promega Wizard Plus SV Minipreps DNA purification system was used for the preparation of plasmid DNA for the *E. coli* and *B. choshinensis* expression systems.

One Top10 colony from the overnight LB agar plate was incubated in 5 mL LB supplemented with 1 X antibiotic corresponding to the antibiotic resistance gene of the plasmid. The culture was incubated overnight at 37°C, 180 rpm. The culture was centrifuged at 10 000 rpm in a bench-top centrifuge for 5 minutes. The supernatant was discarded and any excess media was removed by blotting the tube upside down on a paper towel.

The pellet was thoroughly resuspended in 250 µL of the cell resuspension solution by pipetting and vortexing. 250 µL cell lysis solution was added and the tube inverted four times to gently mix the solutions. The tube was incubated at room temperature for 5 minutes. 10 µL alkaline protease solution was added and mixed by inverting the tube four times. The tube was incubated at room temperature for 5 minutes. 350 µL of the Wizard Plus SV neutralization solution was added and mixed by inverting the tube four times. The bacterial lysate was centrifuged at 14 000 rpm at room temperature in a bench-top centrifuge for 10 minutes.

The cleared lysate was transferred to a spin column (provided in the kit) without disturbing the white precipitate. The solution was centrifuged for 1 minute at room temperature at 14 000 rpm in a bench-top centrifuge. The supernatant was discarded. 750 µL of column wash solution diluted in 95% ethanol was added to the Spin Column and the tube was centrifuged for 1 minute at room temperature at 14 000 rpm. The supernatant was discarded. 250 µL column wash solution was added to the spin column and the tube was centrifuged for 2 minutes at room temperature at 14 000 rpm. The spin column was transferred to a fresh 1.5 mL Eppendorf tube and 100 µL nuclease-free water was added to the spin column.

The nuclease-free water and any excess ethanol from the column wash solution was removed by lyophilising the plasmid DNA. Once dry, the DNA was resuspended in 50 µL nuclease-free water. The purified plasmid DNA was quantified by UV 260 nm measurement and stored at -20°C.

### Plasmid DNA giga prep

The QIAGEN Giga QIAfilter Plasmid Purification Kit was used for the preparation of large quantities of plasmid DNA for use in the cell-free protein expression of 2dCD4-WT.

One Top10 colony from the LB agar plate was used to inoculate three 100 mL cell culture flasks filled with 10 mL LB and supplemented with 1 X ampicillin. The starter cultures were incubated for 8 hours at 37°C, 250 rpm. 5 mL starter culture was used per 500 mL LB in a 2 L baffled cell culture flask to a total volume of 2.5 L (5 X 500 mL), supplemented with 1 X ampicillin. The cultures were incubated at 37°C, 180 rpm overnight.

The cells were harvested by centrifugation at 7 000 rpm for 15 minutes at 4°C. The pellet was resuspended in 125 mL of buffer P1 containing 100 µg/mL RNase A and a 1:1 000 dilution of LyseBlue, in a 1 L Schott bottle. The mixture was vigorously shaken to mix. 125 mL of buffer P2 was added and mixed thoroughly by inverting the bottle six times. The mixture was incubated at room temperature for 5 minutes. 125 mL of buffer P3 was added at 4°C and mixed by inverting the bottle six times.

The lysate was poured into a QIAfilter Mega-Giga cartridge attached to a 1 L Schott bottle and incubated a room temperature for 10 minutes. The vacuum pump was used to filter the plasmid DNA from the lysate. The pump was switched off and 50 mL of buffer FWB2 was added to the QIAfilter cartridge and the precipitate was stirred gently with a sterile metal spatula and the vacuum pump was switched back on again until all the liquid had been drawn through the filter.

A QIAGEN-tip 1 000 was equilibrated with 75 mL buffer QBT and allowed to empty by gravity flow. The filtered lysate was applied to the tip. The tip was washed with 600 mL buffer QC and the flow through was discarded. The DNA was eluted with 100 mL of buffer QF. The DNA was precipitated from the solution by adding 70 mL of room temperature isopropanol, subsequent mixing and centrifugation at 12 000 rpm for 30 minutes at 4°C. The supernatant was carefully removed and the DNA pellet was washed with 10 mL of room temperature ethanol. The DNA was centrifuged at 12 000 rpm for 10 minutes at 4°C. The supernatant was discarded.

The pellet was air-dried under the fume hood and the DNA was dissolved in 10 mL of 10 mM Tris-HCl pH 8.5. The purified plasmid DNA was quantified by UV absorbance at 260 nm



measurement and stored at -20°C.

### 3.3.2 2dCD4 Expression in *Escherichia coli*

#### *E. coli* transformation

*E. coli* BL21\* (DE3) (Invitrogen) cells were thawed on ice. 10 ng pET15b-2dCD4-WT plasmid DNA (provided by Dr N. Cerutti at the HPRU, Johannesburg) was added to the cells and incubated for 30 minutes on ice. The cells were subject to a 30 second heat-shock at 42°C after which 250 µL room temperature SOC media was added to the cells before incubating for 1 hour at 37°C, 200 rpm. 50 µL was plated on 1 X ampicillin supplemented LB agar plate and incubated at 37°C overnight.

#### Expression of 2dCD4 in *E. coli*

25 mL LB, supplemented with 1 X ampicillin, was inoculated with one pET15b-2dCD4-WT transformed BL21\*(DE3) colony. The culture was incubated at 37°C, 250 rpm overnight. 1 L LB, supplemented with 1 X ampicillin, was inoculated with 20 mL of starter culture. The culture was incubated 37°C, 150 rpm until an optical density (OD) at 600 nm of 0.6 was reached and the culture was induced by addition of 0.5 mM IPTG and transferring to 20°C and cultured overnight at 150 rpm.

#### Purification of 2dCD4 expressed in *E. coli*

Cells were collected by centrifugation (Beckman Coulter JLA 9 000) at 5 000 rpm for 20 minutes at 4°C, with fast acceleration, slow deceleration. The supernatant was discarded and the pellet was resuspended in 25 mL 1 X PBS per 1 L cell culture. Lysozyme (hen egg white, Sigma) was added to a final concentration of 0.5 mg/mL and stirred at 4°C for 1 hour. The cell suspension was flash frozen in liquid nitrogen and stored at -80°C.

After three cycles of freeze thawing, the cell suspension was disrupted by sonication (Bioblock Scientific VibraCell 75115) at 80% amplitude for a 1 minute pulse: 1 second on, 5 seconds off with a 1 minute pause and stirring. 3-6 rounds of sonication were required depending on volume and clearing of lysate.

The lysate was centrifuged (Beckman Coulter rotor JA 20) at 16 000 rpm for 45 minutes at 4°C. Immobilised metal affinity chromatography (IMAC) resin was equilibrated in H<sub>2</sub>O and buffer and charged with 100 mM NiSO<sub>4</sub>. The supernatant was discarded and the recovered inclusion bodies were resuspended in wash buffer 1 and homogenised to a fine suspension (Omni Tissue Master 125 homogeniser). The homogenised solution was centrifuged at 16 000 rpm for 45 minutes at 4°C with fast acceleration and slow deceleration. The supernatant was incubated with the Ni IMAC resin, being careful to avoid adding genomic DNA to the resin and incubated with stirring overnight at 4°C.

The resin was collected by centrifugation at 3 000 rpm for 3 minutes and discarding supernatant. The resin was washed five times with 25 mL wash buffer 1 each time (total wash volume of 125 mL). The resin was then washed five times with 25 mL (total wash volume of 125 mL) wash buffer 2. The resin was resuspended in 10 mL elution buffer, the column (Bio-Rad medium pressure 50 mL bed volume) was packed and incubated for 10 minutes. The CD4 eluant was collected and the elution repeated with 5 mL elution buffer.

### **2dCD4-WT expressed in *E. coli* refolding**

The CD4 containing eluent was dialysed against 2 L chilled refolding buffer A overnight. The resulting CD4 was dialysed against 2 L chilled refolding buffer B overnight. The CD4 elution was dialysed against SEC buffer for 2 hours. The dialysis against SEC buffer was repeated three times. The CD4 was concentrated using a 10 kDa molecular weight cut-off (MWCO) Amicon®Ultra-15 centrifugal filter (Merck). The refolded CD4 was separated into 100 µL aliquots, flash frozen in liquid nitrogen and stored at -80°C.

### **Size-exclusion chromatography purification of refolded 2dCD4-WT**

Before use, 2dCD4-WT aliquots were defrosted, pooled and filtered using Spin-X 0.2 µm centrifugal filters (Corning) by centrifuging at 14 000 rpm for 1 minute. Filtered 2dCD4-WT was loaded onto a HiLoad 16/600 Superdex™ 75 preparative grade (S75 pg) column (GE Healthcare) equilibrated in SEC buffer. CD4 was eluted at 1 mL/min into 1 mL fractions. CD4 containing fractions were identified by SDS-PAGE with Coomassie blue staining.

### 3.3.3 Deuteration of 2dCD4 in *E. coli*

#### Adaptation of *E. coli* to deuterated growth conditions

pETm11-2dCD4-WT plasmid DNA (provided by Dr N. Cerutti at the HPRU, Johannesburg) was transformed into BL21\* (DE3) as described in section 3.3.2. One colony was picked to inoculate 2 X 10 mL LB supplemented with 1 X kanamycin. One culture was induced with 0.5 mM IPTG at an OD<sub>600</sub> of 0.6 and cultured at 20°C, 180 rpm overnight. Before the culture was induced a 500 µL sample was taken for analysis by anti-his western blotting. While the other was cultured overnight at 37°C, 180 rpm overnight. After overnight culture the OD<sub>600</sub> was measured and a 500 µL sample was collected for analysis by anti-his western blotting from the induced culture.

When the cells reached the late exponential growth phase, 2 X 1 mL of the uninduced LB culture was used to inoculate 2 X 10 mL H-Enfors medium, supplemented with 1 X kanamycin and the process was repeated. 2 X 1 mL of the uninduced H-Enfors culture was used to inoculate 2 X 10 mL 85% D-Enfors medium, supplemented with 1 X kanamycin and the process was repeated.

1 mL of the uninduced 85% D-Enfors culture was used to inoculate 1 X 10 mL 85% D-Enfors medium, supplemented with 1 X kanamycin and cultured at 37°C, 180 rpm overnight. This was repeated 5 times for a total of seven passages into 85% D-Enfors medium.

#### Fed-batch fermentation to produce a high cell density culture

Fed-batch fermentation was carried out as described by Haertlein *et al.* 2016 [171].

The 10 mL of passage 7 was added to 150 mL 85% D-Enfors in a 500 mL plastic culture flask (Corning) and supplemented with 1 X kanamycin. The starter culture was incubated overnight at 37°C, 180 rpm. The starter culture was used to inoculate 1.2 L 85% D-Enfors in a Labors 2.3 L fermenter which was computer-controlled using IRIS software (both Infors). The starting OD<sub>600</sub> was 0.18. After 24 hours the initial glycerol supply was depleted at the end of the initial batch phase (OD<sub>600</sub> of 3.9). The culture was thus supplied with fresh feed containing 12% deuterated glycerol in the fed-batch phase. The pD was controlled by addition of NaOD to maintain a constant pD of 6.4. The potential oxygen was maintained at 30% by controlling the stirring of the culture.

At an OD<sub>600</sub> of 15.8 (after 48 hours) the culture was induced by addition of 0.5 mM IPTG and the temperature was reduced to 30°C. The culture was stopped after 66 hours at a final OD<sub>600</sub> of 17.8. 74 g of cell paste was collected by centrifugation at 6 000 rpm.

The total expression was tested for the LB culture before and after induction, the H and D-Enfors culture after induction and the fermenter before and after induction with IPTG using anti-His western blotting with colorimetric (nanogram sensitivity) and electrochemiluminescence (ECL) detection (femtogram sensitivity). The colorimetric detection used Western Blue<sup>®</sup> substrate for alkaline phosphatase (Promega). Whereas the ECL detection used Clarity<sup>™</sup> Western ECL blotting substrate (BioRad).

### 3.3.4 2dCD4 Expression in *Brevibacillus choshinensis*

Glycerol stocks of *B. choshinensis* transformed with the pNCM02-2dCD4-WT clone were prepared according to the manufacturer's instructions (TAKARA). Expression was tested by anti-his western blotting and SDS-PAGE with Coomassie blue staining analysis before storage at -80°C.

1 L 2SY media supplemented with 50 µg/mL of neomycin was inoculated with one vial of *B. choshinensis* transformed with pCN-His-2dCD4-WT glycerol stock (1.8 mL). Subsequent, expression was carried out at 30°C, 120 rpm over 4 days.

#### **Immobilised metal affinity chromatography (IMAC) purification of 2dCD4-WT expressed in *B. choshinensis***

After 4 days expression the supernatant was collected by centrifugation at 8 000 rpm for 30 minutes. The supernatant was filtered through a 0.8 µm filter. Post-expression one of two methods were used to concentrate the protein by reducing the volume of the expression medium, prior to purification by nickel affinity chromatography: ammonium sulphate precipitation and viva-flow concentration as outlined hereafter.

**Ammonium sulphate precipitation of 2dCD4 expressed in *B. choshinensis*** Salting out by ammonium precipitation at 55% saturation was carried out at 4°C to precipitate the protein out of solution. The precipitate was collected by centrifuging the solution at 10 000 rpm, 4°C. The

supernatant was carefully discarded and the pellet was resuspended in loading buffer. The suspension was then dialysed against this same buffer.

**Vivaflow® concentration of 2dCD4-WT containing *B. choshinensis* expression media** At 8°C the protein containing expression media supernatant (cells removed by centrifugation) was loaded onto a Vivaflow®200 cassette (Sartorius) and passed through the cassette until a volume of approximately 100 mL was reached. The concentrated protein was dialysed against 2 X 2 L loading buffer, overnight.

**Nickel IMAC** The dialysed suspension was diluted 1:7 in loading buffer and loaded at 1 mL/min onto 5 mL HisTrap NiNTA column (GE Healthcare), equilibrated with loading buffer. After loading, the column was washed with 10 column volumes (CV) wash buffer 1, followed by 10 CV wash buffer 2. The protein was eluted in 2 mL fractions by a 100 mL 0-100% gradient of elution buffer, followed by a further 5 CV of elution buffer. CD4 containing fractions were identified by SDS-PAGE with Coomassie blue staining

#### **Size Exclusion Chromatography**

The protein containing fractions were pooled and dialysed into SEC buffer and concentrated to load by a 5 mL loop into a S75 pg column equilibrated in SEC buffer. The protein was eluted at 1 mL/min in 1 mL fractions. CD4 containing fractions were identified by SDS-PAGE with Coomassie blue staining.

### **3.3.5 Deuteration of 2dCD4 in *B. choshinensis***

#### **Expression tests of 2dCD4-WT in *B. choshinensis* in deuterated media**

90 µL glycerol stock of *B. choshinensis* transformed with the pCNMO2-2dCD4-WT-His plasmid was used to inoculate 50 mL of the following expression media in a 250 mL plastic cell culture flask (Corning), supplemented with 10 X neomycin antibiotic:

- 2SY in 0% D<sub>2</sub>O (100% H<sub>2</sub>O)
- 2SY in 75% D<sub>2</sub>O

- 2SY in 100% D<sub>2</sub>O
- 0% D-Silantes (100% H-Silantes)
- 0% D-Enfors (100% H-Enfors)
- 75% D-Silantes

The cultures were incubated at 20°C, 150 rpm for 72 hours with samples being taken at various time points (4, 7, 24, 28, 31, 49, 53 and 72 hours) to assess the progress of 2dCD4-WT expression by anti-His western blotting and to follow the *B. choshinensis* growth curve by measuring the OD<sub>600</sub>.

The expression test was repeated using: 2SY in 100% H<sub>2</sub>O, 0% D-Silantes, 65% D-Silantes and 75% D-Silantes with anti-his western blotting analysis.

#### **Purification tests of deuterated 2dCD4-WT**

Supernatant containing 2dCD4-WT was dialysed against loading buffer. 10 µL magnetic beads (Promega MagneHis Ni particles) suspension (supernatant removed) was incubated with 1 mL binding buffer at room temperature for 5 minutes with mixing. The binding buffer was removed from the magnetic beads and 1 mL 2dCD4-WT containing supernatant was added to the magnetic beads for 30 minutes at room temperature with mixing. The supernatant was removed and a sample was kept for analysis by western blot and SDS-PAGE.

500 µL wash buffer 1 was added to the magnetic beads and mixed gently. The supernatant was removed, keeping a sample for SDS-PAGE and western blot analysis. The wash was repeated three times for wash buffer 1. 500 µL wash buffer 2 was added and to the magnetic beads and the washing procedure was repeated as for wash buffer 1 (three times in total). The magnetic beads were incubated for 2 minutes with 100 µL elution buffer and gentle mixing. The supernatant was removed and a sample kept for analysis by western blot and SDS-PAGE. The elution process was repeated. The supernatant from each step was analysed by SDS-PAGE with Coomassie blue staining and anti-His western blotting for the presence of 2dCD4-WT.

### 3.3.6 Cell-free protein expression of 2dCD4

#### Plasmid DNA preparation

DNA coding for wild-type two domain CD4 (domains 1 and 2) from *Homo sapiens*, a truncated form of UniProtKB entry P01730 (full-length CD4), was synthesised and cloned into the pIVEX2.3d expression vector between Nde1/Xho1 restriction sites by Proteogenix. Two constructs were synthesised: one with an N-terminal TEV cleavable 6 X His tag and enhancer sequence and the other with a C-terminal TEV cleavable 6 X His tag. Following expression tests (see results), further experiments were carried out using the C-term His tag clone so as to avoid any possible steric hindrance between gp120 and the His tag at the N-terminus.

#### Expression of 2dCD4-WT using CFPE

Analytical scale batch cell-free reactions were set up as per the protocol described by Kim *et al.* 2006 [185] to assess the level of expression of soluble and insoluble 2dCD4-WT based on varying expression conditions:

- Overnight expression: 20°C and 30°C
- 6 X His tag location: N-terminal and C-terminal
- Reaction volume: 50 µL, 1 mL and 10 mL
- Filter pore size: 0.20 µm and 0.45 µm
- Mg<sup>2+</sup> concentration: 5 mM and 10 mM
- Deuterated amino acid concentration: 1, 2, 5, 10 and 20 mg/mL

Soluble and insoluble portions of the reactions were separated by centrifugation in a bench-top centrifuge at 14 000 rpm for 30 minutes. The presence of 2dCD4-WT was determined by western blotting and, in some cases, SDS-PAGE.

Cell lysate was prepared from the *E. coli* SHuffle strain by Vinesh Jugnarain at Synthelisis. The cell-free reaction was set up using the conditions determined from the results (see 3.4.4) and was supplemented with 0.2% sodium azide to prevent bacterial contamination. The cell-free

reaction was then filtered into the expression vessel before incubating for at least 12 hours at 40 rpm.

### **Purification of 2dCD4-WT produced using CFPE**

After incubation overnight, the cell-free reactions were pooled and centrifuged at 5 000 rpm for 20 minutes. The supernatant was then diluted 1:5 with loading buffer and filtered through a 0.22  $\mu\text{m}$  filter to remove any remaining debris prior to purification.

The diluted reaction was loaded at 1 mL/min onto a 5 mL HisTrap NiNTA column (GE Healthcare), equilibrated in loading buffer. The column was washed with 10 CV wash buffer 1, followed by 10 CV wash buffer 2. The protein was then eluted in 1 mL fractions with 30 mL of elution buffer.

Prior to SEC the CD4 containing fractions from the NiNTA purification were pooled and loaded via a 5 mL loop, 4 mL at a time, onto a S75 pg column equilibrated in SEC buffer. The protein was eluted at 1 mL/min in 1 mL fractions.

### **Deuteration of 2dCD4-WT using CFPE**

The deuteration protocol for 2dCD4-WT in the cell free protein expression system was identical to that for production of the protiated 2dCD4-WT with the exception for the addition of 2 mg/mL deuterated amino acids during the cell free reaction set-up.

### **3.3.7 Expression and purification of the 2dCD4 domain 2 disulphide bond knock-out variant (2dCD4-D2A)**

DNA coding for 2dCD4 whose Cys130 and Cys159 codons were replaced by ones coding for Ala was synthesised and cloned into the pIVEX2.3d expression vector between Nde1/Xho1 restriction sites by Proteogenix. The sequence had a C-terminal TEV cleavable 6 X His tag.

A 50 mL CFPE reaction was set-up for expression of 2dCD4-D2A and purified by nickel IMAC and SEC as described for 2dCD4-WT section 3.3.6. 2dCD4-D2A presence was detected by SDS-PAGE with Coomassie blue staining and anti-His western blotting.



### 2dCD4-D2A expression tests

50  $\mu$ L 2dCD4-D2A CFPE reactions were set-up using the following conditions:

- Temperature: 16°C and 20°C
- Agitation at 20°C: 0 rpm and 40 rpm
- n-Dodecyl  $\beta$ -D-maltoside: 0.25%, 0.5% and 1.0%
- Brij-35: 0.25%, 0.5% and 1.0%

The total and soluble expression was analysed using anti-His western blotting.

### 2dCD4-D2A purification test

Purification testing of 2dCD4-D2A was carried out using the same method as described for purification testing of d-2dCD4-WT expressed in *B. choshinensis* (section 3.3.5). The purification buffers were the same as those used for 2dCD4-WT produced using CFPE.

## 3.4 Results

### 3.4.1 Expression of 2dCD4-WT in *E. coli*

Due to the disulphide bonds present in domains 1 and 2 of 2dCD4-WT, it is expressed by *E. coli* into insoluble inclusion bodies. Soluble protein can be recovered from the inclusion bodies by collecting them by centrifugation and denaturing them using 8 M urea. The recombinant protein can be purified in its denatured state because the His-tag, used for the affinity purification of 2dCD4-WT, is a linear tag and does not depend on the tertiary structure of the protein. After affinity purification, the purified protein can then be refolded (with significant losses in yield) using a refolding protocol which involves slow dialysis and hence dilution of the urea, under specific conditions. These include the presence of a chaotrope such as urea which is used to denature the inclusion bodies and a redox system such as reduced and oxidised glutathion (GSH/GSSG) to ensure correct formation of disulphide bonds.

### Purification of 2dCD4-WT from inclusion bodies

2dCD4-WT was purified in its denatured form by the addition of 8 M urea using a manual (no fast-protein liquid chromatography machine) nickel immobilised metal affinity chromatography (IMAC) purification. Figure 3.1 shows the steps of 2dCD4-WT IMAC purification analysed by SDS-PAGE with Coomassie blue staining. The 2dCD4-WT elution samples appear quite pure after just one round of IMAC purification. Most of the contaminating proteins did not bind to the column as many contaminants can be seen in the flow through and wash samples, but not in the 2dCD4-WT elution.

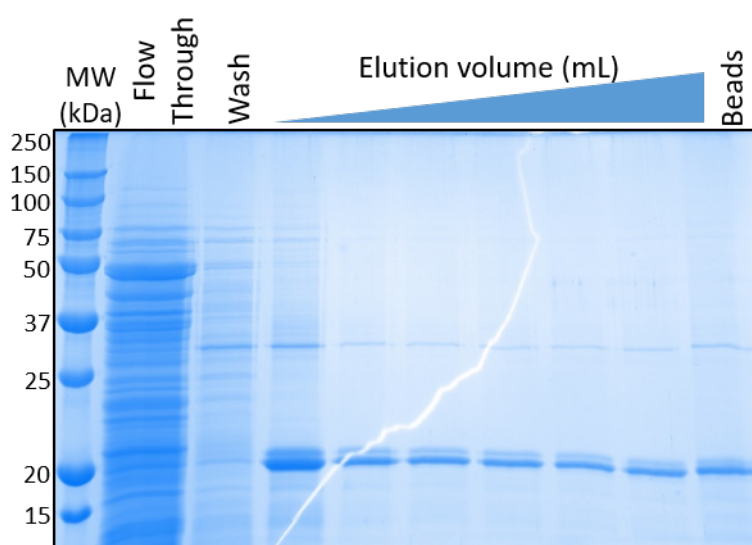
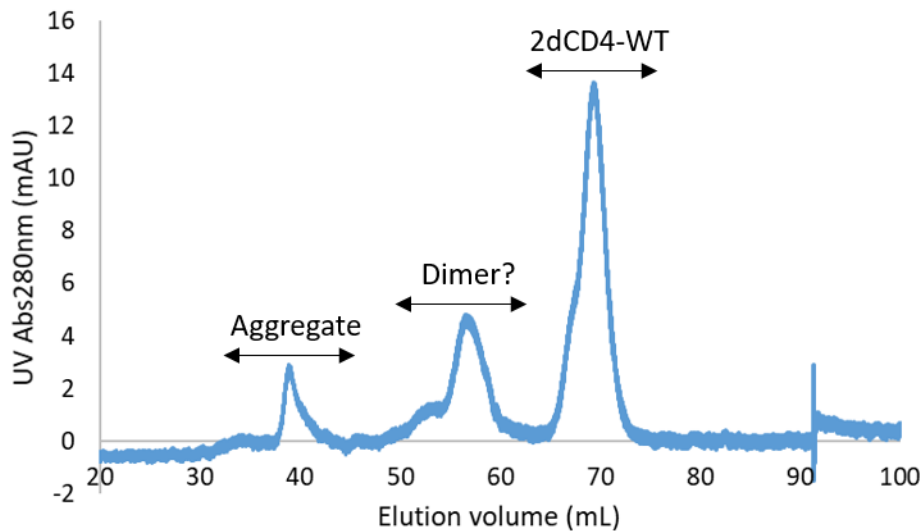


FIGURE 3.1: 2dCD4-WT was purified from inclusion bodies expressed by *E. coli* BL21\* (DE3) by denaturing them in 8 M urea before carrying out nickel immobilised-metal affinity chromatography. Samples were taken from the nickel column flow through, the wash and the elution fractions. No 2dCD4-WT appears to elute in the flow through or wash which suggests that the protein is bound to the nickel column. The elution fractions contain few contaminants which shows that the wash step was effective in removing many contaminants.

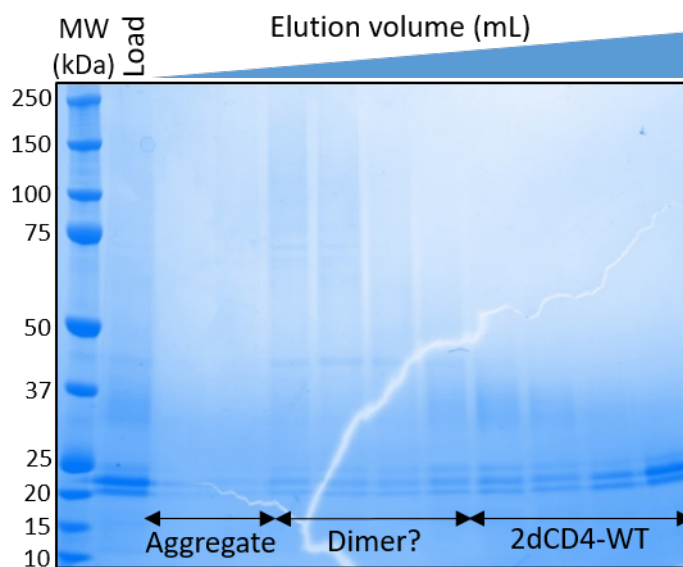
After IMAC purification the protein was refolded by slow dialysis against a lower concentration of urea (4 M) with GSH/GSSG, a buffer with GSH/GSSG and no urea and then against 1 X PBS three times. During the refolding process a large amount of white precipitate was formed. As such the precipitate was separated from the soluble protein by centrifugation, followed by filtration. The protein was then purified by SEC to remove any aggregates. The SEC chromatogram shown in figure 3.2a presents three peaks of which the first elutes just after the dead volume of the S75 pg column (approximately 40 mL). There is a second small peak followed

by a large peak. The contents of the peaks were analysed by non-reducing SDS-PAGE with Coomassie blue staining (fig. 3.2b). 2dCD-WT is shown to be present in all three peaks.

The first peak elutes just after the dead volume of the column indicating that this species is large. Since the SDS-PAGE analysis shows that only 2dCD4-WT is present in this peak this suggests that the first peak is the result of a small amount of aggregated 2dCD4-WT. The second peak contains 2dCD4-WT and a small amount of a protein band which resolves at an apparent molecular weight approximately twice that of 2dCD4-WT. This band may correspond to a disulphide bond linked dimer, since the SDS-PAGE was run in non-reducing conditions. This peak may also correspond to incorrectly folded 2dCD4-WT. The final peak contains only 2dCD4-WT, so the fractions from this peak were pooled and concentrated for subsequent analyses. The three closely spaced bands on the SDS-PAGE correspond from the bottom band up to; the fully oxidised 2dCD4-WT (2dCD4-Ox); 2dCD4-WT reduced in domain 2 (2dCD4-R1) and fully reduced 2dCD4-WT (2dCD4-R2) as shown by Cerutti *et al.* 2014 [63] by disulphide bond knockout reducing and non-reducing SDS-PAGE mobility analysis.



(A)



(B)

FIGURE 3.2: After 2dCD4-WT was refolded by slow dialysis against a 4 M urea buffer followed by a 0 M urea buffer (both containing reduced and oxidised glutathione), it was further purified by size-exclusion chromatography (A). The elution fractions were analysed by SDS-PAGE with Coomassie blue staining (B). Three peaks corresponded to aggregated 2dCD4-WT, a potential 2dCD4-WT dimer and correctly folded 2dCD4-WT.

### Optimising 2dCD4-WT expression in *E. coli*

Several attempts were made to optimise the expression of 2dCD4-WT in *E. coli* as follows:

- *E. coli* expression strain: SHuffle, Arctic Express, Rosetta, Rosetta Gami, Origami B, SoluBL21 and BL21\* (DE3)

- Expression vector: pETm11, pET15b and pMK-T
- Expression temperature: 16°C, 20°C, 25°C, 30°C and 37°C
- IPTG concentration: 0.1, 0.2, 0.3, 0.4 and 0.5 mM.

However 2dCD4-WT was not detected by SDS-PAGE with Coomassie blue staining nor by anti-His western blotting. Eventually it was found that in addition to lysis of the cells by cell disruption and lysozyme treatment, that the protein needed to be solubilised in 8 M urea in order to enter the gel. The results of these tests have not been presented here because 2dCD4-WT was not detected.

#### 3.4.2 Deuteration of 2dCD4-WT using *E. coli*

For deuteration it was necessary to use a plasmid with a kanamycin resistance gene as opposed to an ampicillin resistance gene. The ampicillin metabolising protein, beta-lactamase, is located in the periplasmic space and is therefore at risk of being lost during deuteration which causes the cell membrane to become leaky. In this situation, selectivity of only cells containing the expression plasmid for 2dCD4 with the antibiotic resistance marker would be lost. The protein which inactivates kanamycin however, is located in the cytoplasm and is therefore not at risk of being lost whilst the bacterium is cultured in deuterated conditions. Therefore, while the pET15b expression vector with ampicillin resistance was used for expression of protiated 2dCD4-WT, the cDNA coding for 2dCD4-WT in the pETm11 expression vector was required for production of recombinant deuterated 2dCD4 for subsequent SANS experiments.

##### **Deuteration of 2dCD4-WT in *E. coli***

The BL21\* (DE3) strain transformed with the pETm11-2dCD4-WT clone was used for deuteration of 2dCD4-WT given the need for a kanamycin resistant clone and the inconclusive results from the expression testing using different clones and hosts. From a 10 mL LB starter culture the BL21\* (DE3) cells were passaged into protiated then 85% deuterated minimal media (H and D-Enfors). Seven passages were made to allow the bacteria to adapt to growth under deuterated conditions. The last passaged culture was used to inoculate a 150 mL 85% deuterated minimal media starter culture. Finally the starter culture was used to inoculate 1.2 L 85%

deuterated minimal media for cell culture by fed-batch fermentation over a three day period. Samples were taken to assess expression throughout the adaptation and fermentation process. Expression of 2dCD4-WT was assessed using anti-His western blotting (fig. 3.4.2) with colorimetric (3.3a) and electrochemiluminescence (3.3b).

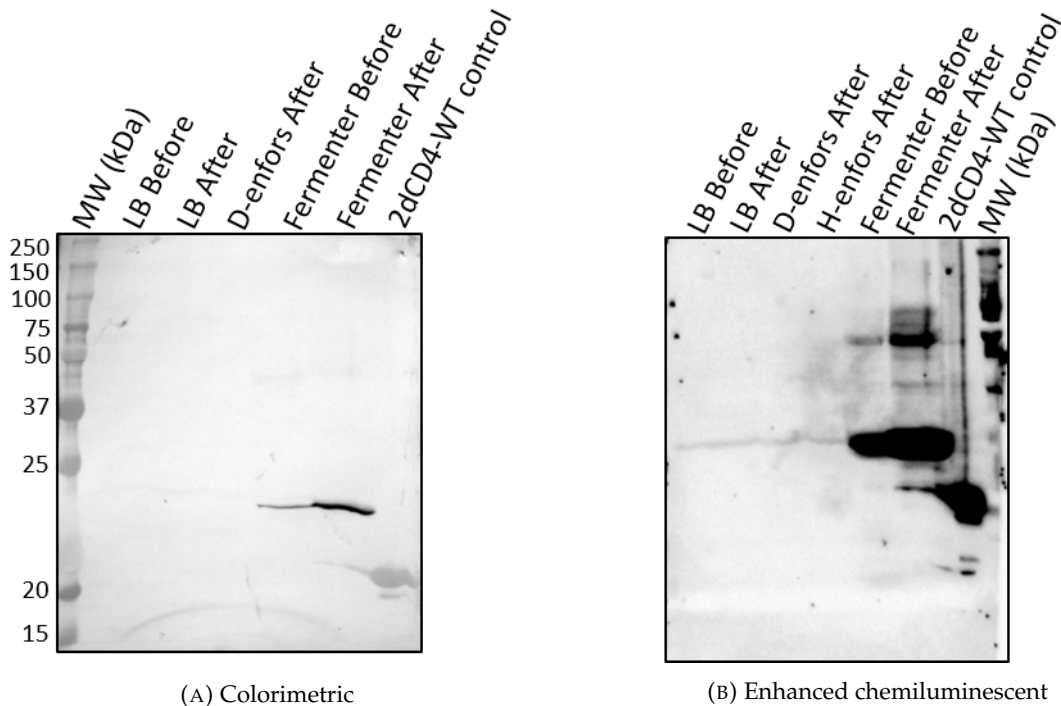


FIGURE 3.3: The expression of 2dCD4-WT was tested before and after fed-batch fermentation. (A) Colorimetric and (B) enhanced chemiluminescent detection was used to identify BL21\* (DE3) expression of 2dCD4-WT in LB, H-Enfors, 85% D-Enfors and the fermenter in 85% D-Enfors.

No expression is observed with colorimetric detection in LB and D-Enfors. There was a small amount of a expression prior to fermentation which increases slightly after fermentation. The expression bands shown resolve to a higher molecular weight than the 2dCD4-WT control because the control is protiated, whereas the before and after fermentation 2dCD4-WT detected is deuterated and is therefore of a higher molecular weight. Electrochemiluminescence detection was also used as it has an enhanced detection level compared to that of colorimetric assays (nanogram versus femtogram) with which there is a line spanning the LB before and after, the 85% D-Enfors and H-Enfors after induction tests. However this "band" could be noise since it resolves to the same apparent molecular weight as the deuterated protein after fermentation. The LB and H-Enfors bands should resolve to the same molecular mass as the control and there should be a difference in expression before and after induction. However, if the control

had suffered some degradation it may resolve at a molecular weight lower than that expected. From the western blot the yield of the deuterated 2dCD4-WT (d-2dCD4-WT) yield was estimated to be just 800 µg. This is an estimate before the refolding process during which much of the protein yield is lost due to aggregation. Therefore the quantity of d-2dCD4-WT was not sufficient for SANS studies which typically requires milligram quantities of d-2dCD4-WT. It was thus necessary to explore other methods to prepare deuterated 2dCD4-WT.

### 3.4.3 Expression and secretion of 2dCD4-WT from *B. choshinensis*

Since development of a deuteration protocol in *E. coli* was unsuccessful, expression of 2dCD4-WT in a novel gram positive expression host called *Brevibacillus choshinensis* was explored. *B. choshinensis* has benefits over *E. coli* as an expression host because the protein is secreted in a soluble form into the oxidising extracellular milieu, allowing the correct formation of disulphide bonds, in disulphide bond containing proteins. The protein can therefore be purified from the expression medium and there is no need to lyse the cells nor to carry out any refolding as was the case with *E. coli*.

#### Purification of protiated 2dCD4-WT from *B. choshinensis* secretion

Since *B. choshinensis* secretes the protein into the expression medium, the main hurdle lied in the processing of large volumes of supernatant (the cells are pelleted and removed from the supernatant by centrifugation) which was rich in contaminants. Large volumes of supernatant containing the protein of interested could be concentrated thereby reducing the volume of supernatant to be processed. Two avenues for supernatant concentration were explored: Vivaflow®cassette concentration and 55% ammonium sulphate  $(\text{NH}_4)_2\text{SO}_4$  precipitation (established previously). Vivaflow®cassettes function by using a membrane to retain the protein contents of the supernatant whilst filtering out some of the aqueous volume of the supernatant. Whereas ammonium sulphate precipitation functions by "salting out" of the protein in which the local ion concentration around the protein increases, thereby decreasing the solubility of the protein until a stable protein-salt precipitant is formed.

Figure 3.4a shows the presence of protein by SDS-PAGE through the steps of nickel IMAC purification, while figures 3.4b and 3.4c show the elution of the protein off of the nickel column for

the Vivaflow®(red) and ammonium sulphate (green) precipitation supernatant preparations, respectively.

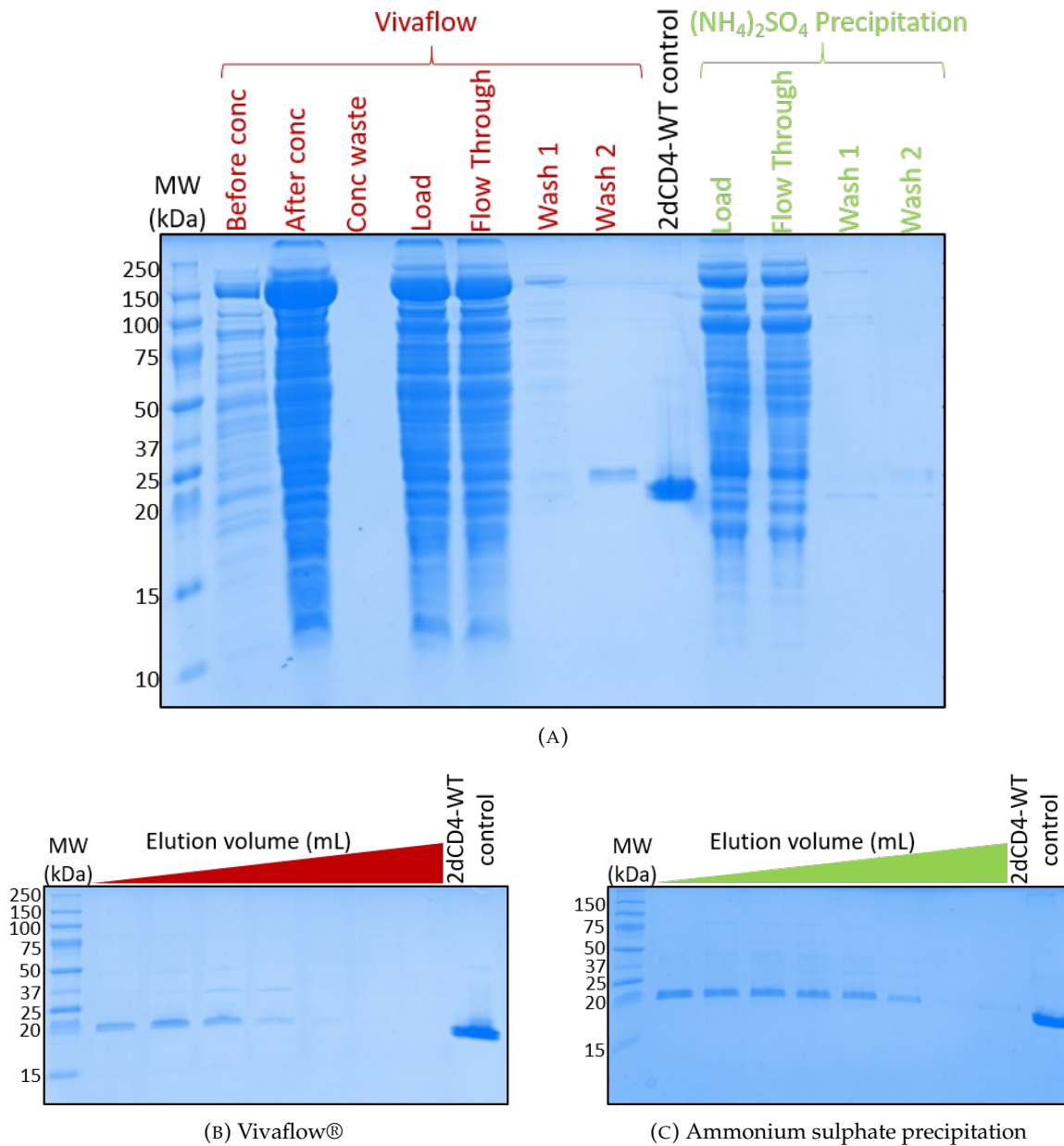


FIGURE 3.4: 2SY supernatant (cells pelleted by centrifugation) was concentrated by either a Vivaflow®cassette (red) or ammonium sulphate precipitation (green). (A) SDS-PAGE analysis of the nickel affinity purification of Vivaflow®cassette concentration and ammonium sulphate precipitation, elution fractions of nickel affinity purification from Vivaflow®concentration (B) and ammonium sulphate precipitation (C).

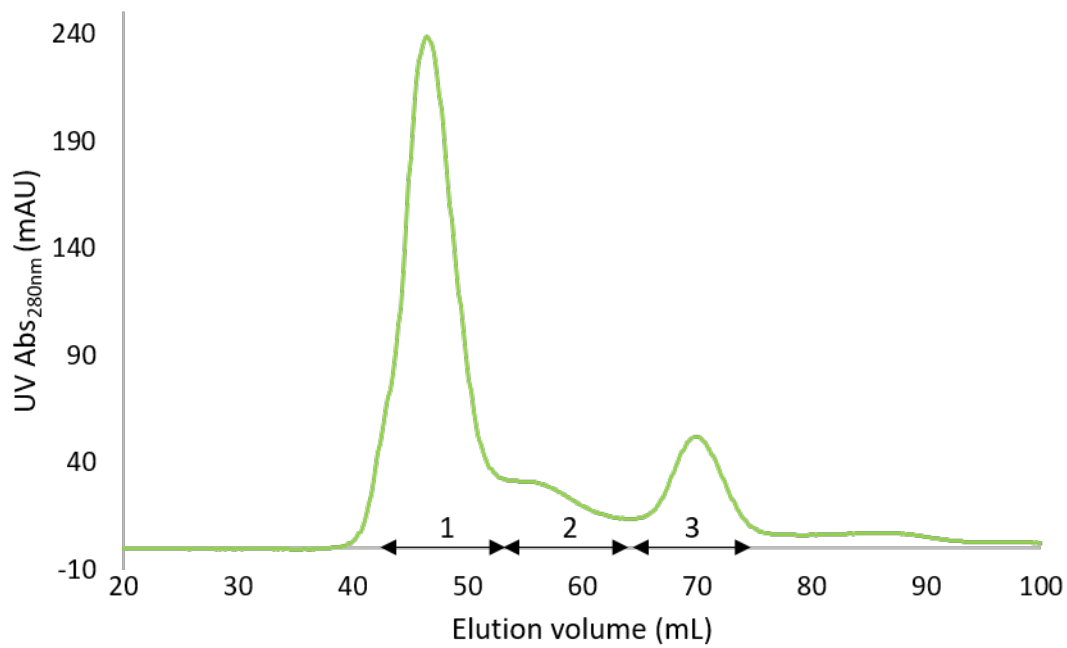
The before and after concentration lanes in figure 3.4a show clearly that the Vivaflow®cassette has concentrated the proteins present in the supernatant and that no protein came out in the



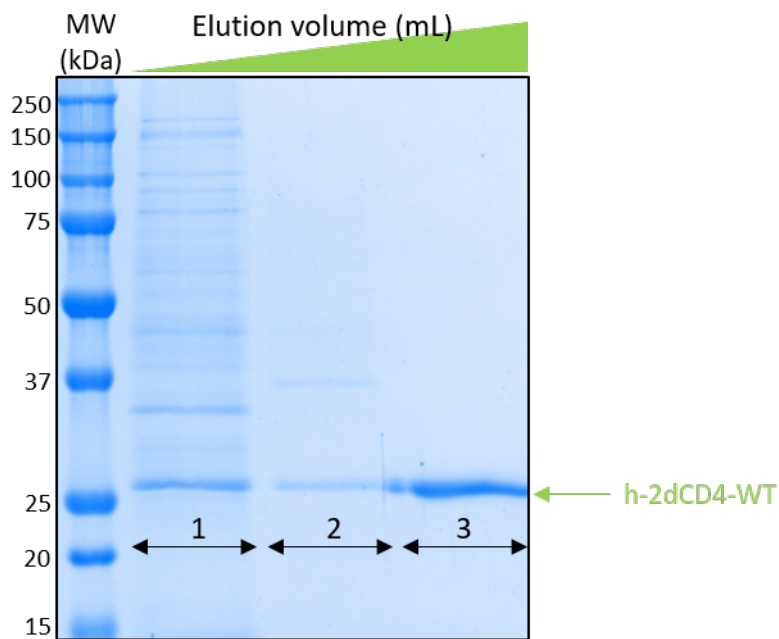
waste supernatant. Compared with the load from the ammonium sulphate precipitation preparation, the Vivaflow® load has more protein contaminants present. The use of 55% saturated ammonium sulphate precipitation allows some of the protein contaminants to be excluded as they are not salted out of solution with the protein of interest since they require higher concentrations of ammonium sulphate to precipitate out of solution. Ammonium sulphate precipitation can therefore act as an additional purification step prior to nickel affinity purification.

Indeed it appears that more 2dCD4-WT eluted off in wash 2 of the Vivaflow® supernatant preparation than in the ammonium sulphate precipitation preparation. This may be due to saturation of the nickel column with non-specific binding of protein contaminants. It also appears from the SDS-PAGE showing the elution fractions, that more 2dCD4 was recovered from the supernatant prepared by ammonium sulphate precipitation than by Vivaflow® concentration (fig. 3.4b and 3.4c, respectively). To this end, ammonium sulphate precipitation was considered the superior method for concentrating the supernatant and allowing maximum 2dCD4-WT recovery and was used in subsequent purifications.

After nickel IMAC the protein prepared by ammonium sulphate precipitation was further purified by SEC (fig. 3.5). The SEC chromatogram (fig. 3.5a) displays three peaks which were analysed by SDS-PAGE (fig. 3.5b). The first peak eluted from the column just after the dead volume which suggests that the contents of the peak are large. SDS-PAGE analysis of peak 1 shows that it contains many protein contaminants as well as 2dCD4-WT. This indicates that the 2dCD4-WT in this peak is aggregated. 2dCD4-WT can also be found in peak 2 with some higher apparent molecular weight contaminants. The third peak contains pure 2dCD4-WT.



(A)



(B)

FIGURE 3.5: 2dCD4-WT expressed and secreted by *B. choshinensis* was purified by SEC after nickel IMAC. (A) The SEC chromatogram shows three peaks whose contents were analysed by (B) SDS-PAGE. Peak 1 contains many protein contaminants as well as 2dCD4-WT. The elution position of peak 1 (after the dead volume of the column) suggests that this is formed of large aggregates. Peak 2 also contains 2dCD4-WT as well as higher molecular weight protein contaminants. Peak 3 contains pure 2dCD4-WT.

### Deuterated 2dCD4-WT expression tests in *B. choshinensis*

Sufficient yields of deuterated 2dCD4-WT using the classical *E. coli* expression system were not produced, so the novel *Brevibacillus choshinensis* expression system was used in an attempt to deuterate 2dCD4-WT. The expression of 2dCD4-WT was tested in several different media types with differing levels of deuteration. The growth of *B. choshinensis* was followed over 3 days by measuring the optical density at 600 nm (OD<sub>600</sub>) at various time points (fig. 3.6), at which point samples were taken to test for expression by western blotting using an anti-His primary antibody (fig. 3.7).

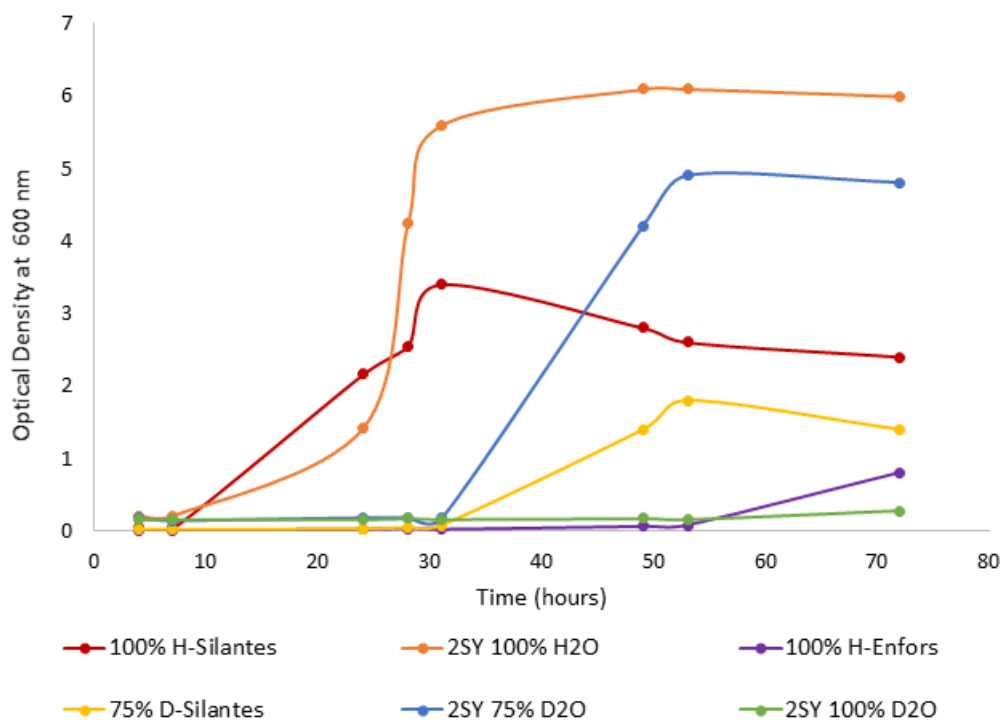


FIGURE 3.6: *B. choshinensis* transformed with the pNCM02-2dCD4-WT clone was cultured over a period of 72 hours in different media with varying levels of deuteration: 100% H-Silantes (red), 2SY in 100% D<sub>2</sub>O (orange), 100% H-Enfors (purple), 75% D-Silantes (yellow), 2SY in 75% D<sub>2</sub>O (blue) and 2SY in 100% D<sub>2</sub>O (green). Samples were taken periodically and their optical density at 600 nm was measured.

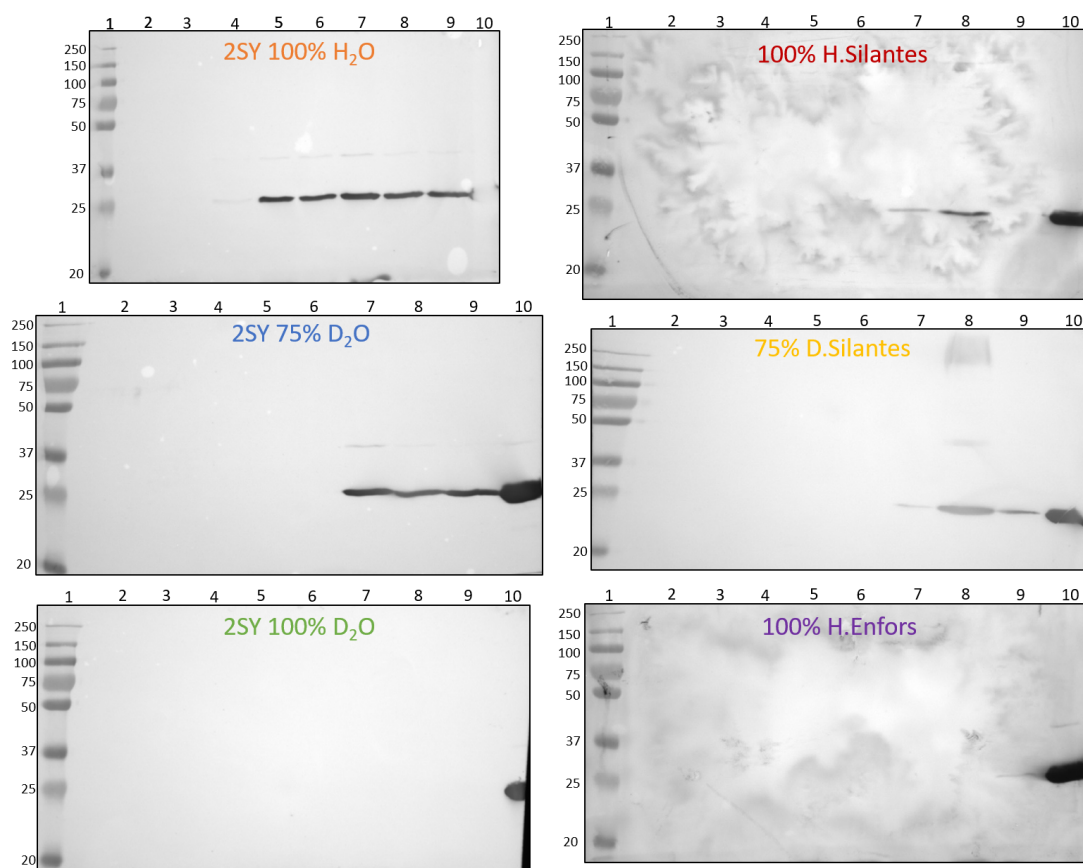


FIGURE 3.7: 2dCD4-WT was expressed in different media with differing levels of deuteration using the *B. choshinensis* expression system: orange = 2SY in 100% H<sub>2</sub>O, red = 100% H-Silantes, blue = 2SY in 75% D<sub>2</sub>O, yellow = 75% D-Silantes, green = 2SY in 100% D<sub>2</sub>O and purple = 100% H-Enfors. Lanes: 1 = Molecular weight (kDa), 2 = 4 h, 3 = 7 h, 4 = 24 h, 5 = 28 h, 6 = 31 h, 7 = 48 h, 8 = 53 h, 9 = 72 h, 10 = 2dCD4-WT control.

The western blot depicting the expression in 2SY media dissolved in 100% H<sub>2</sub>O (light water, orange) is the control western blot as the conditions used were those described by the manufacturer (TAKARA). Between the control western blot and the one in which 2dCD4-WT was expressed in 2SY dissolved in 75% D<sub>2</sub>O (blue) there does not seem to be much difference in the overall level of expression, however expression starts after 31 h for the 2SY 75% D<sub>2</sub>O whereas expression starts after 7 h for the control. For 2SY dissolved in 100% D<sub>2</sub>O (green) there is no expression at all. Relating this result to the growth curves in figure 3.6 it can be seen that in fact the cells did not survive culture in 2SY dissolved in 100% D<sub>2</sub>O.

The expression in 100% H-Silantes (red) is not as high as the 2SY 100% H<sub>2</sub>O and appears to drop off at 72 h. The growth curve shows that the growth begins to drop off after 31 h. While the cells grow to a higher optical density in 100% H-Silantes than in 75% D-Silantes, the expression

appears to be comparable and, interestingly, maybe even higher for the deuterated media. The expression appears to drop off again at 72 h which is reflected in the growth curve which declines between 53 and 72 h. The lag period for the *B. choshinensis* cultured in 100% H-enfors (purple) is long, lasting around 53 h. There is no expression detected in the western blot until 72 hours and is very faint. *B. choshinensis* cannot tolerate minimal media and is therefore is not a suitable expression host for fed-batch fermentation.

Disregarding the 100% D<sub>2</sub>O and 100% H-Enfors conditions for which there was little to no cell growth or 2dCD4-WT expression, it is apparent that *B. choshinensis* is able to reach higher cell densities with the 2SY media conditions and in the absence of D<sub>2</sub>O. The stationary phase is also achieved earlier in the absence of D<sub>2</sub>O. Furthermore, the cells appear to enter the death phase earlier using the Silantes media.

The 2SY media was protiated and just dissolved in water of differing H/D ratios. As such the level of deuteration possible using this approach is limited. Whilst the growth and expression were reduced in the Silantes media, the carbon source as well as the other components of this media are deuterated allowing a higher level of deuteration to be obtained. Silantes medium cannot be easily used for fed-batch fermentation as it is an enriched medium but it can be used for in-flask deuteration. Therefore, further testing was carried out on the D-Silantes condition.

#### **Deuterated 2dCD4-WT expression tests in *B. choshinensis* using D-Silantes**

As the expression of 2dCD4-WT was already low using 75% D-Silantes it was decided that 2dCD4-WT expression would also be tested in 65% D-Silantes. The expression of 2dCD4-WT in 65 and 75% D-Silantes was compared (fig. 3.8) and the expression in 0% D<sub>2</sub>O 2SY and 100% H-Silantes were used as non-deuterated media controls.

Little can be deduced from the SDS-PAGE (fig. 3.8a) as the expression bands of 2dCD4-WT are very faint. However the western blot (fig. 3.8b) paints a clearer picture. The optimal expression media is obviously the 0% D<sub>2</sub>O 2SY which was the control expression condition. After 1 day there is greater expression in 65% D-Silantes than in 75% D-Silantes. In fact the expression band is barely visible for the 75% D-Silantes. However, after 4 days of expression the expression in 75% D-Silantes is only slightly less than that in 65% D-Silantes. There is little difference in the

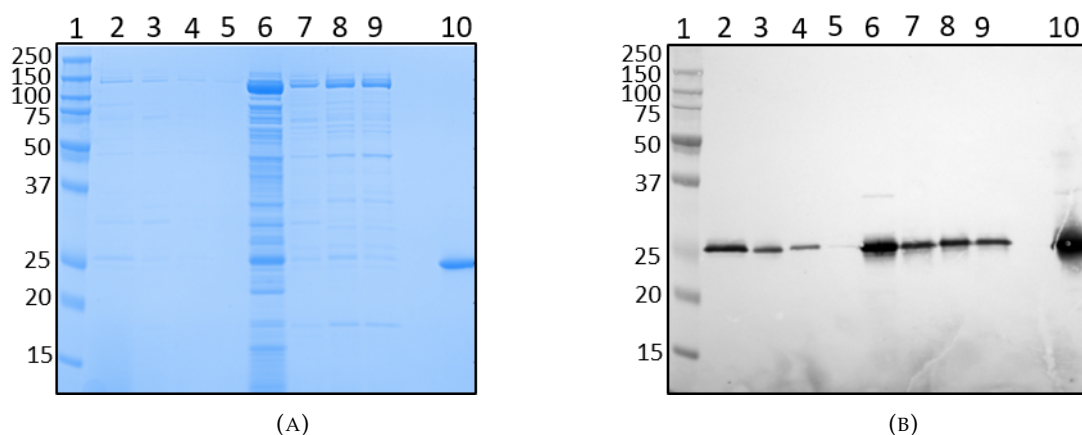


FIGURE 3.8: *B. choshinensis* was cultured in 2SY 0% D<sub>2</sub>O and Silantes media at 0, 65 and 75% deuteration (lanes 2 & 6, 3 & 7, 4 & 8 and 5 & 9, respectively). Expression was tested after 1 and 4 days (lanes 2-5 and 6-9, respectively) and tested by SDS-PAGE (A) and western blot (B). Lane 1 is the molecular weight marker in kDa and lane 10 is a 2dCD4-WT control.

levels of expression at 65 and 75% D-Silantes, therefore both were taken forward in analytical scale purifications.

#### Purification tests of 2dCD4-WT expressed in D-Silantes from *B. choshinensis*

Analytical purification tests were carried out using nickel conjugated magnetic beads (GE Healthcare) which allow microlitre scale purifications of protein with milligram binding capacity. After purification, the intention was to get a better idea for the yields possible from expression in 65 versus 75% D-Silantes. The pure protein could then be tested by mass spectrometry to determine the percentage deuteration possible using 65 versus 75% D-Silantes. 2SY 0% D<sub>2</sub>O (dissolved in 100% H<sub>2</sub>O, fig. 3.9a) was used first to test the analytical scale purification conditions and 0% D-Silantes (100% H-Silantes, fig. 3.9b) was used as a control for the 65 and 75% D-Silantes analytical scale purifications (figs. 3.9c and 3.9d, respectively). The top panels display Coomassie blue stained SDS-PAGE and the bottom panels show anti-His western blots.

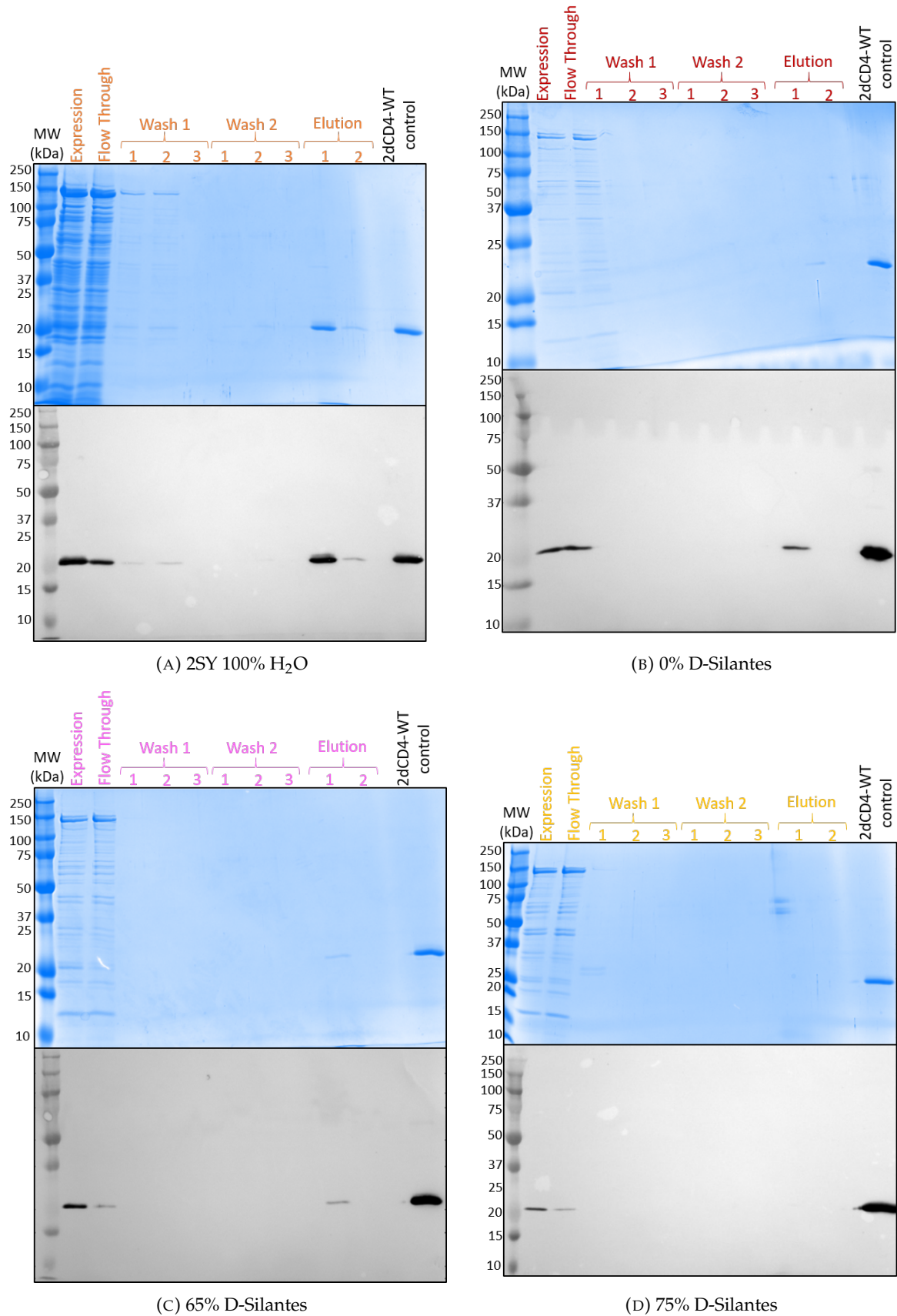


FIGURE 3.9: 2dCD4-WT was expressed by *B. choshinensis* cultured in (A) 2SY dissolved in 100% H<sub>2</sub>O, (B) 0% D-Silantes, (C) 65% D-Silantes and (D) 75% D-Silantes. Analytical scale purifications were carried out using his-conjugated magnetic beads. As the deuteration level increases the presence of 2dCD4-WT in the elution fractions diminishes.

From both the SDS-PAGE and western blots the presence of 2dCD4-WT in the elution is observably weaker in 0% D-Silantes medium, compared to 2SY in 100% H<sub>2</sub>O and then as the deuteration percentage increases less and less 2dCD4-WT can be seen in the elution. Insufficient 2dCD4-WT was present in the elution fraction for analysis by mass-spectrometry. Further deuteration tests in *B. choshinensis* were abandoned due to time constraints.

#### 3.4.4 Cell-free protein expression and purification of 2dCD4-WT

The final expression system used for the production of protiated and deuterated 2dCD4-WT was the cell-free protein expression system. CFPE allows decoupling of cell growth from protein expression and direct access to the chemical environment of recombinant protein synthesis. CFPE of protiated and deuterated 2dCD4-WT was carried out in close collaboration with Vinesh Jugnarain from Synthelis, France. Expression conditions were discussed collaboratively but initial analytical scale cell-free reactions were set-up solely by Vinesh Jugnarain. Analysis by western blot or SDS-PAGE was carried out by the author of this thesis. The cell paste for cell-lysate preparation was either prepared by Vinesh Jugnarain or the author of this thesis but cell-paste post-processing to produce the cell-lysate was solely prepared by Vinesh Jugnarain. Large-scale cell-free protein expression reactions were set-up either by Vinesh Jugnarain or the author of this thesis. The precise protocols are not described because Synthelis is the owner of these protocols and a non-disclosure agreement is in place between Synthelis and the Life Sciences Group of the ILL.

##### Expression of protiated 2dCD4-WT

The first expression test carried out using CFPE was to determine the optimal temperature for soluble 2dCD4-WT expression. Two reactions for an N-terminal tagged and C-terminal tagged clone were set-up with one to be incubated at 20°C, and the other to be incubated at 30°C overnight and the total versus soluble protein expression was assessed by anti-His western blotting (fig. 3.10). At 20°C it can be seen that both the total and soluble 2dCD4-WT expression is higher than that at 30°C for both clones. Further expression was therefore carried out at 20°C.



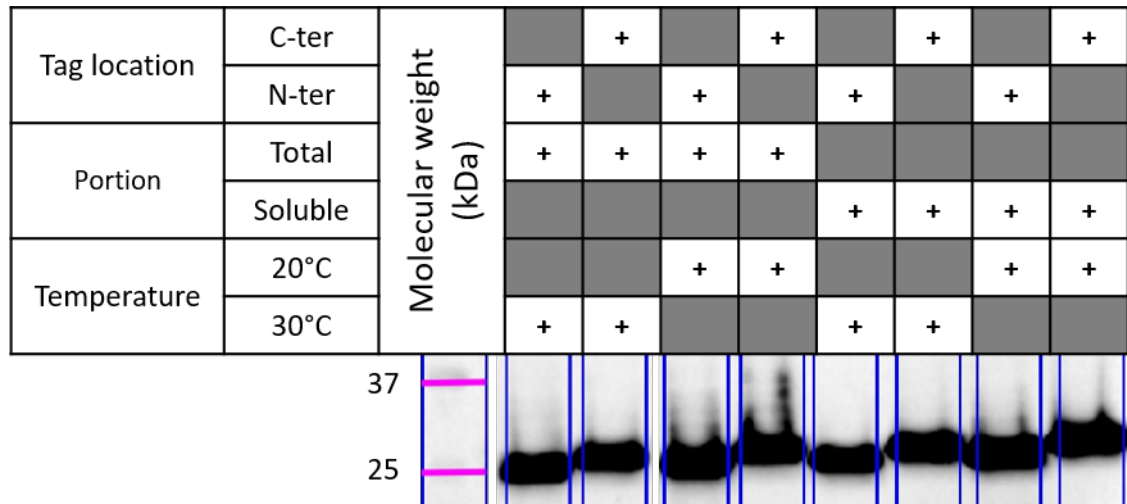


FIGURE 3.10: The total and soluble expression was tested using western blot for the N-terminally and C-terminally TEV-His-Tagged 2dCD4-WT at 20°C and 30°C to determine the optimal temperature giving the highest yield of soluble protein for the cell-free reaction. Expression was higher at 20°C for both clones.

Furthermore, since gp120 binds to the N-terminus of 2dCD4-WT which could result in steric hindrance caused by the tagging the protein here, the expression levels of the two clones were compared. From the western blot in figure 3.11, it is apparent that the expression of the C-terminal tagged clone is better than the N-terminal tagged clone. In addition, it seems that there are 2dCD4-WT degradation products formed with the N-term tagged clone, as indicated by the lower molecular weight bands. All further reactions were therefore carried out using the C-terminal tagged 2dCD4-WT clone.

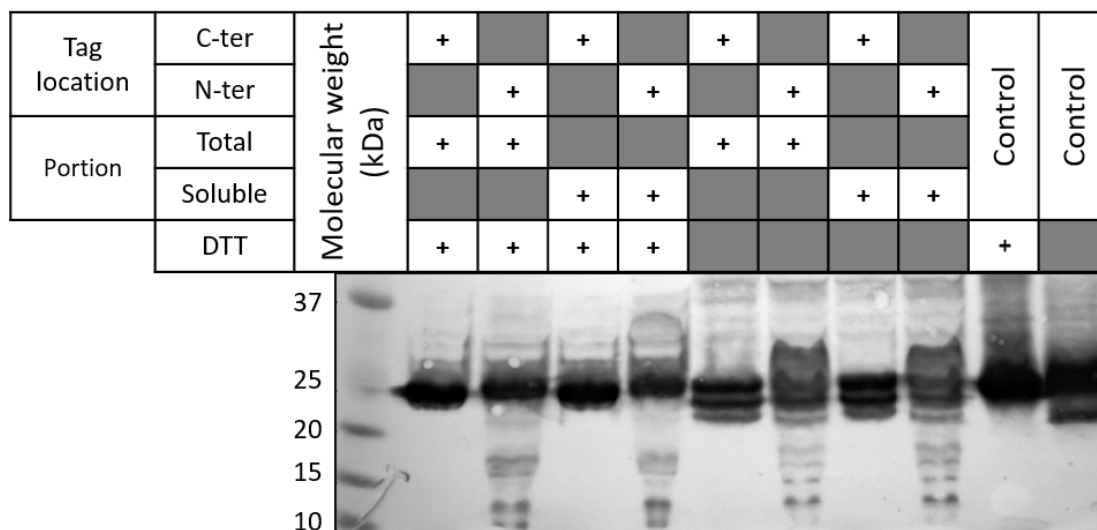


FIGURE 3.11: The total and soluble expression of N- and C-terminally TEV-His-tagged 2dCD4-WT was tested using Western blot under reducing and non-reducing conditions to see whether the location of the His-tag affected the expression and if the triple banding pattern corresponding to the three redox states of 2dCD4-WT were present; 2dCD4-R2 (fully reduced, 28 kDa), 2dCD4-R1 (partially reduced, 25 kDa) and 2dCD4-OX (fully oxidised, 22 kDa). The 2dCD4-WT clone tagged at the C-terminus expresses better. In addition, there are evident degradation products in the N-terminally tagged sample.

Finally, the effect of scaling up and of the filter pore size of the filtration step prior to incubation was tested on the expression of the 2dCD4-WT. Interestingly, the expression of 2dCD4-WT appears to diminish in the 1 mL reaction volume but remains similar for the 50  $\mu$ L and 10 mL reaction volumes (fig. 3.12). This could arise from the ratio of surface to air exposure since the expression vessels were different for each reaction. The 50  $\mu$ L reactions were carried out in a microtube (Eppendorf), whereas the 1 mL reactions were carried out in a 50 mL Falcon tube and the 10 mL reactions were carried out in a 100 mL conical flask (Corning). This result confirms that scaling up is possible without any significant effect on protein expression. In addition, there does not appear to be any difference in the levels of expression using different filter sizes prior to leaving the reaction to incubate overnight. For the ease of scaling up, the 0.20  $\mu$ m filter size was chosen because vacuum 0.20  $\mu$ m filters could be attached to sterile Schott bottles for ease of filtration.

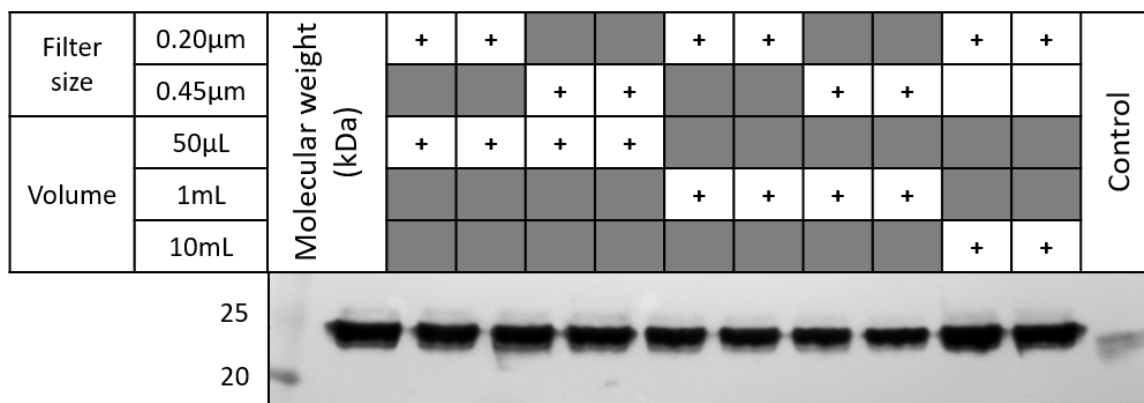


FIGURE 3.12: Western blot was used to test the effect of reaction volume and filter size through which the reaction was filtered before incubation on expression of C-terminally TEV-His-tagged 2dCD4-WT. Three reaction volumes were tested: 50 µL, 1 mL and 10 mL and two filter sizes: 0.20 µm and 0.45 µm. Expression appears similar in 50 µL and 10 mL but weaker in 1 mL. Filter size does not seem to have a significant effect.

### Expression of deuterated 2dCD4-WT

Having established optimal conditions for the expression of soluble protiated 2dCD4-WT, the conditions for deuteration of 2dCD4-WT using CFPE were investigated.  $Mg^{2+}$  ion concentration and the concentration of deuterated amino acids were compared to determine the best conditions for soluble deuterated 2dCD4-WT expression. Figure 3.13 displays the western blot results for the amino acid concentration screen using 5 and 10 mM  $Mg^{2+}$  (top and bottom western blot), respectively. In general, the expression seems relatively constant across the amino acid concentrations with a slight increase in overall protein expression between 1 mg/mL and 2 mg/mL deuterated amino acids. The 5 mM  $Mg^{2+}$  expression test appears to display a very slight increase in the ratio of soluble to total expression suggesting that the use of 5 mM  $Mg^{2+}$  in the reaction mixture either increases the production of soluble protein or reduces the production of insoluble protein.

Therefore, the conditions selected for deuterated protein expression were 5 mM  $Mg^{2+}$  using 2 mg/mL deuterated amino acids to improve the cost efficiency of the reaction, since the expression was similar across the deuterated amino acid concentration screen.

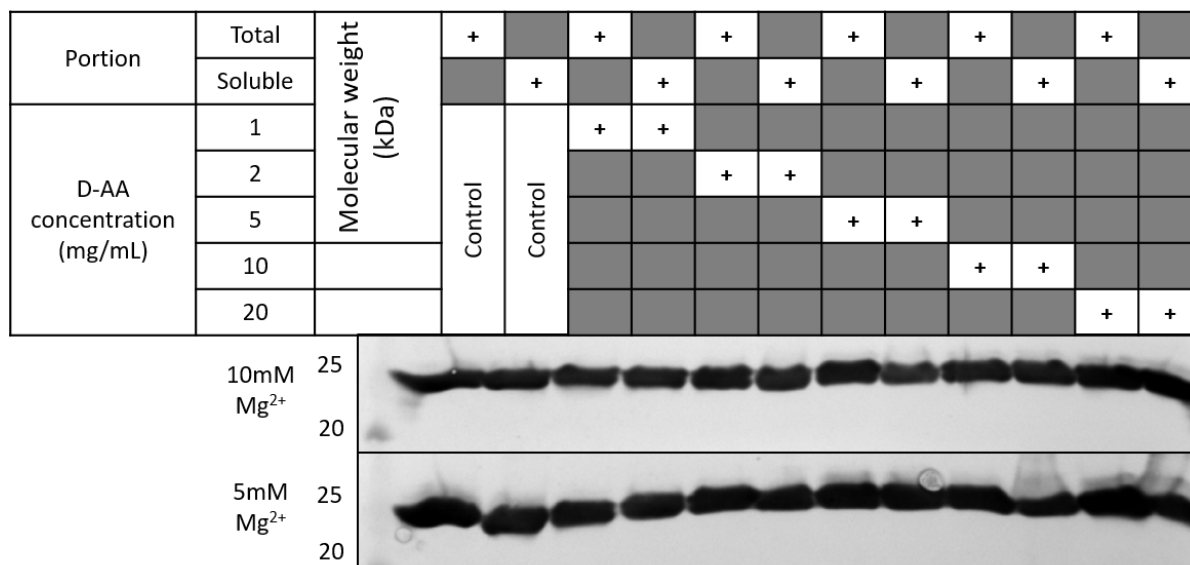


FIGURE 3.13: Two deuterated amino acids concentration screens were carried out; one using 5 mM  $Mg^{2+}$  and the other using 10 mM  $Mg^{2+}$  in the reaction to determine the optimal conditions for soluble deuterated 2dCD4-WT expression. The overall expression is noticeably higher using 5 mM  $Mg^{2+}$  in the reaction set-up. In addition, the portion of soluble protein appears higher using 5 mM  $Mg^{2+}$ . There is a slight increase in expression from 1 to 2 mg/mL but overall the expression is relatively constant, regardless of the amino acid concentration.

### Scaling-up and purification of cell-free protein expression of protiated and deuterated 2dCD4-WT

During the scale-up process, it was noticed that the agitation speed of the reactions has a crucial effect on protein expression. It was found that higher agitation produced cloudy reaction mixtures after incubation producing lower protein expression; this was particularly noticeable for the 10 mL reaction. For scaling-up, slower agitation speeds  $\leq 50$  rpm were therefore used. To minimise the impact of precipitant formation on subsequent purification, the CFPE reactions were centrifuged after the 12 hour incubation and the supernatant diluted five-fold in loading buffer before loading onto the NiNTA column.

Figure 3.14 depicts the chromatogram (3.14a) and SDS-PAGE analysis (3.14b) of the nickel IMAC purification of a 50 mL CFPE h-2dCD4-WT reaction while figure 3.15 depicts the nickel IMAC purification chromatogram (3.15a) and SDS-PAGE analysis (3.15b) of a 50 mL CFPE d-2dCD4-WT reaction. Reaction volumes as high as 50 mL per 250 mL plastic conical flask were used without compromising the yields.

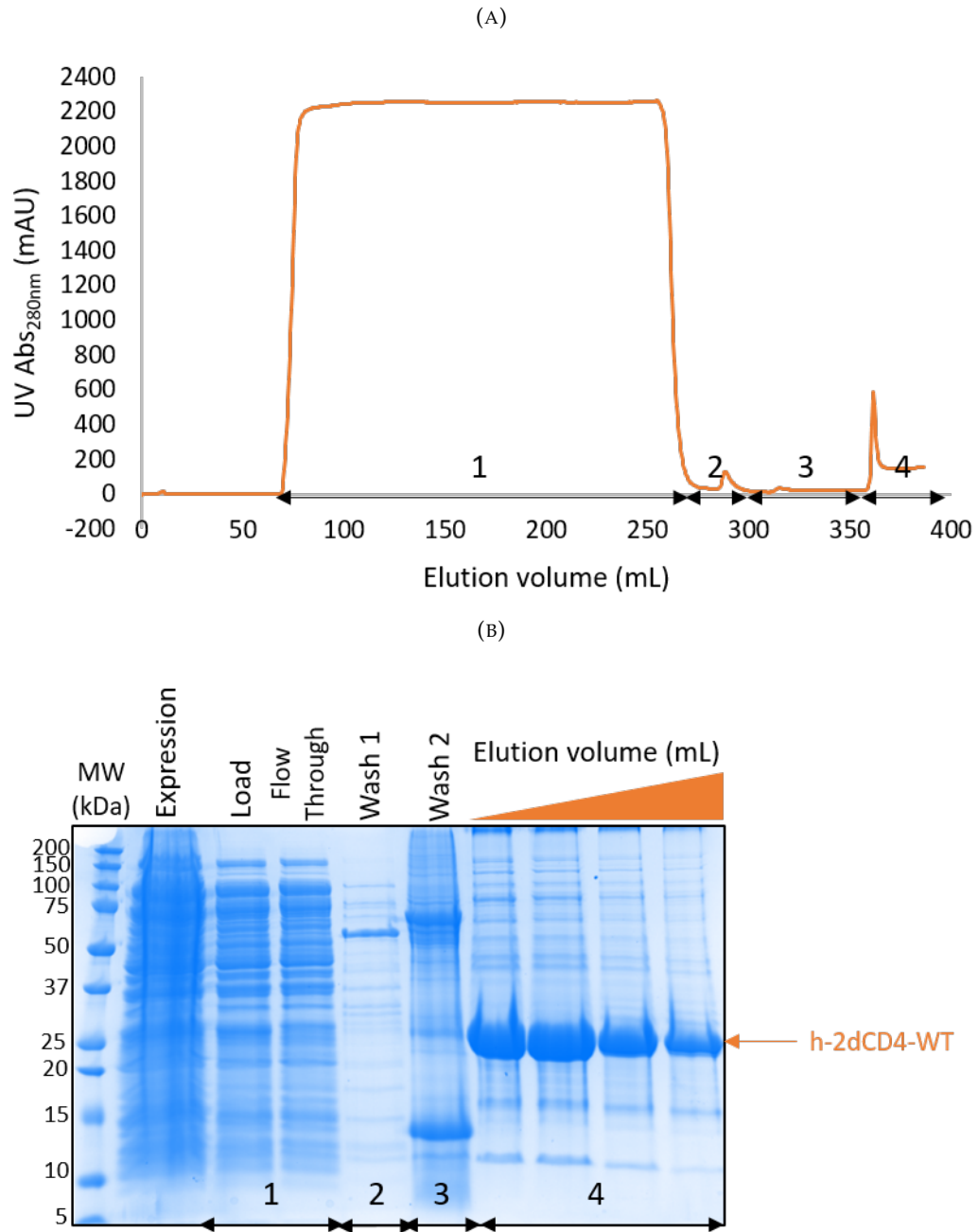


FIGURE 3.14: 50 mL cell-free protein expression reaction of protiated 2dCD4-WT (h-2dCD4-WT) was diluted 1:5 in binding buffer containing 5 mM imidazole and loaded at 1 mL/min onto a 5 mL prepacked Nickel HisTrap column (GE). The column was washed with 10 column volumes (CV) of wash buffer 1 containing 1 M NaCl and 20 mM imidazole and then washed with 10 CV of wash buffer 2 containing 40 mM imidazole. The protein was eluted into 1 mL fractions with gradient of elution buffer containing 300 mM imidazole. (A) Nickel IMAC chromatogram of h-2dCD4-WT. The flow through, washes and elution fractions were tested for the presence of d-2dCD4-WT by Coomassie blue stained SDS-PAGE (B). The numbers 1, 2, 3 and 4 correlate what is seen in the chromatogram with the SDS-PAGE analysis. 1 = load/flow through; 2 = wash 1; 3 = wash 2; 4 = elution.

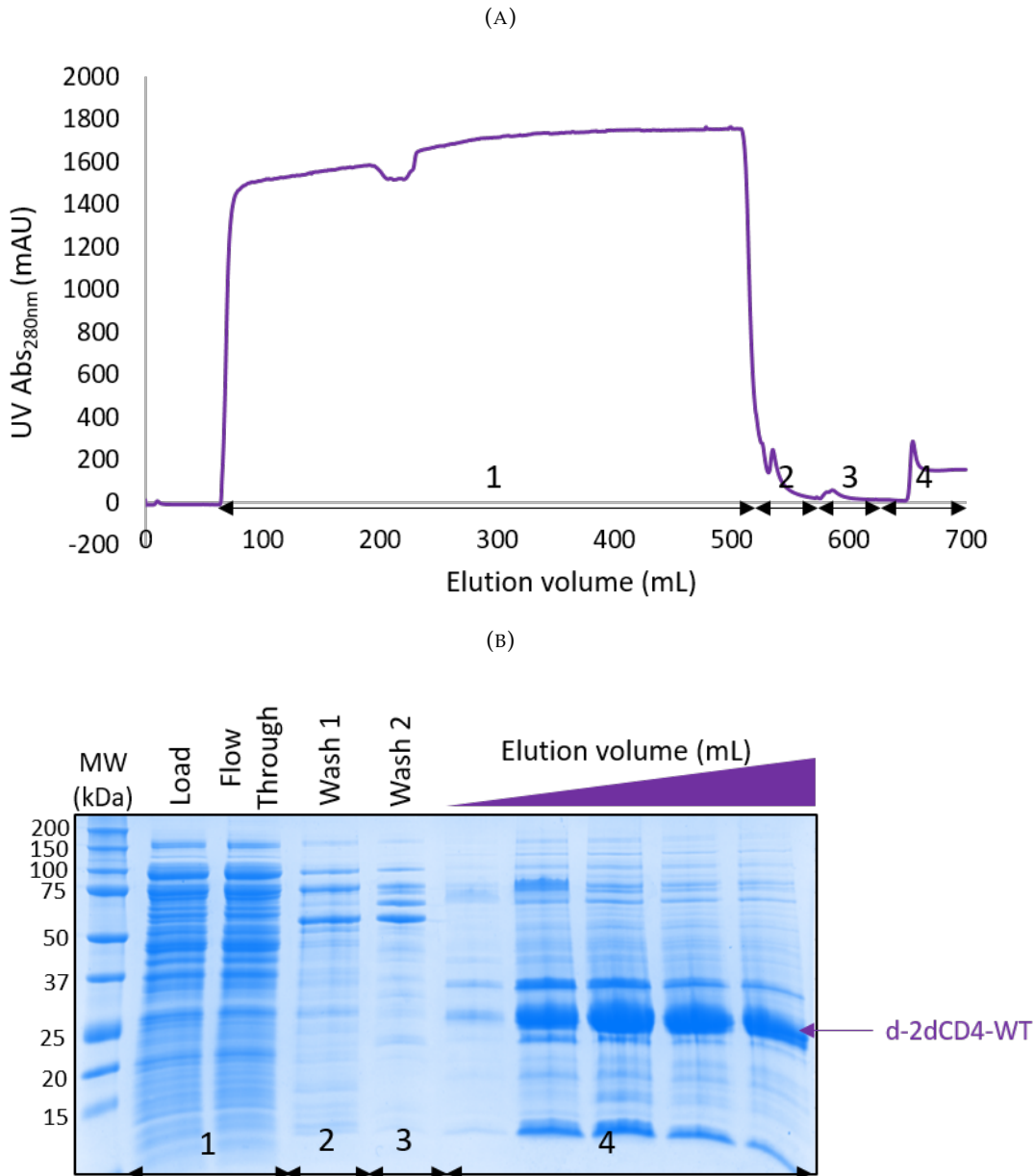


FIGURE 3.15: 50 mL cell-free protein expression reaction of deuterated 2dCD4-WT (d-2dCD4-WT) was diluted 1:5 in binding buffer containing 5 mM imidazole and loaded at 1 mL/min onto a 5 mL prepacked Nickel HisTrap column (GE). The column was washed 10 column volumes (CV) with wash buffer 1 containing 1M NaCl and 20 mM imidazole and then washed with 10 CV of wash buffer 2 containing 40 mM imidazole. The protein was eluted into 1 mL fractions with gradient of elution buffer containing 300 mM imidazole. (A) Nickel IMAC chromatogram of d-2dCD4-WT. The flow through, washes and elution fractions were tested for the presence of d-2dCD4-WT by Coomassie blue stained SDS-PAGE (B). The numbers 1, 2, 3 and 4 correlate what is seen in the chromatogram with the SDS-PAGE analysis. 1 = load/flow through; 2 = wash 1; 3 = wash 2; 4 = elution.

Following nickel IMAC purification, the protein containing fractions were pooled loaded directly onto a S75 pg size-exclusion column to be further purified. Examples of a h-2dCD4-WT

and d-2dCD4-WT SEC chromatogram and subsequent SDS-PAGE analysis of the fractions can be observed in figures 3.16 and 3.17, respectively. Yields between 2 and 3 mg per 50 mL CFPE reaction for both hydrogenated and deuterated 2dCD4-WT were possible which allowed access to further biophysical analyses and SAS experiments.

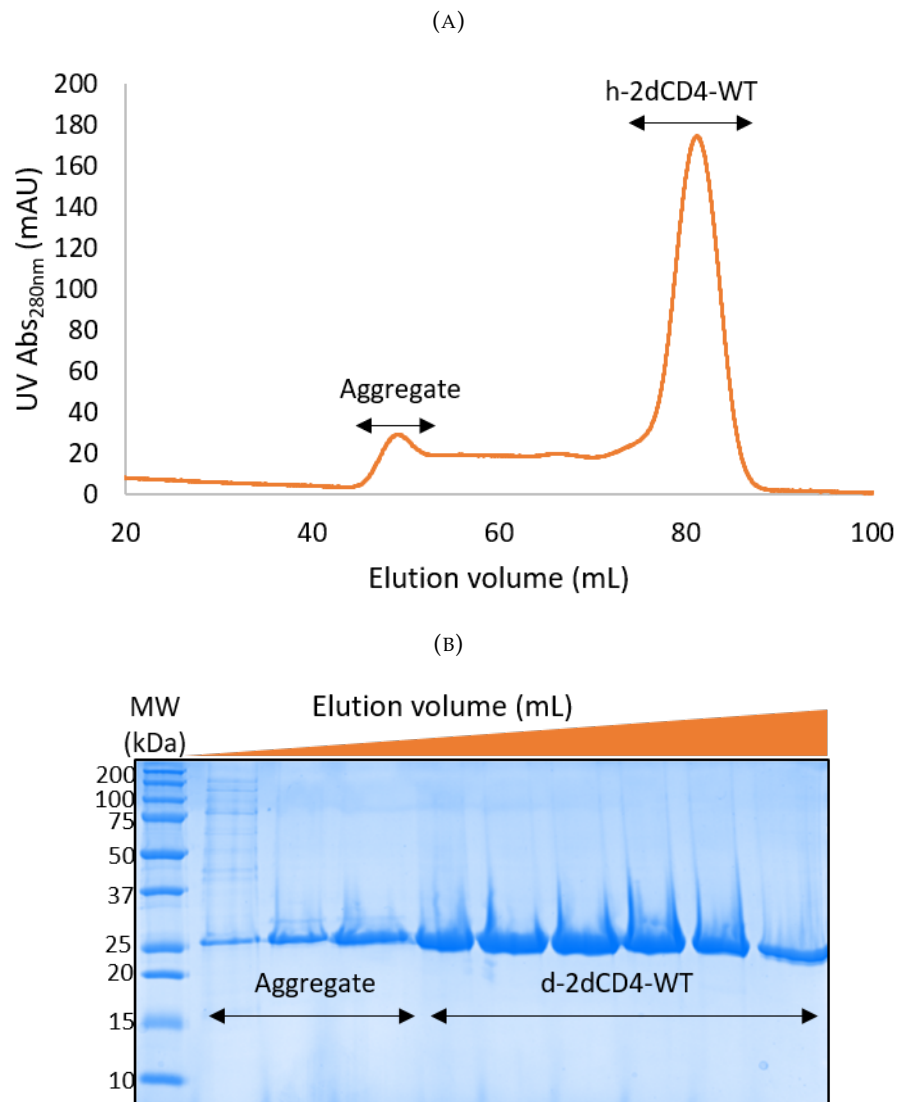


FIGURE 3.16: h-2dCD4-WT post NiNTA was further purified by size-exclusion chromatography on an S75 HiLoad 16/600 column equilibrated in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% sucrose. (A) h-2dCD4-WT SEC chromatogram. The protein eluted at 1 mL/min and was collected in 1 mL fractions. (B) The presence and purity of h-2dCD4-WT was detected by SDS-PAGE with Coomassie staining.

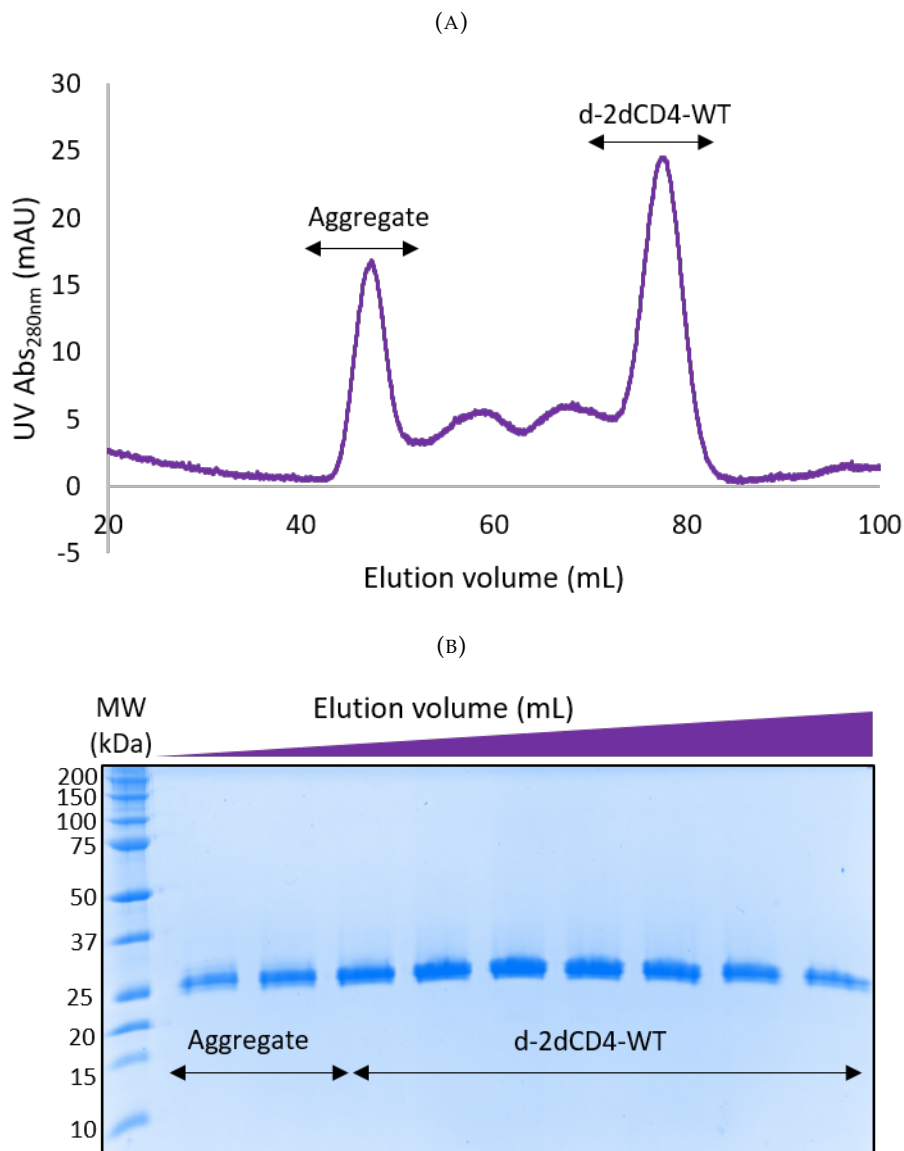


FIGURE 3.17: d-2dCD4-WT post NiNTA was further purified by size-exclusion chromatography on an S75 HiLoad 16/600 column equilibrated in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% sucrose. (A) d-2dCD4-WT SEC chromatogram. The protein eluted at 1 mL/min and was collected in 1 mL fractions. (B) The presence and purity of d-2dCD4-WT was detected by SDS-PAGE with Coomassie staining.

### 3.4.5 Expression of 2dCD4-D2A by CFPE

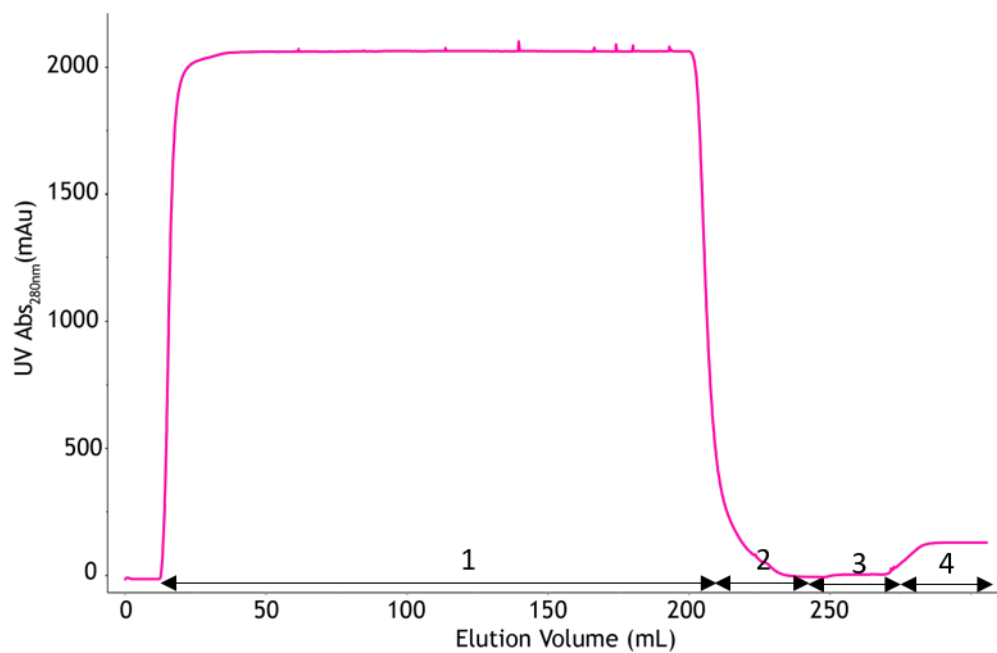
Having successfully established a protocol for the expression of labelled and unlabelled 2dCD4-WT using CFPE, several experiments were carried out to test the expression of the 2dCD4 domain 2 disulphide bond knockout in which the cysteine residues had been replaced by alanine residues (2dCD4-D2A).



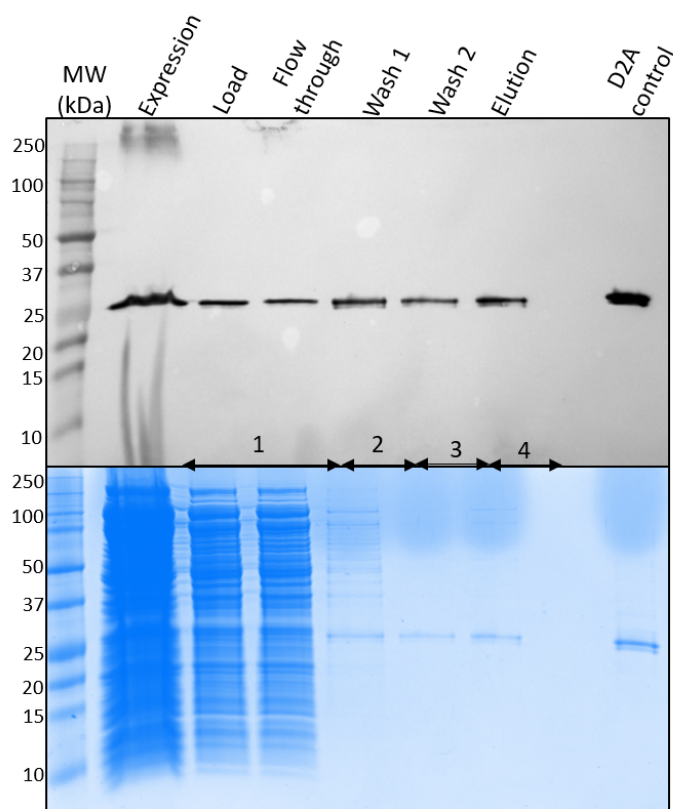
### 2dCD4-D2A purification

Using the same conditions as those described for production of protiated 2dCD4-WT using CFPE (section 3.4.4), a 50 mL CFPE reaction of 2dCD4-D2A was set-up and purified by nickel IMAC (fig. 3.18). Figure 3.18a shows the nickel IMAC UV  $Abs_{280nm}$  chromatogram and figure 3.18b shows anti-His western blotting (top) and Coomassie blue stained SDS-PAGE (bottom) analysis of the steps of the purification. 2dCD4-D2A is detected in the flow through and washes which suggests that it did not bind well to the column.

After nickel IMAC purification 2dCD4-D2A was further purified by SEC (fig. 3.19). Figure 3.19a shows the SEC chromatogram and figure 3.19b shows SDS-PAGE analysis of the peaks identified in the chromatogram. A large peak which elutes just after the dead volume of the column contains the highest concentration of 2dCD4-D2A, according to the SDS-PAGE analysis. The position of the peak with regards to its elution volume and the content from the SDS-PAGE suggests that this peak corresponds to aggregated 2dCD4-D2A. Peak 3 elutes at the expected position of 2dCD4-D2A but also contains a contaminant which resolves at a lower apparent molecular weight by SDS-PAGE analysis which may be a truncation product of the CFPE reaction. Peak 4 contains a lower molecular weight contaminant. The final yield of 2dCD4-D2A was just 1 mL at 0.08 mg/mL as much of the protein was lost due to aggregation.

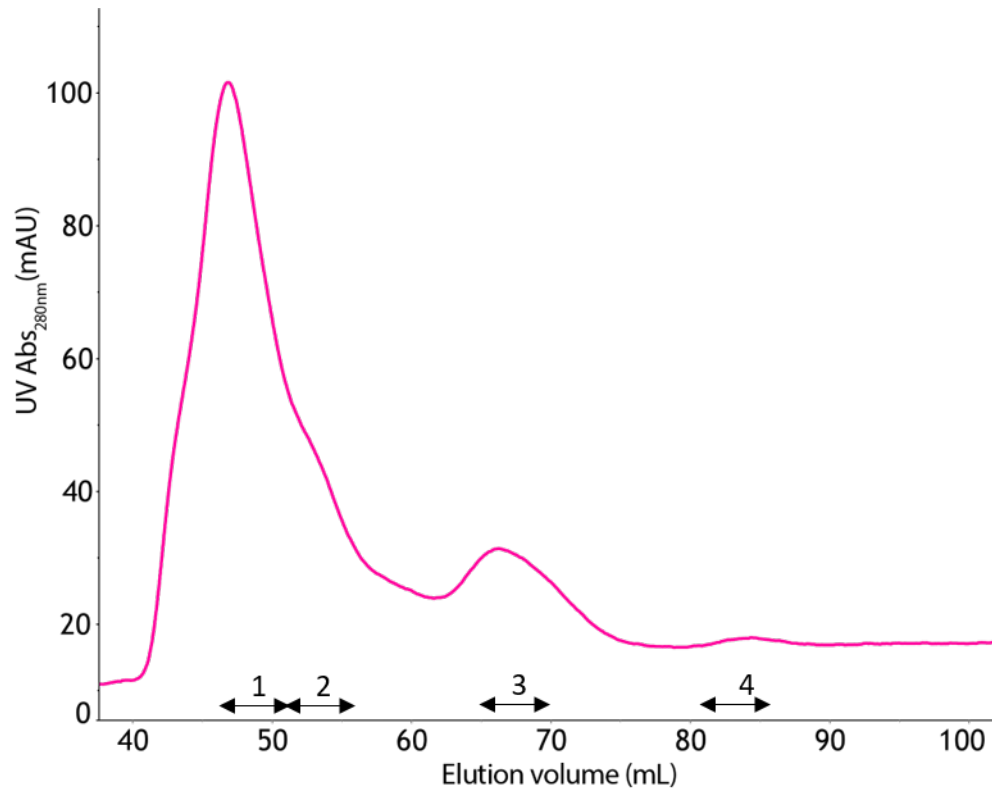


(A)

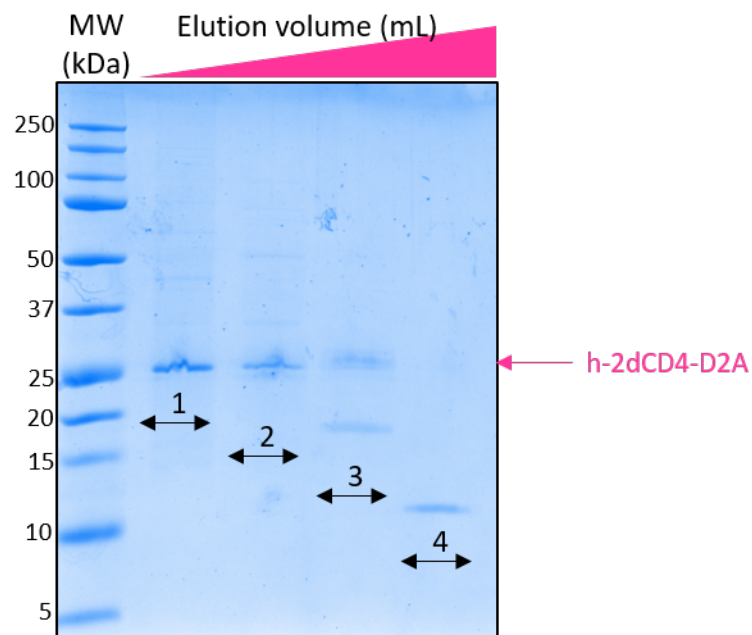


(B)

FIGURE 3.18: A 50mL 2dCD4-D2A CFPE reaction was set-up and purified by nickel IMAC. (A) Shows the UV absorbance at 280nm to follow the protein during the purification. (B) Shows anti-His western blot (top) and Coomassie blue stained SDS-PAGE (bottom) analysis of the steps of the purification. 2dCD4-D2A came off in the flow-through and washes which suggests that it did not bind well to the column.



(A)



(B)

FIGURE 3.19: The 2dCD4-D2A nickel IMAC elution fractions were pooled and purified by SEC on an S75 pg. (A) SEC chromatogram of 2dCD4-D2A. (B) The presence and purity of 2dCD4-D2A was assessed by Coomassie blue stained SDS-PAGE analysis. A large peak eluted just after the dead volume of the column which was shown to contain 2dCD4-D2A. This corresponds to a 2dCD4-D2A aggregate and is the major 2dCD4-D2A peak.

### Stabilising 2dCD4-D2A produced by CFPE

In an attempt to increase the stability of 2dCD4-D2A produced by CFPE, the reaction conditions were revisited. Firstly, the reaction temperatures of 16°C and 20°C at 40 rpm and 20°C at 0 rpm were tested. The total and soluble expression was assessed by anti-his western blot (fig. 3.20). Lack of agitation appears to have a negative effect on soluble expression. However, the temperature did not have any significant effect on the soluble expression.

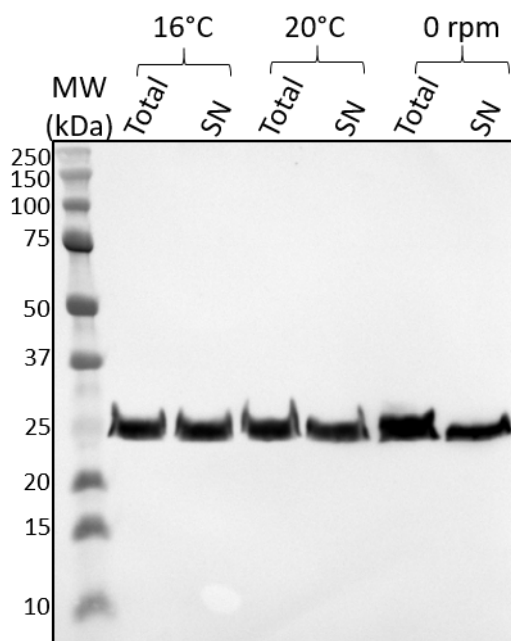


FIGURE 3.20: 50  $\mu$ L analytical scale 2dCD4-D2A CFPE reactions were set-up and incubated at 16°C and 20°C with 40 rpm agitation and at 20°C with no agitation to see if this would affect soluble protein yield. The total and soluble expression was tested using anti-His western blotting. No agitation has a negative effect on the soluble expression. There is no difference in the soluble expression between temperatures.

Secondly, expression in the presence of different concentrations of detergent in the reaction mix was assessed. The addition of detergent was hoped to stabilise the protein by occluding patches of hydrophobicity located on the surface of 2dCD4-D2A, such as that located at the domain 2/domain 3 interface, which may promote aggregation. n-Dodecyl  $\beta$ -D-maltoside (DDM) and Brij35 detergents were added to analytical scale reactions at final concentrations of 0.25%, 0.5% and 1.0%. The total and soluble expression was tested using anti-his western blotting (fig. 3.21). 0.25% DDM gave the best soluble expression of the reactions set-up using DDM. 0.5% Brij35 also gave good soluble expression however, the lane is streaky thus complicating the analysis.

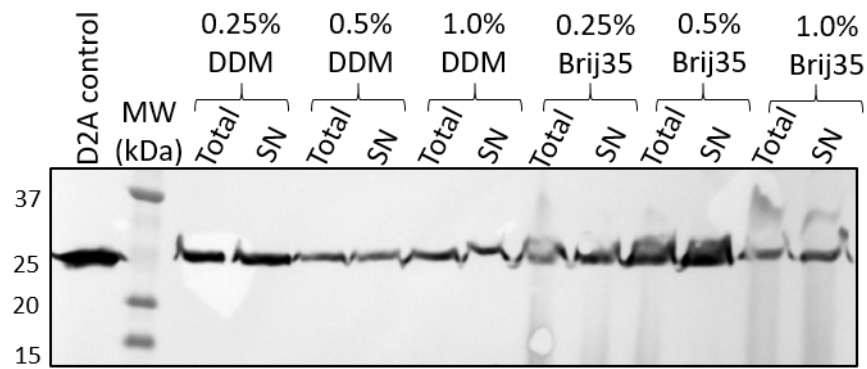


FIGURE 3.21: 50  $\mu$ L analytical scale 2dCD4-D2A CFPE reactions were set-up using different concentrations of DDM and Brij35 detergent to see if this would affect soluble protein yield. The total and soluble expression was tested using anti-His western blotting. 0.25% DDM appears to give the best total and soluble expression. At 0.5% Brij35 the lanes are streaky which confuses the analysis.

Finally, an analytical purification of 2dCD4-D2A with 0.25% DDM in the reaction mix was compared to that without DDM by anti-his western blotting (fig. 3.22). The 2dCD4-D2A content of each purification step appears the same for both purification with and without 0.25% DDM.

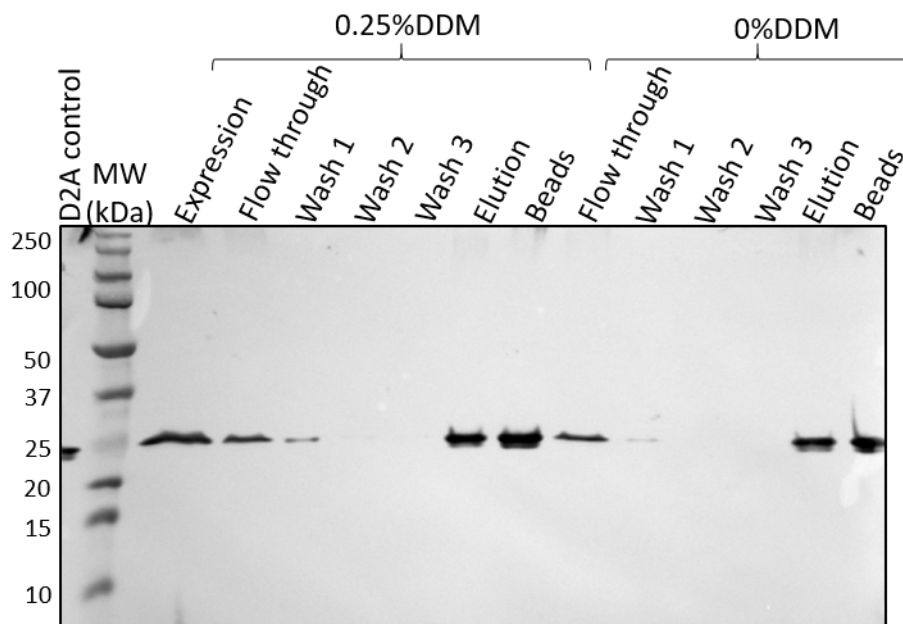


FIGURE 3.22: A 1 mL 2dCD4-D2A analytical scale CFPE reaction was set-up using 0.25% DDM in the reaction mix and purified using his conjugated magnetic beads. This was compared to a reaction on the same scale with no DDM in the reaction mix. Samples were taken at each step of the purification and tested for the presence of 2dCD4-D2A by anti-His western blot analysis. The amount of protein visible at each stage of the purification is comparable between the two conditions.

Due to unfortunate time constraints further analysis and development of 2dCD4-D2A expression and purification was not possible beyond this point.

### 3.5 Discussion

CD4, as the primary receptor of the HIV surface protein gp120, has been and continues to be studied a great deal in both its native and recombinant forms. Recombinant CD4 expresses poorly into insoluble inclusion bodies when produced using *E. coli*, therefore it is typically produced in mammalian cells which allow the correct formation of its disulphide bonds. However, match-out labelling for SANS experiments is not possible using mammalian cells. *E. coli* is the typical expression system used for match-out labelling of recombinant proteins using fed-batch fermentation. Expression, purification and refolding of 2dCD4-WT from insoluble inclusion bodies produced in *E. coli* was found to be extremely problematic. Therefore, expression in the gram positive bacterium *B. choshinensis* was explored, which allowed secretion of correctly folded 2dCD4-WT into the supernatant. However development of a deuteration strategy for 2dCD4 using *B. choshinensis* as an expression host proved challenging and expression levels were insufficient to allow structural studies.

Eventually, through a collaboration with cell-free protein expression experts Synthelis, production of 2dCD4-WT in both its deuterated and protiated forms was optimised and scaled-up using cell-free protein expression from *E. coli* SHuffle cell lysates. CFPE proved able to produce sufficient quantities 2dCD4-WT suitable for biophysical, SAXS and SANS experiments present in the subsequent chapters of this thesis. CFPE was used in an attempt to produce the second domain disulphide bond knock-out protein; 2dCD4-D2A in which the second domain Cys are replaced by Ala residues. Unfortunately, stable production of the disulphide knock-out variant was not possible within the time-frame, given the initial and prolonged difficulties in achieving sufficient yields of the wild-type protein.

#### 3.5.1 Poor 2dCD4-WT yield from refolded *E. coli* expressed inclusion bodies

Several avenues were explored to improve expression of 2dCD4-WT in *E. coli*. While expression of recombinant proteins by *E. coli* into insoluble inclusion bodies requires laborious refolding

protocols to obtain correctly folded bioactive protein, one of the benefits of inclusion body formation is usually that the recombinant protein is hugely over-expressed. However this did not appear to be the case for 2dCD4-WT. Initially no expression was detected during total protein fraction expression tests. Several expression tests were carried out including changing of the expression vector, expression strain, expression temperature and IPTG concentration but no 2dCD4-WT was detected by western blot or SDS-PAGE. And yet 2dCD4-WT was still recovered after isolation, solubilisation, purification and refolding from inclusion bodies expressed by BL21\* (DE3). It was eventually discovered that the expression of 2dCD4-WT into insoluble inclusion bodies necessitated denaturation using 8 M urea .

When it came to preparing deuterated 2dCD4-WT for future SANS experiments fed-batch fermentation was used and it was discovered that the expression samples required solubilisation in 8 M urea in order for 2dCD4-WT to enter the pores of an SDS-PAGE to detect expression by Coomassie blue staining or western blot detection of the His-tag. Once this was established, expression of 2dCD4-WT was found to be very low. The total yield of 2dCD4-WT from the fermenter was estimated to be just 800 µg before solubilisation, purification and refolding which would significantly decrease the yield obtained. Using *E. coli* as an expression host with fed-batch fermentation was therefore not a suitable method for deuterium labelling of 2dCD4-WT. In addition, the protein recovered from inclusion bodies was not very stable as attempts to concentrate the sample above 1.5 mg/mL resulted in aggregation. This meant that even with protiated 2dCD4-WT it was difficult to carry out any biophysical or structural analysis on the protein.

### 3.5.2 Expression and secretion of 2dCD4-WT by *B. choshinensis*

By using *B. choshinensis* as an expression host for the production of recombinant 2dCD4-WT there was no longer any need for a lengthy refolding protocol as the protein was secreted directly into the expression medium. However, with this came issues with the ease of purification as there were large volumes of the expression medium supernatant (cells pelleted by centrifugation) which contained many protein contaminants. This issue was partly resolved by ammonium sulphate precipitation by concentrating 2dCD4-WT into a smaller volume of supernatant during which some of the protein contaminants were removed in the process. In

addition, expression took four days so while the lengthy refolding protocol was not needed this time was lost elsewhere.

While expression of unlabelled 2dCD4-WT was generally good, a large portion of the protein was found to be aggregated during the SEC purification step, post-IMAC purification. A large peak which eluted just after the dead volume of the SEC column which was found to contain 2dCD4-WT was attributed to aggregation. Therefore the final yields of 2dCD4-WT from *B. choshinensis* were low. Additionally, the protein was not particularly stable and concentrations over 2 mg/mL were not attainable.

Several attempts were made to determine conditions in which *B. choshinensis* would grow in deuterated media. *B. choshinensis* does not grow in minimal media [192], which suggests that this bacterium is an amino acid auxotroph and so cannot synthesise all twenty amino acids. It was also shown not to grow in 100% D<sub>2</sub>O but some growth was seen in partially deuterated Silantes media. Some expression of 2dCD4-WT was also seen in D-Silantes but subsequent yields after analytical purification of 2dCD4-WT were too low for determination of the percentage deuteration from growth in 65% and 75% D-Silantes by mass spectrometry as no protein was detected (therefore the results are not shown).

### 3.5.3 Production of protiated and deuterated 2dCD4-WT using cell-free protein expression

Thanks to a collaboration established with Vinesh Jugnarain at Syntheliss, sufficient yields of deuterated and protiated 2dCD4-WT were achieved which allowed biophysical (chapter 4, 5) and small-angle scattering studies (chapter 4, 6) to be conducted on the protein in isolation. Initial expression tests allowed determination of CFPE reaction conditions which would yield the most soluble expression. Thanks to the speed of the CFPE reaction, the optimal soluble protein expression conditions were determined rapidly compared to *B. choshinensis* which needed several days of expression. The lengthy part of CFPE was in production of the lysate (approximately 1 week) but when this was completed the reactions could be set-up and the protein purified within 24 hours, which is a great advantage and made CFPE the fastest expression method of the three used.



Thanks to the openness of the CFPE reaction, match-out labelling of 2dCD4-WT was no more complicated than the production of unlabelled 2dCD4-WT, requiring simply the addition of deuterated amino acids rather than unlabelled ones to the reaction mix. The yield for d-2dCD4-WT was also comparable to that for h-2dCD4-WT. The main drawback of using CFPE for match-out deuteration therefore lies in the cost, as deuterated amino acids are expensive. In general use of CFPE is costly because of the need for the tRNAs and energy regenerating systems in order to carry out the protein synthesis reaction.

#### 3.5.4 2dCD4-D2A variant production using CFPE

Having successfully produced 2dCD4-WT using CFPE, the same protocol was used in production of 2dCD4-D2A. However, while the expression levels were good, the protein was found to be unstable during nickel IMAC. This was surprising given the success had with the production of 2dCD4-WT using CFPE. It is difficult to rationalise why this variant protein was so unstable. The fact that the expression of soluble 2dCD4-D2A appears similar to that of 2dCD4-WT using CFPE but that the protein is found to aggregate severely (as shown by SEC analysis) suggests that the issue lies in the folding of 2dCD4-D2A (or lack thereof) when produced using CFPE.

The addition of detergents to the reaction mix was used in an attempt to decrease any aggregation which may have arisen through hydrophobic interactions. The addition of 0.25% DDM had little effect on the yield of 2dCD4-D2A in an analytical scale purification. However, subsequent analysis was not carried due to insufficient time, which prevented any further experiments being carried out. The intention was to attempt to scale up the reaction with 0.25% DDM and assess the protein using analytical scale SEC to see if the present of the aggregation peak seen in the 50 mL SEC purification was diminished. The resulting protein with and without 0.25% protein would then have been concentrated to see if the protein in 0.25% DDM concentrated to a higher concentration than the 2dCD4-D2A without DDM. Mass spectrometry analysis could then have been carried out to assess the purity.

### 3.5.5 Future work

A protocol was established for production of deuterium labelled and unlabelled 2dCD4-WT using CFPE, however the same cannot be said for 2dCD4-D2A. Unfortunately, due to time constraints, a protocol for the expression of correctly folded 2dCD4-D2A was not developed. Therefore the result of the presence of 0.25% DDM detergent on the stability of the protein was not tested. However if the protein was initially incorrectly folded the presence of detergent would not necessarily alter the stability of the protein. Use of CFPE for production of 2dCD4-D2A does not provide any advantage over *E. coli* expression into inclusion bodies, if the protein is produced in an incorrectly folded form. In future experiments it would be pertinent to assess the presence of a chaperone protein system on the production of stable 2dCD4-D2A. Kang *et al.* 2005 [193] used CFPE with over expression of either the GroEL/ES and DnaK/J-GrpE chaperone system to produce soluble, active forms of aggregation-prone proteins. Introduction of a chaperone protein system could help 2dCD4-D2A to fold correctly. If this was successful then subsequent biochemical, biophysical and structural analyses could be carried out.

While CFPE was used in the context of match-out labelling for SANS contrast variation studies, CFPE could also be used for per-deuteration for neutron crystallography. However this would require setting up of the CFPE in a deuterated reaction buffer as well as using deuterated amino acids. Alternatively, triple labelled amino acids could be used for NMR studies. Use of a high resolution technique such as neutron crystallography or NMR would allow determination of local, small structural realignments as a function of the redox state of 2dCD4-WT.

## 3.6 Conclusions

2dCD4-WT was expressed as insoluble inclusion bodies by *E. coli* which required solubilisation in 8 M urea before refolding, during which most of the protein yield was lost. Fed-batch fermentation of an *E. coli* culture was carried out in 85% deuterated minimal media to express deuterated 2dCD4-WT but the yield before refolding was under one milligram, which is insufficient for small-angle scattering studies. *B. choshinensis* is a gram-positive bacterium which secreted soluble 2dCD4-WT into the expression medium. However there were large volumes of expression medium to process and the resultant yield was low as there was a significant

---

amount of aggregation identified by size-exclusion chromatography. In addition, *B. choshinensis* does not grow in minimal media and expression of 2dCD4-WT in rich deuterated media was poor. Eventually, milligram quantities of both deuterated and protiated 2dCD4-WT were produced using cell-free protein expression. Cell-free protein expression, although expensive, was also the quickest of the three expression techniques used. Expression of the 2dCD4 domain 2 disulphide knockout variant (2dCD4-D2A) was comparable to that of 2dCD4-WT but size-exclusion chromatography showed that the protein was aggregated. Further tests would need to be conducted to establish a protocol for the production of correctly folded 2dCD4-D2A.



## Chapter 4

# Biophysical characterisation of cell-free expressed 2dCD4

### 4.1 Abstract

This chapter explores the characterisation of recombinant protiated and deuterated wild-type two-domain CD4 (h- and d-2dCD4-WT) produced using the cell-free protein expression (CFPE) system compared to protiated 2dCD4-WT expressed using *Escherichia coli* and refolded from inclusion bodies and 2dCD4-WT expressed and secreted by *Brevibacillus choshinensis* as described in chapter 3, for use in subsequent small angle scattering studies described in chapters 5 and 6.

h- and d-2dCD4-WT produced by CFPE were shown to be functional using enzyme-linked immunosorbant assay and a pseudo-viral neutralisation assay, with h-2dCD4-WT produced using the *E. coli* and *B. choshinensis* expression systems as functional references. In addition, standard and HETSOFAST one-dimensional proton nuclear magnetic resonance experiments were carried out to assess whether the CFPE produced h-2dCD4-WT was correctly folded. Size-exclusion chromatography couple to multi-angle laser light scattering and refractive index measurements, and mass-spectrometry were employed to analyse the monodispersity and absence of proteolytic degradation of the protein after its purification. Mass spectrometry was also used to determine the percentage deuterium incorporation and N-terminal protein sequencing to confirm that the correct protein was produced and that it was correctly processed at its N-terminus.

## 4.2 Introduction

As addressed in chapter 3, CD4 is typically expressed in mammalian cells for structural studies in complex with gp120 [34] or the engineered SOSIP trimers [43] (stable trimeric envelope glycoprotein complex mimic). However for this thesis work it was necessary to produce match-out labelled 2dCD4 with deuterium so that it could be studied using contrast variation SANS studies (chapter 6). Attempts were made to deuterate the protein in both *E. coli* and *B. choshinensis* expression systems but, unfortunately, the yields were low (approximately 0.25 mg/L) making structural analysis difficult. Through a collaboration with Syntheliss, La Tronche, France, it was possible to establish recombinant production protocols for both protiated and deuterated 2dCD4-WT as described in chapter 3 which allowed structural analyses by both SAXS and SANS. This chapter is concerned with the biochemical and biophysical characterisation of the protiated and deuterated cell-free protein expressed (CFPE) 2dCD4-WT (hereafter referred to as h-2dCD4-WT and d-2dCD4-WT) prior to its use for structural studies and; where possible, they are compared to refolded h-2dCD4-WT expressed in *E. coli* as inclusion bodies and secreted *B. choshinensis* h-2dCD4-WT.

The functionality of CFPE produced h- and d-2dCD4-WT was compared to that of h-2dCD4-WT expressed by *E. coli* and *B. choshinensis* using ELISA and pseudo-viral neutralisation assays. Indirect ELISAs were used, whereby the detection event is of the primary antibody by a secondary antibody. The indirect ELISA method was first applied to an indirect-direct detection of 2dCD4-WT, in which the primary antibody directly bound to 2dCD4-WT. For this ELISA the antibody ibalizumab was used which binds to an epitope located opposite the gp120 and MHC-II binding sites at the interface between domains 1 and 2 [97]. Binding of ibalizumab to 2dCD4-WT would directly indicate that the ibalizumab epitope is intact.

The second ELISA used an indirect-indirect detection of 2dCD4-WT, in which the primary antibody detected a 2dCD4-WT induced epitope on gp120. The indirect-indirect ELISA used the 17b antibody which is an anti-gp120 antibody, binding to the CD4 induced (CD4i) epitope on gp120. The CD4i epitope is only exposed on gp120 after CD4 has bound [35], therefore binding of the 17b antibody to gp120 indicates that the CD4 is able to bind to the CD4 binding site (CD4bs) on gp120 thereby indirectly suggesting that the gp120 binding site on CD4 is intact. With an increase in CD4 concentration it was expected that there would be an increase in the

absorbance at 450 nm of the electrochemiluminescent substrate, corresponding to an increase in primary antibody binding to their specific epitopes.

A second method used for testing the functionality of the 2dCD4-WT analogues was the pseudo-viral neutralisation assay. During this assay the 2dCD4-WT analogues were used to inhibit the binding of native CD4 from TZM-bl cells (containing a luciferase reporter gene) to a series of tier 1 and 2 HIV's with a non-HIV control [146]. With an increase in the 2dCD4-WT concentration it was expected that there would be a decrease in the absorbance at 450 nm as there would be a reduction in HIV-1 tat mediated transcription of the *luc* reporter gene, corresponding to 2dCD4-WT mediated inhibition of native CD4 binding to the viral Env protein (see Chapter 2 fig. 2.4). Also, it was predicted that the tier 1 viruses should be neutralised at lower concentrations of CD4 than the tier 2 viruses.

A series of biophysical analyses were subsequently carried out on the CFPE-produced 2dCD4-WT as a means of controlling for the quality of the protein produced. In-line with the functional assays, standard and HETSOFAST one-dimensional proton nuclear magnetic resonance (1D NMR) experiments were used as a biophysical assessment of the folding of the protein. Liquid chromatography, electro-spray ionisation, time-of-flight mass-spectrometry (LC-ESI-TOF MS) were used both as a means of testing the purity of the 2dCD4-WT analogues produced in the different expression systems and to determine the percentage deuteration achieved using the CFPE system. N-terminal sequencing was used to confirm that the protein sequence of the sample identified by mass spectrometry corresponded to that of 2dCD4-WT. Finally, size-exclusion chromatography coupled to multi-angle laser light scattering and refractive index (SEC-MALLS-RI) was used to assess the polydispersity of the 2dCD4-WT and its stability during freezing. Unfortunately, the NMR and SEC-MALLS-RI experiments could not be conducted for comparison with the *E. coli* and *B. choshinensis* analogues because the sample yield requirements were too high. Crystallisation of the 2dCD4-WT produced using CFPE was attempted, however no crystals were produced and optimisation of the crystallisation process was not carried out due to time constraints.

## 4.3 Materials and methods

### 4.3.1 2dCD4 expression and purification

Protiated and deuterium-labelled 2dCD4-WT analogues were expressed and purified using the *E. coli*, *B. choshinensis* and CFPE expression systems as outlined in chapter 3.

### 4.3.2 Gp120 expression and purification

Gp120 was also expressed and purified according to the protocols described in chapter 6.

### 4.3.3 Functional assays

#### Enzyme-linked immunosorbant assay

One maxisorb 96-well flat-bottomed plate (Nunc) per antibody (17b & ibalizumab) was coated overnight at 4°C with lectin (200 ng/well). The plates were blocked with 1% BSA (bovine serum albumin) in PBS-T (phosphate buffered saline-tween) for 1 hour at room temperature. The CD4 variants were added to the plate (100-0.032 ng/well). 100 ng/well gp120 was added to the plate. The CD4 was left to incubate with the gp120 for 1 hour at room temperature. The CD4-induced epitope on gp120 was detected with the 17b antibody and the second domain of 2dCD4 with ibalizumab (30 ng/well). The primary antibodies were detected with 100 µL/well anti-human ECL (enhanced chemiluminescence) HRP-linked (horse radish peroxidase) antibody from sheep (Sigma) at a starting concentration of 1:2 000. Washes were carried out using 1 X PBS-T. 100 µL TMB (3,3',5,5'-tetramethylbenzidine, ThermoScientific) antibody was added per well. After addition of 50 µL 1M HCl per well to stop the reaction, the absorbance at 450 nm was measured by an iMark™ plate reader (BioRad).

#### Pseudo-viral neutralisation assay

A 1:3 serial dilution of the various 2dCD4-WT proteins was set up, in duplicate, in a 96-well flat bottomed plate to a final volume of 195 µL with DMEM (Dulbecco's modified eagle media) growth media and 10% foetal bovine serum (FBS) (both GIBCO). The highest concentration



used was 100 µg/mL and the lowest used concentration was 0.046 µg/mL. 25 µL was transferred from the dilution plate to each duplicate well across five 96-well flat bottomed plates, each of which was used for one of five viruses. The following viruses were used for the assay: 6535, ZM53, DU156, SF162 and VSV-G, of which the latter is a non-HIV control. The viruses were used at 4 000 TCID<sub>50</sub>/mL (tissue culture infectious dose 50%) with dilution in the growth media stated above as necessary. 25 µL of diluted virus (200 TCID<sub>50</sub>/mL) was added to the second lane of the plate, which acts as the virus -only control, and to all subsequent wells, except for the cell-only control lane (first lane) of the plate.

The plates were incubated for 1 hour at 37°C, 5% CO<sub>2</sub>. Meanwhile, the TZM-bl cells (NIH AIDS reagents) were diluted to a cell density of 5x10<sup>5</sup> cells/mL in growth media at 70 µg/mL DEAE dextran (Sigma). To prepare the cells, the growth media was discarded from the cell-culture flask and the cells were washed with 10 mL 1 X PBS. The 1 X PBS was discarded and the cells were treated with trypsin-EDTA (GIBCO) for 2 minutes at 37°C, 5% CO<sub>2</sub>. 10 mL of growth media was added to neutralise the trypsin and the cells were removed from the surface of the cell-culture flask by pipetting. 10 µL of cell suspension was added to 10 µL countess dye (Life Tech), of which 10 µL of this mixture was added to a countess cell counting slide (Life Tech). The cells were counted and prepared to a cell count of 1x10<sup>6</sup> in the growth media. 186 µL DEAE dextran at a concentration of 7.5 mg/mL was added per 20 mL cell suspension. 20 µL of cell suspension was added to the cell-only control and then to all subsequent wells.

The plates were incubated at 37°C, 5% CO<sub>2</sub> for 24 hours. 130 µL growth media (DMEM supplemented with 10% FBS) was added to each well and the plates were incubated for a further 24 hours at 37°C, 5% CO<sub>2</sub>. To assess the luciferase activity and thus the inhibition of viral entry into the TZM-bl cells 100 µL growth media was removed from each well, replaced by 100 µL Bright-Glo reagent (Promega) and incubated at room temperature for 2 minutes. The Bright-Glo was then vigorously mixed by pipetting and 100 µL transferred to a white 96-well flat bottomed plate (Corning). The luminescence of the reaction mixture in each white 96-well plate was then measured using a GloMax Explorer (Promega).

#### 4.3.4 1D nuclear magnetic resonance

0.125 mM 2dCD-WT in 50 mM Tris-HCl pH 7.5, 300 mM NaCl with 5% sucrose was tested by standard 1D NMR and 1D Heterogeneity-Selective Optimised-Flip-Angle Short-Transient (HET-SOFAST) NMR using a 600 MHz spectrometer (Bruker) equipped with a standard HCN probe, whose specifications are calibrated quarterly.

#### 4.3.5 LC-ESI-TOF Mass-spectrometry

20  $\mu$ L samples of hydrogenated 2dCD4-WT and deuterated 2dCD4-WT produced by CFPE were prepared at concentrations of 2.10 and 2.26 mg/mL with and without 50mM DTT treatment, respectively. 20  $\mu$ L of *E. coli* expressed and refolded 2dCD4-WT and *B. choshinensis* expressed and secreted 2dCD4-WT were also provided with and without 50 mM DTT treatment at concentrations of 1 mg/mL. The samples were provided to the IBS Mass Spectrometry platform, Grenoble for LC-ESI-TOF MS analysis.

A 6210 LC/ESI-TOF mass spectrometer with an inline HPLC binary pump system (Agilent Technologies) was calibrated in the mass-to-charge ( $m/z$ ) 300-3 000 range with standard calibrants (ESI-L, Low concentration tuning mix, Agilent Technologies) before measurements and mass spectra were recorded in the 300-3 200  $m/z$  range.

Prior to analysis the samples were diluted in acidic denaturing conditions to a final concentration of 4  $\mu$ M with solution A (0.03% trifluoroacetic acid [TFA] in water, Acros Organics). The temperature of the samples was adjusted to 10°C and the analysis was run by injecting 4  $\mu$ L of each sample. They were first trapped and desalted on a reverse phase-C8 cartridge (Zorbax 300SB-C8, 5  $\mu$ m, 300  $\mu$ m ID X 5 mm, Agilent Technologies) for 3 minutes at a flow rate of 50  $\mu$ L/min with 100% solution A and then eluted with 70% solution B (95% acetonitrile, 5% water, 0.03% TFA) at a flow rate of 50  $\mu$ L/min for MS detection. The RP-C8 cartridge was then re-equilibrated for 4 min with 100% solvent A at a flow rate of 50  $\mu$ L/min.

MS acquisition was carried out in the positive ion mode with spectra in the profile mode. The MS instrument was operated with the following experimental settings: ESI source temperature was set at 300°C; nitrogen was used as drying gas (7 l/min) and as nebulizer gas (10 psi); the capillary needle voltage was set at 4 000 V. The spectra acquisition rate was 1.03 spectra/s. All

solvents used were HPLC grade (Chromasolv, Sigma [apart from TFA, as mentioned previously]). The MS spectra were acquired and the data processed with MassHunter workstation software (v. B.02.00, Agilent Technologies) and with GPMAW software (v. 7.00b2, Lighthouse Data, Denmark).

#### **4.3.6 N-terminal protein sequencing**

200 picomoles of protilated 2dCD4-WT was supplied to the protein sequencing platform whereby amino acid sequence determination based on Edman degradation was performed using an Applied Biosystems gas-phase sequencer model 492 (s/n: 9510287J). Phenylthiohydantoin amino acid derivatives generated at each sequence cycle were identified and quantified on-line with an Applied Biosystems Model 140C HPLC system using the data analysis system for protein sequencing from Applied Biosystems (software Procise PC v2.1). The PTH-amino acid standard kit (Perkin-Elmer P/N 4340968) was used and reconstituted according to the manufacturer's instructions (900776 Rev D).

The procedures and reagents used were as recommended by the manufacturer. Chromatography was used to identify and quantify the derivatised amino acid removed at each sequence cycle. Retention times and integration values of peaks were compared to the chromatographic profile obtained for a standard mixture of derivatised amino acids.

#### **4.3.7 SEC-MALLS-RI**

50  $\mu$ L deuterated 2dCD4-WT at a concentration of 2.55 mg/mL was loaded via a pump (L2130, Elite LaChrom) onto a Superdex 75 10/300 size-exclusion column (GE Healthcare) equilibrated in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% sucrose buffer, with an inline spectrophotometer (L2400, Elite LaChrom). As the protein eluted from the column it passed through a flow cell where the multi-angle laser light scattering signal was measured (Dawn Helios-u, Wyatt). Finally, the refractive index was recorded by a refractometer (Optilab T-rEX, Wyatt) and the protein collected by a fraction collector (BioRad). The data were analysed using the software ASTRA6 (Wyatt Technology).

## 4.4 Results

### 4.4.1 Enzyme-linked immunosorbant assay (ELISA)

The ELISA measurements show that all 2dCD4-WT variants are functional as demonstrated by the CD4 concentration dependence on absorbance at 450 nm (fig. 4.1). Using the ibalizumab antibody it was possible to directly assess the conformational activity of 2dCD4-WT analogues. The ibalizumab epitope is located at the interface between the first and second domain of CD4, therefore requiring conformational integrity in order for the epitope to exist. From the absorbance at 450 nm curves (figure 4.1a) it can be seen that there is a CD4 concentration dependence on ibalizumab binding, since the ibalizumab concentration was kept constant. As the CD4 concentration increases, the absorbance at 450 nm increases because there are more epitopes available for ibalizumab binding. This trend is apparent for all 2dCD4 variants.

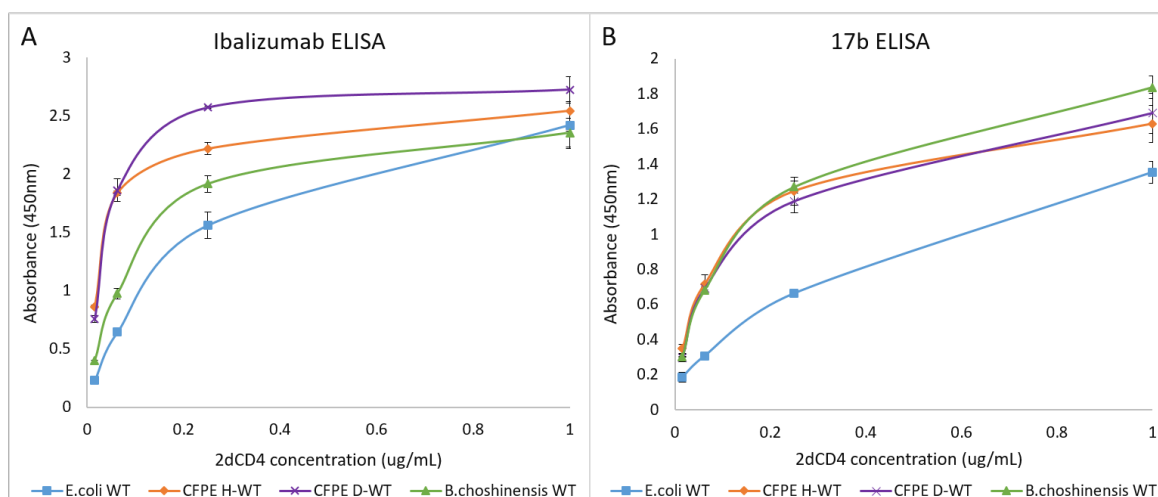


FIGURE 4.1: Enzyme-linked immunosorbant assay of 2dCD4-WT produced in different expression systems using a direct approach with the anti-CD4 antibody ibalizumab which binds directly to 2dCD4-WT (a) and an indirect approach with the anti-gp120 CD4 antibody 17b which binds to the CD4-induced epitope (b). Blue = *E. coli* expressed and refolded protiated 2dCD4-WT (h-2dCD4-WT), orange = deuterated 2dCD4-WT produced using cell-free protein expression (CFPE), purple = h-2dCD4-WT from CFPE and green = *B. choshinensis* expressed and secreted h-2dCD4-WT. The absorbance of the reaction product between the secondary antibody conjugated horse radish peroxidase enzyme and the TMB substrate at 450 nm is recorded as a function of the 2dCD4 concentration. All proteins are functional shown by their 2dCD4-WT concentration dependence on the absorbance at 450 nm.

The 17b ELISA was used to determine whether the gp120 binding site of the CD4 variants was intact with an indirect approach. CD4 binding to gp120 should induce a conformational

	Ibalizumab	17b
No CD4	0.065	0.038
No primary antibody	0.043	0.039
No lectin	0.076	0.035

TABLE 4.1: No CD4, no primary antibody and no lectin control values for the ibalizumab and 17b enzyme-linked immunosorbant assays.

change resulting in exposure of the CD4 induced site, detected by the 17b monoclonal antibody. Figure 4.1b shows the average absorbance at 450 nm for duplicate samples as a function of 2dCD4-WT analogue concentration. The CFPE H-WT, D-WT and *B. choshinensis* WT variants (orange, purple and green, respectively) exhibit very similar CD4 concentration dependence on the absorbance at 450 nm (fig. 4.1b) and therefore the binding of 17b to the CD4i epitope. Since both the 17b and gp120 concentrations were constant it can be said that the higher the CD4 concentration, the higher the absorbance at 450 nm as more CD4i sites became available for 17b binding. Measurements were also taken to control for the CD4 variants alone, gp120 alone, the lectin coating and the primary antibody.

While it appears that the 2dCD4-WT isoform produced in *E. coli* is slightly less active in both ELISAs, the samples were only prepared in duplicate and therefore some caution is required in extrapolating a correlation between the real binding efficiency of each 2dCD4-WT analogue. It would not be appropriate to try to quantitatively analyse the difference in activity between the various active 2dCD4 isoforms using statistics, given the qualitative way in which the ELISAs were conducted i.e. with few repeats.

Table 4.1 lists the absorbance values for the control measurements without CD4, lectin and primary antibody (for both the 17b and ibalizumab ELISA) all of which are considerably lower than the absorbance values for the lowest CD4 concentration and can therefore be considered as background absorbance values.

#### 4.4.2 Pseudo-viral neutralisation assay

A pseudo-viral neutralisation assay was used to test the ability of the 2dCD4-WT analogues to bind to viral Env, thereby inhibiting binding of native CD4 on TZM-bl cells. Five different viruses were used: 6535, ZM53, DU156, SF162 and, a non-HIV control, VSV-G.

As expected, the 2dCD4-WT analogues display weaker inhibition against the tier 2 viruses (DU156 and ZM53), as shown both by their IC<sub>50</sub> titers (table 4.2) and the neutralisation curves (fig. 4.2b and c, respectively). This show that a higher concentration of CD4 is required to neutralise 50% of the virus. For the ZM53 neutralisation assay all CD4 variants display similar CD4 concentration dependent inhibition of the virus. However for the DU156 neutralisation assay the *E. coli* and *B. choshinensis* 2dCD4-WT variants appear to be slightly less effective at neutralising the virus.

	Tier 1		Tier 2		Control
	6535	SF162	ZM53	DU156	VSV-G
<i>E. coli</i> 1dCD4-D1	0.110	0.560	0.390	0.650	nil
<i>E. coli</i> 2dCD4-WT	0.160	0.020	1.110	1.880	nil
CFPE d-2dCD4-WT	0.088	0.014	0.532	0.538	nil
CFPE h-2dCD4-WT	0.089	0.012	0.404	0.592	nil
<i>B. choshinensis</i> 2dCD4-WT	0.170	0.020	1.040	2.090	nil

TABLE 4.2: The concentration ( $\mu\text{M}$ ) at which 50% of the viruses were neutralised by protiated 2dCD4-WT produced in *E. coli*, *B. choshinensis*, CFPE and deuterated 2dCD4-WT produced using CFPE. *E. coli* expressed domain 1 of 2dCD4-WT was used as a control as the minimal binding domain for viral gp120 and VSV-G was used as a non-HIV control.

The 2dCD4-WT analogues display stronger inhibition against the tier 1 viruses (SF162 and 6535) (fig. 4.2d and 4.2e, respectively), which was also expected. Compared to the tier 2 viruses, a lower concentration of 2dCD4-WT is required to neutralise the virus (100% inhibition). This can be observed in the neutralisation curves (fig. 4.2) as the curves plateau to 100% inhibition at lower CD4 concentrations. There is an anomalous point in the *E. coli* 2dCD4-D1 curve for the SF162 virus (fig. 4.2d) at a concentration of approximately 0.3  $\mu\text{M}$  at which 0% inhibition was observed. Since the rest of the 2dCD4-D1 curve follows the same trend line, it is likely that this anomaly was due to pipetting error. The wells for this 2dCD4-D1 concentration may have been missed when adding one of the reagents.

The VSV-G control virus shows residual neutralisation effects at higher CD4 concentrations (fig. 4.2a). The IC<sub>50</sub> values have not been extrapolated from the neutralisation data, had they been, the values would be much higher than the highest concentration of the CD4 variants used as not even 50% neutralisation was achieved with this non-HIV control. The small residual "neutralisation" could be as a result of leaky expression of the luciferase gene.

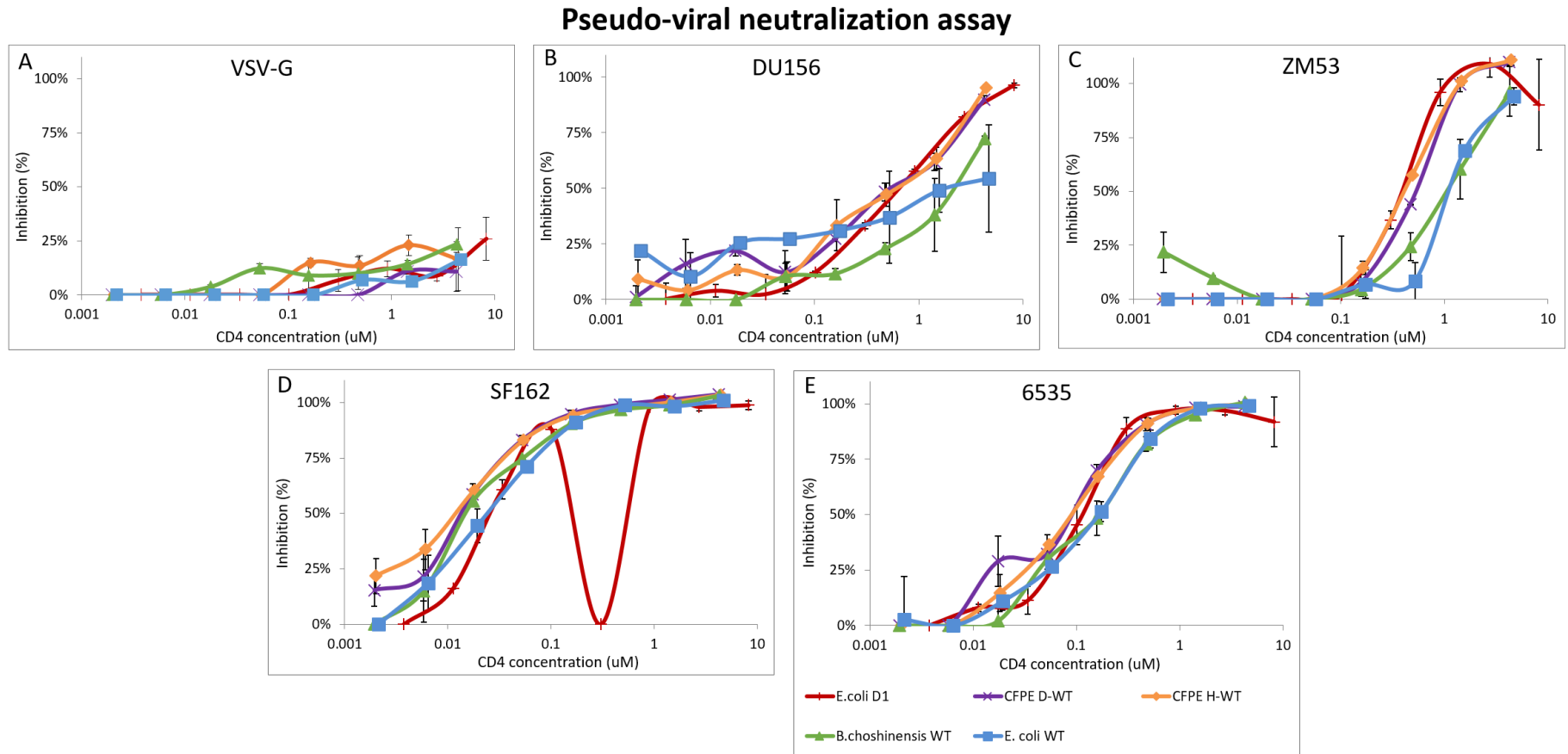


FIGURE 4.2: Pseudo-viral neutralisation assay was used to test the functionality of the recombinant 2dCD4-WT analogues produced in different expression systems. Domain 1 of CD4 was used as a control (red line) as it is the minimal binding domain of viral gp120. Protiated and deuterated CFPE 2dCD4-WT are shown in orange and purple, respectively. *B.choshinensis* expressed 2dCD4-WT is shown in green and *E. coli* expressed 2dCD4-WT, in blue. CD4 concentration in  $\mu\text{M}$  was plotted against percentage inhibition of a series of viruses: A. Control non-HIV virus VSV-G, B & C. Tier 2 viruses DU156 and ZM53, respectively. D & E. Tier 1 viruses SF162 and 6535. All analogues are shown to be functional.

### 4.4.3 One dimensional proton nuclear magnetic resonance (1D NMR)

1D NMR analysis was carried out to structurally assess the folding of 2dCD4-WT expressed in the cell-free system. From the 1D NMR spectra in figure 4.3 important chemical shift dispersions can be observed in both the methyl proton region (0 to 1ppm) and the amide proton region (8 to 10ppm) which are characteristic of a correctly folded protein. If 2dCD4-WT was not correctly folded by CFPE, the 1D NMR spectra would have fewer features in the methyl and amide proton regions. However, this is not the case as there are clear features in both proton regions for 2dCD4-WT.

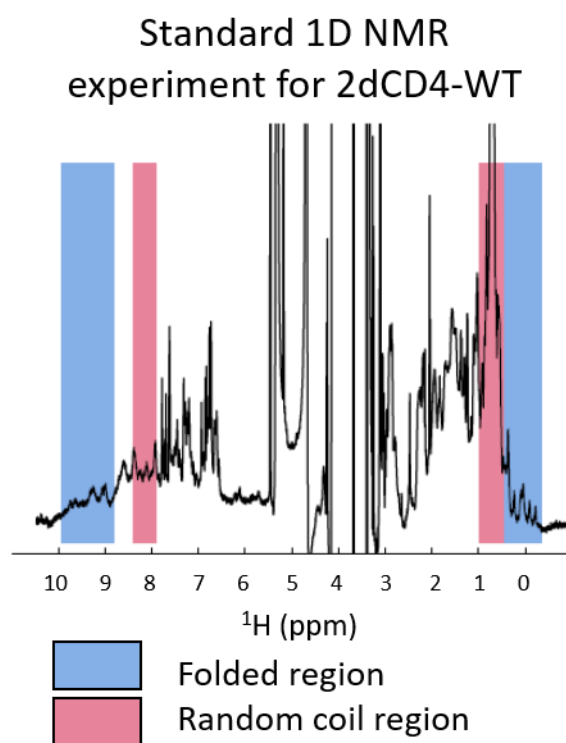


FIGURE 4.3: 2dCD4-WT produced using CFPE was subject to a standard 1D NMR experiment to assess its folding. Important chemical shift dispersions can be seen in the methyl and amide proton regions (0 to 1ppm and 8 to 10ppm, respectively), indicative of a well-folded protein.

Additionally, 1D HETSOFAST analysis was carried out whereby the "compactness" of the protein is assessed. From the 1D HETSOFAST spectra in figure 4.4, the derived  $\gamma_{NOE}$  parameter is calculated to be 0.40, revealing a dense proton network which corresponds to a correctly folded protein conformation with a few flexible loops. According to the crystal structure [66, 67] these flexible loops could be the C'C" loop of domain 1 and the AB and EF loops of domain 2.



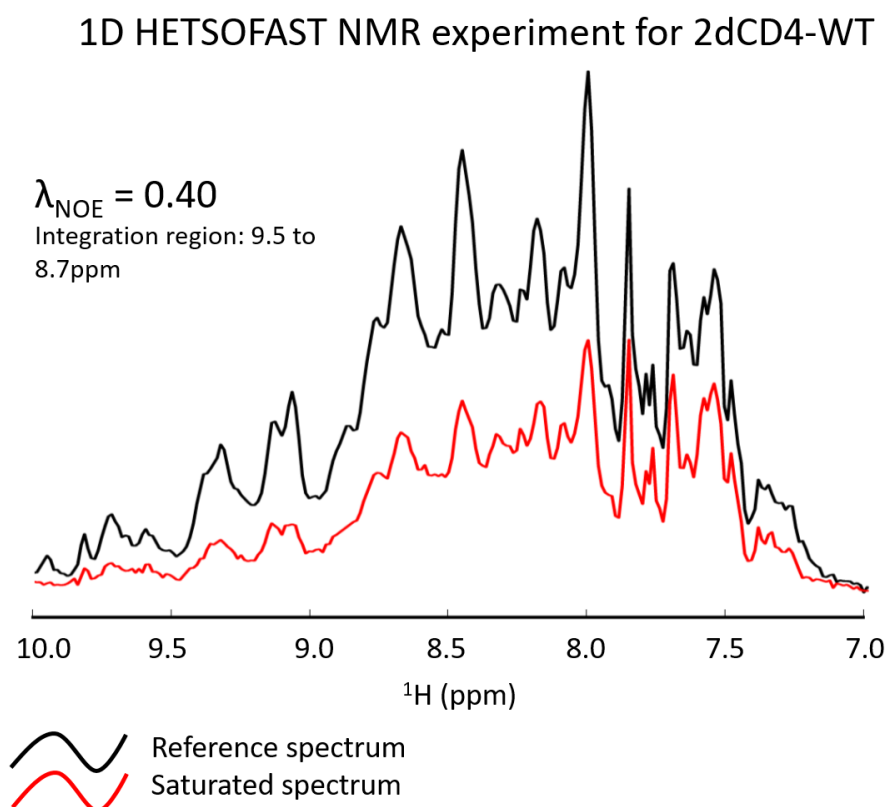


FIGURE 4.4: A 1D HETSOFAST experiment was carried out on 2dCD4-WT produced using CFPE to determine the compactness of the protein. The  $\lambda_{\text{NOE}}$  value of 0.40 is indicative of a well-folded protein with a few flexible loops.

#### 4.4.4 Mass spectrometry analysis of 2dCD4-WT

The expected molecular masses (as determined from the amino acid sequence by the ProtParam tool available online from ExPASy [<http://web.expasy.org/protparam/>]) versus the experimentally determined molecular masses of the 2dCD4-WT analogues produced in the different expression systems are listed in table 4.3.

The molecular weight of the *E. coli* expressed and refolded 2dCD4-WT in the presence and absence of 50 mM DTT are within the correct range for the fully oxidised and fully reduced redox isomers at 21 296 vs 21 927 Da and 21 300 vs 21 299 Da for the expected and experimental molecular mass in the absence and presence of DTT, respectively. There also appears to be a dimer species present in both samples with and without the presence of 50 mM DTT which is subject to the same reducing pattern. Interestingly, the DTT untreated sample has a molecular mass 1 Da greater than that of the expected molecular mass for the fully oxidised isoform and

Expression system	Conditions	Expected MW (Da)	Experimental MW (Da)
<i>E. coli</i>	protiated -DTT	21 296	21 297 & 42 593
	protiated +DTT	21 300	21 299 & 42 597
<i>B. choshinensis</i>	protiated -DTT	23 353	23 354
	protiated +DTT	23 357	23 358
Cell-free	protiated -DTT	22 687	57 197 & 22 697
	protiated +DTT	22 691	22 569 & 22 700 & 22 727
	deuterated -DTT	22 687 - 24 298	23 578 & 16 660
	deuterated +DTT	22 691 - 24 302	-

TABLE 4.3: Expected molecular masses for the 2dCD4-WT analogues was determined using the ExPASy ProtParam online tool. The expected molecular mass for the deuterated 2dCD4-WT is between that of the unlabelled analogue and the expected perdeuterated mass in which all of the protiums are replaced by deuterium. The samples were tested with and without 50 mM DTT to control for the contribution of the additional H to the molecular mass in the reduced form.

the DTT treated sample has a molecular mass 1 Da less than that of the expected molecular mass for the fully oxidised isoform. This might suggest that one disulphide bond is oxidised and that only one of the thiols in the other cysteine pair is reduced. However since this is chemically very unlikely it is maybe an average of two forms and it is difficult to determine the exact redox state of the sample. The three redox isomers are present on SDS-PAGE in the *E. coli* expressed 2dCD4-WT so it may be as a result of smearing or averaging of the signal. In any case, the addition of 50 mM DTT has a reducing effect on the protein. The same pattern is observed in the dimeric species.

For the *B. choshinensis* expressed and secreted 2dCD4-WT, there is a similar pattern observed as for the *E. coli* sample as the addition of DTT is causing a reduction event to occur. There is a difference of 1 Da between the expected and experimental molecular masses, but a difference of 4 Da between the DTT treated and un-treated samples, suggesting a shift from a completely oxidised 2dCD4-WT to a completely reduced 2dCD4-WT. On SDS-PAGE the *B. choshinensis* expressed 2dCD4-WT does not display the three bands corresponding to the fully oxidised, partially reduced and fully reduced 2dCD4-WT bands (chapter 3), suggesting that the protein is expressed in its fully oxidised form.

The situation is less clear for the cell-free produced 2dCD4-WT. Looking first at the protiated 2dCD4-WT, the molecular weight of the DTT untreated sample is 10 Da higher than that predicted by the ProtParam programme. This difference is too large to be accounted for by  $^1\text{H}$  and too small to be accounted for by a missing amino acid. What is more likely is that there is an amino acid substitution somewhere in the sequence of the protein. For example the difference

in the molecular weight between Pro and Ser is 10 Da. Upon reduction by 50 mM DTT the species increases in molecular mass by 3 Da corresponding to three protons. Since the cell-free produced 2dCD4-WT displays the triple banding pattern observed by SDS-PAGE (chapter 3), this 3 Da increase may be due to a smearing of the peak due to the presence of partially reduced and fully reduced isoforms. In addition, there appears to be a higher molecular weight species present at 57 197 Da which is too large to be a dimer and too small to be a trimer. Interestingly, when treated with 50 mM DTT the 57 197 Da species disappears which suggests that it may be a higher oligomeric structure associated through CD4's disulphide bonds. In its place there are the appearance of 22 727 Da and 22 569 Da species.

The experimental molecular mass obtained for the deuterated 2dCD4-WT produced in the cell-free system in the absence of DTT was determined to be 23 578 Da. However, a second smaller species at 16 660 Da was also identified. This second species could be a truncation product produced during the cell-free reaction. With the addition of 50 mM DTT, the mass spectrometry yielded no clear results. The raw mass-spectra are presented in appendix C.2.

#### Percentage deuteration of 2dCD4-WT produced using CFPE

Following the mass spectrometry results for the protiated and deuterated 2dCD4-WT produced by the cell-free system, the data given in appendix C.2 was used to calculate the percentage deuteration from the mass-spectrometry results. This was done by first calculating the number of non-labile H by multiplying the number of non-labile H per amino acid present in the protein e.g. 2dCD4-WT has six alanine residues and per alanine residue four of the hydrogens are non-labile. This was done for all twenty amino acids and the result was summed. For 2dCD4-WT expressed in the cell-free system there are 1 214 non-labile H. The repeat calculation was carried out for the number of labile H of which there are 399 in 2dCD4-WT expressed in the cell-free system. The protiated mass, the non-labile hydrogen deuterated mass and the per-deuterated mass are therefore calculated to be 22 690 Da, 23 904 Da and 24 302 Da, respectively.

The percentage deuteration of 2dCD4-WT as calculated by mass-spectrometry realised in H<sub>2</sub>O is thus 73% and is determined by the following formula:

$$\left[ 1 - N_{non-exH} - \frac{MM_{H_2O} - {}^1HMM}{N_{non-exH}} \right] \times 100 \quad (4.1)$$

Where  $N_{non-exH}$  is the number of non-labile H,  $MM_{H_2O}$  is the experimental molecular mass realised in  $H_2O$  and  ${}^1HMM$  is the protiated molecular mass.

#### 4.4.5 N-terminal protein sequencing of 2dCD4-WT

N-terminal protein sequencing was used to determine whether the truncated form identified by mass-spectrometry at 22 569 Da was as a function of the methionine loss. There were two sequences found: MAHMKKVV and AHMKKVVL of which both are in accordance with the 2dCD4-WT sequence and the second of which indicates that the methionine, coded for by the start codon at the beginning of the sequence, is missing.

#### 4.4.6 Size-exclusion chromatography coupled to multiple-angle laser light scattering and refractive index (SEC-MALLS-RI)

SEC-MALLS-RI was used as a quality control step to determine the molecular mass, polydispersity and stability after freezing of the deuterated 2dCD4-WT produced in the cell-free system. The SEC step prior to analysis by the UV, MALLS and RI detectors allows separation of particles in the sample solution, by size. The subsequent UV absorbance at 280 nm allows the elution of the deuterated 2dCD4-WT to be followed, while the light scattering can provide information about the size of the particle as well as non-protein contaminations that would not be picked up in the UV 280 nm trace. In addition, the refractive index provides information on the polydispersity of the sample.

The light scattering spectrum of the d-2dCD4-WT prior to freezing (fig. 4.5a, green curve) shows a large peak which elutes early, suggesting the particle is large in size. The UV and refractive index peaks (grey and blue curves, respectively) in this region are small compared to the main 2dCD4-WT peak at 11 mL. This suggests that this large species eluting early is probably a small amount of protein aggregate as large particles scatter skew the scattering of light disproportionately compared smaller particles. The large refractive index peak eluting at around 17 mL (blue) is likely to correspond to the sucrose in the buffer, which has a large refractive index.

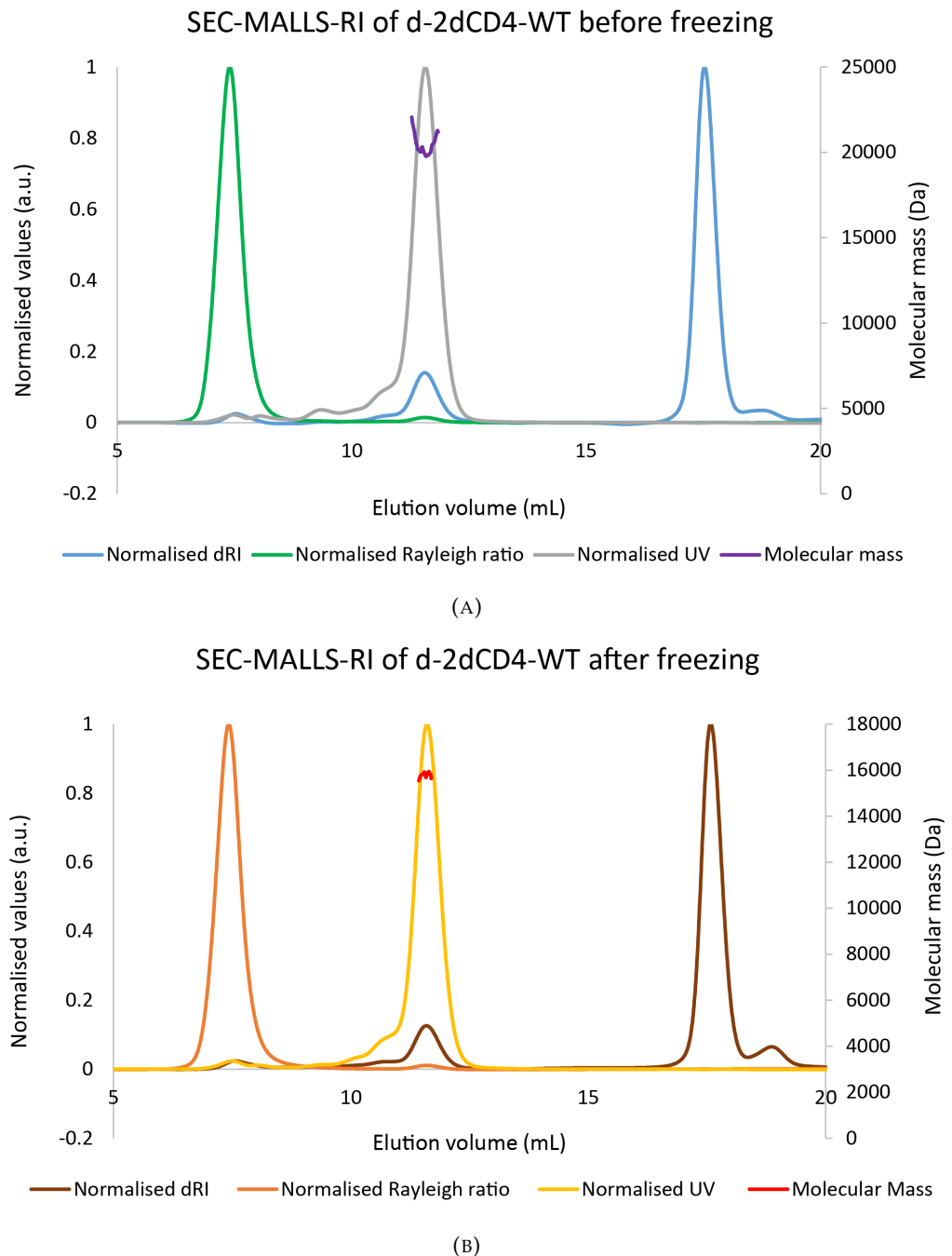


FIGURE 4.5: d-2dCD4-WT was tested using SEC-MALLS-RI before (A) and after (B) flash freezing in liquid nitrogen and storage at  $-80^{\circ}\text{C}$ . An observed molecular mass (purple curve) of 20 370 Da ( $\pm 1.7\%$ ) was determined from the refractive index for the sample before freezing (blue curve). After freezing an observed molecular mass (red curve) of 15 800 Da ( $\pm 2.4\%$ ) was determined from the refractive index (brown curve). The light scattering for the before (green curve) and after (orange curve) freezing indicate the presence of a small amount of large particles. The UV absorbance at 280 nm allowed the elution of d-2dCD4-WT before (grey) and after (yellow curve) freezing to be followed.

Analysis of the main protein peak at 11mL elution volume gives a molecular mass of 20 370 Da ( $\pm 1.7\%$ ) which including the  $\pm 5\%$  error of the machine gives a molecular mass close to but not

exactly the 23 578 Da expected for the deuterated 2dCD4-WT species. The observed molecular mass is shown across the peak in purple.

The SEC-MALLS-RI experiment was repeated on a sample from the sample purification of d-2dCD4-WT that had been flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  for 1 hour to test the freezing capacity of the protein for storage purposes. The UV, light scattering and RI spectra (fig. 4.5 (b), yellow, orange and brown curves, respectively) look very similar to that of the sample which has not been frozen. There is still the presence of a small amount of aggregation; however the molecular weight determined from the UV peak has a significantly lower molecular weight of 15 800 Da (+2.4%) (red line across the peak) compared to the 23 578 Da measured by mass spectrometry for the full length, deuterated 2dCD4-WT. Figure 4.6 depicts a zoomed-in view of the main d-2dCD4-WT peaks before and after freezing which shows that the UV peak for the d-2dCD4-WT after freezing (yellow) is slightly displaced to the right with respect to the UV peak for the d-2dCD4-WT before freezing (grey).

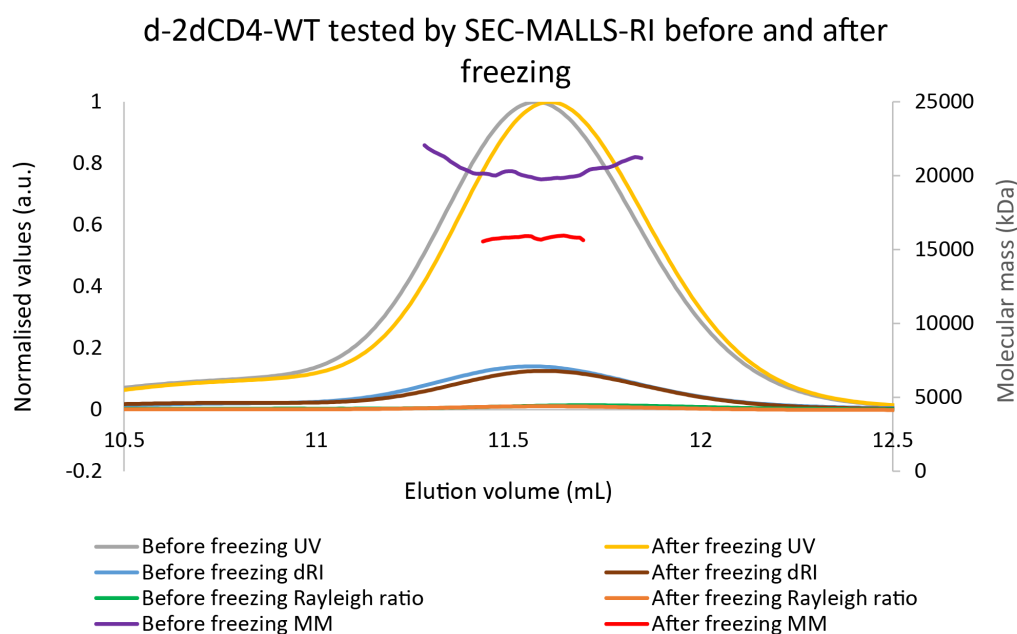


FIGURE 4.6: A zoomed-in view of the overlaid d-2dCD4-WT peaks. The after freezing (yellow) UV peak is slightly displaced to the right with respect to the before freezing (grey) UV peak which suggests that the protein is smaller. This is reflected in the molecular mass as the d-2dCD4-WT before freezing (purple line) has a higher molecular mass than d-2dCD4-WT after freezing (red line).

The mass spectrometry analysis of deuterated 2dCD4-WT also showed that there was a truncated form produced which has a molecular weight of 16 660 Da, which is in agreement with

the molecular mass determined by SEC-MALLS-RI. The mass spectrometry samples had previously been flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . This suggests that d-2dCD4-WT does not store well as a frozen sample.

## 4.5 Discussion

2dCD4-WT was produced using *E. coli*, *B. choshinensis* and cell-free protein expression systems as described in chapter 3. The yields of 2dCD4-WT produced in the *in vivo* bacterial expression systems were insufficient for the small-angle scattering studies to be carried out (chapter 6). This chapter is therefore restricted to the results of biochemical and biophysical analyses conducted on the protiated 2dCD4-WT (h-2dCD4-WT) and deuterated 2dCD4-WT (d-2dCD4-WT) produced from cell-free synthesis and where possible compared these to h-2dCD4-WT expressed using bacterial expression systems.

### 4.5.1 2dCD4-WT produced using cell-free synthesis is functional and correctly folded

In order to provide an extra layer of validity to the physiological conclusions drawn from structural data it was necessary to assess the functionality, and folding, of the purified recombinant proteins used in the study. If the protein in question is not functional or improperly folded, then the meaningfulness of the structural conclusions drawn are questionable. This becomes particularly important when using low resolution structural techniques such as small angle X-ray and neutron scattering where there are risks in interpretation. To this end, ELISA and pseudo-viral neutralisation assays were used to demonstrate that the purified, recombinant 2dCD4 analogues produced using the different expression systems are functional. Additionally 1D-NMR was used to measure the extent of folding of cell-free produced 2dCD4-WT.

The ELISA results show that all 2dCD4-WT analogues are functional as demonstrated by the CD4 concentration dependence on absorbance at 450 nm. For the ibalizumab ELISA this implies that as the concentration of 2dCD4 increases more ibalizumab epitopes become available, increasing ibalizumab binding, as one would expect. This results in more of the secondary antibody being able to bind as there is a higher concentration of ibalizumab present and therefore there is greater amplification of the reporter signal by the TMB substrate reaction with the

HRP conjugated to the secondary antibody. The 17b ELISA indirectly assesses the functionality of CD4 because the antibody binds to the CD4 induced epitope on gp120, and not the CD4 itself, however the implications for the avidity of the reporter signal are the same as for the ibalizumab ELISA.

Whilst it appears that the 2dCD4-WT isoform produced in *E. coli* is slightly less functional in both ELISAs, the samples were only prepared in duplicate and therefore the sample size is not large enough to draw accurate conclusions from statistics which could be employed with a larger sample pool. It would not be appropriate to try to quantitatively analyse the difference in activity between the various, active, 2dCD4 isoforms due to the qualitative nature with which the ELISAs were conducted (see 4.5.3 for how this experiment could be developed).

Similarly to the ELISA data the pseudo-viral neutralisation assay shows that all 2dCD4 analogues are functional as indicated by the CD4 concentration dependence on the absorbance at 450 nm. As the CD4 concentration increases there is a decrease in the reporter signal. As expected, the tier 1 viruses: 6535 and SF162 are more easily neutralised than the tier 2 viruses: ZM53 and DU156 shown both by the shift to the right in the curves from the tier 1 viruses to the tier 2 viruses and thus the resulting IC<sub>50</sub> values which demonstrate that a higher concentration of CD4 is necessary to neutralise 50% of the virus. As with the ELISA data it is inappropriate to attempt to draw statistical significance from the data as the sample size is too small and the assays were conducted in a qualitative rather than quantitative manner (see 4.5.3 for development of this experiment).

Due to the sample requirements for 1D-NMR (600  $\mu$ L of  $\geq 100 \mu$ M) it was not possible to test the *E. coli* and *B. choshinensis* expressed 2dCD4-WT using this technique. The classical 1D-NMR experiment showed that the h-2dCD4-WT produced important chemical shifts in both the methyl and amide proton regions which are characteristic of a well-folded protein. Moreover, the 1D HETSOFAST experiment yield a  $\lambda_{NOE}$  value of 0.4 which is indicative of a well-folded, compact protein with a few flexible loops which are possibly attributable to the C'C" loop in D1 and the AB, EF loops in D2 of 2dCD4-WT.

The ELISA and pseudo-viral assays are essentially measures of folding as well as the 1D-NMR technique, as the ibalizumab and 17b antibodies rely on structural epitopes at the D1D2 interface of CD4 and the CD4 induced epitope of gp120, respectively. Taken with the 1D-NMR



results, it can be concluded that the h-2dCD4-WT and d-2dCD4-WT analogues are folded and functional.

#### 4.5.2 Truncation products are produced using cell-free synthesis

Whilst the functional and folding studies of 2dCD4-WT gave a clear picture of the functionality and compactness of 2dCD4-WT. The SEC-MALLS-RI, mass-spectrometry and N-terminal sequencing imply that the cell-free system produced truncated forms of 2dCD4-WT along with a correctly sized 2dCD4-WT. Due to the sample requirements of SEC-MALLS-RI ( $\geq 2$  mg/mL) it was not possible to carry out experiments on the *E. coli* and *B. choshinensis* expressed 2dCD4-WT using this technique.

The mass-spectrometry results for the 2dCD4-WT analogues produced by bacterial expression systems in the absence of DTT yielded molecular masses which corresponded to the expected values as determined using the online ExPASy programme. With the addition of DTT, reduction events can be observed for both analogues. The results are suggestive of partial reduction for the *E. coli* analogue which may be caused by the three redoximers which are present in this sample as seen by SDS-PAGE (chapter 3), introducing a broadening of the peak and a smearing of the signal. This is not seen for the *B. choshinensis* expressed 2dCD4-WT suggesting that the protein is expressed in its completely oxidised form and completely reduced by the addition of 50 mM DTT. The *E. coli* expressed 2dCD-4WT presents a dimeric species in addition to the monomeric species in both the presence and absence of 50 mM DTT.

A reduction event similar to that observed for the *E. coli* expressed 2dCD4-WT is seen for the protiated analogue produced using CFPE. However the unreduced sample also contains a 57 kDa species which does not correspond to a dimer nor a trimer. This 57 kDa species could be a non-CD4 contaminant or an oligomerisation of smaller truncated forms of the protein. Upon reduction the higher molecular weight species disappears but two additional species close to the expected molecular mass of 2dCD4-WT appear. One of these species looked as if it could be caused by loss of the N-terminal methionine.

To test this hypothesis, N-terminal sequencing of the first 6 N-terminal amino acids was carried out which demonstrated that there were indeed two N-terminal sequences of which one was missing the first methionine residue. This loss of methionine could be explained by the action

of the methionine aminopeptidase enzyme (MetAP) which performs the hydrolytic cleavage of N-terminal methionine from newly synthesised polypeptides. MetAP1, present in eubacteria such as *E. coli*, performs this hydrolytic cleavage co-translationally as the newly synthesised polypeptide is produced from the ribosome. In addition, the 2dCD4-WT sequence fulfills the sequence requirement of MetAP1 in which the P1' position should be an amino acid residue with a short R group such as alanine [194, 195]. *E. coli* derived MetAP1 enzyme which may have been present in the CFPE lysate preparation since the cell-free lysates were made from *E. coli* cultures.

Under non-reducing conditions there are two species present in the d-2dCD4-WT sample, of which one corresponds to a deuterated full-length monomeric 2dCD4-WT species and the second appears to be a truncated form resolving at 16 660 Da. From the difference between the molecular mass between the protiated and deuterated forms, the percentage deuteration of the labile-H was calculated to be 73%. Proteins that are match-out labelled using the standard deuteration protocols developed by the Deuteration Laboratory (D-Lab) at the ILL's Life Sciences Group are approximately 75% deuterated at their labile-H positions and can be match-out in 100% D<sub>2</sub>O [171]. Therefore, 73% deuteration indicated that the match-out point of the CFPE produced d-2dCD4-WT was likely just under 100% (section 6.4.2). Upon reduction, there were no clear peaks (appendix C.2) from which to determine the molecular mass of the species present in the reduced d-2dCD4-WT sample.

The SEC-MALLS-RI experiment showed that there was a small amount of large particles eluting early from the column which was attributed to protein aggregates. The main d-2dCD4-WT peak of the sample before freezing presented a molecular mass under that expected for the deuterated protein. Whereas the main peak of the d-2dCD4-WT after flash freezing in liquid nitrogen and storage at -80°C presented a molecular mass of approximately 16 kDa which correlates with the mass-spectrometry result showing a truncation product at 16 kDa. The molecular mass of the d-2dCD4-WT before freezing may be of a slightly lower molecular mass than expected because the resolution of the column may not have allowed sufficient separation of full-length d-2dCD4-WT at 23 kDa and the 16 kDa truncated species or the 16 kDa species may be co-purifying with the 23 kDa species.

*E. coli* derived cell-free lysates have been shown to be susceptible to the translation of 3'-degraded mRNA transcripts, the release of incomplete translation products and accumulation of cleaved or partially degraded proteins [196] due to the lack of chaperone and degradation mechanisms. The range of molecular weight species of CFPE produced 2dCD4-WT would suggest that a number of truncation products had accumulated during the cell-free reaction, which may contribute to the aggregation seen in the SEC-MALLS-RI. Additionally, as the mass-spectrometry results were collected on a sample which had also been flash-frozen in liquid nitrogen, these results suggest that the protein cannot withstand freeze-thaw treatment.

#### 4.5.3 Future work

The ELISA and pseudo-viral neutralisation assays could be repeated with a larger sample size in order to be able to use statistical testing to draw meaningful conclusions about the specific activity of the different 2dCD4-WT analogues. By conducting the experiments in a more quantitative way it would be possible to determine which of the analogues has the highest specific activity and whether the difference between the activity of the analogues is significantly different or not. It would also be interesting to repeat this experiment using the 2dCD4-D2A mutant produced using the different expression systems. Since this isoform is the supposed gp120 binding protein it should display higher functionality than 2dCD4-WT, which is a mixture of the three redoximers, at the same concentration. In addition to assessing the function by ELISA and pseudo-viral neutralisation assay, future studies could involve determination of the affinity of the specific 2dCD4 isoforms for gp120. The determination of the  $K_d$  (dissociation constant) for each isoform could also provide information as to whether the domain 2 knockout (D2A) variant has a higher affinity for gp120 than the WT.

The mass-spectrometry experiment could be repeated using CFPE produced 2dCD4-WT that had not first been frozen to determine whether production of CFPE truncation products is truly aggravated by the freezing or freeze-thaw process. The protein could be sequenced entirely rather than just sequencing the 6 N-terminal residues to determine at what point in the sequence the truncation occurs at or whether there are many more truncation products. It would be useful to assess whether co-expression of chaperone proteins would reduce the production of truncated proteins, or if use of CFPE lysates derived from an *E. coli* strain such as BL21 Star™(DE3) would improve mRNA stability.

Finally, if higher yields of the *E. coli* and *B. choshinensis* analogues could be obtained, it would be interesting to see how the SEC-MALLS-RI and NMR experiments differed between the 2dCD4-WT analogues produced using the different expression systems. The instability of the 2dCD4-WT analogues produced by the bacterial expression systems is not rationalised by their performance in the assays to test the functionality; so it would be interesting to see whether the NMR yields a less folded structure compared to the CFPE produced protein and whether there are more aggregates by SEC-MALLS-RI.

## 4.6 Conclusions

Cell-free protein synthesis produces functional protiated and deuterated 2dCD4-WT as determined by ELISA, pseudo-viral neutralisation assay and 1D NMR. However, mass-spectrometry and SEC-MALLS-RI results suggest that the protein is produced in a truncated 16.6 kDa form in addition to the full-length 2dCD4-WT 22.6 kDa form which may be aggravated by freezing the protein for storage. Using mass-spectrometry the percentage deuteration achieved using CFPE was calculated to be 73% meaning that the match-out point for the d-2dCD4-WT is expected to be under 100% D<sub>2</sub>O.

## Chapter 5

# The effect of disulphide bond ablation on 2dCD4-WT structure

### 5.1 Abstract

Wild-type two domain CD4 is a redox active protein which resolves at different apparent molecular weights when analysed by non-reducing SDS-PAGE depending on its redox state. This chapter describes how the redox state of 2dCD4-WT affects its hydrodynamic volume by size-exclusion chromatography (SEC) and small-angle X-ray scattering (SAXS). By analysing SEC UV absorbance at 280 nm spectra, fully reduced 2dCD4-WT (2dCD4-R2) was found to have a larger hydrodynamic volume than the redoximer mixture of fully oxidised 2dCD4-WT (2dCD4-Ox) and partially reduced 2dCD4 in which the second domain disulphide bond is reduced (2dCD4-R1). Contrasting with this, 2dCD4-Ox was found to have a larger hydrodynamic volume than 2dCD4-R1 which alludes to a difference in the role between the structural, stabilising disulphide bond in domain 1, and the allosteric, destabilising disulphide bond in domain 2.

A shift in the distribution of the redoximers was observed by SDS-PAGE analysis of a DTT titration on 2dCD4-WT expressed by *E. coli* and refolded from inclusion bodies and cell-free protein expressed (CFPE) 2dCD4-WT. It was found that 2dCD4-WT produced using *E. coli* and CFPE systems gave different initial starting distributions of the redoximers. CFPE produced 2dCD4-WT seeming to favour the R1 redoximer with *E. coli* producing a more even distribution of 2dCD4-Ox and -R1.

The DTT titration and SEC approaches to the isolation of the 2dCD4-WT redoximers were then applied to a series of SAXS experiments to determine whether the shift in the SEC elution volume as a function of disulphide bond reduction of 2dCD4-WT, translates to a difference in the radius of gyration ( $R_g$ ) and maximum dimension ( $D_{max}$ ) as determined by SAXS.

## 5.2 Introduction

As described in Chapter 1, 2dCD4 has been shown to be a redox active protein due to its allosteric disulphide bond in domains 2. Cerutti *et al.* 2014[63] showed that the different redox isomers (redoximers) of 2dCD4-WT resolve at different apparent molecular weights as shown by non-reducing SDS-PAGE (fig. 5.1). They correlated the apparent molecular weights of the redox isoforms of 2dCD4-WT to the redox state of the disulphide bonds by generating a panel of 2dCD4 disulphide knockout variant proteins in which the cysteine residues of the disulphide bonds were substituted for alanine residues. In doing so they found that viral gp120 binds to both the 2dCD4-D1A and -D2A variants (domain 1 Cys to Ala and domain 2 Cys to Ala mutation) which mimic 2dCD4 reduced in its first and second domains, respectively (fig. 5.2). With the metastable allosteric disulphide bond in domain 2, physiologically, only 2dCD4 reduced in its second domain (rather than its first domain) is likely to exist. Table 5.1 gives a graphical explanation of the 2dCD4-WT redox isomers and the cysteine to alanine variant proteins which mimic these.

During recombinant expression, 2dCD4-WT is produced in a mixture of redox isoforms: fully reduced (2dCD4-R2), domain 2 reduced (2dCD4-R1) and fully oxidised (2dCD4-Ox). In order to be able to study the 2dCD4-R2 redoximer, which is the isoform uniquely binding the HIV-1 gp120 surface receptor, it is necessary to isolate it from the other isoforms of 2dCD4-WT. As the disulphide bonds in 2dCD4-WT are sensitive to treatment with DTT [63], an initial DTT titration was tested to assess the effect of DTT concentration on the distribution of the redoximers of *E. coli* expressed and refolded 2dCD4-WT. It was expected that as the DTT concentration increased that there would be a shift in the distribution of the redox isomers in favour of the 2dCD4-R1 and -R2 forms.

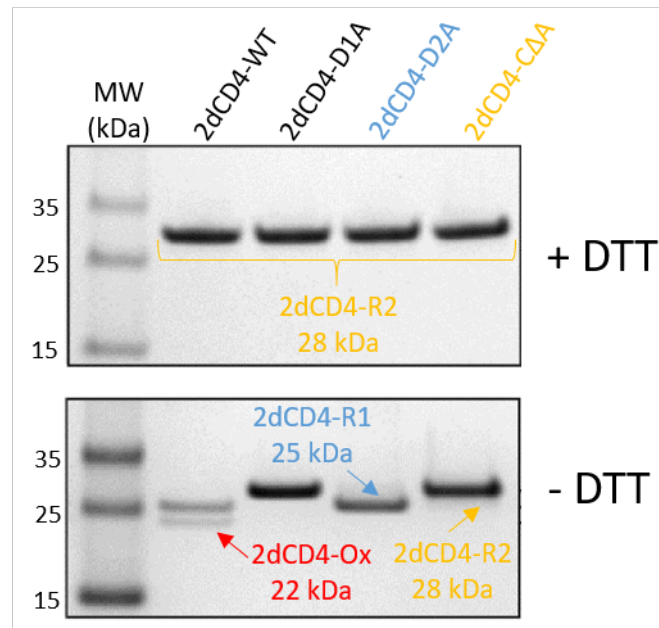


FIGURE 5.1: Reducing (top panel) and non-reducing (bottom panel) SDS-PAGE with Coomassie blue staining analysis of 2dCD4-WT, 2dCD4-D1A (domain 1 Cys to Ala substitution), 2dCD4-D2A (domain 2 Cys to Ala substitution) and 2dCD4-CΔA (domain 1 and 2 Cys to Ala substitution). Oxidised 2dCD4-WT (2dCD4-Ox, red) resolves to an apparent molecular weight of 22 kDa under non-reducing conditions and 28 kDa when fully reduced. 2dCD4-D1A mimics 2dCD4-WT reduced in domain 1 and resolves to an apparent molecular weight of 28 kDa both with and without 50 mM DTT treatment. 2dCD4-D2A mimics 2dCD4-WT reduced in domain 2 and resolves to an apparent molecular weight of 25 kDa under non-reducing conditions and 28 kDa when fully reduced. 2dCD4-CΔA mimics fully reduced 2dCD4-WT (2dCD4-R2, yellow) and resolves to 28 kDa. Therefore, the band at 25 kDa above that of 2dCD4-Ox in 2dCD4-WT under non-reducing conditions (bottom panel) can be attributed to 2dCD4-R1 (blue) since 2dCD4-D2A resolves to the same apparent molecular weight and not 2dCD4-D1A. Adapted from: [63]

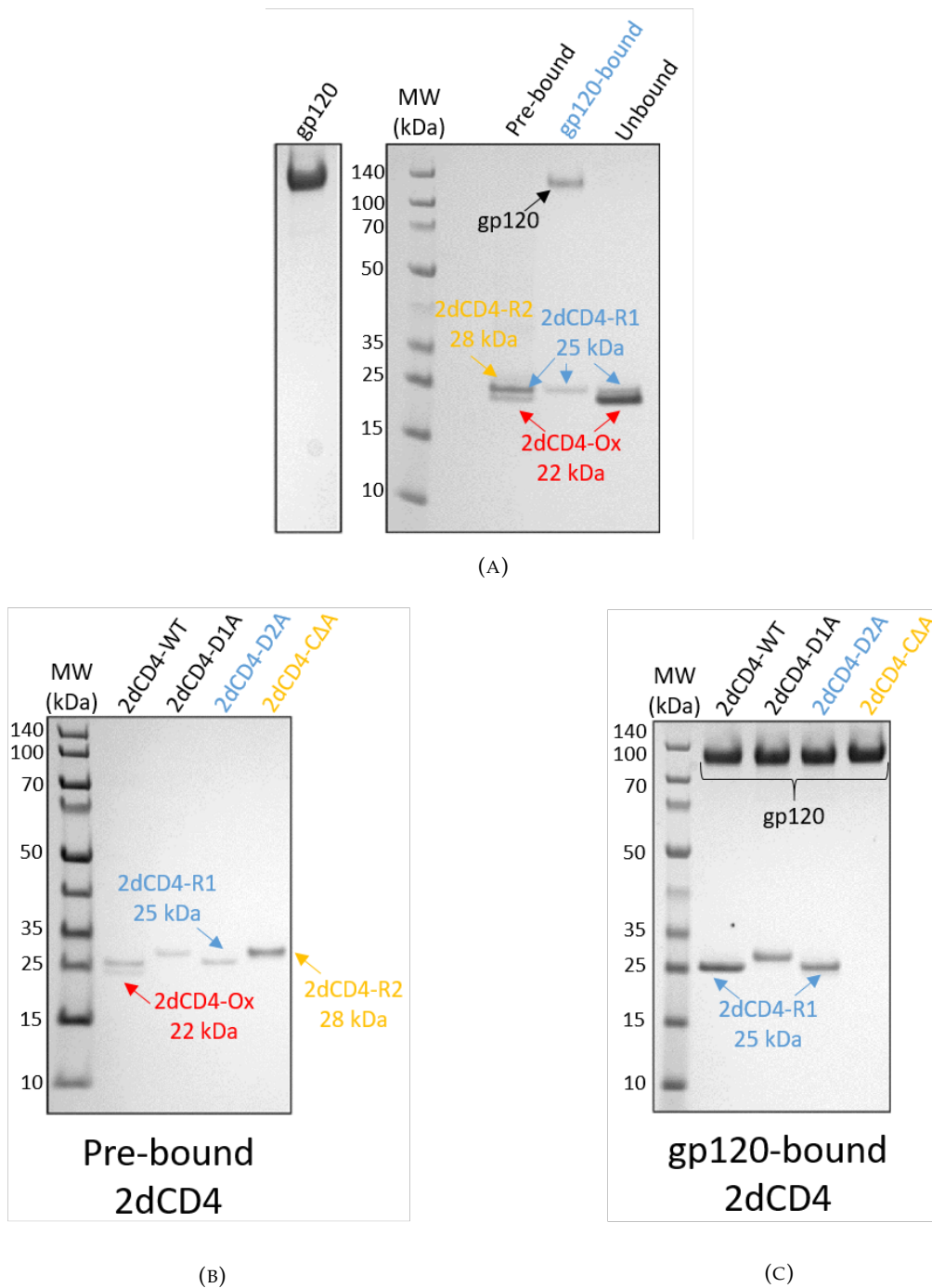


FIGURE 5.2: 2dCD4-WT and Cys to Ala variant protein binding to gp120 was analysed by SDS-PAGE. (A) gp120 pre-bound, bound and unbound 2dCD4-WT was analysed by non-reducing SDS-PAGE. In the pre-bound and unbound fractions 2dCD4-Ox and 2dCD4-R1 are observed. In the gp120-bound fraction only 2dCD4-R1 can be observed. 2dCD4-WT and the Cys to Ala variant proteins were analysed by SDS-PAGE before (B) and after (C) gp120 binding. gp120 binds to both 2dCD4-D1A and 2dCD4-D2A but only binds to 2dCD4-R1 in wild-type 2dCD4. Since 2dCD4-D1A (2dCD4 reduced in its first domain) is a non-physiological redox isoform due to the instability induced by reduction of the structural disulphide bond in domain 1, it was deduced that 2dCD4-WT binds uniquely to 2dCD4-R1 in which the second domain disulphide bond is reduced.

Adapted from: [63]



Function	2dCD4-WT redox isoform	2dCD4 disulphide bond knockout variant
Resting T-cell	2dCD4-Ox 	
Binds gp120	2dCD4-R1 	2dCD4-D2A 
Unphysiological	2dCD4-R2 	2dCD4-CΔA 
Binds gp120 (unphysiological)		2dCD4-D1A 

TABLE 5.1: 2dCD4-WT exists in three different redox forms: fully oxidised (2dCD4-Ox), reduced in domain 2 (2dCD4-R1) and fully reduced (2dCD4-R2). The disulphide bonds can be knocked out by substituting the Cys residues for Ala residues to produce a: domain 2 disulphide bond knockout (2dCD4-D2A), a domain 1 and 2 disulphide bond knockout (2dCD4-CΔA) and a domain 1 disulphide bond knockout (2dCD4-D1A). The table shows their corresponding 2dCD4-WT redox isoform equivalents and a schematic to help visualise the redox state of the Cysteines and the disulphide bond knockout variant. Domain 1 is shown in green and domain 2 is shown in blue.

### 5.2.1 Previous studies of 2dCD4-WT by SAXS

2dCD4-WT has previously been studied by SAXS in experiments carried out by Guttman *et al.* 2012 [72], Ashish *et al.* 2008 [118] and Ashish *et al.* 2006 [197]. Ashish *et al.* 2006 [197] measured 2dCD4-WT and gp120 alone and in complex to gain an understanding as to how the conformation of gp120 changes upon 2dCD4 binding. Their later study in 2008 involved looking at how the structure of 2dCD4 and 4dCD4 changes as a function of its ligand: gp120 and the tick saliva immunosuppressor Salp15. Guttman *et al.* 2012 [72] studied fully glycosylated gp120 and 2dCD4 in isolation and in complex to look for conformational changes in gp120. In each experiment the 2dCD4 was acquired from the NIH AIDS reagents programme, which may be testament to the difficulty in producing good quality 2dCD4 for SAXS, by the authors. The resulting size parameters from these experiments reported for 2dCD4 are listed in table 5.2.

Publication	$R_g$ ((Å) from Guinier)	$R_g$ ((Å) from P(r))	$D_{max}$ (Å)
Ashish et al. 2006 and 2008	22.3	–	75
Guttman et al. 2012	20.9	21.7	73

TABLE 5.2: Previously reported size parameters  $R_g$  and  $D_{max}$  for 2dCD4-WT as determined by SAXS.

In this study, SAXS data were collected from 2dCD4-WT produced using the CFPE to identify how reduction of the disulphide bonds in domain 2 affects the structure of the protein. It was also interesting to compare the size parameters obtained with those already published.

### 5.2.2 Experiments to test the effect of reduction on 2dCD4-WT structure

Since the different redoximers resolve at different apparent molecular weights during non-reducing SDS-PAGE, it was reasoned that it should be possible to separate the redox isomers of *E. coli* expressed 2dCD4 by size-exclusion chromatography. SDS-PAGE functions by denaturing proteins and separating them based on their mobility through a polyacrylamide matrix. SDS is used to provide a charge to the protein. The separation is thus based on their charge which corresponds directly to their mass. However, it has been shown that under non-reducing conditions, the presence of disulphide bonds can increase the mobility of a protein

through the matrix which is attributed to the protein having a smaller (denatured) hydrodynamic volume [198]. SEC also functions by separating proteins based on their mobility through a polymer matrix and thus hydrodynamic volume, although this time in their native configuration. The size-exclusion column contains pores which smaller proteins are able to enter, thus larger proteins elute first as they are not able to enter the pores. Applying these principles to the 2dCD4-WT redoximers it was expected that the 2dCD4-Ox form would have the smallest hydrodynamic volume.

Following the biochemical analyses carried out on *E. coli* expressed 2dCD4-WT, SAXS analyses were carried out on 2dCD4-WT produced using the cell-free protein expression system to assess the effect of DTT, and therefore reduction, on the size parameters  $R_g$  and  $D_{max}$  of 2dCD4-WT. CFPE produced 2dCD4-WT was used for the SAXS analyses as higher yields were obtainable using this system, thus allowing for structural analyses.

In a first experiment, an initial concentration series of 2dCD4-WT in the absence of DTT was tested. This would allow limited sample supply to be conserved for subsequent experiments by avoiding repeating a concentration series (necessary for SAXS analysis to extrapolate to zero concentration) for each DTT concentration of the DTT titration experiment. The optimal concentration selected was a compromise between signal-to-noise ratio and mitigating inter-particle effects. After which, the DTT titration was carried out using the robotic sample changer on BM29 at the ESRF and it was predicted that with an increase in the DTT concentration, the size parameters ( $R_g$  and  $D_{max}$ ) would increase.

Online SEC-SAXS was carried out on a 2dCD4-WT sample, produced using the cell-free system, in the presence and absence of 50 mM DTT to see if the reduction by DTT caused a shift in the presence of the UV 280 nm spectra peaks and a change in the size parameters  $R_g$  and  $D_{max}$  as derived from the Guinier analysis and pair-distance distribution function. Finally, SEC-MALLS was used to assess the quality of the samples for aggregation, to observe how the position of the UV 280 nm and light scattering peak shifts with respect to the elution volume and how the molecular weight across the peaks changed upon reduction of 2dCD4-WT.

## 5.3 Materials and Methods

### 5.3.1 Recombinant protein expression and purification

2dCD4-WT for the DTT titration and SEC redoximer purification was expressed in *E. coli* strain BL21\*(DE3) (Invitrogen) and purified from insoluble exclusion bodies by a denaturing nickel IMAC protocol followed by a dialysis refolding protocol as detailed in Chapter 3 section 3.3.2.

2dCD4-WT for the SAXS DTT titration and redox SEC-SAXS was expressed and purified from the cell-free protein expression system as described in Chapter 3 section 3.3.6.

### 5.3.2 DTT titration

7.25 µg 2dCD4-WT was incubated with the following concentrations of DTT for 1 hour at room temperature: 0.1, 0.25, 0.5, 1.0, 5.0 and 10.0 mM with a 0 mM DTT control. Each sample was then incubated with 25 mM iodoacetamide for 30 minutes at room temperature. 4 X non-reducing LDS-loading buffer (Life Technologies) was added to each reaction mix and loaded onto a 12% bis-Tris SDS polyacrylamide gel (Life Technologies).

### 5.3.3 SEC Purification of 2dCD4-WT redoximers

#### 2dCD4-WT SEC Purification

A 120 ml preparative grade Superdex<sup>TM</sup> 75 16/600 column (GE Healthcare) was equilibrated in 1 X PBS and 2 mL of 1.4 mg/mL 2dCD4-WT was loaded onto the column and run at a flow rate of 0.25 mL/min. 1 mL fractions were collected for analysis on non-reducing SDS-PAGE.

#### Superdex 75 16/600 calibration

The Superdex<sup>TM</sup> 75 16/600 column was calibrated using the following protein standards (GE Healthcare): 3 mg/mL apoprotinin (6.5 kDa), 3 mg/mL ribonuclease (13.7 kDa), 3 mg/mL carbonic anhydrase (29 kDa), 4 mg/mL ovalbumin (43 kDa) and 3 mg/mL conalbumin (75 kDa).

The protein standard mix was prepared in 1 X PBS. 500  $\mu$ L was injected onto the column using a 1 mL loop and the UV 280nm trace was followed as the proteins eluted at a flow-rate of 1 mL/min.

#### **SDS-PAGE analysis of fractions**

The 2dCD4-WT elution fractions from across the peak of the size-exclusion spectrum were analysed by SDS-PAGE using a 12% Bis-Tris gel (Novex).

#### **5.3.4 SAXS analysis of the effect of DTT on the size of 2dCD4-WT**

SAXS was carried out on the small-angle scattering beamline BM29 dedicated to biological samples at the ESRF.

#### **Concentration series**

Protiated 2dCD4-WT was purified using the CFPE method described in Chapter 3 section 3.4.4. 2dCD4-WT was concentrated to 8.1 mg/mL and a dilution series was prepared in 50 mM Tris-HCl, 300 mM NaCl, 5% sucrose. The final concentrations measured were: 8.1 mg/mL, 3.9 mg/mL, 1.8 mg/mL, 1.0 mg/mL. 55  $\mu$ L of each sample was added to PCR tubes which were loaded into the robotic sample changer. 50  $\mu$ L sample was flowed through the capillary and exposed to the X-ray beam. 10 frames were measured at a rate of 1 s/frame.

#### **DTT titration**

The sample at 3.9 mg/mL produced data with a good compromise between signal to noise ratio and inter-particle effects (see section 5.4.3) and therefore this concentration was chosen to move forward with the DTT titration experiment. 247.5  $\mu$ g 2dCD4-WT was used for each DTT concentration. DTT was added to a final concentration of: 0 mM, 1 mM, 2.5 mM, 5 mM, 10 mM, 25 mM, 50 mM and the volume topped up to 55  $\mu$ L with 50 mM Tris-HCl, 300 mM NaCl, 5% sucrose (to increase the stability of the 2dCD4-WT). The samples were transferred to PCR tubes and loaded into the robotic sample changer. 50  $\mu$ L sample was flowed through the capillary

and exposed to the X-ray beam. 10 frames were measured at a rate of 1 s/frame, as described previously.

A small amount of sample was conserved to run on a 4-12% Bis-Tris SDS-PAGE gel. 18  $\mu\text{L}$  was loaded with 6  $\mu\text{L}$  4 X non-reducing loading buffer and 20  $\mu\text{L}$  was loaded onto the gel. The 2dCD4-WT was visualised by Coomassie blue staining.

### Redox SEC-SAXS

A Superdex<sup>TM</sup> 75 Increase 3.2/300 (GE Healthcare) was equilibrated in 50 mM Tris-HCl, 300 mM NaCl, 5% sucrose. 55  $\mu\text{L}$  (at 4.5 mg/mL) 2dCD4-WT was loaded onto the column and the protein eluted at a flow rate of 0.1 mL/min into the capillary. The SEC was run for 30 minutes and the frames were taken every second. The column was then re-equilibrated in the same buffer containing 50 mM DTT. 55  $\mu\text{L}$  4.5 mg/mL 2dCD4-WT which had been pre-treated with 50 mM DTT for 30 minutes was then loaded onto the column and the measuring process was repeated as previously stated.

### Treatment of the data

Initial data treatment was done using the program PRIMUS [199] of the ATSAS suite of software, whereby the multiple frames taken during the experiment per samples were averaged after inspection for radiation damage using the data comparison function. Frames with a p-value  $\leq 0.05$  were excluded from the averaging, as the difference was found to be statistically significant. The p-value refers to the probability of the null hypothesis (the SAS curves are the same) being true. At p-values of 0.05 or less the null-hypothesis is rejected and therefore the curves are considered to be significantly different. Had the p-value been larger than 0.05 then there would be less evidence to reject the null-hypothesis so the data would not be considered significantly different. Post-averaging, the buffer scattering curves were subtracted from those of the sample to yield the scattering arising from the protein samples, only. After subtraction, the resulting curves from the concentration series were merged. Merging of the curves allows the low-q region to be relieved of any inter-particle effects driven by high concentrations and to increase the signal to noise ratio at high-q where the signal is low and the noise is high at low concentrations.

Buffer subtraction of SEC-SAXS data was carried out using the SCATTER software [200]. The scattering intensity  $I(q)$  was plotted against the frame number which gives a similar looking profile to a SEC chromatogram. 50-100 buffer curves were selected from a flat section of the profile corresponding to buffer scattering. Auto- $R_g$  was used to plot the  $R_g$  across the scattering peak of the sample. Frames were selected where the  $R_g$  was constant, indicating the similarity in particle size.

After the initial data treatment, the PRIMUS [201] and GNOM [202] programmes were used to determine the Guinier derived  $I(0)$ ,  $R_g$  and the real space  $R_g$ , respectively.

### SEC-MALLS-RI

55  $\mu$ L 4.5 mg/mL 2dCD4-WT was measured by SEC-MALLS-RI in 50 mM Tris-HCl, 300 mM NaCl, 5% sucrose. The instrument set-up comprised: a pump (L2130, Elite LaChrom) connected to a Superdex 75 10/300 size-exclusion column (GE Healthcare), with a spectrophotometer (L2400, Elite LaChrom) to follow the absorbance at 280 nm, a multi-angle laser scattering device (Dawn Helios-u, Wyatt) and a refractometer (Optilab T-rEX, Wyatt). The software used for analysis was ASTRA6 (Wyatt Technology). 2dCD4-WT in 0 mM DTT was run at a flow rate of 0.5 mL/min and as it eluted off of the column the light scattering, refractive index and UV absorbance at 280 nm was recorded. The column was then equilibrated in the same buffer with the addition of 50 mM DTT. 55  $\mu$ L 4.5 mg/mL 2dCD4-WT treated with 50 mM DTT was loaded onto the column to repeat the experiment.

## 5.4 Results

### 5.4.1 DTT titration causes a shift in *E. coli* expressed and refolded 2dCD4-WT redoximer distribution

A dithiothreitol (DTT) titration was used to observe the influence of an increase in reducing agent concentration on the band intensity and therefore the isoform of 2dCD4-WT present. The SDS-PAGE analysis (fig. 5.3) shows that with a titration of increasing concentration of DTT there is a shift in band intensity from the 22 kDa band to the 25 kDa band and to the 28 kDa band. These bands correspond to the fully oxidised (OX, red arrow), partially reduced (R1,

blue arrow) and fully reduced isoforms (R2, yellow arrow), respectively. This suggests that a redox shuffling event has occurred during which a disulphide bond between a cysteine thiol pair within 2dCD4-WT is reduced in the presence of DTT. With higher concentrations of DTT it is expected that the 2dCD4-WT could be entirely reduced in which case only the 2dCD4-R2 form would be present and only one band at 28 kDa would be observed.

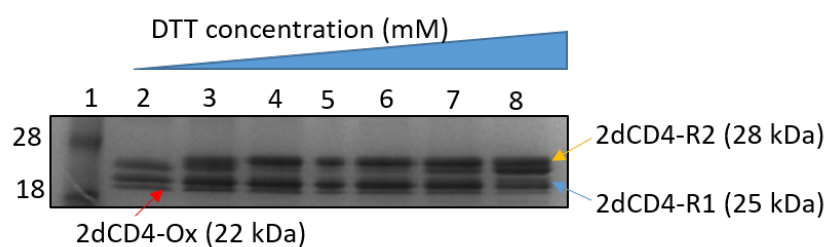


FIGURE 5.3: *E. coli* expressed 2dCD4-WT was treated with a 0, 0.1, 0.25, 0.5, 1.0, 5.0 and 10.0 mM DTT (lanes 2-8) to observe the effect of DTT on the distribution of redox isomers. As the DTT concentration increases there is a shift in the band intensity away from the 2dCD4-OX (red arrow) species in favour of the 2dCD4-R1 and 2dCD4-R2 species (blue and orange arrows, respectively). Lane 1 shows the molecular weight marker with the molecular weight in kDa.

#### 5.4.2 Partially reduced 2dCD4-WT has a smaller hydrodynamic volume than oxidised 2dCD4-WT

SEC was used to purify the redoximers of 2dCD4-WT according to their apparent molecular weight, as observed by non-reducing SDS PAGE. Figure 5.4a shows the size-exclusion chromatogram for the 2dCD4-WT redoximer mixture with the calibration curve overlaid and figure 5.4b shows the elution of two different 2dCD4-WT isoforms (red and blue boxes). The corresponding non-reducing SDS-PAGE (5.4c) analysis across these two peaks shows that the smaller peak (red box - eluting first) corresponds to a 2dCD4-WT isoform which resolves at 22 kDa whereas the larger peak (blue box - eluting second) corresponds to a species of 2dCD4-WT which resolves at 25 kDa.



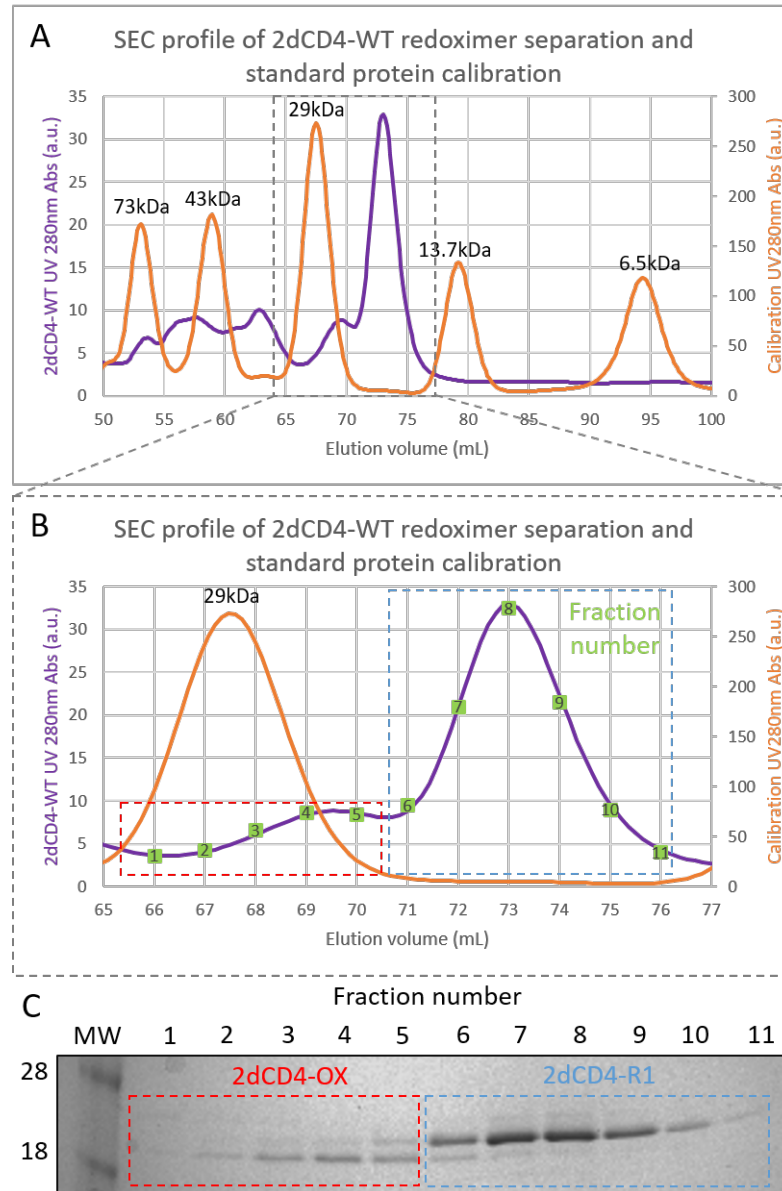


FIGURE 5.4: 2dCD4-WT was separated into its 2dCD4-Ox and 2dCD4-R1 forms by SEC. (A) 2dCD4-WT was purified using size-exclusion chromatography. (B) The S75 column was calibrated using known molecular weight protein standards and the two UV 280nm spectra were overlaid. There were two 2dCD4 redoximers present of which the one in the red box has a **larger native** hydrodynamic volume than the one in the blue box as it elutes from the column first. The green squares denote the positions across the peaks from which fractions were collected and run on the non-reducing SDS-PAGE in C. (C) The nature of the species and their apparent molecular weight was analysed by SDS-PAGE. 2dCD4-OX (red box) was found to have a **smaller unfolded** hydrodynamic volume than 2dCD4-R1 (blue box).

The elution position of the two 2dCD4 species is unexpected given their resolution by non-reducing SDS-PAGE for which one would expect the 25 kDa isoform to elute before the 22 kDa isoform. The S75 pg column was calibrated using standard globular proteins of known

molecular mass before running the 2dCD4-WT. The peak corresponding to the 22 kDa species resolves at a molecular weight lower than carbonic anhydrase at 29 kDa and therefore could not be attributed to a 2dCD4-WT dimer.

Both SEC and SDS-PAGE separate proteins based on their hydrodynamic volume and mobility through a polymer matrix. Whilst SDS-PAGE is a denaturing technique in which the tertiary structure of proteins is disrupted to form a random coil, SEC is a native technique in which the integrity of the protein structure is not compromised. If this is applied to the results observed here, it suggests that the partially reduced 2dCD4-WT has a smaller native hydrodynamic volume than that of the fully oxidised form as determined by SEC. Additionally, applying the same theory to the SDS-PAGE results this would suggest that the 25 kDa isoform has a larger denatured hydrodynamic volume than that of the 22 kDa isoform.

From both the SEC and the (importantly) non-reducing SDS-PAGE data, the results show that the integrity of the disulphide bonds is maintained suggesting that they have an impact on the hydrodynamic volume of the protein. It might be expected that disulphide bond reduction would release physical restraints on the random coil denatured protein so that it can adopt a more flexible, open structure. In line with this, the fully oxidised form has both disulphide bonds intact and therefore the denatured structure is more compact as the disulphide bonds physically hold the random coil in a more compact structure. Conversely the opposite pattern is observed for the native SEC technique. These results are summarised in table 5.3.

Technique	2dCD4-Ox $V_H$	2dCD4-R1 $V_H$
SEC (native)	Larger native	Smaller native
non-reducing SDS-PAGE (denaturing)	Smaller denatured	Larger denatured

TABLE 5.3: 2dCD4-Ox has a larger native hydrodynamic volume ( $V_H$ ) and a smaller denatured hydrodynamic volume than 2dCD4-R1 as determined by SEC and non-reducing SDS-PAGE, respectively.

### 5.4.3 Choosing a 2dCD4-WT concentration

The same experiments as carried out above were to be applied to SAXS experiments but in order to conserve precious 2dCD4-WT sample, an initial concentration series was carried out to determine which concentration of 2dCD4-WT resulted in an optimal signal-to-noise ratio, and having the least interference from inter-particle effects. Subsequently, 2dCD4-WT was

measured at: 1.0 , 1.8, 3.9 and 8.1 mg/mL using the robotic sample changer on BM29 at the ESRF.

### 1D scattering curves of a 2dCD4-WT concentration series

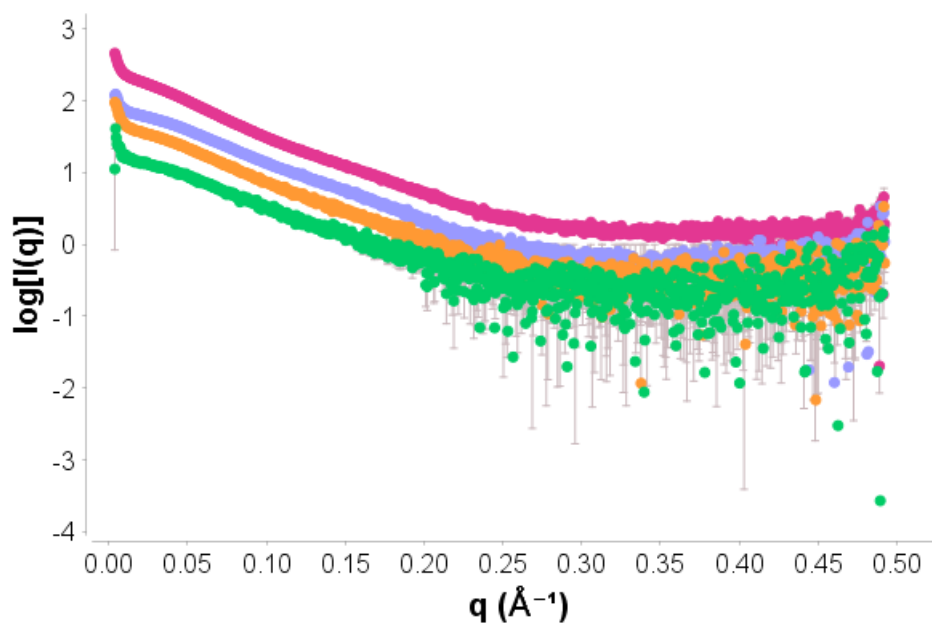


FIGURE 5.5: A concentration series of CFPE produced 2dCD4-WT was tested by SAXS to determine the optimal concentration to use for a DTT titration. At low- $q$  there is a sharp inflection of the curve towards  $x = 0$  which is characteristic of aggregation. The 3.9 mg/mL curve gave the best compromise between signal:noise and inter-particle effects. Green = 1.0 mg/mL, orange = 1.8 mg/mL, purple = 3.9 mg/mL, pink = 8.1 mg/mL.

The 1D scattering profile for the 2dCD4-WT concentration series after buffer subtraction is shown in figure 5.5. It is evident that the samples are all subject to some inter-particle effect, most likely to be aggregation considering the upward turn visible at low- $q$ . Regardless of the inter-particle effects the sample at 3.9 mg/mL appears to give the best compromise between signal:noise and these inter-particle effects and therefore a concentration around 3.9 mg/mL was chosen for the following DTT titration experiment.

#### 5.4.4 Increasing DTT concentration causes a loss of 2dCD4-WT structure

The 1D curves for the DTT titration after buffer subtraction are shown in figure 5.6. Again, it is evident from the upward turn of the curve at low- $q$  that there are inter-particle effects which indeed seem to worsen with increasing DTT concentration. Nevertheless, size parameters were

DTT concentration (mM)	Absolute I(0) (cm <sup>-1</sup> )	MM (kDa)	R <sub>g</sub> (Å)	R <sub>g</sub> (real) (Å)	D <sub>max</sub> (Å)
0	1.45×10 <sup>-2</sup>	22.4	26.0(+0.3)	26.2	92.5
1	1.41×10 <sup>-2</sup>	21.8	26.1(+0.4)	26.2	101
2.5	1.44×10 <sup>-2</sup>	22.2	26.3(+0.3)	26.1	96.5
5	1.38×10 <sup>-2</sup>	21.3	25.3(+0.4)	26.3	99.0
10	1.48×10 <sup>-2</sup>	22.9	27.0(+0.7)	28.2	109
25	1.30×10 <sup>-2</sup>	20.1	28.2(+0.7)	28.5	116
50	8.69×10 <sup>-3</sup>	13.4	28.9(+0.8)	25.0	78.0

TABLE 5.4: The size parameters, R<sub>g</sub> and D<sub>max</sub>, of 2dCD4-WT treated with increasing concentrations of DTT were determined by SAXS. The R<sub>g</sub> was calculated from both the Guinier and P(r) analysis. The MM was calculated from the absolute I(0) (corrected for concentration and scaled against the absolute scattering of water). The D<sub>max</sub> was also determined from the P(r) analysis. With increasing DTT concentration there is a general increase in R<sub>g</sub> and D<sub>max</sub>.

extracted from the data by Guinier and P(r) analyses and the MM was calculated from the I(0) and these are listed in table 5.4.

### 1D scattering of a 2dCD4-WT DTT titration

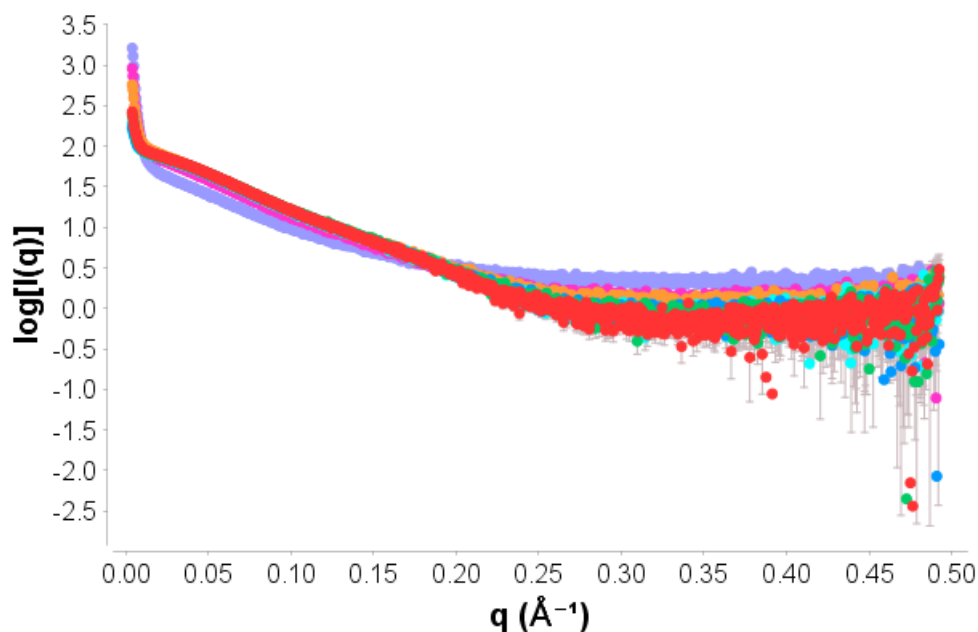


FIGURE 5.6: CFPE produced 2dCD4-WT was treated with increasing concentrations of DTT to observe the effect of DTT on its size and structure. At low- $q$  there is a sharp inflection of the curve towards  $x = 0$  which is characteristic of aggregation. Red = 0 mM, green = 1 mM, blue = 2.5 mM, cyan = 5 mM, orange = 10 mM, pink = 25 mM, purple = 50 mM.

The Guinier-derived R<sub>g</sub> shows a steady increase in the R<sub>g</sub> with the increase in DTT concentration except for the sample measured at 5 mM DTT which shows a decrease in 1 Å from the 2.5 mM sample. The pattern then resumes and the R<sub>g</sub> increases to 27.0 Å for the sample treated

with 10 mM DTT. The real space determined  $R_g$  are generally in good agreement with the Guinier-derived  $R_g$  for each sample ( $\pm 1 \text{ \AA}$ ), except for the sample measured at 50  $\text{\AA}$  which displays a large discrepancy between the two  $R_g$  of 28.9  $\text{\AA}$  and 25.0  $\text{\AA}$  for the Guinier and real space derived parameters, respectively. The trend for the real space  $R_g$  shows a steady increase in the  $R_g$  from 0 to 5 mM DTT, followed by a 2  $\text{\AA}$  increase from 5 to 10 mM DTT and then a sharp decrease to 25.0  $\text{\AA}$  for the 50 mM DTT sample which is neither in agreement with the Guinier derived  $R_g$  nor the pattern of the real space derived  $R_g$ .

The same general pattern can be seen for the  $D_{max}$ : with an increase in DTT concentration there is an increase in the  $D_{max}$ . However, the  $D_{max}$  measurement at 1 mM DTT appears to be anomalous as there is a large jump of 9  $\text{\AA}$  between the  $D_{max}$  measured for the 0 mM DTT sample and the 1 mM DTT sample. Again, the 50 mM sample is determined to be the smallest with a  $D_{max}$  of just 78  $\text{\AA}$  compared to the 92  $\text{\AA}$  of the 0 mM sample. The MM for 2dCD4-WT in 0 to 10 mM DTT are within 10% of the expected molecular mass (22.7 kDa) [163]. Upwards of 10 mM DTT the calculated MM are out of this range.

Whilst the size parameters show a general increase in the size of 2dCD4-WT as a function of DTT concentration, it is particularly interesting to look at the Kratky plot for these data (fig. 5.7). As the DTT concentration increases it appears that the curves become more open and tend less towards  $y=0$ . Relating this pattern back to the Kratky description in Chapter 2 section 2.4.4, it can be seen that this could be as a result of unfolding of the protein. At 0 mM DTT (red curve) the Kratky curve is characteristic of a partially folded protein. As the DTT concentration increases there does not appear to be much change in the Kratky curves between 0 and 10 mM DTT. However there is a distinct change in the shape of the Kratky curve at 25 mM DTT, which suggests protein unfolding. There is another distinct change at 50 mM DTT where it appears that the protein is completely unfolded. Taken with the MM data this suggests that there is something significant happening after 10 mM DTT. However it is difficult to decipher whether this is due to loss of structure due to disulphide bond reduction or whether this is due to aggregation.

### Kratky plot of a 2dCD4-WT DTT titration

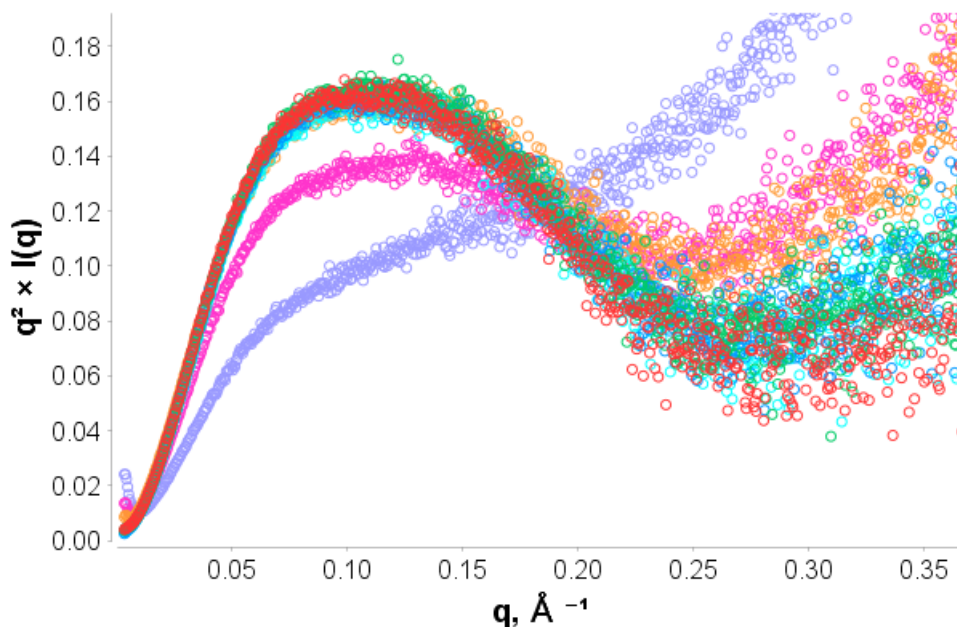


FIGURE 5.7: CFPE produced 2dCD4-WT was treated with increasing concentrations of DTT to observe the effect of DTT on its size and structure. The Kratky plot shows a gradual loss of foldedness with an increase in DTT concentration from 0 to 50 mM DTT. Red = 0 mM, green = 1 mM, blue = 2.5 mM, cyan = 5 mM, orange = 10 mM, pink = 25 mM, purple = 50 mM.

Due to the aggregation of the samples, the size parameters cannot be accurately extracted since the aggregation is quite significant and influences a large portion of the scattering curve. Regardless, it appears that the increase in DTT concentration has an effect on the size and structure of 2dCD4-WT. Relating this to the SDS-PAGE of the DTT titration (fig. 5.8), a shift in the band pattern is observed in favour of 2dCD4-R2, as the DTT concentration increases from 0 to 1 mM DTT, after which there appears to be little further change. However, according to the SAXS data there appears to be a significant change as the DTT concentration increases above 1mM DTT. The increase in DTT concentration appears to be causing the samples to aggregate more (fig. 5.6) and unfold (fig 5.7), so it is not possible to decipher whether the pattern of an increasing  $R_g$  and  $D_{max}$  with DTT concentration is a real feature or an artefact of the aggregation.

Interestingly, the *E. coli* expressed refolded 2dCD4-WT DTT titration (fig. 5.3) shows that 100% reduction was not achieved with the range of DTT concentrations used. Therefore, higher DTT concentrations were used for the CFPE 2dCD4-WT SAXS DTT titration. However, it appears from the shift in the band pattern shown by the SDS-PAGE of CFPE 2dCD4-WT that full reduction was possible at a concentration of 1 mM DTT. This may be because the *E. coli* and

CFPE protocols produce different distributions of the 2dCD4-WT redoximers. From the CFPE preparation it would appear that the partially reduced form is the major band whereas in the *E. coli* preparation the oxidised and reduced forms appear to be present in almost equivalent quantities.

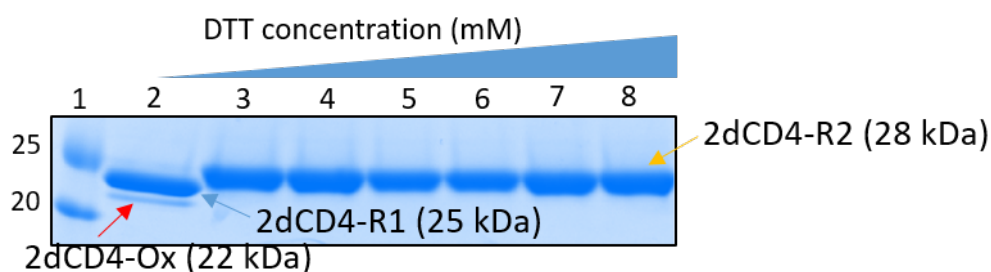


FIGURE 5.8: DTT titration of CFPE produced 2dCD4-WT at 0, 1, 2.5, 5, 10, 25 and 50 mM DTT (lanes 2-8). Lane 1 shows the molecular weight marker in kDa. The samples were treated with DTT and then measured using SAXS to observe how increasing DTT concentration affects the size parameters  $R_g$  and  $D_{max}$ .

#### 5.4.5 2dCD4-R2 is larger than 2dCD4-R1

Following the DTT titration SAXS experiment which produced low quality data, SEC-SAXS was carried out on 2dCD4-WT at 0 mM and 50 mM DTT to determine whether the features seen in the DTT titration were representative of what was happening to 2dCD4-WT in solution with increasing DTT concentration, or whether this arose from an artefact of the aggregation. The intention was to look for a shift in the SEC peak, which would indicate a change in the hydrodynamic volume, as well as a change in the size parameters. SEC-SAXS was used to minimise the inter-particle effects observed in the DTT titration. Any aggregation would be eliminated by the size-exclusion step, allowing an accurate analysis of the size parameters and the effect of complete reduction of the two disulphide bonds on the structure of 2dCD4-WT.

Firstly, by comparing the UV 280nm spectra in figure 5.9, a slight shift in the position of the peak can be observed. The peak for the 2dCD4-WT treated with 50 mM DTT is shifted to the left with respect to the untreated 2dCD4-WT, indicating that the hydrodynamic volume of the fully reduced 2dCD4-WT is larger than that of the 2dCD4-WT in which the partially reduced isoform is the majority species.

The curves to be included for analysis were first subject to auto $R_g$  to determine which curves contained species of a similar  $R_g$ . The frame number was plotted against the UV trace and the

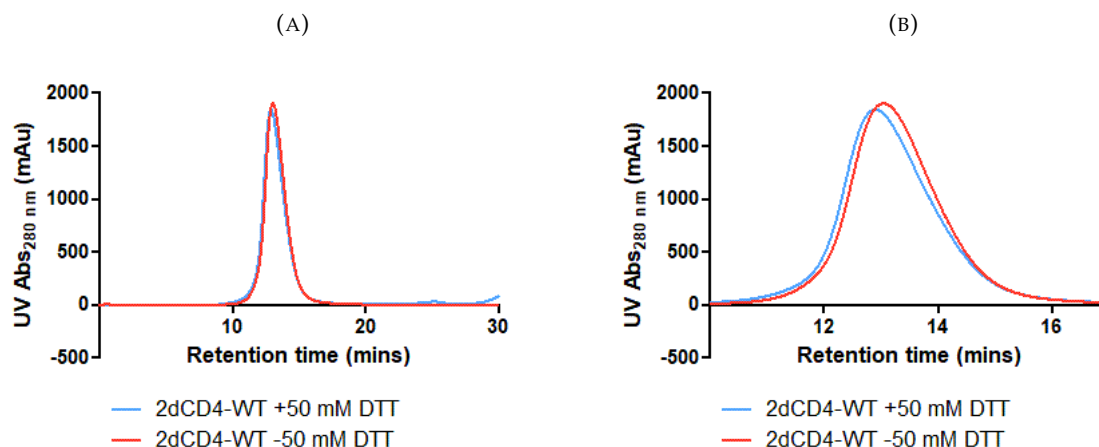


FIGURE 5.9: 2dCD4-WT with and without 50 mM DTT was analysed by SEC-SAXS. (A) 2dCD4-WT with (blue) and without (red) 50 mM DTT was analysed by SEC-SAXS. (B) A zoomed in view of the UV Abs<sub>280nm</sub> spectra show that 2dCD4-WT treated with 50 mM DTT elutes slightly before 2dCD4-WT without DTT treatment. This suggests that the reduced 2dCD4-WT (with 50 mM DTT) has a larger hydrodynamic volume than the unreduced 2dCD4-WT (without 0 mM DTT).

$R_g$  was determined across the 2dCD4-WT peak (fig. 5.10). The  $R_g$  are not that flat across the peak so the mid selection of the peak from frames 1017 to 1072 and from 1042 to 1091 for the 50 mM and 0 mM DTT treated samples were selected. These curves were therefore chosen for buffer subtraction and further analysis.

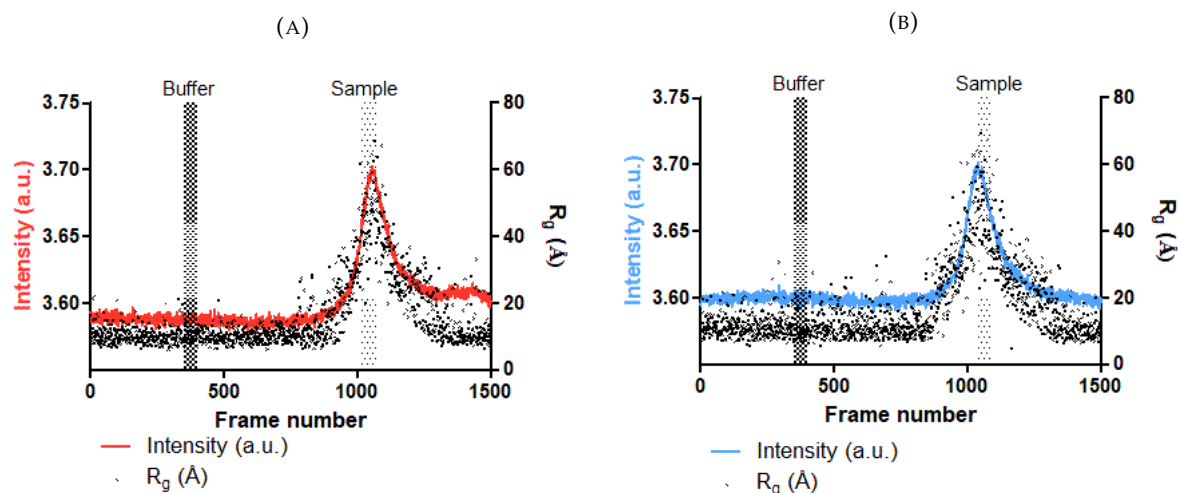


FIGURE 5.10: 2dCD4-WT with (A) and without (B) 50 mM DTT SEC-SAXS intensity and auto $R_g$  data were plotted against the frame number. The buffer frames were taken from a flat section of the data before the protein peak. The  $R_g$  was not flat across the peak so the sample frames were taken from the mid-section of the peak.

From the 1D scattering curves after buffer subtraction in figure 5.11 the use of SEC-SAXS appears to have eliminated the inter-particle effects which were observed in the DTT titration 1D



### 1D scattering of 2dCD4-WT with and without 50mM DTT

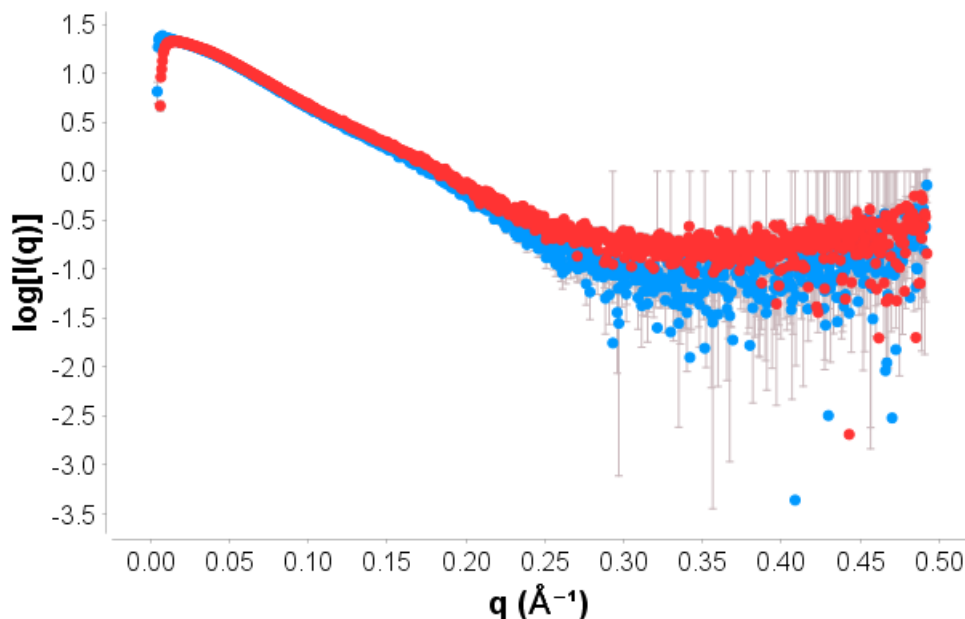


FIGURE 5.11: CFPE produced 2dCD4-WT was tested using SEC-SAXS for the effect of 50 mM DTT on the structure of 2dCD4-WT. The red curve corresponds to 2dCD4-WT without DTT treatment (0 mM DTT) and the blue curve corresponds to 2dCD4-WT with DTT treatment (50 mM).

scattering curves. Therefore, the size parameters extracted from the data are a more accurate representation of the size of the 2dCD4-WT. The  $R_g$  calculated by both the Guinier and  $P(r)$  analysis (table 5.5) are within good agreement of each other for each data set, which is an indicator of the good quality of the data. An increase in the  $R_g$  and  $D_{max}$  of 2dCD4-WT is observed with the addition of 50 mM DTT. From the  $P(r)$  in figure 5.5, the curves tend smoothly towards  $P(r) = 0$  which also indicates that the data is of good quality. The curves are of a similar shape, with the bump at around  $50\text{\AA}$  corresponding to the two Ig domains of 2dCD4-WT. At  $P(r) = 0$  the curves have a tail which could be caused by the His-tag which was uncleaved from the protein during purification to avoid sample loss.

DTT concentration (mM)	$R_g$ ( $\text{\AA}$ )	$R_g$ (real $\text{\AA}$ )	$D_{max}$ ( $\text{\AA}$ )
0	24.6(+0.2)	25.2	94
50	25.1(+0.2)	26.2	105

TABLE 5.5: The SEC-SAXS  $R_g$  determined from the Guinier analysis and the  $R_g$  and  $D_{max}$  determined from the  $P(r)$  analysis of 2dCD4-WT with and without 50 mM DTT treatment. 2dCD4-WT post-50 mM DTT treatment is larger than without DTT treatment.

Interestingly the Kratky plot for the SEC-SAXS data (fig. 5.13) tells a different story to that of the DTT titration. In this instance the curves are both indicative of a well-folded protein, as

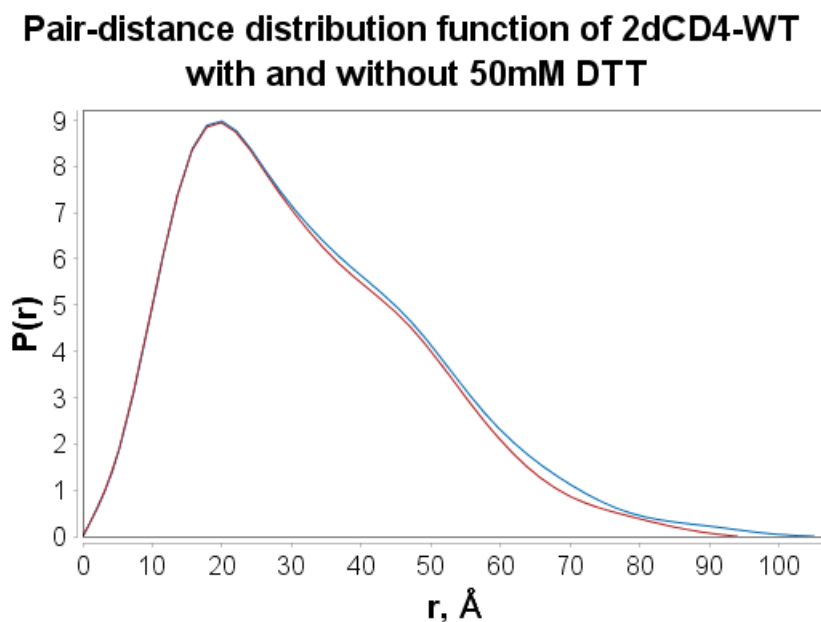


FIGURE 5.12: 2dCD4-WT with (blue) and without (red) 50 mM DTT treatment was analysed by SEC-SAXS. The  $P(r)$  function shows that the sample treated with 50 mM DTT has a larger  $D_{max}$  than the sample without DTT treatment. The shape of the  $P(r)$  function corresponds to the two Ig domains of 2dCD4-WT.

they approach the x-axis. In the DTT titration the 50 mM sample was indicative of an unfolded protein. Comparing this to the SEC-SAXS data, this would suggest that the Kratky results were a function of the increase in the aggregation with the DTT concentration, causing the protein to denature. This could also account for the aberrant real space  $D_{max}$  and  $R_g$  which were abnormally small compared to the trend observed, in which there was an increase in 2dCD4-WT size with DTT concentration. If the protein was unfolded, combined with aggregation, then the size parameter analysis is inaccurate and these values are not representative of what was actually happening in solution.

### Kratky plot of 2dCD4-WT with and without 50mM DTT

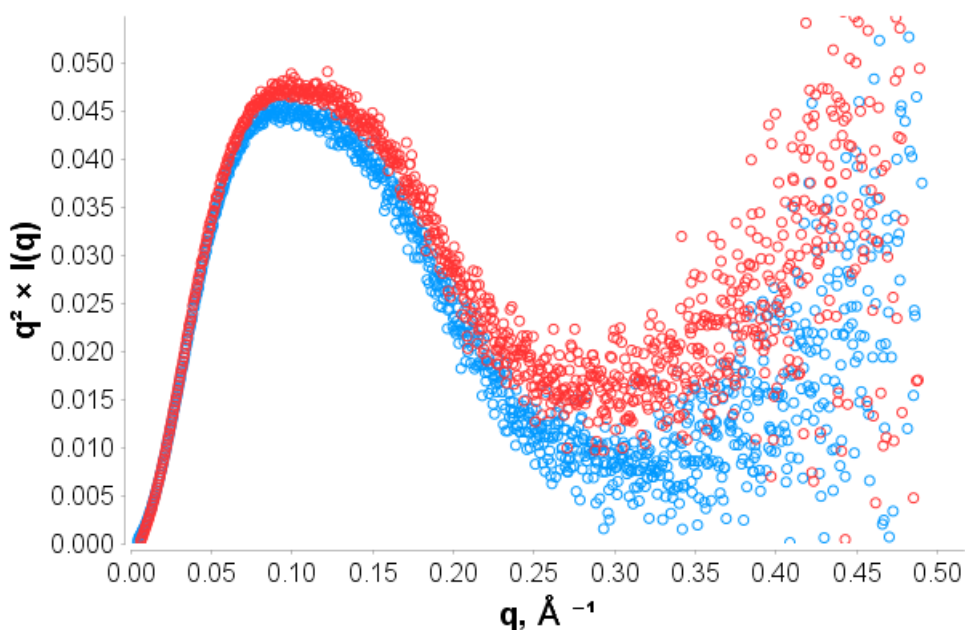


FIGURE 5.13: 2dCD4-WT with (blue) and without (red) 50 mM DTT treatment was analysed by SEC-SAXS. The Kratky curves for both the 0 and 50 mM DTT treated samples are characteristic of a well-folded protein.

The SEC-MALLS-RI data (fig. 5.14) show a high refractive index and light scattering signal at an elution volume of around 11 mL for the 0 mM DTT (5.14a dark purple and lilac, respectively) and 50 mM DTT (5.14b dark green and light green, respectively) treated 2dC4-WT. The UV 280 nm spectra and molecular mass data in figure 5.14c show a slight off-set between the two spectra. The curve recorded from 2dCD4-WT treated with 50 mM DTT (yellow) curve is shifted slightly to the left with respect to the 0mM DTT treated 2dCD4-WT curve (blue).

The orange and blue scatter curve across the peak of the two UV 280 nm spectra correspond to the molecular mass in Daltons for the 50 mM and 0 mM DTT treated 2dCD4-WT, respectively. The molecular mass of 2dCD4-WT in 0 mM DTT appears less stable across the peak than that of the protein in 50 mM DTT. This suggests that the sample without DTT may be less polydisperse. Indeed, the calculated molecular mass across the peaks were 22 020 Da (+10.8%) and 22 050 Da (+7.1%) with polydispersity indexes of 1.007(+6.9%) and 1.001(+4.4%) for the 0 mM and 50 mM DTT, respectively. The molecular weight is within agreement of the expected value of 22 691 Da, however the errors are high and the polydispersity indexes are indicative of polydispersity as they are both greater than 1.000.

There is also a large light scattering signal at an elution volume of around 7 mL, which is likely

to be a very small amount of a large particle. The particle is large because it elutes at a low elution volume of the column however, as large particles disproportionately scatter more light, it is likely that the contaminating particle is present in a small proportion (i.e. less than 5%) regardless of the large scattering signal it produces. The light scattering of the contaminant in 0 mM DTT is more significant than the contaminant in the 50 mM DTT sample. This contaminant is likely to be aggregation as the samples were tested via SEC-MALLS-RI 48 hours after the SAXS experiment and, as was observed in the DTT titration SAXS experiment, 2dCD4-WT has a tendency to aggregate.

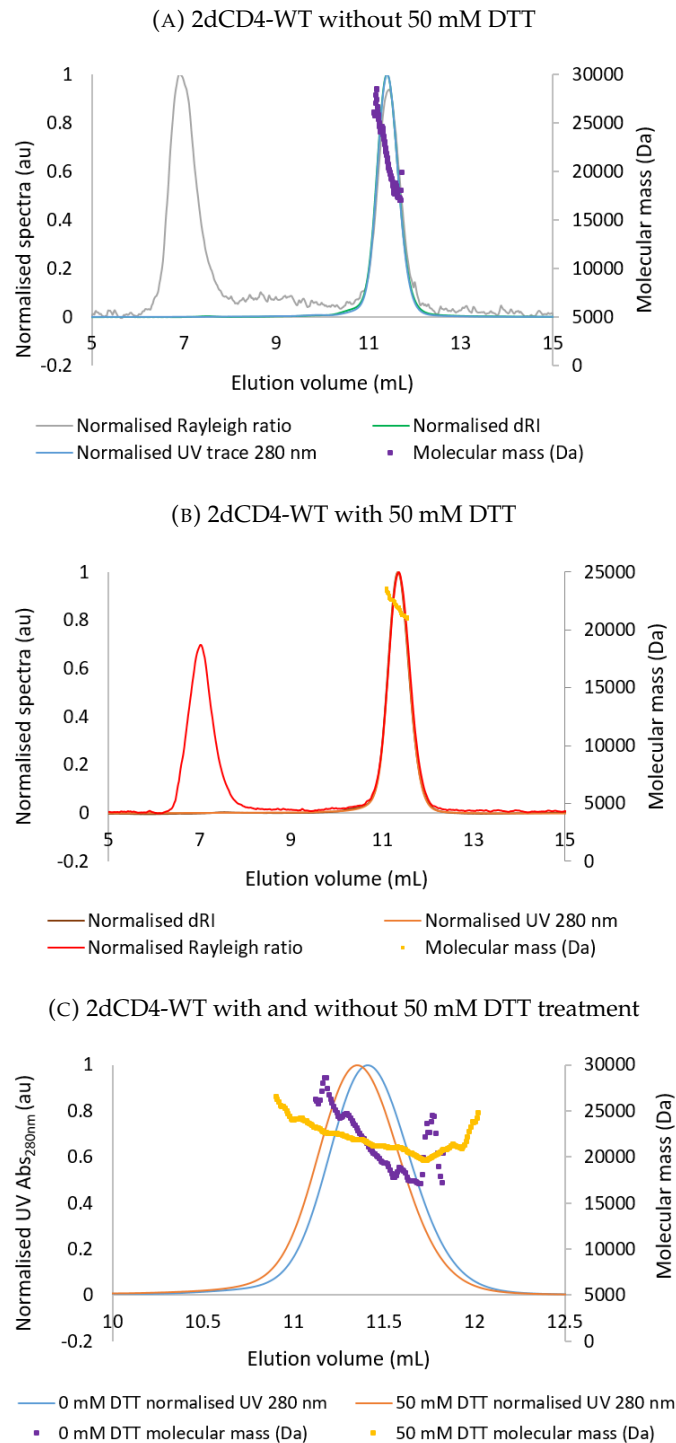


FIGURE 5.14: 2dCD4-WT in 0 mM and 50 mM DTT was tested using SEC-MALLS-RI to assess the quality of the samples tested using SEC-SAXS. (A) The normalised light scattering, refractive index and UV  $Abs_{280nm}$  spectra for 2dCD4-WT in 0 mM DTT (grey, green and blue, respectively) are shown. The molecular mass is shown across the main protein peak in Daltons in purple. (B) The normalised light scattering, refractive index and UV  $Abs_{280nm}$  spectra for 2dCD4-WT in 50 mM DTT (red, brown and orange, respectively) are shown. The molecular mass is shown across the main protein peak in Daltons in yellow. (C) The UV absorbance spectra at 280 nm and molecular weight in Daltons of 2dCD4-WT in 0 mM DTT (blue and purple, respectively) and 50 mM DTT (orange and yellow, respectively). The UV spectrum for the 50 mM DTT treated samples is shifted to the left with respect to the spectrum for 2dCD4-WT in 0 mM DTT suggesting that the protein in 50 mM DTT has a smaller hydrodynamic volume.

## 5.5 Discussion

CD4 is the primary receptor for HIV-1, binding the gp120 component of the viral envelope complex. CD4 possesses a metastable, allosteric disulphide bond in its second domain whose redox state may impact on its biological function as indicated in the literature [63, 113]. gp120 binds a novel redox isomer of 2dCD4 in which the second domain, allosteric disulphide bond is reduced suggesting that this protein may have a unique structure and is therefore a target for HIV-1 entry inhibitor and anti-HIV-1 immunogen design. The wild-type two-domain CD4, consisting of domains 1 and 2, exists in 3 different redox states due to the resident disulphide bond of each domain, as shown by knocking out of the disulphide bond and assessing the mobility on non-reducing SDS-PAGE. The disulphide bonds of recombinant 2dCD4-WT have also been shown to be sensitive to reduction by DTT [63]. Recombinant 2dCD4-WT produced using *E. coli* and refolded from inclusion bodies and cell-free protein expression systems has been shown to contain a mixture of the oxidised (Ox), reduced in domain 2 (R1) and fully reduced (R2) redox species. A combination of manipulation of the redox state using DTT titrations and SEC was used in combination with SAXS to try to characterise the behaviour of the 2dCD4-WT redox isomers.

Using a DTT titration against 2dCD4-WT expressed in *E. coli*, a shift in the band intensity was observed from the lower apparent molecular weight species to the higher apparent molecular weight species with an increase in DTT concentration. This can be attributed to a reduction in the disulphide bonds of 2dCD4 which causes a decrease in the gel mobility [198]. Full reduction of *E. coli* expressed 2dCD4-WT was not achieved at the maximum concentration of 10 mM DTT, since all three bands were present, nor was there a DTT concentration at which the 2dCD4-R1 species was isolated. Conversely, for the 2dCD4-WT produced by CFPE, full reduction was possible at just 1 mM DTT. This could be explained by the fact that the initial distribution of redoximers was different, with the partially reduced isoform being the major species present, therefore less DTT was required to fully reduce both disulphide bonds. A higher initial starting concentration of DTT was used with CFPE produced 2dCD4-WT as full reduction had not been achieved with the *E. coli* expressed 2dCD4-WT.

When the CFPE produced 2dCD4-WT DTT titrated samples were measured by SAXS, an increase in the size parameters was observed with increasing DTT concentration. An unfolding

of the structure was also indicated from the Kratky plot. This reduction event is mostly of the first domain structural disulphide bond, since the SDS-PAGE of the DTT titration indicated that the primary species present in the CFPE 2dCD4-WT in 0 mM DTT is the 2dCD4-R1, where the second domain allosteric disulphide bond is already reduced. The increase in the size parameters  $R_g$  and  $D_{max}$  was corroborated by the SEC-SAXS experiment in which CFPE treated 2dCD4-WT was tested with and without 50 mM DTT. However, whilst the 50 mM DTT concentration showed anomalous values in the DTT titration that are likely to be due to aggregation, the SEC-SAXS samples were free from aggregation and therefore provide a more accurate picture of the state of fully reduced 2dCD4-WT. It was unclear whether the unfolding seen in the Kratky plot of the DTT titration was an accurate representation of what was happening in solution or whether it was an artefact of the aggregation. After testing the samples using SEC-SAXS it appears that the unfolding observed was due to denaturation of the protein because of the relatively high DTT concentration. The results observed for the DTT titration SAXS experiment are therefore not representative of native 2dCD4-WT behaviour.

In addition to the increase in the size parameters  $R_g$  and  $D_{max}$  observed by SEC-SAXS, a difference in the hydrodynamic volume of the 2dCD4-WT with and without 50 mM DTT treatment was observed. The UV 280 nm absorbance peak was shifted to the left with the addition of 50 mM DTT at which point the protein is completely reduced. This indicates an increase in the hydrodynamic volume of the protein as it elutes more quickly from the column. Since the CFPE produced 2dCD4-WT in 0 mM DTT was primarily in the R1 form, as seen from the SDS-PAGE, this indicates that the 2dCD4-R2 has a larger hydrodynamic volume than the 2dCD4-R1 form, as expected. However, the inverse pattern is seen by the size-exclusion purification of the 2dCD4-OX and 2dCD4-R1 redoximers, whereby the oxidised form of the protein elutes first and the partially reduced form elutes second. This suggests that the oxidised form of the protein has a larger hydrodynamic radius than the partially reduced form.

The SEC-MALLS-RI data showed the presence of aggregation but given that any aggregation was eliminated from the 2dCD4-WT signal by the SEC step, this poses no problem for the SAXS data analysis as SEC-SAXS was used and would eliminate any such contaminants. The monodispersity was also questionable across the protein peaks. However, the SEC-SAXS samples were freshly prepared and only the scattering curves in the  $R_g$  vs  $I(0)$  plot with a stable  $R_g$  indicating their similarity were used for the SEC-SAXS analysis. The same UV spectra shift

was seen for the SEC-MALLS-RI data as for the SEC-SAXS data. Considering the columns used in the two techniques have differing resolution, this reinforces the idea that this is a real feature observed and not a result of aggregation.

### 5.5.1 Relating the data to the disulphide bonds in 2dCD4-WT

The SEC behavior of the 2dCD4-WT redox isomers suggests that the R2 species has a larger hydrodynamic volume than the R1 form and that the OX form also has a larger hydrodynamic volume than the R1 form. Given the findings by Owen *et al.* 2016 [203] suggesting that domain 1 has a stabilising effect, it is predicted that the R2 form has a larger hydrodynamic volume than the OX form, with the R1 having the smallest hydrodynamic volume as outlined in figure 5.15. Further work is required to confirm this. The difference in the hydrodynamic volume and the surprising finding that R1 has the smallest hydrodynamic volume can be reasoned by the difference in the properties of the disulphide bond in domains 1 and 2.

The disulphide bond in domain 1 is a stable, structural disulphide bond which serves to provide structural integrity to the domain. The torsional strain is low because of the geometry of the bond in which the disulphide bond is formed between the two beta sheets of the Ig-like domain. On the other hand, the disulphide bond in domain 2 is a metastable allosteric disulphide bond which can undergo redox shuffling events to affect the function of 2dCD4-WT ([112]). The bond is formed between two cysteines in neighbouring strands of the same beta sheet, which is a geometrically unusual configuration, causing the disulphide bond to have a high dihedral energy and be under high torsional strain.

Due to this high dihedral energy and torsional strain of the disulphide in domain 2, reduction of this highly strained bond could cause an energetically favourable relaxation of the beta-strands involved in the bond. The allosteric disulphide bond in domain 2 causes puckering of beta-strands C and F, thus reduction of this disulphide bond may allow relaxation of this puckering effect. This relaxation could result in a slight shrinking of the domain as the beta strands packed against this disulphide bond may shift inwards, allowing the domain to adopt an energy minimal conformation. Conversely it is possible to imagine that reduction of the structural disulphide bond in domain 1 would cause this domain to lose its structural integrity which would result in an increase in the size.



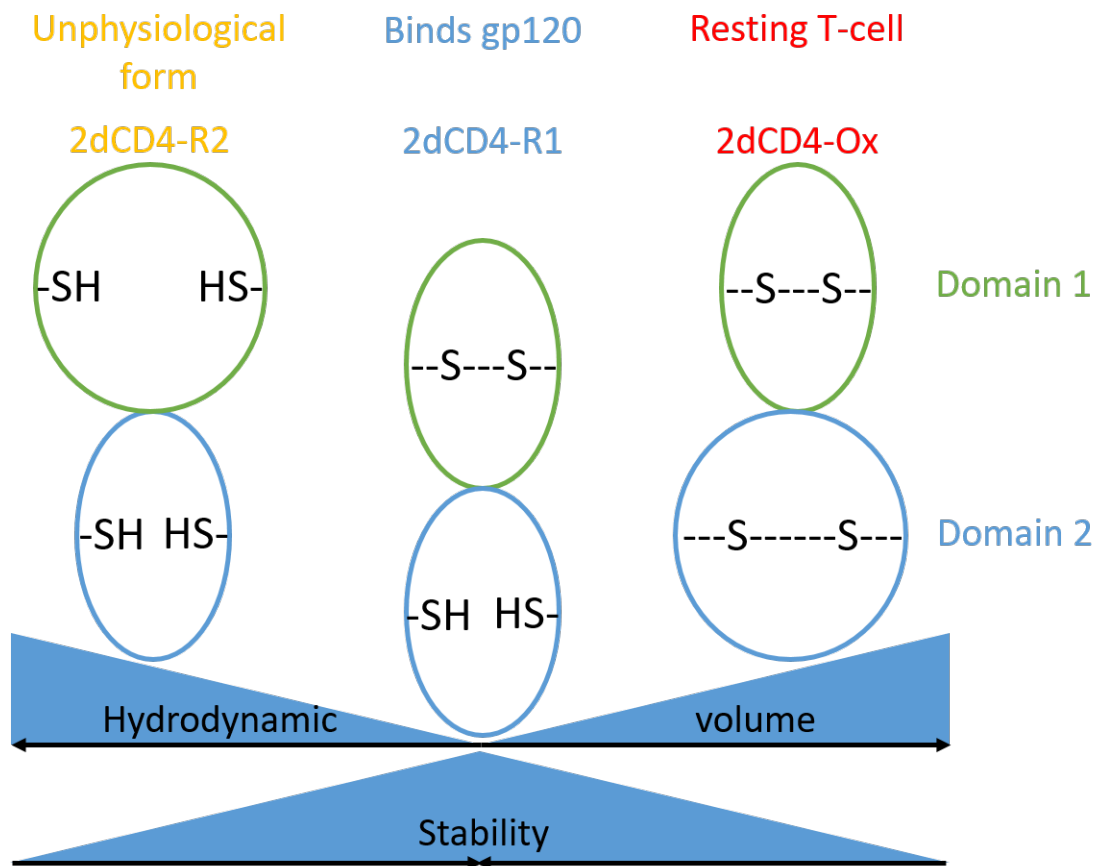


FIGURE 5.15: 2dCD4-R1 (middle, blue) which binds gp120, was found to have the smallest hydrodynamic volume of the three redox isomers. 2dCD4-Ox (right, red) which is found on resting T-cells and 2dCD4-R2 (left, yellow) which does not have a physiological function, were both found to have a larger hydrodynamic volume than 2dCD4-R1. The data do not provide information as to whether 2dCD4-R2 has larger, smaller or the same hydrodynamic volume as 2dCD4-Ox. 2dCD4-R1 has previously been shown to be the most stable redox isoform [203]. Domain 1 is shown in green and domain 2 is shown in blue. Reduced disulphide bonds are denoted by -SH or HS- and oxidised disulphide bonds are denoted by -S-S-.

As stated above, the SEC and SEC-SAXS experiments suggest that the order in hydrodynamic volume of the redoximers is:  $R_2 \geq R_1$ . This can be related to the disulphide properties above. 2dCD4-R1 has its domain 1 disulphide bond intact which is the lowest energy configuration for this domain and the domain is at its smallest hydrodynamic volume. The domain 2, allosteric disulphide bond is reduced which is the lowest energy state for this domain and it is also at its smallest hydrodynamic volume [203]. 2dCD4-Ox, which also has domain 1 its energetically favourable configuration, has its domain 2 disulphide bond intact, which is energetically unfavourable, causing the domain to have a larger hydrodynamic volume as the disulphide bond causes the beta strands to be pulled together which may cause displacements of the sheets further away from the allosteric disulphide bond. Finally, 2dCD4-R2 has 2dCD4-WT in its most energetically favourable form but the domain 1 disulphide bond is reduced, so the structural integrity of this domain is compromised the hydrodynamic volume of this domain increases.

This pattern is not seen in the non-reducing SDS-PAGE data because the proteins are denatured forming a random coil. So those disulphide bonds which are intact serve to increase the electrophoretic mobility of the random coils through the polyacrylamide matrix of the gel as they hold the random coil in a tighter structure. Therefore the denatured hydrodynamic volume of the redox species is not the same as the native hydrodynamic volume. This trend is reflected by the size parameters derived from the SEC-SAXS data as the  $R_g$  and  $D_{max}$  increase with the addition of 50 mM DTT.

### 5.5.2 2dCD4-WT previously measured by SAXS

2dCD4-WT has previously been studied by SAXS in isolation and in complex with two of its ligands: gp120 and Salp15. The reported values for the Guinier  $R_g$  are 22.3 and 20.9, the real space  $R_g$  was 21.7 and the  $D_{max}$  were 73 and 75 Å (table 5.2). Whilst these values are not in line with the  $R_g$  and  $D_{max}$  found in this study (fig. 5.5) the redox state of the 2dCD4 used in the previous studies was not stated. The 2dCD4 used previously was not produced in the cell-free system, it was expressed in *E. coli* and refolded from the inclusions bodies. In addition, the 2dCD4-WT used in this study had an uncleaved His-tag with an intact TEV cleavage site which probably caused the sample to appear larger. This is also suggested by the  $P(r)$  data from the

SEC-SAXS experiment since there appears to be a tail as the curve approaches  $r=D_{max}$ ,  $P(r)=0$ . Cleaving the His tag may result in smaller size parameters for the CFPE produced 2dCD4-WT.

### 5.5.3 Future work

The DTT titration should be repeated using lower concentrations of DTT to observe how the distribution of the redoximers changes for the CFPE produced 2dCD4-WT. This could also be altered by changing the redox potential during the CFPE reaction. In order to avoid aggregation which badly affected the analysis of the DTT titration SAXS data, SEC-SAXS should be employed with the lower DTT concentrations to see how the size parameters  $R_g$  and  $D_{max}$  change during the transition between the R1 and R2 redox species. If it would be possible to isolate the fully oxidised form by using oxidising reagents then the hydrodynamic volume of the oxidised and reduced forms could be compared, as before, using SEC to complete the model shown in figure 5.15.

During this thesis work it was not possible to purify a stable 2dCD4-D2A mutant form using the cell-free system. Although the CFPE 2dCD4-WT in the absence of reducing agent appears to be primarily present in the R1 form there was still a small amount of the R2 and OX forms present as seen by the SDS-PAGE. Producing a soluble, stable 2dCD4-D2A variant protein would allow effective isolation of the 2dCD4-R1 species for analysis by SAXS to determine its structure. It would be useful to cleave the His-tag from the 2dCD4 to see if the size parameters fall in line with those previously reported. It would also be interesting to look at the SEC behaviour of the individual domains in isolation in the presence and absence of DTT.

Ultimately, SAXS is a low resolution technique allowing the determination of the global size and shape of macromolecules in solution and does not provide the capability to look at the local structure around the disulphide bond. It would be interesting to look at the high resolution structure of the oxidised and reduced forms by either X-ray crystallography or high-resolution NMR without inducing radiation damage.

Finally, neutron scattering techniques could be used. Neutron crystallography would allow unambiguous determination of the redox states of the disulphide bonds since the hydrogens can be viewed using this technique. In addition, the samples would not be at risk of radiation damage which is particularly worrisome for samples containing disulphide bonds. However,

due to the necessity for perdeuteration, followed by crystallisation, neutron crystallography is a non-trivial technique for difficult samples such as 2dCD4-WT. SANS experiments were carried out in chapter 6, however, due to the long measuring times required because of the lower flux and the tendency of 2dCD4-WT to aggregate (as seen in the concentration series and DTT titration experiments), it was difficult to obtain reliable data for 2dCD4-WT let alone studying the effect of the modulation of the disulphide bond redox states with DTT. SEC-SANS [204] has recently been developed and might be an interesting supplement to the SAXS study.

## 5.6 Conclusion

2dCD4-R1 has been shown to have the smallest hydrodynamic volume of the three redox isomers. 2dCD4-Ox and 2dCD4-R2 have both been shown to have a larger hydrodynamic volume than 2dCD4-R1 but it has not been determined whether 2dCD4-R2 has the same or a larger hydrodynamic volume than 2dCD4-Ox. 2dCD4-R1 has also been shown to have a smaller  $R_g$  and  $D_{max}$  than 2dCD4-R2. The size difference maybe explained in terms of the reduction of the allosteric disulphide bond in domain 2, causing a relaxation of the domain by releasing the high-torsional strain in the bond, which is of unusual geometry. 2dCD4-WT has a tendency to aggregate and so is most appropriately studied by size-exclusion chromatography coupled to small-angle scattering. While, the importance of the reduction of the domain 2 disulphide bond for gp120 binding still remains unclear, this work has shown that there is a change in the size of 2dCD4-WT as a function of its redox state, which alludes to the presence of conformational dynamics which may be necessary for interaction with gp120.

## Chapter 6

# Characterisation of gp120 and small-angle scattering analysis of the complex

### 6.1 Abstract

A significant effort was required to establish an expression protocol for 2dCD4-WT that produced sufficient yields of recombinant, unlabelled and deuterium labelled protein for small-angle neutron and X-ray scattering (SANS and SAXS) studies in this thesis work, as described in Chapter 3. The resulting SANS experiments have allowed determination of the size and shape of 2dCD4-WT and gp120 in isolation and whilst in complex, through the exploitation of protium/deuterium contrast and match-out labelling. Whilst size-exclusion chromatography coupled to SAXS (SEC-SAXS) has allowed determination of the global envelope structure of 2dCD4-WT analogues produced from different expression systems in complex with gp120.

The biophysical characterisation of a gp120 monomer/dimer by mass-spectrometry and SEC coupled to multi-angle laser light scattering and refractive index is also addressed in this chapter. Following which the results of several SANS experiments carried out on the D22 diffractometer at the ILL are presented. The first of which was to determine the global match-out point of the highly glycosylated protiated gp120 viral protein (h-gp120). The aim of the second experiment was to determine the match-out point for the deuterium labelled 2dCD4-WT (d-2dCD4-WT) protein produced using the cell-free system. Then the final experiment utilised

the match-out points determined in the first two experiments to determine the structure of the h-gp120/d-2dCD4-WT complex and the d-2dCD4-WT bound to matched out h-gp120. In the final experiment, SAXS was also carried out on the same samples on the BM29 bioSAXS beamline at the ESRF in an attempt to corroborate the SANS data obtained.

Finally, a series of SEC-SAXS experiments were conducted on the gp120/2dCD4-WT analogue complexes with the 2dCD4-WT produced from three different expression systems: *E. coli* re-folded, *B. choshinensis* secreted and cell-free protein expressed (CFPE). SAXS data has already been published for the gp120/2dCD4 and 4dCD4 complexes, so the complex data collected in this thesis served as a way of ensuring that the data correlated with that already published to validate the *B. choshinensis* and CFPE expression systems as means of producing 2dCD4-WT for small-angle X-ray scattering studies.

## 6.2 Introduction

The atomic resolution X-ray crystal structure of the 2dCD4-WT/gp120 core complex was first determined in 1998 and demonstrated the interactions between key CD4 residues: Phe43 and Arg59 and numerous gp120 residues including Asp368, Glu370 and Trp427 [34]. However, in this study, gp120 was truncated at its N- and C-termini, in its V1-V3 loops and deglycosylated. In addition, the complex was bound with the 17b broadly neutralising antibody which binds the CD4 induced epitope on gp120 to stabilise the highly flexible gp120 core, thus fixing the complex in the 17b bound form. The complex was therefore missing parts of the vital V3-loop necessary for co-receptor binding and the glycan shield which is essential to viral immune evasion.

Since then, there have been several small-angle X-ray scattering studies of 2dCD4-WT (D1 and D2), 4dCD4-WT (D1-D4), gp120 and their respective complexes [72, 118, 197]. These SAXS experiments have allowed investigation of the complex of the full-length gp120 monomer in its glycosylated form. Ashish *et al.* 2006 examined the complex with *E. coli* expressed 2dCD4-WT [197] and again in 2008 with 4dCD4-WT expressed in *E. coli* [118]. Size parameters were determined from both experiments but only the experiment with 4dCD4-WT suggested that there is a conformational change upon gp120 binding. This conformational change was suggested to occur at the D2/D3 interface, whereby the extended rod-like shape of CD4 would bend

in this region to facilitate host-cell/viral membrane fusion, by bringing the viral and host-cell membranes into proximity for subsequent V3-loop/co-receptor driven structural realignments. Guttman *et al.* 2012 found that there were extensive rearrangements within the V1/V2 loop region of gp120 upon 2dCD4-WT binding but they did not draw attention to the 2dCD4-WT component of the complex [72]. The previously published size parameters determined from SAXS experiments on 2dCD4-WT, gp120 and the complex are presented in table 6.1.

Publication	Sample	$R_g$ (Å)	$D_{max}$ (Å)
Ashish <i>et al.</i> 2006	2dCD4-WT	22.3	75
	gp120 dimer	43.6	166
	2dCD4-WT-gp120(dimer) complex	44.4	165
Guttman <i>et al.</i> 2012	2dCD4-WT	20.9	73
	gp120 monomer	37.1	130
	2dCD4-WT-gp120(monomer) complex	37.0	135

TABLE 6.1: Published size parameters  $R_g$  and  $D_{max}$  for 2dCD4-WT, gp120 monomer and dimer and their respective complexes as determined by SAXS.

There are therefore no SAS data which explore the relationship between the structure of CD4 as a function of its redox state and its ability to bind gp120. In addition, there are no small-angle neutron scattering (SANS) studies to describe the gp120/CD4 complex. SANS lends itself to the study of protein-protein complexes, such as the gp120/CD4 complex, as contrast variation studies can be performed. In SANS it is possible to exploit contrast variation by moderating the scattering length density (SLD) of both the solvent (by altering the heavy/light water ratio) and the protein (by deuterium labelling) to observe the scattering of just one of the protein partners within the formed complex (see Chapter 2 for a detailed explanation).

In this chapter, contrast variation SANS studies of a protiated-gp120 and deuterium-labelled 2dCD4-WT complex (h-gp120/d-2dCD4-WT) were conducted. The aim was to determine the low-resolution envelope structure as well as the extent of folding and size parameters, such as the radius of gyration of gp120-bound and un-bound d-2dCD4-WT. Since only the partially reduced isoform of 2dCD4-WT (2dCD4-R1) can form a complex with gp120 (as addressed in chapters 1, 3 and 5), it was expected that differences would be observed between the 2dCD4-WT in isolation and in complex with gp120 which could suggest why ablation of the second

domain disulphide is essential for gp120 binding to CD4.

Prior to study of the structure of the h-gp120/d-2dCD4-WT complex by SAS, the h-gp120 component was first characterised. gp120 from two different HIV-1 subtypes (or clades) was compared: gp120 BAL which is from a subtype B HIV-1 and gp120<sub>ZA</sub>CAP45 which is from a subtype C HIV-1. Subtype B affects mostly western communities, whereas subtype C is predominantly found in sub-Saharan Africa and parts of Asia. The two gp120 subtypes were then characterised biophysically using mass spectrometry to determine the level of glycosylation. After this the gp120<sub>ZA</sub>CAP45 subtype was further analysed using size-exclusion chromatography coupled to multi-angle laser light scattering and refractive index (SEC-MALLS-RI).

The contrast match point (CMP) of gp120<sub>ZA</sub>CAP45 was determined by measuring the SANS signal in a series of D<sub>2</sub>O solvent contrasts. This was repeated for the cell-free protein expressed (CFPE) d-2dCD4-WT to determine its CMP. Subsequently the h-gp120, d-2dCD4-WT and the respective complex was studied using contrast variation SANS.

Following SANS, the samples were also tested by SAXS in an attempt to cross validate the data obtained by SANS. Finally, the gp120/2dCD4 complex was studied using size-exclusion chromatography coupled to SAXS (SEC-SAXS) with 2dCD4-WT produced in the *E. coli*, cell-free protein expression (CFPE) and *B. choshinensis* expression systems (described in chapter 3). The SEC-SAXS aimed to corroborate both the SANS data obtained and the description of the gp120/2dCD4-WT complex in the literature to validate the CFPE and *B. choshinensis* expression systems as means of producing proteins for SAS experiments.

This chapter is split into three sections: (I) biophysical characterisation of gp120 (section 6.3), (II) small-angle neutron scattering (section 6.4) and (III) small-angle X-ray scattering (section 6.5.2), in order to facilitate the explanation of the methods and results, before the discussion in section 6.6.



## 6.3 Part I: Biophysical characterisation of gp120

### 6.3.1 Materials and methods

#### Stably transfected HEK293 cell culture

HEK293 FreeStyle cells which had been stably transfected by the pCIneo-gp120BAL or pcDNA 3.1(-)A012-gp120<sub>ZA</sub>CAP45 were provided by Dr Mark Killick at the HIV Pathogenesis Research Unit in Johannesburg, South Africa. 1 mL of human embryonic kidney 293 FreeStyle (HEK293FS, Life Technologies) cell suspension, stably transfected with the BAL or CAP45 plasmid, was defrosted and diluted 1:10 in 9 mL of room temperature 293 FreeStyle expression media (Life Technologies) in a T75 flask (Falcon) and supplemented with 1 X G418 antibiotic (Life Technologies). The cells were incubated at 37°C, 5% CO<sub>2</sub>. After 4 days the cells were collected by gentle centrifugation at 500 X g and gently resuspended in 15 mL fresh expression medium supplemented with 1 X G418 antibiotic until the cells were homogeneously distributed in the media. The resuspended cells were gently pipetted into a T250 cell culture flask (Falcon).

When the cells formed clumps they were centrifuged and the expression media was changed as described above. When the cells became adherent, the media was changed by gently pouring off the old media and gently pipetting 15 mL of fresh expression media into the flask. For expression in roller bottles: three T250 flasks worth of confluent cells were required and for hyperflask expression: eight T250 flasks worth of confluent cells were required. When the cells became confluent, the cells from the appropriate number of T250 flasks were resuspended in 15 mL of the old expression media before transfer to the expression vessel.

#### gp120 expression

The stably transfected HEK293FS cells were cultured in 100 mL 293 FreeStyle media (Life Technologies), supplemented with 0.5 X G418 (Life Technologies) in roller bottle flasks (Corning) at 37°C, 5% CO<sub>2</sub> and in 560 mL 293 FreeStyle media, supplemented with 0.5 X G418. The supernatant was collected by centrifuging at room temperature for 5 minutes at 1 000 rpm, filtered through a 0.8 µm filter and stored at -20°C. The supernatant was harvested every 2 days for 1 month.

**gp120 purification**

The gp120<sub>ZACAP45</sub> clone was not His-tagged, however due to the extent of glycosylation of the protein and the affinity of glycans for lectin, purification by lectin affinity was exploited, using agarose immobilised lectin (Sigma). Whilst gp120 BAL was His-tagged, lectin-affinity purification was used for its purification.

The total harvested supernatant was defrosted and incubated with 2 mL lectin overnight at 4°C with stirring. A 50 mL bed volume medium-pressure column (BioRad) was packed manually and the flow through collected. The resin was washed with 4 column volumes (CV = 50 mL) 0.5 M NaCl in 1 X phosphate buffered saline (PBS). The column was then washed with 4 CV 1 M NaCl in 1 X PBS. The column was washed with 2 CV 1 X PBS. Finally gp120 was eluted with 25 mL 1 M  $\alpha$ -D-methylmannopyranoside (MMP) (Sigma) and the column washed with 1 X PBS into the eluent.

The resin was incubated overnight at 4°C with the flow through from the column packing to repeat the purification to ensure maximum gp120 recovery from the supernatant. The purified gp120 was concentrated using an Amicon®Ultra-15 centrifugal filter (Merck) with 1 X PSB washing to remove the MMP, flash frozen in liquid nitrogen and stored at -80°C.

**Matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS)**

MALDI TOF mass spectra of gp120 BAL and gp120<sub>ZACAP45</sub> samples were measured with an Autoflex mass spectrometer (Bruker Daltonics, Bremen, Germany) operated in linear positive ion mode. External mass calibration of the instrument was performed using protein calibration standards (Bruker Daltonics). The gp120 samples were mixed in a ratio 1:1 (v:v) with sinapinic acid matrix (Sigma, 10 mg/mL in water/acetonitrile/trifluoroacetic acid, 50/50/0.1, v/v/v) and 1-2  $\mu$ L of this mixture was deposited on the target and allowed to air dry. Mass spectra data were processed with Flexanalysis software (v.3.0, Bruker Daltonics).

### **Size-exclusion chromatography coupled to multi-angle laser light scattering and refractive index (SEC-MALLS-RI)**

50  $\mu\text{L}$  gp120<sub>ZA</sub>CAP45 at a concentration of 13 mg/mL was loaded via a pump (L2130, Elite LaChrom) onto an analytical Superdex<sup>TM</sup> 200 10/300 size-exclusion column (S200 ag, GE Healthcare) equilibrated in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% sucrose buffer, with an inline spectrophotometer (L2400, Elite LaChrom). As the protein eluted from the column it passed through a flow cell where the multi-angle laser light scattering signal was measured (Dawn Helios-u, Wyatt). Finally, the refractive index was recorded by a refractometer (Optilab T-rEX, Wyatt) and the protein collected by a fraction collector (BioRad). The data were analysed using the software ASTRA6 (Wyatt Technology).

#### **gp120 monomer/dimer separation**

500  $\mu\text{g}$  gp120<sub>ZA</sub>CAP45 was incubated with 5 mM DTT for 1 hour at room temperature. 500  $\mu\text{L}$  was injected via a 1 mL loop onto a S200 ag column equilibrated in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% sucrose buffer. The protein was pumped through the column at a flow rate of 0.5 mL/min and the elution was collected in 500  $\mu\text{L}$  fractions. The fractions were analysed for the presence of gp120 by SDS-PAGE analysis with Coomassie blue staining. The same procedure was repeated without DTT treatment to compare the UV<sub>280nm</sub> spectra.

### **6.3.2 Results**

#### **gp120 purification**

2 L of FreeStyle expression media from gp120 BAL and <sub>ZA</sub>CAP45 preparations (referred to as BAL and CAP45 henceforth), from which HEK293FS cells had been removed by centrifugation and filtration before freezing, were defrosted and purified using lectin affinity purification. After purification, 16  $\mu\text{L}$  of 1.0 mg/mL BAL and 1.2 mg/mL CAP45 were added to 4  $\mu\text{L}$  loading buffer and analysed by SDS-PAGE with Coomassie blue staining (fig. 6.1).

The BAL sample appeared less pure than the CAP45 sample, as can be seen from the greater number of contaminating bands of different apparent molecular weight as compared to CAP45. The presence of the contaminating bands would have skewed the concentration determination

of BAL to appear higher than the actual concentration of just BAL within sample. Therefore, the CAP45 sample also displays greater yield as well as purity. The BAL band resolves at a molecular weight between 100 and 150 kDa whereas the CAP45 band resolves to a molecular weight between 75 and 150 kDa. This is because the CAP45 band is very large due of overloading of the sample onto the gel.

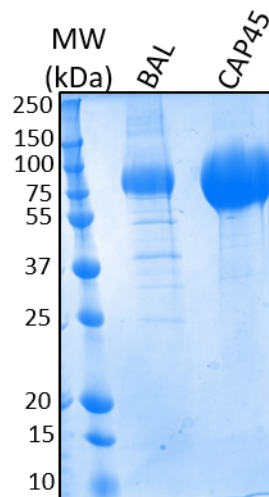


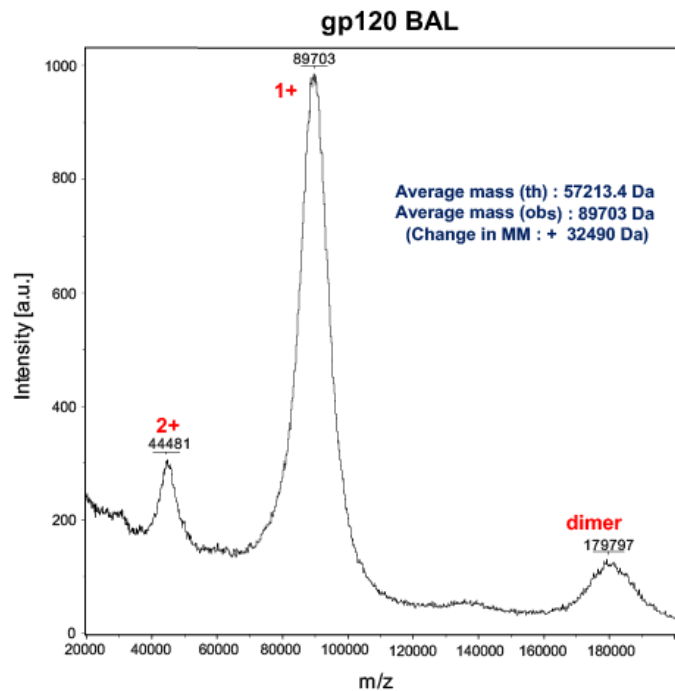
FIGURE 6.1: The purified and concentrated gp120 BAL and gp120<sub>ZA</sub>CAP45 strains were compared by SDS-PAGE to determine which preparation yielded the highest purity. The BAL sample appears to have more contaminating bands at different molecular weights than does the CAP45 sample. The molecular weight marker of standard proteins on the left-hand side of the gel displays the molecular mass in kDa.

### Determination of percentage glycosylation of gp120 by MALDI-TOF-MS

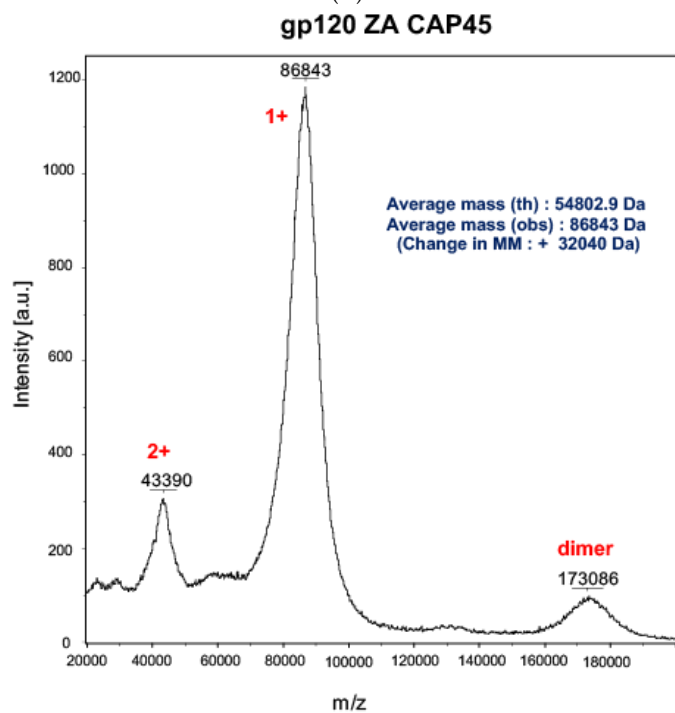
MALDI-TOF mass-spectrometry was used as a way to assess the precise molecular mass of the two gp120 strains: BAL (fig. 6.2a) and CAP45 (fig. 6.2b) and, as a result, the extent of their glycosylation. Figure 6.2a presented a main peak with a molecular mass of 89 703 Da and two smaller peaks at 44 481 and 179 797 Da for BAL. The predicted molecular mass for the protein portion of gp120, determined from the amino acid sequence using the ProtParam tool from ExPASy (<https://web.expasy.org/protparam/>), is 57 213.4 Da. Therefore, the molecular mass of the glycosylation was calculated to be 32 490 Da from the difference in the mass. CAP45 6.2b displayed a main peak at 86 843 Da with smaller peaks at 43 390 and 173 086 Da. Using the same method, the molecular mass of the protein portion was calculated to be 54 802.9 Da, therefore the glycan portion was determined to be 32 040 Da. BAL is therefore 36.2% glycosylated and

CAP45 is 36.9% glycosylated. However, the percentage glycosylation is only an estimate as the total molecular mass of the protein is shown to be very heterogeneous by the broadness of the peaks of the MALDI spectra. This is because the HEK293FS cells do not uniformly glycosylate the proteins.

In both gp120 preparations the presence of a smaller peak at approximately half the molecular mass of the main peak can be seen. These correspond to a fragmented 2+ ion as indicated on the figure whereas the main peak corresponds to an unfragmented 1+ ion. Finally, the presence of a dimeric species is also observed which corresponds to the small peak of a larger molecular mass of 179 797 and 170 386 Da for the BAL and CAP45 species, respectively. The dimeric molecular mass is not precisely twice that of the monomer due to the heterogeneity of the glycosylation.



(A)



(B)

FIGURE 6.2: The (A) gp120 BAL and (B) gp120<sub>ZA</sub> CAP45 strains were analysed by matrix assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry to determine the molecular mass of the glycans. From the mass difference between the protein portion determined using the ProtParam tool and the experimentally observed molecular mass, the molecular mass of the glycans was determined to be (A) 32 490 Da and (B) 32 040 Da. Dimeric species were also observed for both strains.

### Estimation of the composition of the gp120 glycans

Panico *et al.* 2016 mapped the entire glycoprofile of HIV-1 BAL and found that of the 24 glycosylation sites approximately 20 are high mannose glycans, 3 are complex glycans and 1 is a hybrid glycan [205]. High mannose glycans usually contain between 4 and 9 mannose moieties so 6 was taken as an average. The complex glycans frequently consist of a varying number of hexose, N-acetylglucosamine and a fucose moiety and so these were approximated to a ratio of 3:5:1. Finally the hybrid glycans often contain hexose and N-acetylhexosamine which were approximated to a ratio of 5:3. This information was applied to the gp120<sub>ZA</sub>CAP45 used in this study. This gave a total molecular mass of 30 237 Da which left a difference of 1 802 Da according to the mass spectrometry results which showed a mass of 32 040 Da worth of glycans. 1 802.86 Da equates to roughly 10 additional mannose moieties which is possible given the 4-9 potential mannose moieties per high mannose glycosylation site. The total molecular mass was thus 31 710.42 Da which is inline with the mass spectrometry results and therefore a good estimate of the sugar composition with a molecular formula of C<sub>1068</sub>H<sub>2118</sub>N<sub>18</sub>O<sub>1031</sub>. Table 6.2 summarises the sugar composition estimation applied to gp120<sub>ZA</sub>CAP45 expressed by HEK293FS cells.

Sugar	Molecular formula	MM (Da)	Repeats/gp120 monomer
Man	C <sub>6</sub> H <sub>12</sub> O <sub>6</sub>	180.16	130
Hex	C <sub>6</sub> H <sub>12</sub> O <sub>6</sub>	180.16	23
Fuc	C <sub>6</sub> H <sub>12</sub> O <sub>5</sub>	164.16	1
HexNAc	C <sub>8</sub> H <sub>15</sub> NO <sub>6</sub>	221.21	18
<b>TOTAL</b>	<b>C<sub>1068</sub>H<sub>2118</sub>N<sub>18</sub>O<sub>1031</sub></b>	<b>31 710.42</b>	

TABLE 6.2: The glycan composition of gp120 was estimated using data obtained from glycoprofiling of gp120 BAL by Panico *et al.* 2016. gp120 has 24 glycosylation sites of which 20 display high-mannose, 3 display complex and 1 displays hybrid glycans. Man = mannose, Hex = hexose, Fuc = fucose and HexNAc = N-Acetylhexosamine.

### SEC-MALLS-RI shows a monomer/dimer distribution

SEC-MALLS-RI (fig. 6.3) was used following mass-spectrometry to confirm the presence and distribution of the supposed gp120 monomer and dimer of the CAP45 subtype. gp120<sub>ZA</sub>CAP45 was chosen to take forwards in further work, as the SDS-PAGE analysis (fig. 6.1) showed that higher yields of greater purity gp120 were possible using the CAP45 versus the BAL subtype.

The light scattering (orange curve) shows the presence of large particles at 6 mL. Since the UV<sub>280nm</sub> trace (grey curve) and refractive index (blue curve) show a very small peak at this elution volume this suggests that this peak corresponds to a small amount of aggregation. There are two main protein peaks at around 10 and 12 mL of which the first, larger peak is approximately twice the size of the second, smaller peak suggesting that there is twice the concentration of the larger species than the smaller one. The molecular masses were calculated for the two peaks of which the first, larger peak has a molecular mass of around 155 200 Da (+1.4%, yellow crosses) and the second, smaller peak has a molecular mass of around 75 990 Da (+1.5%, green asterisks), corresponding to a gp120 dimer and monomer, respectively.

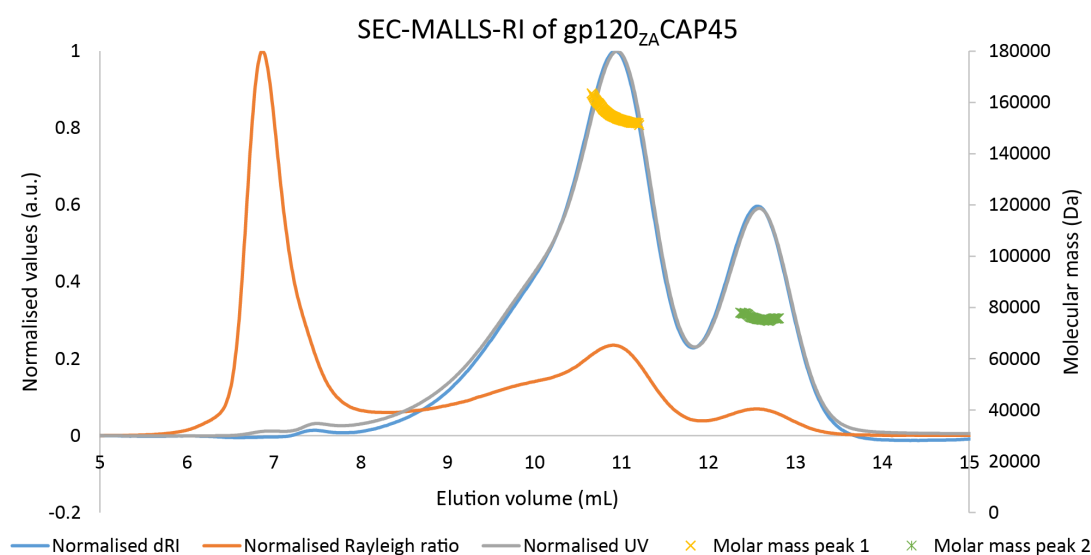


FIGURE 6.3: gp120<sub>ZA</sub>CAP45 was tested using size-exclusion chromatography coupled to multi-angle laser light scattering and refractive index to test the polydispersity of the sample. The light scattering (orange curve) shows a large particle which elutes at 6 mL and two smaller peaks eluting at 10 and 12 mL. The UV trace at 280 nm (grey curve) and refractive index (blue curve) show a very small peak at 6 mL, a large peak at 10 mL and a smaller peak at 12 mL. The molecular mass for the peaks at 10 and 12 mL (yellow crosses and green asterisks, respectively) were in the range of 155 200 Da (+1.4%) and 75 990 Da (+1.5%) which correspond to a gp120 dimer and gp120 monomer, respectively. The column used was an analytical grade Superdex<sup>TM</sup> 200 10/300 (GE Healthcare).

### gp120 dimers can be separated by treatment with DTT and SEC

The monomer and dimer peaks from the SEC-MALLS-RI experiment were pooled separately and re-run on a S200 ag size-exclusion column to observe the separation. The monomeric peak (green box) is shown in the chromatogram in figure 6.4a and it can be seen that there is still some dimer remaining in the pooled monomer sample which suggests that the resolution of



the S200 ag column is not good enough to fully separate the two species. The process was repeated for the dimeric species (fig. 6.4b) from which it can be seen that whilst the dimeric peak (red box) is the main species, there is contamination from the monomer.

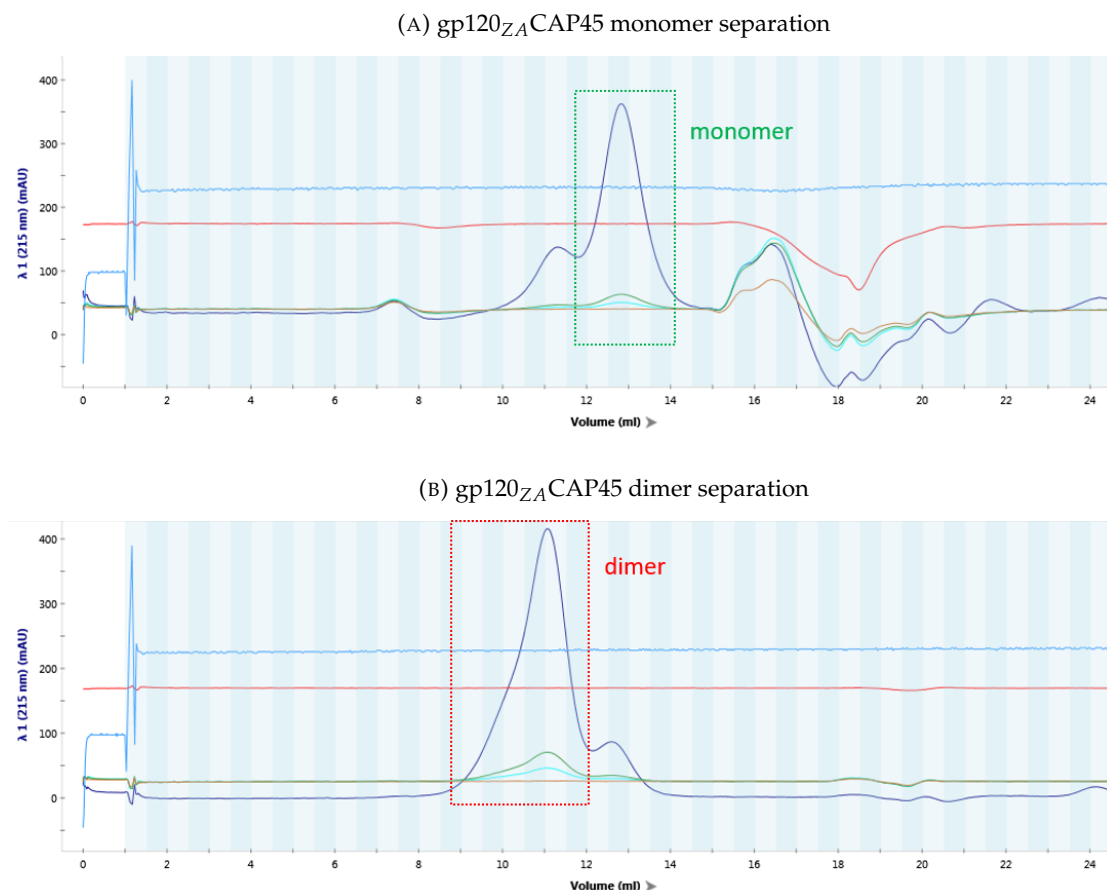


FIGURE 6.4: Following the SEC-MALLS-RI experiment, the gp120 fractions across the monomer and dimer peak were pooled separately and re-loaded onto an analytical grade Superdex™ 200 10/300 (GE Healthcare) size-exclusion column to further purify the monomeric (A, green box) and dimeric (B, red box) species. Incomplete separation is observed in the chromatograms. Dark blue curve = UV absorbance at 215 nm, green curve = UV absorbance at 280 nm, cyan curve = UV absorbance at 260 nm, red curve = conductivity, light blue line = temperature in degrees Celsius.

In an attempt to improve the monomer:dimer ratio, CAP45 was incubated with 5 mM DTT for 1 hour prior to separation on an analytical grade S200 size-exclusion column (fig. 6.5a). This was compared to SEC purification of CAP45 in the absence of DTT (fig. 6.5b). With the addition of DTT the proportion of CAP45 monomer increases, shown by the increase in the size of the monomeric peak. However, this causes further "blurring" of the monomeric and dimeric peaks into one another due to the limited resolution of the column. Therefore, whilst it is possible to enrich the sample with monomer with the addition of DTT, it appears to make the subsequent separation by SEC more complicated.

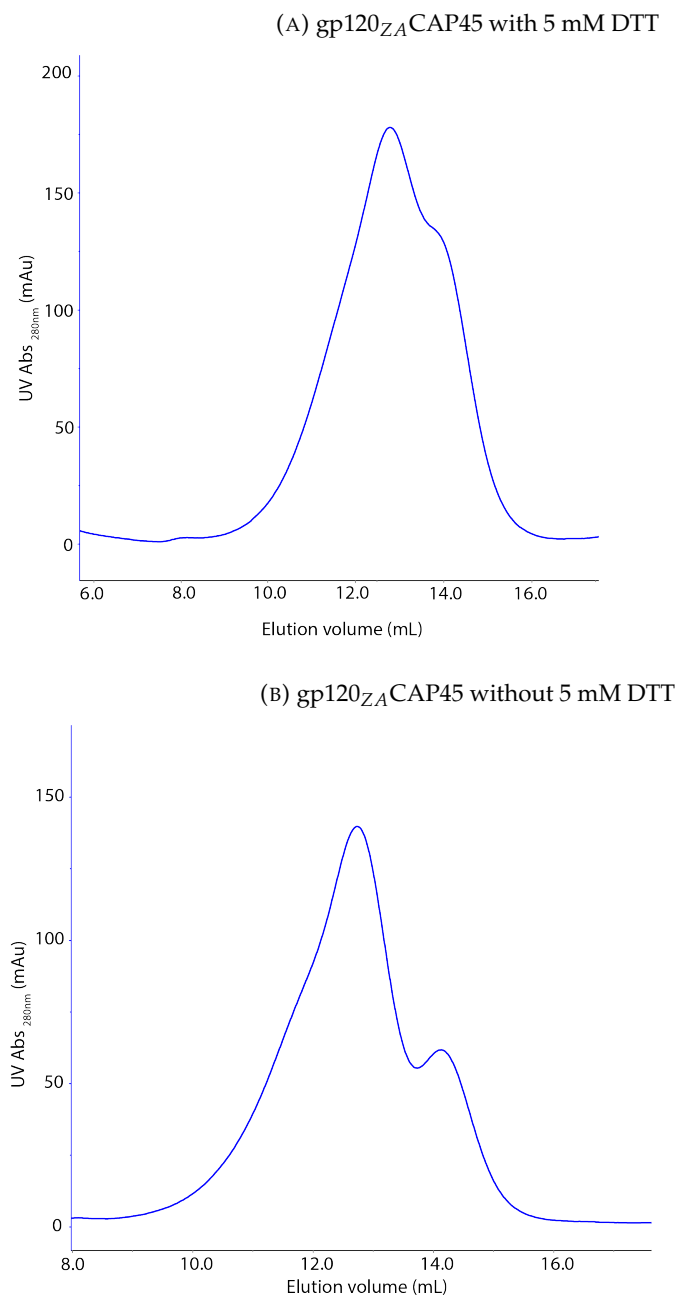


FIGURE 6.5: Treatment with 5 mM DTT followed by purification by size-exclusion chromatography using an analytical grade Superdex<sup>TM</sup> 200 10/300 (GE Healthcare) column was used to increase the yield of gp120 CAP45 monomers. (A) With treatment of 5 mM DTT there is an increase in the proportion of monomeric species present, however the separation resolution of the column is not sufficient. (B) In the absence of DTT there is significantly more dimer present than monomer and the peaks overlap.

## 6.4 Part II: SANS analysis of d-2dCD4-WT, h-gp120 and the complex

Three SANS experiments were carried out:

1. Determination of the protiated glycoprotein 120 (h-gp120) match-out point
2. Determination of the deuterated wild-type two-domain CD4 (d-2dCD4-WT) match-out point
3. Use of the match-out point data to exploit contrast variation to determine the scattering structure of the complex and its individual components.

The experimental parameters are detailed in table 6.3 and the details of each experiment are explained thereafter.

Parameter	Value
Wavelength (Å) ( $\frac{\Delta\lambda}{\lambda}$ )	6 (+-10%)
Detector distances (m)	1.4, [5.6 (gp120 CMP only)] & 8.0
Collimation (m)	2.0, [2.8 (gp120 CMP only)] & 8.0
Q-range (Å <sup>-1</sup> )	0.036 - 0.62
Temperature	Ambient
Cuvette path length (mm)	1
Beam shape	Rectangular

TABLE 6.3: The experimental set-up for the D22 instrument at the ILL.

### 6.4.1 Match-out point determination of protiated gp120

Native gp120 is highly glycosylated; however the glycosylation pattern is dependent on several factors including the HIV clade and subtype from which the gp120 originates and the mammalian cell line in which it is expressed [206–208]. The protiated gp120 used in this experiment (gp120<sub>ZA</sub> CAP45, referred to henceforth as h-gp120) is derived from HIV-1 clade C and recombinantly expressed using HEK293FS cells.

Due to the extent of glycosylation (36.9%), h-gp120 has two phases (components): one of protein and one of carbohydrate. It was therefore expected that the global match-out point of h-gp120 would fall somewhere between the match-out point for that of the protein phase and that of the sugar phase. The protein phase of h-gp120 should have a match-out point of around 42% D<sub>2</sub>O, as is characteristic for non-glycosylated proteins, whereas the sugar phase should have a higher match-out point due to a higher density of hydrogens of which many are bound

to oxygen, are therefore labile and can exchange with the deuterium in the solvent. In order to determine the global match-out point of h-gp120 (including the protein and sugar portions), the SANS signal was measured for h-gp120 in a series of D<sub>2</sub>O contrasts.

### **Sample preparation**

1 mL purified gp120<sub>ZA</sub> CAP45 (as described in section 6.3.1) was split into two equal volumes (500  $\mu$ L) and dialysed into 0% D<sub>2</sub>O and 100% D<sub>2</sub>O 20 mM Tris pH 7.5, 50 mM NaCl buffer, respectively. The 500  $\mu$ L 0% and 500  $\mu$ L 100% D<sub>2</sub>O CAP45 were then mixed in different ratios to give the following contrast series: 0, 20, 40, 60, 80 and 100% D<sub>2</sub>O. Due to low sample volumes the 0% and 100% contrasts were measured first and the samples mixed in different ratios to make 190  $\mu$ L sample volumes, thereafter.

### **Experimental contrast match-point determination of h-gp120**

After data reduction, merging of the data collected at 5.6 m and 2.0 m detector distances and subtraction of the scattering contribution from the buffer, the square root of the forward intensity at 0 angle ( $I(q=0)$ ) for each of the contrasts (corrected for concentration, cuvette thickness and transmission) were plotted against D<sub>2</sub>O% to determine the global match-out point for h-gp120 (fig. 6.6).

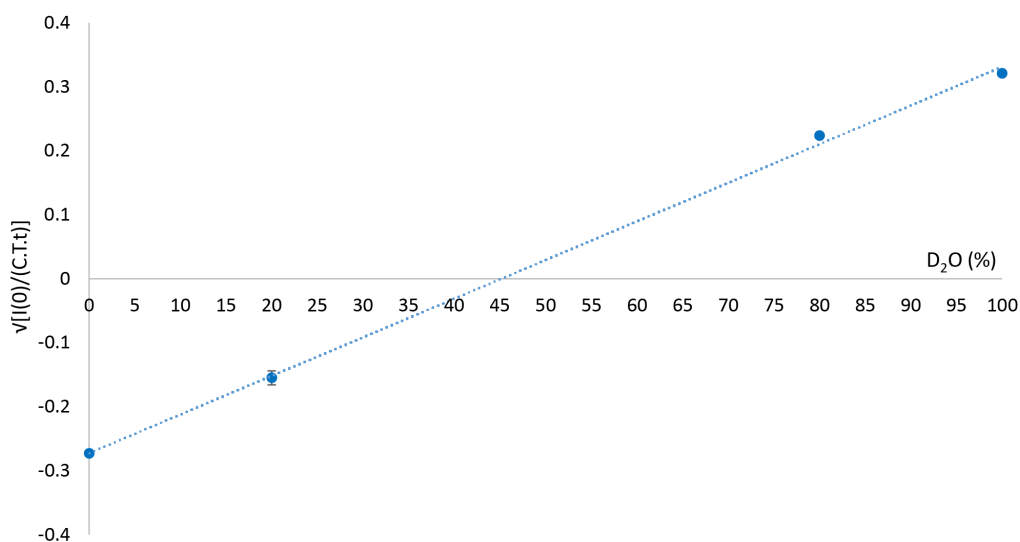


FIGURE 6.6: Protiated gp120<sub>ZA</sub> CAP45 was measured by SANS in a series of D<sub>2</sub>O contrasts. The square root of the  $I(0)$  determined from the Guinier region of each contrast, corrected for the concentration and transmission of the sample and the thickness of the cuvette, was plotted against the percentage of D<sub>2</sub>O. The match-out point of protiated gp120<sub>ZA</sub> CAP45 is found at the x-axis intercept ( $y=0$ ) at 45.2% D<sub>2</sub>O.

The linear fit shows a contrast match-point of 45.2% for h-gp120.

#### 6.4.2 Determination of the contrast-match point of deuterated 2dCD4-WT

The Deuteration Laboratory, through extensive experimentation, has developed a series of standard protocols for perdeuteration and match-out labelling of biological macromolecules necessary for neutron crystallography, and small-angle neutron scattering or neutron reflectivity experiments, respectively. These standard protocols are predominantly based on the use of the *Escherichia coli* and *Pichia pastoris* expression systems [171].

During a standard contrast variation SANS experiment of a protein-protein complex in which a hydrogenated protein is complexed with a deuterated protein produced in *E. coli*; the hydrogenated protein is usually matched-out at approximately 42% D<sub>2</sub>O and the recombinant protein is deuterated to a level of approximately 75% giving a match-out point at or close to 100% D<sub>2</sub>O. This 75% deuterium incorporation is possible by using a minimal growth media containing 85% D<sub>2</sub>O and protiated glycerol. Perdeuterated protein can be used for SANS studies of a protein in order to increase the contrast of the perdeuterated protein relative to the

solvent. However contrast-matching of the protein is not possible when it is perdeuterated because the scattering length density is higher than that of 100% D<sub>2</sub>O. In other words the SLD of the solvent is lower than that of the protein and it cannot be matched-out.

CFPE, as used during this experiment, functions by setting up an *in vitro* reaction for protein expression whereby the protein production machinery (ribosomes) is harvested from *E. coli* grown in hydrogenated conditions and the remaining 'ingredients' for protein synthesis are added (DNA, tRNA, amino acids etc. as detailed in chapter 3). By substituting hydrogenated amino acids for deuterated ones the resulting protein should also be deuterated, with some back protonation due to the use of light water solvents. Therefore, the extent to which the 2dCD4-WT would be deuterated and the SLD and the resulting CMP was unknown. The percentage deuteration of d-2dCD4-WT was calculated to be in the region of 73% D<sub>2</sub>O, as shown by mass spectrometry (see chapter 4), so the CMP was predicted to be slightly lower than 100% D<sub>2</sub>O solvent. Perdeuteration of 2dCD4-WT would have added additional complexity to the deuteration process using CFPE as the reaction would need to have been dialysed against a deuterated buffer to avoid back-protonation. Since perdeuteration was not necessary this avenue was not explored.

### Sample preparation

In order to test the experimental match-out point, 1 mL deuterated 2dCD4-WT produced using CFPE at 1.5 mg/mL was divided equally between two midi-GEBA flex (6-8 MWCO, Millipore-Merck) and dialysed into 100 mL 0% and 100% D<sub>2</sub>O 20 mM Tris pH 7.5, 150 mM NaCl, respectively. Dialysis was carried out at 8°C, changing the dialysis buffer once and dialysing for at least 2 hours each time. After dialysis, the concentrations of 2dCD4-WT in 0 and 100% D<sub>2</sub>O were 1.8 and 1.2 mg/mL, respectively. Due to sample availability limitations, the d-2dCD4-WT at the 0 and 100% D<sub>2</sub>O contrasts were measured first and the subsequent D<sub>2</sub>O contrasts (20, 40, 60 and 80%) were made by mixing the 0 and 100% D<sub>2</sub>O d-2dCD4-WT in the correct ratios thereafter.

### Experimental contrast-match point determination of deuterated 2dCD4-WT

The neutron scattering data were measured for d-2dCD4-WT at detector distances of 1.4 and 8 m for 0, 20, 80 and 100% D<sub>2</sub>O. Due to an error in the experimental sequence only 8 m data were collected for 40 and 60% D<sub>2</sub>O. After the data reduction, the 1.4 and 8 m data were merged and the buffer contribution was subtracted from that of the sample. The I(0) data were extracted from the Guinier analysis for the 0, 20, 80 and 100% data. The square root of I(0) corrected for the thickness of the cuvette and the concentration and transmission of the sample was plotted against the percentage D<sub>2</sub>O as shown in figure 6.7. From the match-out point plot the y=0 intercept is found at 90.5% D<sub>2</sub>O. Data treatment was carried out using the PRIMUS programme of the ATSAS suite of software.

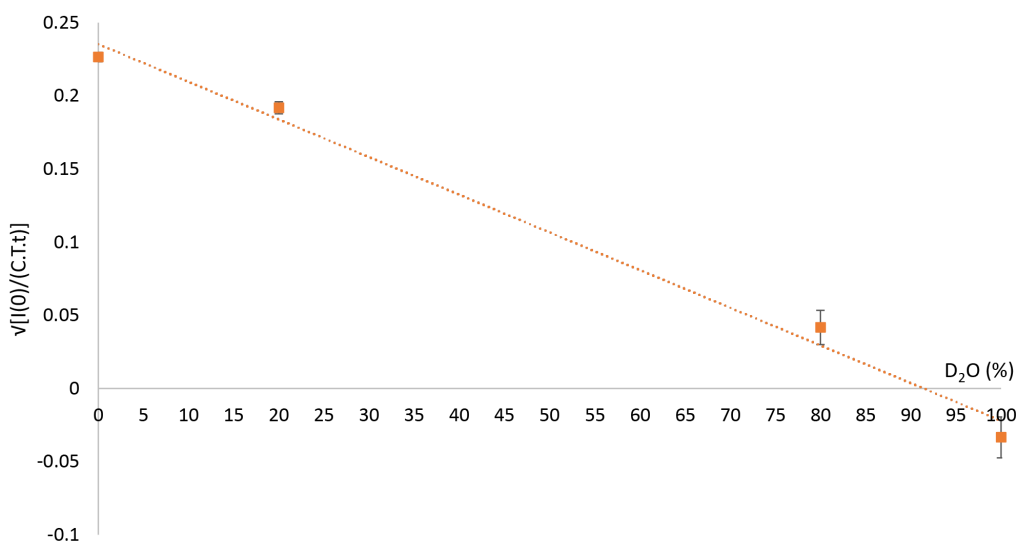


FIGURE 6.7: Deuterated wild-type two-domain CD4 (d-2dCD4-WT) was measured by SANS in a series of D<sub>2</sub>O contrasts. The square root of the I(0) determined from the Guinier region of each contrast, corrected for the concentration and transmission of the sample and the thickness of the cuvette, was plotted against the percentage D<sub>2</sub>O. The match-out point of d-2dCD4-WT is found at the x-axis intercept (y=0) at 90.5% D<sub>2</sub>O.

### 6.4.3 Contrast variation analysis of the h-gp120/d-2dCD4-WT complex

Having experimentally determined the match-out points for the deuterated 2dCD4-WT expressed in the CFPE system (90.5% D<sub>2</sub>O) and the hydrogenated, fully glycosylated gp120 expressed by HEK293 mammalian cells (45.2%); the d-2dCD4-WT/h-gp120 complex could be analysed using contrast variation.

## Sample preparation

### Preparation of h-gp120 monomer/dimers and gp120-unbound d-2dCD4-WT in different D<sub>2</sub>O

**contrasts:** h-gp120 monomers and dimers were prepared in 0% D<sub>2</sub>O buffer by SEC as described in section 6.3.1 and half was dialysed into 90.5% D<sub>2</sub>O buffer. d-2dCD4-WT was prepared in 0% D<sub>2</sub>O buffer by SEC. After measurement by SANS, the d-2dCD4-WT was dialysed into 45.2% D<sub>2</sub>O buffer.

### Preparation of the d-2dCD4-WT/h-gp120 complex in different D<sub>2</sub>O contrasts:

The d-2dCD4-WT/h-gp120 complex was formed by incubation of d-2dCD4-WT with monomer enriched h-gp120 at a ratio of 4:1 based upon molecular mass for 1 hour at room temperature. Monomer enriched h-gp120 was prepared as described in section 6.3.1 by addition of 5 mM DTT to the monomer/dimer mixture prior to purification by SEC.

The CD4-gp120 complex was then purified from the heterogeneous mixture by size-exclusion chromatography using a preparative grade Superdex<sup>TM</sup> 200 16/600 column attached to an ÅKTA purifier (both GE Healthcare), equilibrated in 50 mM Tris-Cl pH 7.5, 150 mM NaCl, 5% sucrose, 45.2% D<sub>2</sub>O. The protein was eluted at a flow rate of 1 mL/min into 1 mL fractions which were subsequently analysed by SDS-PAGE.

After measurement in the 45.2% D<sub>2</sub>O buffer, the complex was re-purified by SEC to eliminate any aggregation that may have accumulated during the SANS measurement and simultaneously exchange the buffer into the 0% D<sub>2</sub>O buffer. Unfortunately, there was not enough measurement time to measure the complex in the 90.5% D<sub>2</sub>O buffer.

## Data treatment

The buffer and sample curves measured at 1.4 m and 8 m were merged after which the background scattering was subtracted from that of the protein by subtracting the buffer scattering curve from that of the sample. The Guinier analysis and the Kratky analysis normalised against the  $R_g$  was carried out using the SCATTER software.  $P(r)$  analysis was carried out using the GNOM programme [202] from the ATSAS package [199]. For the MM calculation of d-2dCD4-WT the SLD was calculated using the MULCH online calculator (<http://smb-research.smb.usyd.edu.au/NCVWeb/input.jsp>).



### Analysis of the size parameters from the SANS data

The samples measured are summarised in table 6.4 and includes the size parameters determined from the SANS data. The  $R_g$  determined from the Guinier plot and the  $R_g$  determined from the  $P(r)$  function for each data set are mostly consistent ( $\pm 2\text{\AA}$ ) which is indicative of an accurate determination of the  $R_g$ .

The discrepancy shown by the  $R_g$  for d-2dCD4-WT in 45.2% D<sub>2</sub>O could be due to aggregation which skew the data at low- $q$  and thus affects accurate determination of the  $R_g$ . Alternatively this could be because the contrast between the d-2dCD4-WT and 45.2% D<sub>2</sub>O solvent is low, therefore the signal is weak. The error is also large for the complex in 45.2% D<sub>2</sub>O. Since the gp120 is matched out at this contrast, the scattering in this sample should be from the d-2dCD4-WT alone. Therefore, the data is also noisy since the relative concentration of d-2dCD4-WT is low and the contrast between the d-2dCD4-WT and 45.2% D<sub>2</sub>O solvent is low as mentioned previously.

In addition, the size parameters for each of the samples are mostly consistent with respect to the published data in table 6.1. The results obtained for the complex fall between those of the two complexes published with monomeric and dimeric gp120 and therefore may represent an average of the two. Since the complex was in a monomer/dimer mixture enriched with monomer, this could explain why the size parameters determined in this SANS experiment fall between those published by Ashish *et al.* 2006 and Guttman *et al.* 2012 who studied the gp120 dimeric and monomeric complexes, respectively [72, 197].

Finally, the size parameters for each sample at their respective contrasts are also consistent ( $\pm 2\text{\AA}$ ), except for the complex. This is to be expected since the gp120 is matched out at the 45.2% D<sub>2</sub>O contrast, as mentioned previously. Thus the size parameters for the complex in 45.2% D<sub>2</sub>O should be comparative to those for d-2dCD4-WT. Since the  $R_g$  and  $D_{max}$  of the gp120-bound d-2dCD4-WT in 45.2% D<sub>2</sub>O are smaller than that for the d-2dCD4-WT in isolation. This could be indicative of a change in the compactness or conformation of d-2dCD4-WT upon gp120 binding.

Sample	D <sub>2</sub> O (%)	Absolute I(0) (I[0]/c) (cm <sup>-1</sup> )	MM (kDa)	R <sub>g</sub> (error) (Å)	R <sub>g</sub> real (Å)	D <sub>max</sub> (Å)
d-2dCD4-WT	0	0.058	20.2	20.2 (+-1.7)	22.1	77
d-2dCD4-WT	45.2	0.013	13.7	25.5 (+-3.0)	21.9	78.5
gp120 monomer	0	0.058	144.1	33.9 (+-1.8)	34.4	123
gp120 monomer	90.5	0.050	94.8	32.7 (+-0.5)	33.6	121.5
gp120 dimer	0	0.122	299.9	44.6 (+-0.8)	46.6	158
gp120 dimer	90.5	0.114	217.5	44.8 (+-1.4)	45.3	160
Complex	0	0.098	134.1	40.4 (+-0.9)	41.8	141
Complex	45.2	0.025	13.4	21.6 (+-4.1)	19.4	70

TABLE 6.4: SANS data were collected on deuterated wild-type two-domain CD4 (d-2dCD4-WT) in isolation and in complex with unlabelled gp120 (h-gp120) in 0 and 45.2% D<sub>2</sub>O contrasts of which the latter is the match-out point of the h-gp120. SANS data were also collected on the h-gp120 monomers and dimers in isolation in 0 and 90.5% D<sub>2</sub>O contrasts of which the latter is the match-out point of the d-2dCD4-WT. The R<sub>g</sub> and D<sub>max</sub> size parameters were calculated from the Guinier plot and P(r) function. The R<sub>g</sub> all within the limit  $qR_g \leq 1.3$ . Finally the molecular mass was calculated from the I(0) extrapolated from the Guinier plot.

### Molecular mass analysis from the SANS data

The MM was calculated from the I(0) obtained by SANS using equation 2.4.4 as described in Chapter 2 and is also shown for each of the samples in 6.4. The online contrast variation calculator: MULCh [209] was used to determine the  $\Delta\rho_M^2$  term of equation 2.4.4 for gp120 using the glycan composition estimate shown earlier. The MM for d-2dCD4-WT in 0% is close to the expected value of 23.7 kDa for the deuterated protein but this is not the case for gp120-bound and -unbound 2dCD4-WT in 45.2% D<sub>2</sub>O for which the MM is 20 kDa smaller than expected. Conversely the MM for the gp120 monomer and dimer in 0% D<sub>2</sub>O are significantly larger than the 86.8 kDa and 173.6 kDa expected. Finally the complex in 0% D<sub>2</sub>O is 24 kDa larger than the 110.5 kDa.

Given the broadness of the mass spectrometry peak for gp120 and the subsequent estimation of the nature of the sugars, it is unsurprising that there are large discrepancies in the theoretical and experimental molecular mass as determined by SANS in samples containing gp120. For d-2dCD4-WT the difference in the experimental and theoretical mass at 0% D<sub>2</sub>O may be down to the poor signal-to-noise ratio which made I(0) determination from the Guinier region non-trivial. For d-2dCD4-WT in 45.2% D<sub>2</sub>O and when bound to gp120 in 45.2% D<sub>2</sub>O the signal-to-noise ratio was even poorer due to the reduced contrasts between a deuterated protein with a CMP of 90.5% D<sub>2</sub>O measured at 45.2% D<sub>2</sub>O. The accuracy of the MM calculation from SAS data is approximately 10% in any case because of the limitations of concentration determination by

UV absorbance at 280 nm [163].

### Analysis of the graphical representation of the d-2dCD4 bound and unbound SANS data

The graphical representation of the data for d-2dCD4-WT and the complex is shown in figure 6.8. Figures 6.8a and 6.8b show a double-logarithmic representation of the one-dimensional (1D) SANS curves. In this format it is clear that the signal from the d-2dCD4-WT is low and that the sample suffers from a small amount of aggregation, as indicated by the up-turn of the data at low- $q$ . The scattering from the complex in 45.2% D<sub>2</sub>O (fig. 6.8b, purple) mimics that of the d-2dCD4-WT (fig. 6.8a) which shows that the gp120 has been successfully matched-out at this contrast.

The Guinier plots in figures 6.8c and 6.8d show a relatively flat Guinier region, after truncation of the first 20 data points which were indicative of aggregation, with oscillation above and below the Guinier fit (black line) due to the noisiness of the data. The signal was stronger for the complex in 0% D<sub>2</sub>O (fig. 6.8b, orange) as the contrast between the d-2dCD4-WT and the 0% D<sub>2</sub>O was stronger, therefore there is less fluctuation and the Guinier fit passes through the points with greater linear correlation than for the complex in 45.2% D<sub>2</sub>O (purple) and the d-2dCD4-WT in figure 6.8a.

Whilst their  $D_{max}$  are similar, the shapes of the  $P(r)$  functions (fig. 6.8a) for d-2dCD4-WT in 0% (pink) and 45.2% (grey) D<sub>2</sub>O are different. The  $P(r)$  function for d-2dCD4-WT in 45.2% D<sub>2</sub>O shows a large distribution ( $P[r]$ ) of smaller distances ( $r$ ) with a second smaller distribution of larger distances which is indicative of a bi-lobular type structure. However the clear difference seen in the two size distributions present the 45.2% d-2dCD4-WT is far less evident for the 0% d-2dCD4-WT, as there is only a slight bump present as the distribution tends towards  $D_{max}$ . This could be an artefact of poor buffer subtraction due to buffer mismatching, the poor signal to noise ratio or aggregation rather than an actual structural feature of the d-2dCD4-WT.

The shape of the 45.2% complex  $P(r)$  function (fig. 6.8d, purple) is similar to that of the 45.2% d-2CD4-WT. There is an additional bump in the  $P(r)$  of the complex at 45.2% D<sub>2</sub>O which is likely caused by the aggregation, as seen in the double-logarithmic plot. The  $P(r)$  function for the 0% complex shows no obvious features, a fairly even distribution across all distances and

a slightly less steep slope towards  $D_{max}$ . The  $P(r)$  function of the 0% complex is suggestive of a globular structure close to that of a sphere.

Finally, Kratky plots normalised against the  $R_g$  determined from the Guinier range are displayed in figures 6.8g and 6.8h for d-2dCD4-WT and the complex, respectively. As with the correlation between the  $P(r)$  functions there are again clear similarities for the 45.2% d-2dCD4-WT (grey) and 45.2% complex (purple) with an unusual difference compared to the 0% d-2dCD4-WT. The shape of the normalised Kratky of d-2dCD4-WT and the 45.2% complex is indicative of an unfolded or disordered protein as the curves do not form an asymptote to the x-axis at higher  $qR_g$  values. The shape of the normalised Kratky for the 0% complex, however, is indicative of a folded protein as the curve is asymptotic to the x-axis at higher- $q$  values.

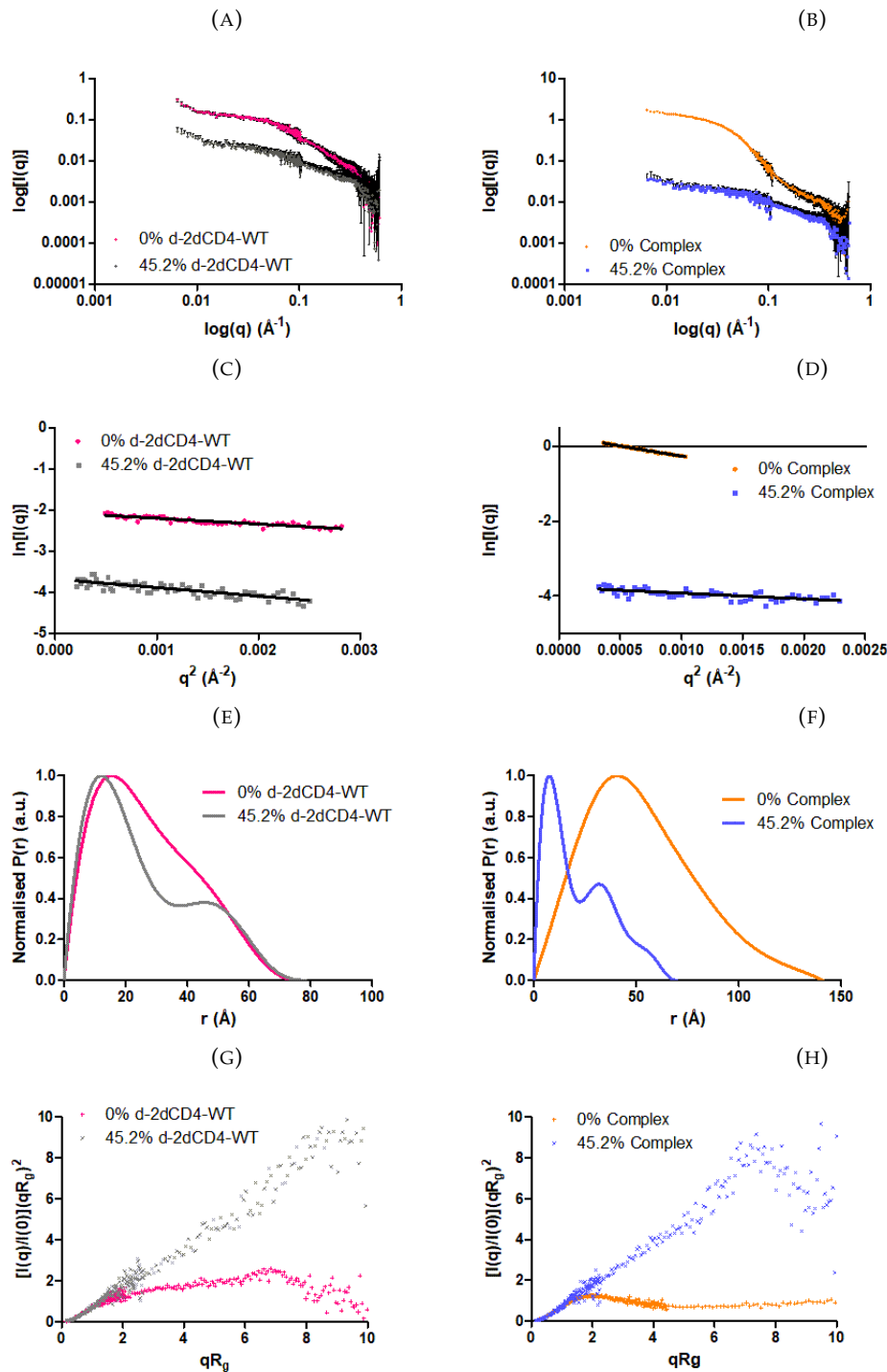


FIGURE 6.8: Graphical representation of the gp120-bound and un-bound d-2dCD4-WT SANS data. (A) and (B) show a double-logarithmic representation of the 1D SANS scattering curves of d-2dCD4-WT and the complex, respectively. (C) and (D) show the Guinier regions for the d-2dCD4-WT (deuterated wild-type two domain CD4) and complex (h-gp120/d-2dCD4-WT), respectively. The  $P(r)$  functions are shown in (E) for d-2dCD4-WT and (F) for the complex. Finally, the Kratky plots normalised against the  $R_g$  determined from the Guinier plot are shown in (G) (d-2dCD4-WT) and (H) (complex). d-2dCD4-WT and the complex in 0%  $D_2O$  solvent are shown in pink and orange, respectively. Whereas d-2dCD4-WT and the complex in 45.2%  $D_2O$  solvent are shown in grey and purple, respectively.

### Analysis of the graphical representation of the gp120 monomer and dimer SANS data

The graphical representation of the SANS data for the gp120 monomer and dimer can be found in figure 6.9. The double logarithmic plots for the monomer and dimer in figures 6.9a and 6.9b indicate that there was a good signal to noise ratio and that there was no aggregation as the curve plateaus and there is no upwards turn of the curve at low- $q$ . The signal is weaker because there is more noise for the samples in 0% D<sub>2</sub>O for both the monomer (yellow) and the dimer (red) due to the large incoherent scattering contribution from the high light water content of the solvent. The linearity of the Guinier fits (figs. 6.9c and 6.9d, black line) are indicative of monodisperse samples absent of aggregation. However, the 0% contrasts are noisier for the aforementioned reason.

The shape of the  $P(r)$  functions are similar for the gp120 monomer (fig. 6.9e) and dimer (fig. 6.9f), although the  $D_{max}$  is larger for the dimer so there is a larger distribution of distances for the dimer than the monomer. The  $P(r)$  functions are indicative of a globular shape close to that of a sphere. The normalised Kratky plots display an almost bell shaped curve that forms an asymptote to the x-axis which is indicative of a well-folded, globular protein.

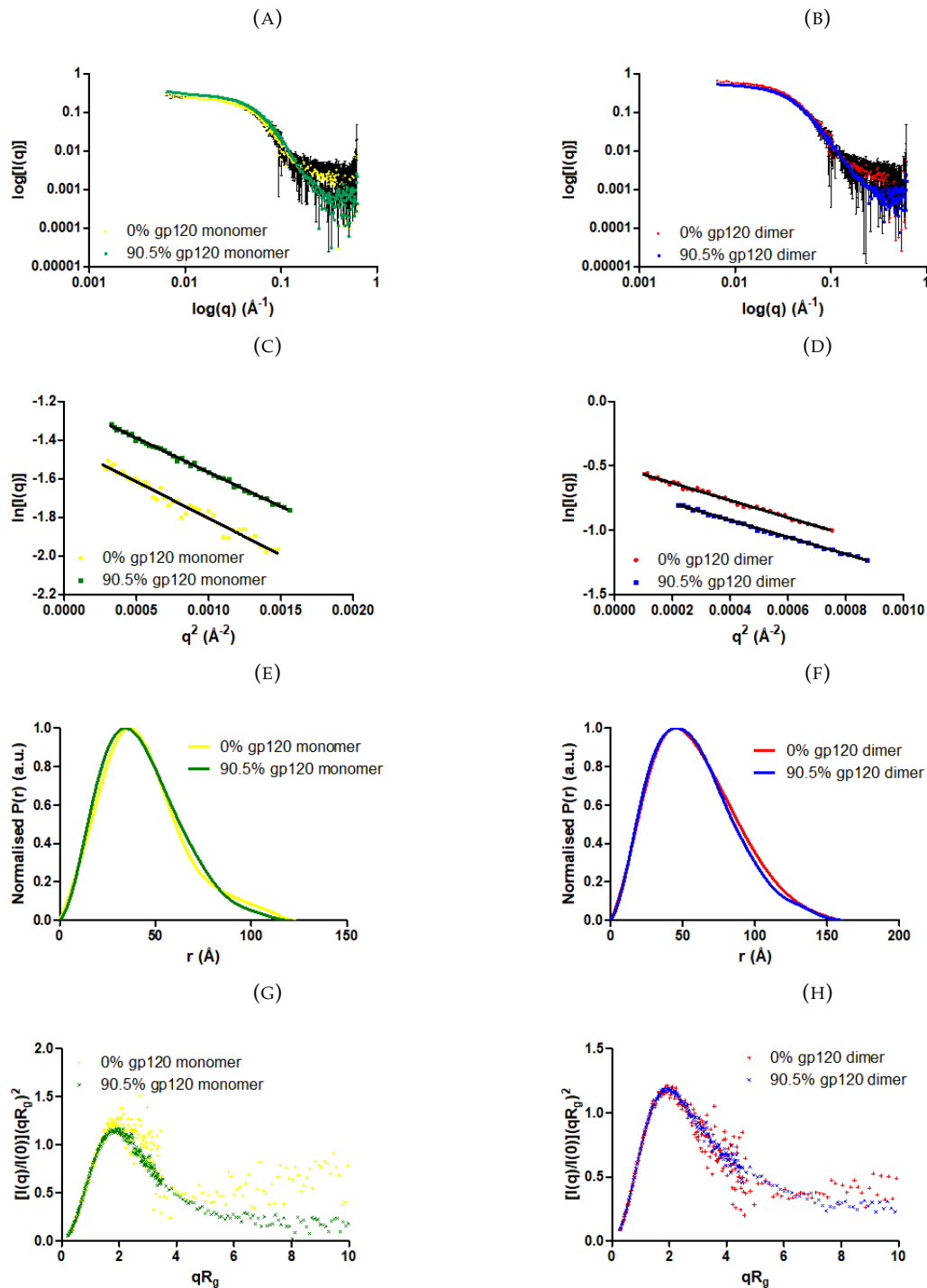


FIGURE 6.9: Graphical representation of the gp120 monomer and dimer SANS data. (A) and (B) show a double-logarithmic representation of the 1D SANS scattering curves of the gp120 monomer and dimer, respectively. (C) and (D) show the Guinier regions for the gp120 monomer and dimer, respectively. The  $P(r)$  functions are shown in (E) for the gp120 monomer and (F) for the gp120 dimer. Finally, the Kratky plots normalised against the  $R_g$  determined from the Guinier plot are shown in (G) (gp120 monomer) and (H) (gp120 dimer). The gp120 monomer and the dimer in 0%  $D_2O$  solvent are shown in yellow and red, respectively. Whereas the gp120 monomer and the dimer in 90.5%  $D_2O$  solvent are shown in green and blue, respectively.

### Comparison of the SANS and X-ray crystal structure data of 2dCD4-WT and the complex

The CRYSON programme [210] from the ATSAS suite of software for SANS data analysis was used to compare the published X-ray crystal structures of 2dCD4-WT and the 2dCD4-WT/gp120 complex to the SANS data collected for d-2dCD4-WT and the d-2dCD4-WT/h-gp120 complex. For d-2dCD4-WT in isolation the PDB entry 3CD4 (2dCD4-WT) was used and for the d-2dCD4-WT/h-gp120 complex the PDB entry 1GC1 (gp120/2dCD4-WT/17b antibody complex) was used with the structural information for the 17b antibody removed.

CRYSON functions by first calculating the theoretical SAS pattern from the crystal structure and then comparing this to the experimental SANS curve. Their similarity is modelled by the  $\chi^2$  value which is calculated by the software. A value close to 1 indicates that the theoretical scattering curve generated from the crystal structure and the experimental SANS curve are identical. The result of the CRYSON comparisons are shown in figure 6.10. The SANS curves for 0% D<sub>2</sub>O (fig. 6.10a, pink) and 45.2% D<sub>2</sub>O (fig. 6.10b, grey) appear to fit relatively well at low- $q$  to the theoretical scattering curve (black line) calculated from PDB entry 3CD4 for d-2dCD4-WT. Above approximately  $0.1 \text{ \AA}^{-1}$  the experimental and theoretical scattering curves diverge. The  $\chi^2$  values are 2.101 and 1.395 for the data at 0% and 45.2% D<sub>2</sub>O, respectively, indicating that the SANS data collected at 45.2% D<sub>2</sub>O is a better fit to the X-ray crystal structure of 2dCD4-WT than the SANS data collected at 0% D<sub>2</sub>O.

The SANS data for the complex in 0% D<sub>2</sub>O (fig. 6.10c, orange) does not fit the theoretical scattering curve calculated from PDB entry 1GC1 and as such a  $\chi^2$  value could not be calculated. This result is unsurprising given the fact that the gp120 measured by SANS was glycosylated with no truncations in any of its variable regions. In addition the 1GC1 PDB entry corresponds to the complex in the presence of the 17b antibody and so, even though the X-ray diffraction coordinates from the 17b had been deleted from 1GC1, the 2dCD4-WT/gp120 complex was locked in the 17b bound conformation. The SANS data collected on the complex at 45.2% D<sub>2</sub>O, which corresponds to the gp120 bound conformation of 2dCD4-WT, was compared uniquely against the 2dCD4-WT component of the 1GC1 PDB entry rather than the 2dCD4-WT/gp120 data of 1GC1 because gp120 was matched out at this D<sub>2</sub>O contrast, as mentioned previously.



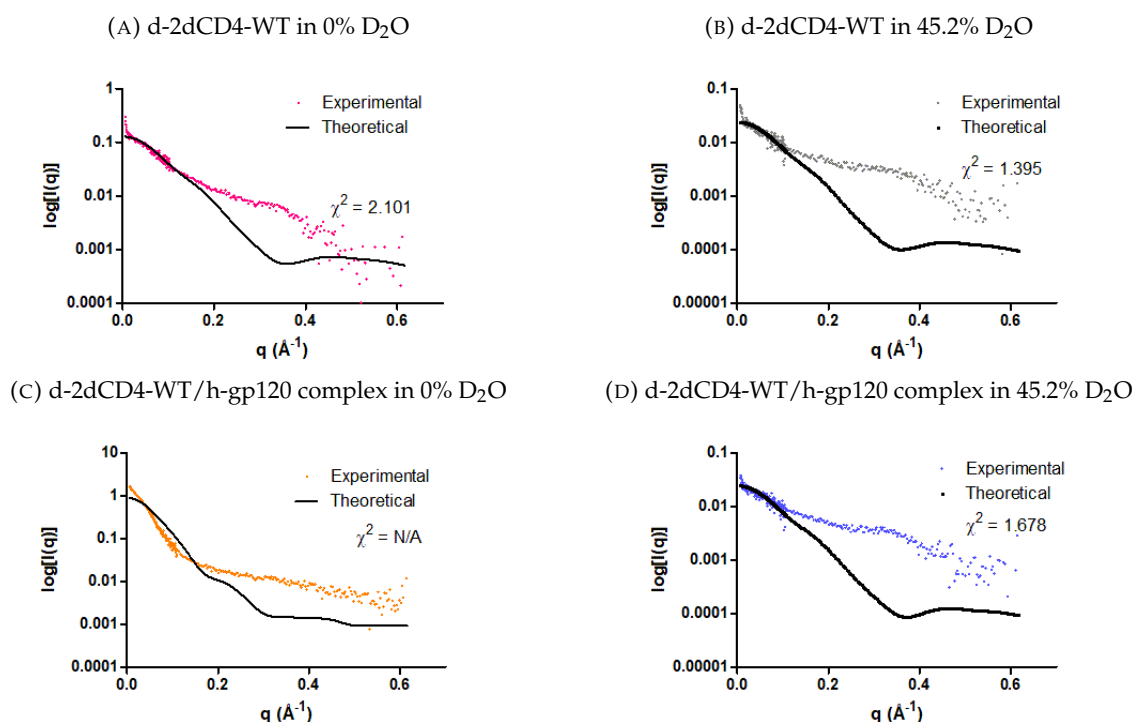


FIGURE 6.10: CRYSON [210] analysis was carried out to compare the SANS data obtained for d-2dCD4-WT in isolation and in complex with gp120 to the theoretical SANS curve generated from the X-ray crystal structure.  $\chi^2$  values demonstrate the closeness-of-fit of the theoretical and experimental data. The experimental SANS data is shown in coloured crosses and the theoretical SANS data is shown as a black line.

Given that the SANS data for gp120-unbound d-2dCD4-WT in 0% and 45.2%  $\text{D}_2\text{O}$  and gp120-bound d-2dCD4-WT in 45.2%  $\text{D}_2\text{O}$  match the theoretical small-angle scattering of the corresponding 2dCD4-WT and gp120-bound 2dCD4-WT X-ray crystal structures at low- $q$ , this suggests that the size and shape of the d-2dCD4-WT with and without gp120 is similar to that of the 2dCD4-WT in the crystal structure (since the low- $q$  data contains information on the size and shape of the particle).

#### *Ab initio* modelling of gp120-bound and -unbound 2dCD4-WT

Given the interesting CRYSON results *ab initio* modelling was carried out to calculate a dummy atom model from the SANS data to which the X-ray crystal structure of 2dCD4-WT could be compared visually. However, considering the uncertainty of the quality of the data from the Kratky plot (which suggests mis-folding or poor buffer subtraction), any results from further processing of the data using *ab initio* modelling should be treated with **extreme** caution when

attempting to draw meaningful conclusions from the data (see section 6.6.3 for a detailed discussion).

The DAMMIF programme [201] from the ATSAS package was used to determine the envelope structure from the  $P(r)$  function which contains the information on the real space distances within the particle. DAMMIF was run in slow mode to improve the reconstruction of the *ab initio* models. 50 iterative repeats were used to produce 50 *ab initio* models which were compared using the DAMAVER [211] suite of programmes to align and average the models and determine their goodness-of-fit. The normalised spatial discrepancy (NSD) value was calculated for pairs of models to find the best alignment in three-dimensions. Any models with an NSD of less than 2 were excluded from the average. This generates a damaver model which is an average of the models with an NSD below 2 and a damfilt model which is filtered to a specific cut-off volume to remove low occupancy and loosely connected dummy atoms.

The *ab initio* model of d-2dCD4-WT in 0% D<sub>2</sub>O is presented in figure 6.11 with the X-ray crystal structure of 2dCD4-WT (PDB entry 3CD4) superimposed using the SUPALM [212] function of the SASpy plugin for PyMOL. The beta-strands are shown in yellow, alpha-helices in red, flexible loops in green and the domain 1 and 2 disulphide bonds are shown as spheres. The grey envelope represents the damaver model. The smaller pink mesh envelope represents the damfilt model. In 0% D<sub>2</sub>O, gp120-unbound d-2dCD4-WT comprises two spherical lobes joined by a bridging region, from which the flexible loop joining beta-strands F and G in domain 1 protrudes from the model (6.11a and 6.11b).

Compared to gp120-unbound d-2dCD4-WT in 0% D<sub>2</sub>O, the gp120-unbound d-2dCD4-WT in 45.2% D<sub>2</sub>O appears similar in shape and size (fig. 6.12). The damaver model is shown again in grey with the damfilt model shown in grey mesh and the X-ray crystal structure is superimposed. The flexible loop between beta-strands F and G in domain 1 is shown to jut out again from the model as with the d-2dCD4-WT in 0% D<sub>2</sub>O. The main difference is that domain 2 appears significantly smaller (fig. 6.12c and 6.12d).

While the bi-lobular structure of the model of gp120-bound d-2dCD4-WT in 45.2% D<sub>2</sub>O (fig. 6.13, damaver model = grey, damfilt model = purple mesh, the crystal structure of the complex is superimposed) generally resembles that of the gp120-unbound d-2dCD4-WT model in 0 and 45.2% D<sub>2</sub>O, there appears to be a slight conformational change on gp120 binding (fig. 6.13b).

Where the two domains (domains 1 and 2) appear linearly stacked in the gp120-unbound models of d-2dCD4-WT, there appears to be a hinge or elbow action in the gp120-bound d-2dCD4-WT which results in domain 2 being slightly rotated around the z-axis with respect to domain 1. Also of note is that domain 2 of the X-ray crystal structure does not fit as well within the dammif model with a significant part of the A and B beta-strands protruding from the model in domain 2 (fig. 6.13a and 6.13b). With a CRYSON calculated  $\chi^2$  value larger than that for gp120-unbound d-2dCD4-WT in 45.2% D<sub>2</sub>O, this supports the suggestion that there is a change on gp120 binding.

The diameter of each lobe in the x and z planes were measured as well as the total length in the y plane of each of the damaver models in WinCoot as summarised in table 6.5.

Model	Domain	Plane	Distance (Å)
0% d-2dCD4-WT	1	x	42
		z	42
	2	x	40
		z	39
	1&2	y	97
	45.2% d-2dCD4-WT	1	x
z			37
2		x	28
		z	28
1&2		y	92
45.2% gp120-bound d-2dCD4-WT		1	x
	z		25
	2	x	18
		z	22
	1&2	y	69

TABLE 6.5: WinCoot was used to measure the maximum distances of the x and z planes for domains 1 and 2 individually and the distance along the y plane spanning domain 1 and domain 2 for each of the damaver *ab initio* models of gp120-unbound and -bound d-2dCD4-WT in 0% and 45.2% D<sub>2</sub>O contrasts, generated using DAMMIF. The *ab initio* model of gp120-bound d-2dCD4-WT was shown to be the smallest in all planes.

From these approximate distances it would indeed appear that d-2dCD4-WT is larger when gp120-unbound and that on gp120 binding the bend observed in figure 6.13b causes the length of the protein to shrink in size. Additionally, there appears to be a shrinking of the width of domain 2, first in 45.2% D<sub>2</sub>O and then on gp120 binding. This is supported by the size parameter determination in table 6.4 which shows that, on gp120-binding, d-2dCD4-WT becomes more compact indicated by a smaller  $R_g$  and  $D_{max}$ .

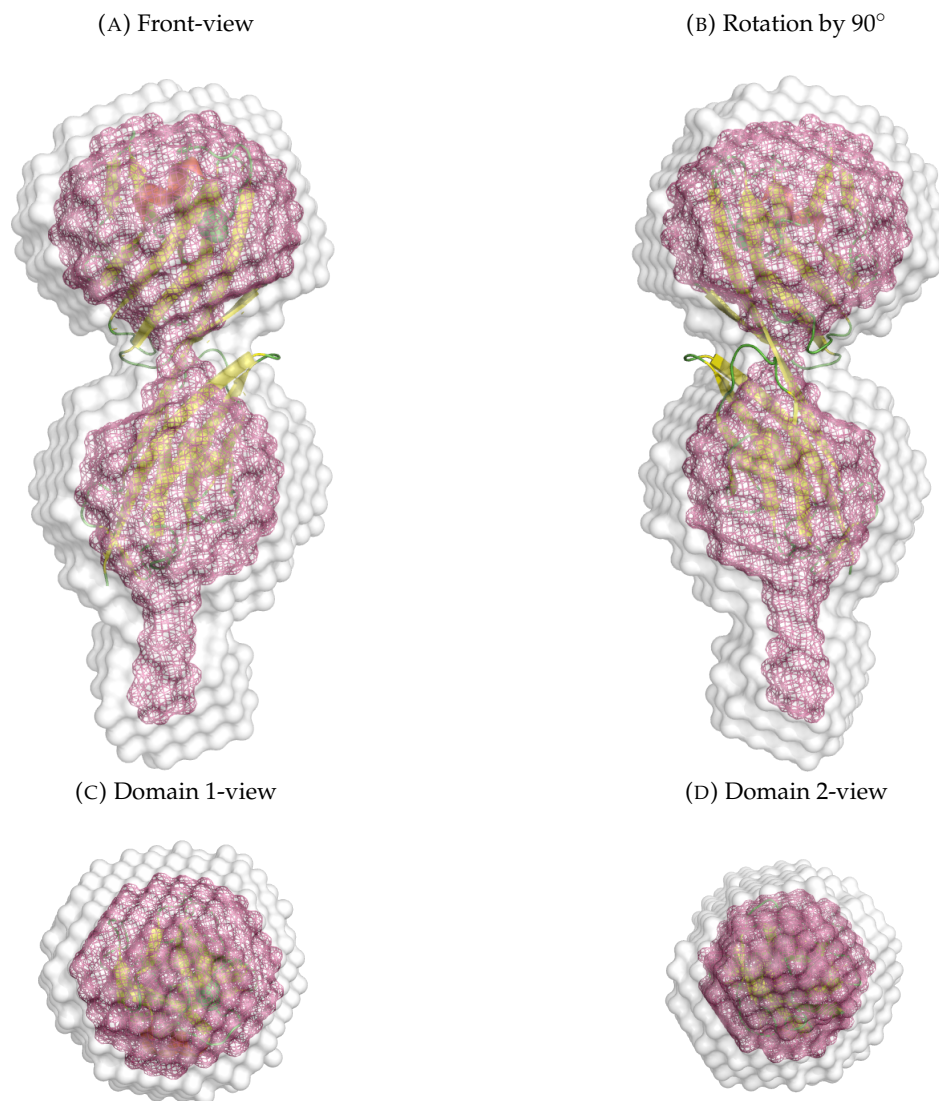


FIGURE 6.11: *ab initio* model of d-2dCD4-WT in 0% D<sub>2</sub>O. A) Front-view, B) rotation by 90°, C) domain 1-view and D) domain 2-view. 50 *ab initio* models were generated using the DAMMIF programme in slow mode and the best-fit models were compared, averaged and filtered by assessing their goodness-of-fit using the DAMAVER suite of programmes. The dammfilt (pink mesh) and damaver (grey surface) models were aligned with the X-ray crystal structure of 2dCD4-WT (PDB entry 3CD4) in PyMOL using supalm of the SASPy plugin. The 2dCD4-WT *ab initio* model shows two lobes of which one comprises domain 1 and a second lobe comprises domain 2. The two lobes are joined by a narrower bridging region which comprises the long G beta-strand in domain 1 which becomes strand A in domain 2. Finally, there is a protrusion from the domain 3 face of domain 2 which is attributable to the uncleaved His-tag. Beta sheets are shown in yellow, alpha helices are shown in red, flexible loops are shown in green and the disulphide bonds are shown as balls.

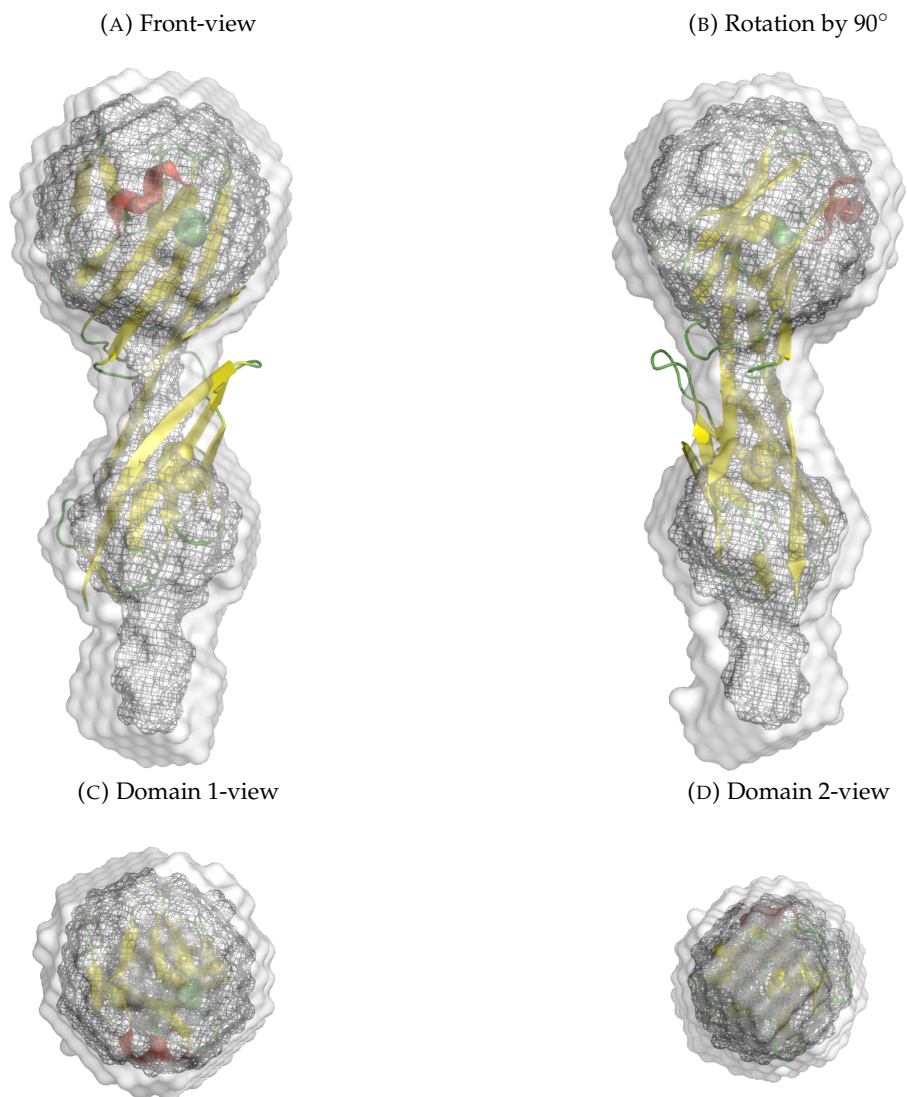


FIGURE 6.12: *ab initio* model of d-2dCD4-WT in 45.2% D<sub>2</sub>O. A) Front-view, B) rotation by 90°, C) domain-1 view and D) domain 2-view. 50 *ab initio* models were generated using the DAMMIF programme in slow mode and the best-fit models were compared, averaged and filtered by assessing their goodness-of-fit using the DAMAVAR suite of programmes. The dammifilt (grey mesh) and damaver (grey surface) models were aligned with the 2dCD4-WT portion of the X-ray crystal structure of the 2dCD4-WT/gp120 complex (PDB entry 1GC1) in PyMOL using supalm of the SASPy plugin. The 2dCD4-WT *ab initio* model shows two lobes joined by a narrower bridging region. One of the lobes comprises domain 1 and a second smaller lobe represents domain 2. The bridging region between the two lobes comprises the long G beta-strand in domain 1 which becomes strand A in domain 2. Finally there is a protrusion at the domain 3 face of domain 2 attributable to the uncleaved His-tag. Beta sheets are shown in yellow, alpha helices are shown in red, flexible loops are shown in green and the disulphide bonds are shown as balls.

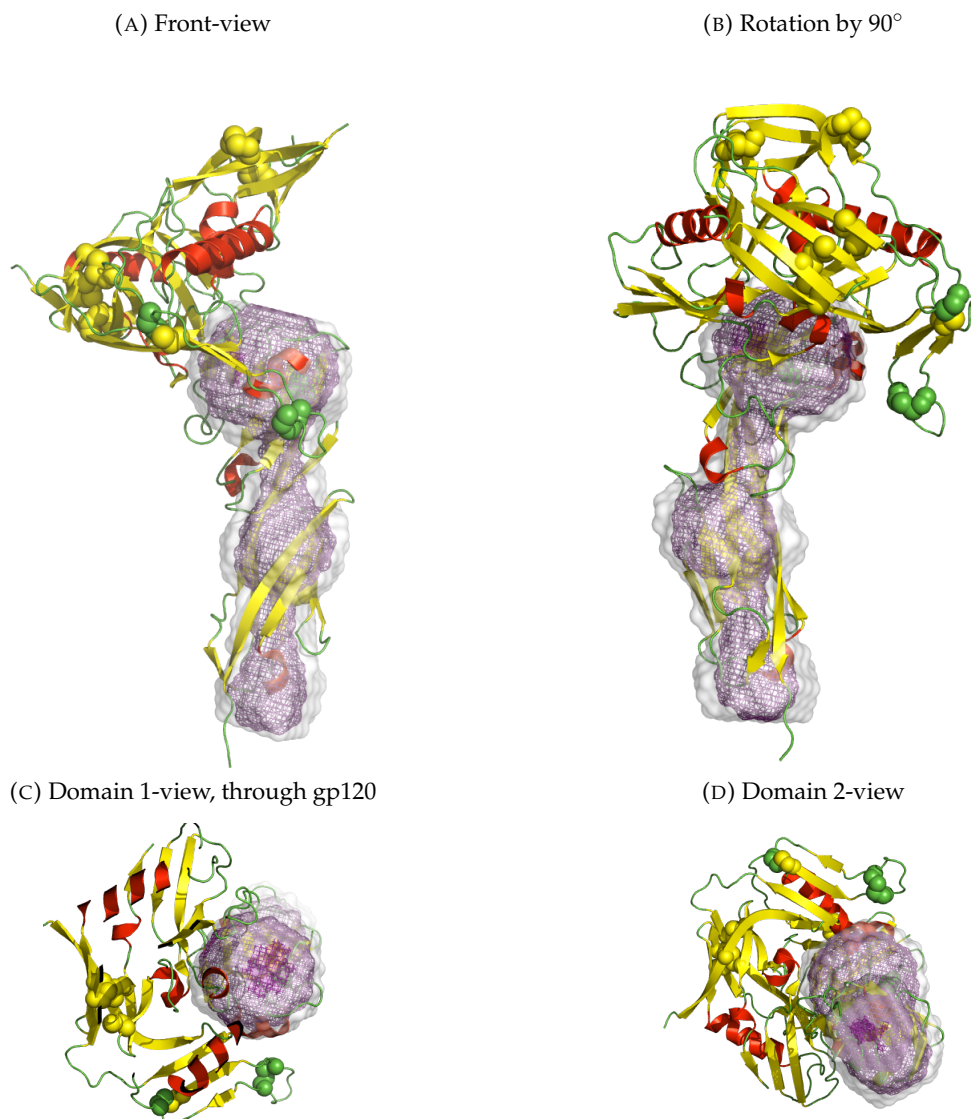


FIGURE 6.13: At 45.2% D<sub>2</sub>O which is the contrast match point of the hydrogenated gp120 (h-gp120), only the neutron scattering from the deuterated 2dCD4-WT (d-2dCD4-WT) can be seen. A) Front-view, B) rotation by 90°, C) top-view through gp120 and D) bottom-view of the *ab initio* model of d-2dCD4-WT in complex with gp120. 50 *ab initio* models were generated using the DAMMIF programme in slow mode and the best-fit models were compared, averaged and filtered by assessing their goodness-of-fit using the DAMAVER suite of programmes. The dammifilt (purple mesh) and damaver (grey surface) models were aligned with the 2dCD4-WT portion of the X-ray crystal structure of the 2dCD4-WT/gp120 complex (PDB entry 1GC1) in PyMOL using supalm of the SASpy plugin. The 2dCD4-WT *ab initio* model shows two lobes of which one comprises domain 1 and is proximal to gp120 and the second smaller lobe comprises domain 2. A wide stalk joins the two lobes and comprises the long G beta-strand in domain 1 which becomes strand A in domain 2. Beta sheets are shown in yellow, alpha helices are shown in red, flexible loops are shown in green and the disulphide bonds are shown as balls.

## 6.5 Part III: Small-angle X-ray scattering analysis of 2dCD4-WT, gp120 and the complex

### 6.5.1 SAXS analysis of samples previously measured by SANS

24 hours after the end of the SANS experiment the samples were transferred to BM29 at the ESRF where they were measured using SAXS, with the intention of corroborating the SANS data that had been collected.

#### Sample preparation

A dilution series of three concentrations was prepared for each sample. The highest concentration was that measured in the SANS experiment was diluted 1:2 and 1:4 in the corresponding buffer to produce the middle and lowest concentrations. The samples were centrifuge filtered using 0.22  $\mu\text{m}$  Spin-X filters (Corning) prior to placing them in the robotic sample changer to remove any precipitant from the sample.

#### Sample measurement

10 frames were measured per sample as it flowed through the capillary exposed to the X-ray beam. After radial averaging, each frame was compared against all others within that sample measurement using the data comparison function of Primus [213] from the ATSAS package [199]. Any frames which differed significantly ( $p \leq 0.05$ ) were excluded from the curve averaging.

The low- $q$  data of the low-concentration data was merged with the mid- $q$  data of the middle concentration and the high- $q$  data of the highest concentration in an attempt to mitigate the effects from concentration-induced aggregation and poor signal from the lowest concentration sample. The background scattering was subtracted from that of the protein by subtracting the buffer curve from the corresponding sample. The resultant curves were analysed using SCATTER to determine the  $R_g$ ,  $I(0)$  and  $R_g$  normalised Krakty plot [200]. GNOM was used to determine the  $P(r)$  function [202].

The experimental parameters are listed in table 6.6.

Parameter	Value
Energy	12.5keV
Detector distance (m)	2.867
Q-range ( $\text{\AA}^{-1}$ )	0.025-60
Temperature	Ambient

TABLE 6.6: The experimental set-up for the BM29 beamline at the ESRF.

### Analysis of the size parameters of the SANS samples determined by SAXS

Table 6.7 outlines the parameters obtained. The  $R_g$  for each sample is relatively consistent between the two D<sub>2</sub>O contrasts and between the  $R_g$  determined from the Guinier region and that determined from the  $P(r)$  function, except for the d-2dCD4-WT in 45.2% D<sub>2</sub>O. The data are consistent between the SANS and SAXS derived size parameters for d-2dCD4-WT and gp120 monomer in 0% D<sub>2</sub>O. For the other samples, the size parameters determined by SAXS are significantly larger than those that were determined from the SANS experiment (table 6.4).

The MM was calculated for the samples using equation 2.4.4 from the  $I(0)$  determined from the Guinier analysis. The MM calculations for the complex are within the 10% accuracy limit [163]. However the MM calculations for all other samples are significantly lower than expected. This indicates that the data is of poor quality. For gp120 this is, again, unsurprising given the estimations made with regards to the sugar content, the incomplete dimer/monomer separation and the heterogeneous distribution of the sugars as shown by mass spectrometry (fig. 6.2). The graphical representation of the d-2dCD4-WT data (below) show that the data quality is very poor and therefore not at all representative of d-2dCD4-WT.

Compared to the published data, the results for d-2dCD4-WT, the gp120 monomer in 0% D<sub>2</sub>O and the complex in 0 and 45.2% D<sub>2</sub>O are similar. For the other samples the size parameters derived from the data collected are larger. The discrepancy between the SANS and SAXS derived size parameters and between the published data can be partly understood by looking at the graphical representations of the data.

### Analysis of the graphical representation of the SAXS data of d-2dCD4-WT

It is apparent from the double-logarithmic representation of the d-2dCD4-WT SAXS data in figure 6.14a that the samples are heavily aggregated. Both the 0% (pink) and 45.2% D<sub>2</sub>O (grey) gp120-unbound d-2dCD4-WT curves are bi-phasic which is suggestive of two populations.



Sample	D <sub>2</sub> O (%)	Absolute I(0) (cm <sup>-1</sup> )	MM (kDa)	R <sub>g</sub> (error) (Å)	R <sub>g</sub> real (Å)	D <sub>max</sub> (Å)
CFPE d-2dCD4-WT	0	3.841x10 <sup>-3</sup>	5.9	23.1 (+-0.7)	24.4	78
CFPE d-2dCD4-WT	45	1.026x10 <sup>-2</sup>	15.8	44.1 (+-1.3)	52.4	240
gp120 monomer	0	2.674x10 <sup>-2</sup>	41.2	36.2 (+-0.2)	36.7	125
gp120 monomer	90.5	2.681x10 <sup>-2</sup>	41.4	38.4 (+-0.3)	38.7	140
gp120 dimer	0	4.803x10 <sup>-2</sup>	74.1	49.9 (+-0.3)	50.4	190
gp120 dimer	90.5	4.831x10 <sup>-2</sup>	74.5	48.8 (+-0.3)	49.4	178
complex	0	6.661x10 <sup>-2</sup>	102.7	46.8 (+-0.3)	45.9	155
complex	45	1.470x10 <sup>-1</sup>	226.6	44.8 (+-0.4)	46.2	158

TABLE 6.7: SAXS data was collected on BM29 at the ESRF samples after measurement by SANS on D22 at the ILL.

The R<sub>g</sub> presented in the Guinier plot (fig. 6.14b) show two clearly different Guinier regions of different gradient which explains the difference in R<sub>g</sub> observed between d-2dCD4-WT in 0 and 45.2% D<sub>2</sub>O. The steepness of the slope of the Guinier region for d-2dCD4-WT in 45.2% D<sub>2</sub>O is indicative of a larger particle, reflected by the larger R<sub>g</sub> of 44.1Å, which is due to the presence of an aggregate of d-2dCD4-WT. Whereas the shallow slope of the Guinier region for d-2dCD4-WT in 0% D<sub>2</sub>O is indicative of a smaller particle reflected by the smaller R<sub>g</sub> of 23.1Å, which could be representative of d-2dCD4-WT as this is close to the R<sub>g</sub> determined by SANS and the published R<sub>g</sub> for 2dCD4.

The P(r) functions in figure 6.14c show the same pattern of a much smaller D<sub>max</sub> for the d-2dCD4-WT in 0% compared to 45.2% D<sub>2</sub>O. The P(r) functions presented are not accurate representations of the d-2dCD4-WT due to the evident polydispersity of the sample shown by the double-logarithmic and Guinier plots (figs. 6.14a and 6.14b). The P(r) function of d-2dCD4-WT in 0% D<sub>2</sub>O (pink) resembles the shape of the P(r) function found for d-2dCD4-WT by SANS (fig. 6.8e), however the curve does not tend smoothly towards the x-axis and ends abruptly suggesting the D<sub>max</sub> may have been underestimated. The P(r) function for d-2dCD4-WT in 45.2% D<sub>2</sub>O (grey) is very bumpy and shows a very large D<sub>max</sub> due to the aggregation. The Kratky plot in figure 6.14d also suggests that the proteins are unfolded. Thus the SAXS results for d-2dCD4-WT are indicative of heavy aggregation and cannot be taken as an accurate representation of the structure of d-2dCD4-WT. Rather, combined with their difference to the SANS results (which were already suggestive of a small amount of aggregation), these results show that d-2dCD4-WT is prone to aggregation over time. There was a 48 hour time delay between the end of the SANS experiment and the beginning of the SAXS experiment therefore, the proteins had time to aggregate.

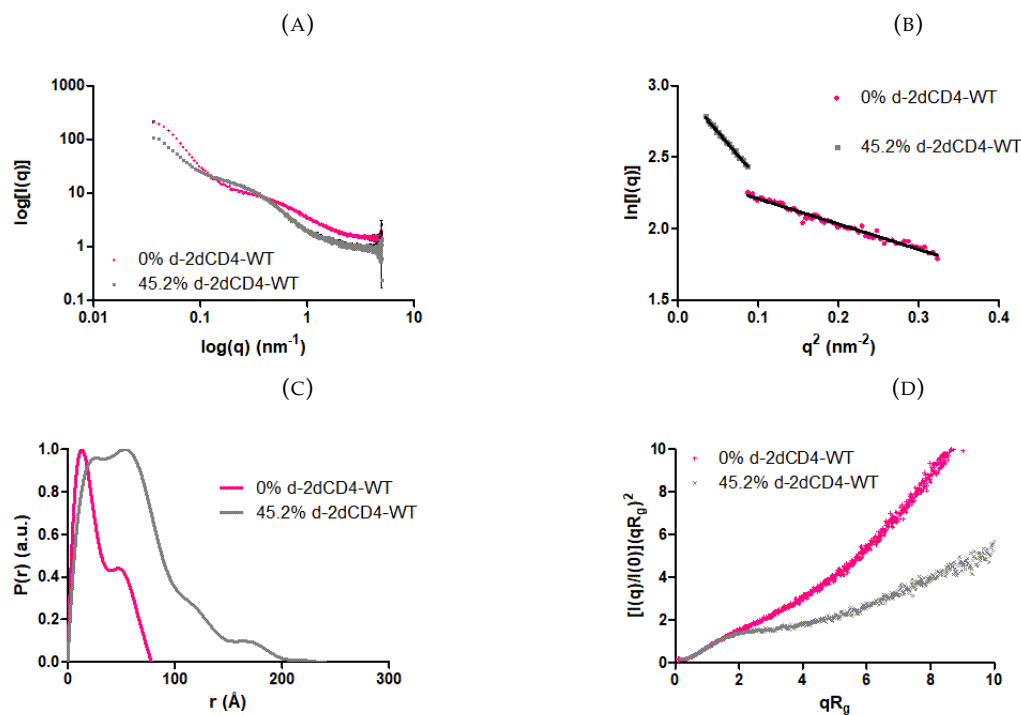


FIGURE 6.14: Graphical representation of the d-2dCD4-WT data measured by SAXS. (A) shows a double-logarithmic representation of the 1D SAXS scattering curves of d-2dCD4-WT (deuterated wild-type two domain CD4). (B), (C) and (D) show the Guinier regions,  $P(r)$  functions and dimensionless Kratky plots, respectively. d-2dCD4-WT in 0 and 45.2%  $D_2O$  solvent are shown in pink and grey, respectively.

### Analysis of the graphical representation of the SAXS data of the d-2dCD4-WT/h-gp120 complex

A small amount of aggregation is observed in the double-logarithmic plot for the complex in 0 and 45.2%  $D_2O$  (fig. 6.15a, orange and purple, respectively) by the up-turn of the data at low- $q$  but not to the same extent as was observed for d-2dCD4-WT in isolation. The Guinier regions (fig. 6.15b) are flat and of the same gradient which is reflected by the similarity in  $R_g$  for the two  $D_2O$  contrasts. The  $P(r)$  functions are the same shape and tend smoothly towards the maximum dimension at  $D_{max}$ . Their shape is indicative of a globular structure, similar to a sphere. Whilst the  $D_{max}$  determined by SAXS is slightly larger than that determined by SANS the shape of the curves is the same. Finally, the dimensionless Kratky plot suggests that both proteins are relatively well folded with what could be a small amount of flexibility, although this could also be an artefact of the aggregation, as the curves tend towards the x-axis but do not asymptote it.

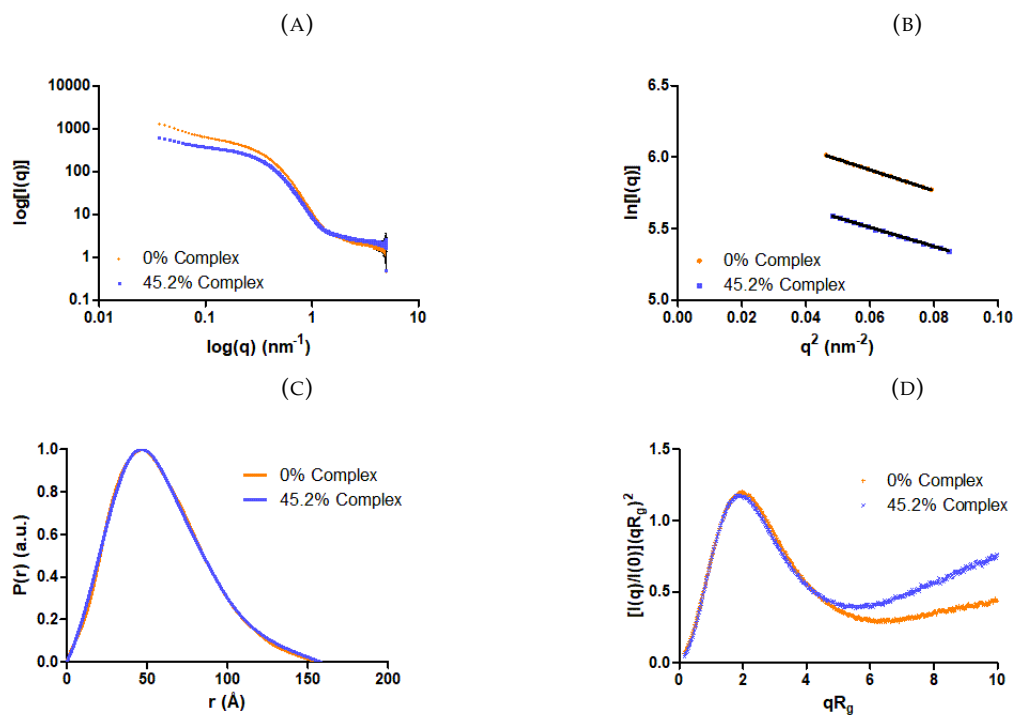


FIGURE 6.15: Graphical representation of the d-2dCD4-WT/h-gp120 complex data measured by SAXS. (A) shows a double-logarithmic representation of the 1D SAXS scattering curves of the d-2dCD4-WT/h-gp120 complex (deuterated wild-type two domain CD4 in complex with protiated gp120). (B), (C) and (D) show the Guinier regions,  $P(r)$  functions and dimensionless Kratky plots, respectively. The complex in 0 and 45.2%  $D_2O$  solvent are shown in orange and purple, respectively.

### Analysis of the graphical representation of the SAXS data of the gp120 monomer and dimer

Again, a small amount of aggregation is suggested by the upwards turn of the double-logarithmic plot of the gp120 monomer in 0 (yellow) and 90.5% (green)  $D_2O$  in figure 6.16a and for the dimer in 0 (red) and 90.5%  $D_2O$  (blue) in figure 6.16b. When compared to the smoothness of the double-logarithmic plots of the 1D SANS data (fig. 6.9a and 6.9b, for the monomer and dimer, respectively) at low- $q$ , the presence of this slight upwards turn is clear. The Guinier regions for both the monomer (fig. 6.16c) and the dimer (fig. 6.16d) are flat and of relatively consistent gradient, reflected by the similarity in the  $R_g$  determined for both  $D_2O$  contrasts for the monomer and dimer.

Whilst the  $P(r)$  for both the monomer and dimer (figs. 6.16e and 6.16f, respectively) appear very similar in shape as those obtained by SANS (figs. 6.9e and 6.9f, respectively) the  $D_{max}$  is globally larger for both the monomer and dimer as determined by SAXS. Finally, the Kratky plots in figures 6.16g and 6.16h are indicative of globular, folded proteins with a small amount

of flexibility as the curves tend towards the x-axis but do not form an asymptote with it. Again this could also be an artefact of the small amount of aggregation seen in the double-logarithmic plots.

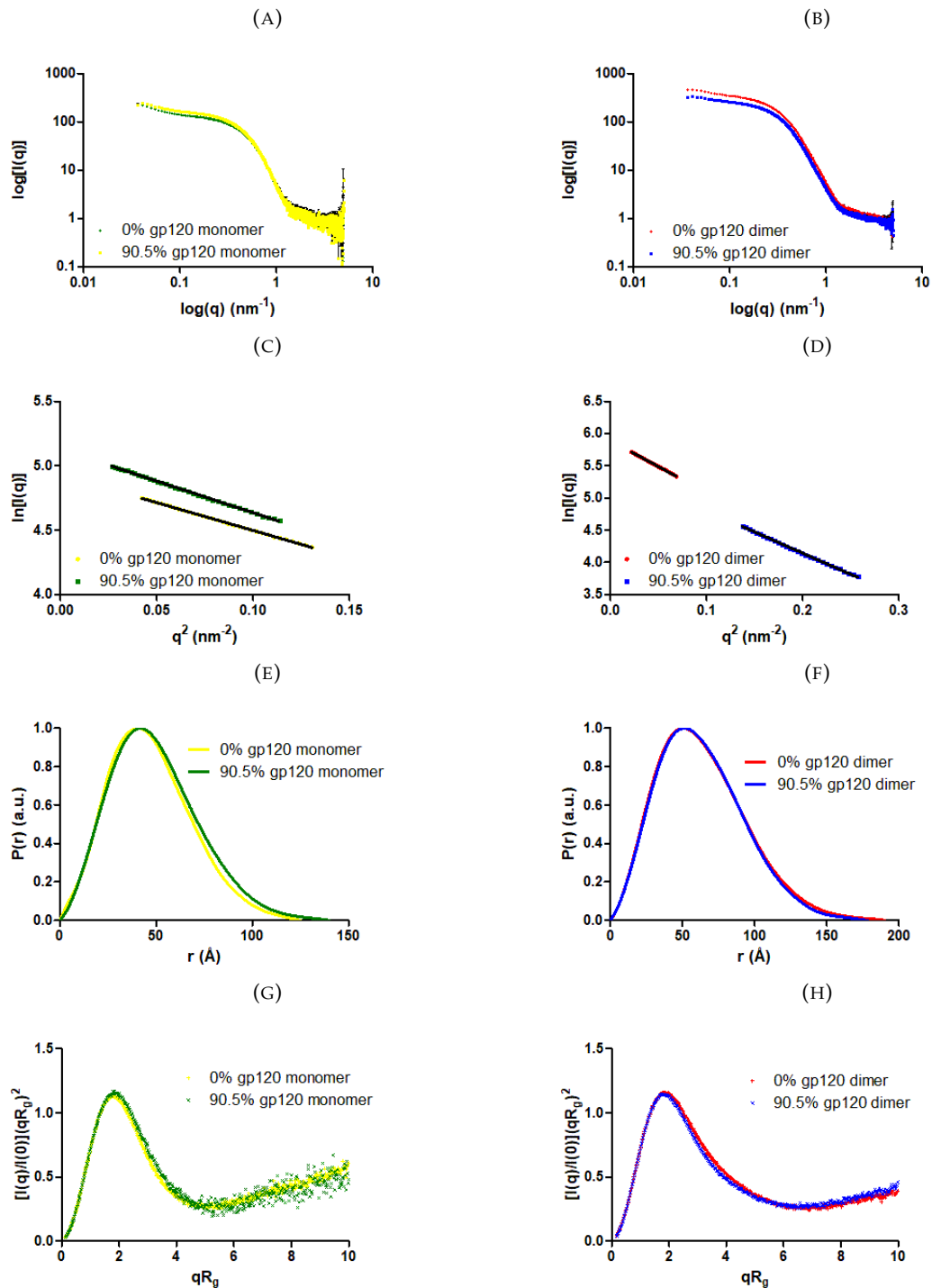


FIGURE 6.16: Graphical representation of the gp120 monomer and dimer data measured by SAXS. (A) and (B) show a double-logarithmic representation of the 1D SAXS scattering curves of the gp120 monomer and dimer, respectively. (C) and (D) show the Guinier regions for the gp120 monomer and dimer, respectively. The  $P(r)$  functions are shown in (E) for the gp120 monomer and (F) for the gp120 dimer. Finally, the Kratky plots normalised against the  $R_g$  determined from the Guinier plot are shown in (G) (gp120 monomer) and (H) (gp120 dimer). The gp120 monomer and the dimer in 0% D<sub>2</sub>O solvent are shown in yellow and red, respectively. Whereas the gp120 monomer and the dimer in 90.5% D<sub>2</sub>O solvent are shown in green and blue, respectively.

### 6.5.2 Size-exclusion chromatography coupled to small angle X-ray scattering analysis of the 2dCD4-WT/gp120 complex

Since there was evidence of aggregation in the d-2dCD4-WT/h-gp120 complex measured by SAXS using the robotic sample changer, SEC-SAXS was used as a way to ensure the monodispersity of the sample by adding an *in situ* SEC purification step, prior to passing directly into the sample capillary where it is exposed to the X-ray beam. This helps to ensure both homogeneity and ideality of the sample which cannot be guaranteed when using the robotic sample changer as the samples remain static for a period of time before measurement.

Therefore SEC-SAXS measurements of d-2dCD4-WT produced using the CFPE system in complex with h-gp120 were performed and compared to SEC-SAXS data collected on the complex formed with h-2dCD4-WT expressed in the *E. coli* and *B. choshinensis* expression systems.

#### Sample preparation and measurement

The 2dCD4-WT expressed using *E. coli* and refolded from inclusion bodies (referred to henceforth as E-2dCD4-WT) and gp120<sub>ZA</sub>CAP45 were used to form the complex as described above. A Superdex™ 200 increase 3.2/300 (GE Healthcare), equilibrated in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, was used to reduce dilution of the sample during the SEC process. The complex was concentrated to 7.8 mg/mL and 50  $\mu$ L was loaded onto the column. The protein was eluted at a flow rate of 0.1 ml/min and SAXS frames were measured every second.

In a second experiment gp120<sub>ZA</sub>CAP45 and 2dCD4-WT expressed and secreted by *Brevibacillus choshinensis* (referred to hereinafter as B-2dCD4-WT) was tested using SEC-SAXS to verify the *B. choshinensis* produced 2dCD4-WT behaves the same way in solution as the *E. coli* analogue. The Superdex™ 200 increase 3.2/300 column was configured in the same set-up using 20 mM Tris-HCl pH 7.5, 300 mM NaCl. The complex was concentrated to 5.3 mg/mL and 50  $\mu$ L was loaded onto the column. As the protein was eluted at 0.1 mL/min, SAXS frames were measured every second.

In a final experiment, deuterated 2dCD4-WT produced in the cell-free protein expression system (referred to henceforth as CFPE-d-2dCD4-WT) was complexed with hydrogenated gp120 and concentrated to 13.7 mg/mL. The Superdex™ 200 increase 3.2/300 was set-up in the same

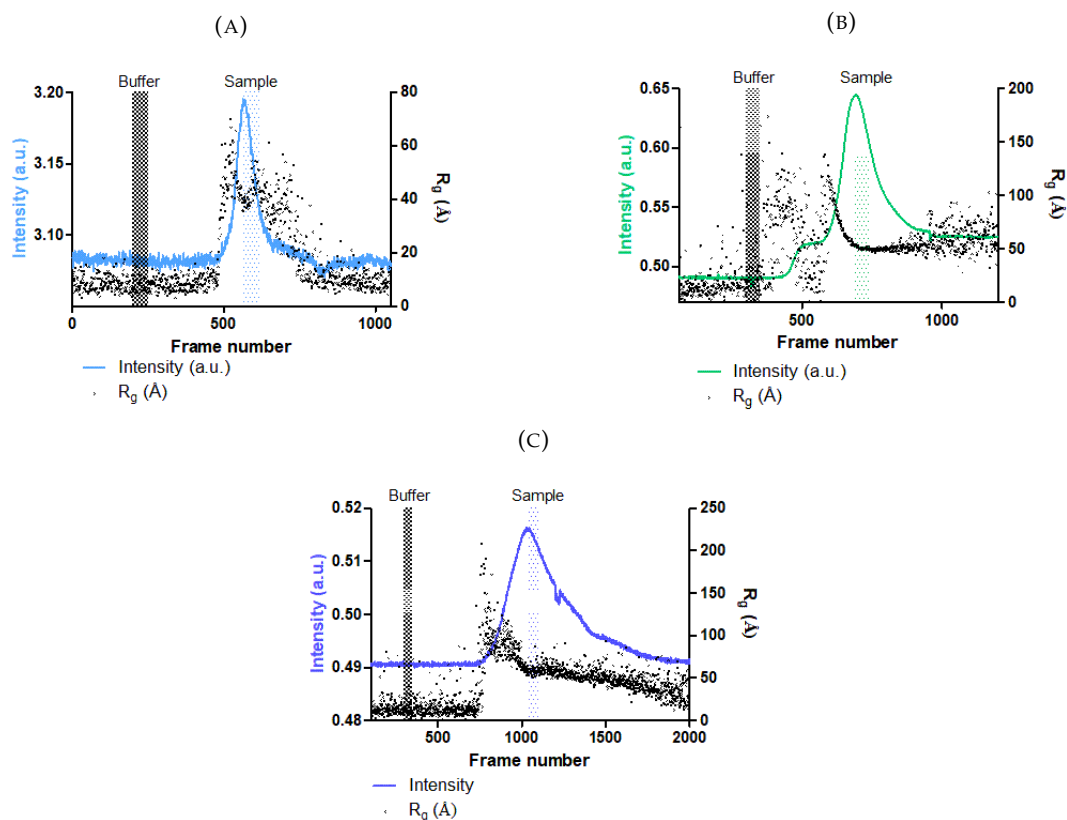


FIGURE 6.17: The SAXS intensity (coloured curve) was plotted against the  $R_g$  determined for each frame. (A) *E. coli* expressed 2dCD4-WT in complex with gp120<sub>ZA</sub>CAP45. (B) *B. choshinensis* expressed 2dCD4-WT in complex with gp120<sub>ZA</sub>CAP45. (C) Cell-free protein expressed deuterated 2dCD4-WT in complex with gp120. The selection of the buffer frames is indicated by the black and white hashed column where the intensity and  $R_g$  were flat. The sample frames were selected across the intensity peak where the  $R_g$  remained relatively flat, indicating a consistent size (coloured column).

way using 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% sucrose. 50  $\mu$ L was loaded onto the column and the protein was eluted at 0.1 mL/min, measuring SAXS frames every second.

### Selecting the buffer and sample frames

In order to select which frames correspond to the buffer and which correspond to the complex, the intensity and  $R_g$  was plotted for each frame number (fig. 6.17). Where the intensity and  $R_g$  are flat and low before the peak the scattering from these frames comes only from the background. Fifty frames approximately in the middle of this flat range were selected and averaged to represent the background scattering. To select the sample frames, frames across the middle of the peak which showed a flat  $R_g$  were selected. The flat  $R_g$  suggests a consistent size and that the scattering curves were of a complex of an approximately similar size.

After selection of the background and sample SAXS frames, the background curves were averaged and subtracted from the averaged sample frame curves using the SCATTER software [200]. Following averaging and subtraction, the 1D scattering curves of the gp120/2dCD4-WT analogue complexes could be analysed as described above for the robotic sample changer measured samples (section 6.5.1).

### Size parameter analysis of the complex determined from SEC-SAXS data

Table 6.8 lists the parameters extracted from the data (fig. 6.18). The  $R_g$  determined from the Guinier region and from the  $P(r)$  function are consistent within  $\pm 2\text{\AA}$  which suggests that the quality of the data is good for each of the data sets. The E-2dCD4-WT/gp120 complex is the smallest of the three with an  $R_g$  of  $37.3\text{\AA}$  as determined from the Guinier region, which is in accordance with the data published on the monomeric gp120 complex [72] (table. 6.1). The B-2dCD4-WT/gp120 and the CFPE-d-2dCD4-WT/gp120 complexes have an  $R_g$  of  $48.8$  and  $46.6\text{\AA}$ , respectively and as such, are both larger than the dimeric gp120/2dCD4-WT complex [197] (table. 6.1).

The CFPE-d-2dCD4-WT/gp120 complex has the largest  $D_{max}$  of the three complexes. The B-2dCD4-WT and CFPE-d-2dCD4-WT appear to have smaller  $D_{max}$  than those published for the gp120 dimer/2dCD4-WT complex. Whereas the E-2dCD4-WT has a  $D_{max}$   $10\text{\AA}$  larger than that published for the monomeric gp120/2dCD4-WT complex. These observed differences could be as a result of the presence of the gp120 monomer/dimer mixture as discussed previously.

Sample	Concentration (mg/mL)	$R_g$ ( $\text{\AA}$ )	$I(0)$ ( $\text{cm}^{-1}$ )	$R_g$ (real, $\text{\AA}$ )	$D_{max}$ ( $\text{\AA}$ )
<i>E. coli</i> 2dCD4-WT/gp120	7.8	37.3	$5.78 \times 10^{-3}$	37.7	145
<i>B. choshinensis</i> 2dCD4-WT/gp120	5.3	48.8	$1.9 \times 10^{-2}$	49.0	155
CFPE d-2dCD4-WT/h-gp120	13.7	46.6	$3.74 \times 10^{-3}$	48.9	159

TABLE 6.8: gp120 in complex with 2dCD4-WT produced in *E. coli*, *B. choshinensis* and CFPE expression systems was measured using SEC-SAXS. The  $R_g$ ,  $D_{max}$  and Porod volume size parameters and their corresponding  $I(0)$  are listed here.



### Analysis of the graphical representation of the SEC-SAXS data of the complex

The SEC-SAXS data is graphically represented in figure 6.18. The double-logarithmic plot in figure 6.18a shows that the CFPE-d-2dCD4-WT/gp120 complex (purple curve) displays a slight upwards turn at low- $q$  which is characteristic of aggregation in samples measured using the robotic sample changer. However, since these samples were measured using the SEC-SAXS set-up, it is expected that any aggregation would have been separated from the sample, eluting at a low elution volume due to its large size. The upwards turn seen here at low- $q$  could also be demonstrative of radiation damage. The B-2dCD4-WT/gp120 complex (green) and the E-2dCD4-WT/gp120 complex (blue) do not exhibit the same inflection at low- $q$  but this cannot be attributed to the deuteration of 2dCD4-WT as deuteration has been shown to provide some protection against radiation damage [214].

Since deuterated protein is more prone to aggregation it could be that the protein aggregated as it eluted from the column. The quartz capillary in which the sample is measured by the SAXS beam is not washed during a SEC-SAXS run and the flow rate of the SEC-SAXS run is determined by the maximum flow rate supported by the column (0.1 mL/min). For a 50  $\mu$ L sample measured using the robotic sample changer with 10 frames measured every 1 s, the flow rate can be estimated to be approximately 0.3 mL/min. This means that the flow rate achievable is higher using the robotic sample changer than by SEC-SAXS. With these factors taken into account the CFPE-d-2dCD4-WT/h-gp120 complex may have aggregated along the capillary before measurement by the SAXS beam.

The Guinier regions are flat but of differing gradients, hence the difference in the  $R_g$  observed in table 6.8. The shapes of the  $P(r)$  functions in figure 6.18c are similar, especially for the B-2dCD4-WT/gp120 and CFPE-d-2dCD4-WT/gp120 complexes. The  $D_{max}$  for the E-2dCD4-WT/gp120 complex is 10-14  $\text{\AA}$  smaller than that for the other two complexes and has a more gradual tail off at the  $x$ -axis towards  $D_{max}$ . Finally, the shape of the dimensionless Kratky plots (fig. 6.18d) are characteristic of well-folded, globular proteins as the curves return towards and are asymptotic to the  $x$ -axis.

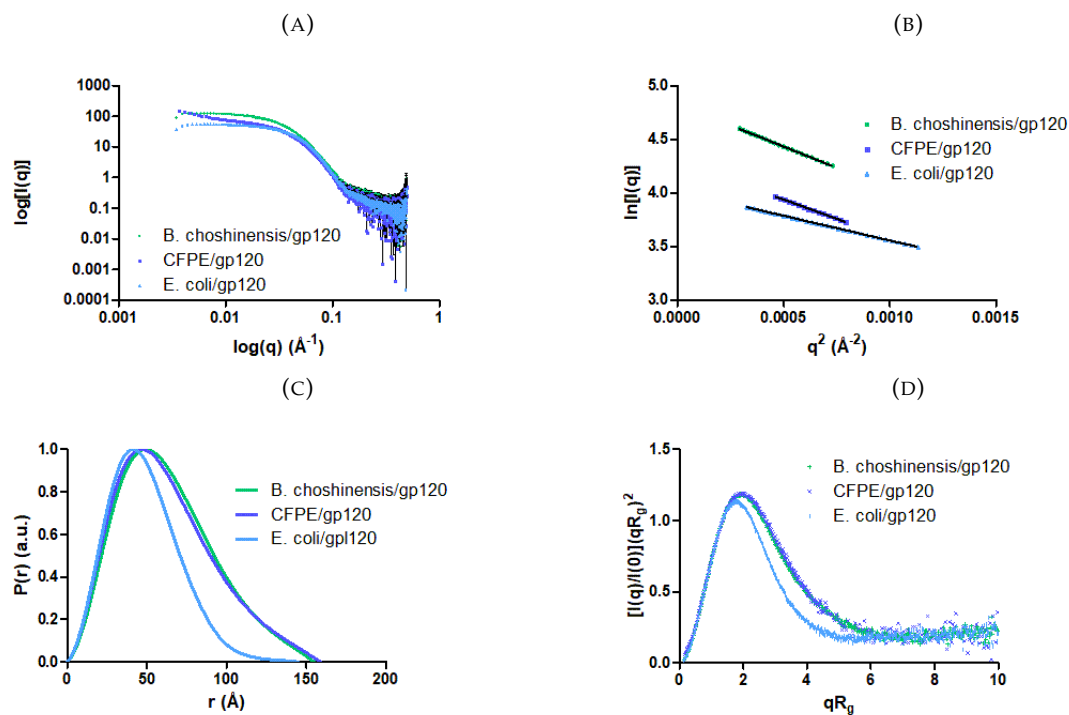


FIGURE 6.18: Graphical representation of 2dCD4-WT (wild-type two domain CD4) analogues in complex with gp120. (A) shows a double-logarithmic representation of the 1D SAXS scattering curves. (B), (C) and (D) show the Guinier regions,  $P(r)$  functions and dimensionless Kratky plots, respectively. blue = *E. coli* h-2dCD4-WT/h-gp120, green = *B. choshinensis* h-2dCD4-WT/h-gp120, purple = CFPE d-2dCD4-WT/h-gp120.

## 6.6 Discussion

Small-angle neutron and X-ray scattering was used to obtain structural information on 2dCD4-WT and gp120 in isolation and when complexed. Where the biophysical characterisation of 2dCD4-WT was previously addressed in chapter 4, prior to SAS experiments, the characterisation of gp120 by MALDI-TOF mass-septrometry followed by SEC-MALLS-RI is addressed here. While both the gp120 BAL and gp120<sub>ZA</sub>CAP45 analogues were shown to form monomers and dimers, the CAP45 species was selected due to superior yield and purity for further experiments. The contrast-match point of protiated gp120<sub>ZA</sub>CAP45 and deuterated 2dCD4-WT were then calculated to be 45.2% and 90.5% D<sub>2</sub>O, respectively, by measuring their neutron scattering signal in a series of D<sub>2</sub>O contrasts to allow contrast variation studies of the complex. SAXS measurements were then carried out on the sample which had been measured by SANS in an attempt to corroborate the data obtained. Finally, SEC-SAXS data were collected on the 2dCD4-WT/gp120 complex using deuterated 2dCD4-WT expressed using CFPE as well as protiated

2dCD4-WT produced using the *E. coli* and *B. choshinensis* expression systems.

### 6.6.1 Analysis of gp120

#### **gp120 is expressed as a monomer and a non-physiological dimer**

The presence of gp120 monomers and non-physiological dimers was confirmed by SEC-MALLS-RI and suggested that the monomer dimer ratio was approximately 1:2. Finzi *et al.* 2010 [215], showed that gp120 can be expressed as abnormal disulphide-linked dimers when produced by over-expressing cells, and dimer formation is exacerbated when the V2 loop is intact. These dimers are described as non-physiological as they are not produced in the physiological HIV-1 infection cycle. They were able to decrease dimer formation by introducing a point mutation at Leu 111 to Ala 111, as well as separating the gp120 dimer by use of reducing agents. By the addition of 5 mM DTT it was possible to enrich the gp120 sample with monomer. However subsequent separation by SEC was non-trivial due to increased overlapping of the monomer/dimer peaks. Full-separation would have resulted in significant sample loss, therefore for SAS studies the gp120 was enriched with monomer prior to complex formation. In addition, monomeric and dimeric gp120 species were measured by both SANS and SAXS and the size parameters determined corroborated with the published data.

#### **gp120 glycosylation complicates analysis of SAS data**

SAS data analysis of gp120 was non-trivial due to the glycosylation of the protein. The SLD of sugars is different to that of protein because the atomic composition of a sugar molecule is different to that of a protein (few nitrogen atoms) and therefore their neutron scattering behaviour is different. Since recombinant gp120 glycosylation patterns are heavily dependent on the HIV-1 subtype from which they are derived and the cellular expression system used to produce them the identity of the complex glycans was unknown [205, 207]. A labour intensive glycoprofiling analysis using mass-spectrometry would have been necessary to accurately determine the nature of the complex glycans on the surface of the protein. Given the extent of glycosylation determined by mass-spectrometry (36.9% total mass) and the large heterogeneity shown by the broadness of the peaks, this would have been a mammoth task which was outside

the scope of this thesis given the difficulties in developing a deuteration protocol for 2dCD4-WT. For an example of the analysis required for this, the reader is directed to work carried out by Cao *et al.* 2017 [216] who determined the global glycoprofile of gp120 BAL expressed by 293F cells. From this work it is clear that the glycoprofile is complex and heterogeneous. An estimate of the sugar composition was made using data from Panico *et al.* 2016 so an estimate of the MM could be made [205]. However the analysis was not stretched as far as *ab initio* modelling on any samples containing gp120 given the ambiguity around the sugars.

Since this thesis work was more concerned with the structure of 2dCD4-WT and it was possible to match-out the gp120 component of the complex using contrast variation, this allowed the study of d-2dCD4-WT by SANS without too much complication from the gp120 as a result of the monomer/dimer state, or the glycan phase.

## 6.6.2 gp120-bound and un-bound 2dCD4-WT analysis

### 2dCD4-WT is more stable when in complex with gp120

Whilst there is evidence of severe aggregation of the d-2dCD4-WT measured by SAXS, shown by the bi-phasic double-logarithmic plot suggestive of two populations, there is only a small amount of aggregation present in the d-2dCD4-WT/gp120 complex sample. Similarly some aggregation can be seen for the SANS measurements of the d-2dCD4-WT in isolation which is not seen in the measurement of the complex. The presence of aggregation was also shown by SEC-MALLS-RI in chapter 4. In addition, concentration of 2dCD4-WT was non-trivial as the protein often crashed out of solution to form an insoluble precipitant at concentrations above 1 mg/mL regardless of the expression system used and whether it was deuterium labelled or not. Conversely when bound to gp120, concentrations upwards of 10 mg/mL were achieved for the complex with no obvious signs of aggregation in the low- $q$  region of the SANS and SAXS data. Combined, this suggests that 2dCD4-WT is more stable when bound to the complex.

The reason for this increase in stability may be due to the redox biology at play within 2dCD4-WT in relation to gp120 binding. The 2dCD4-WT produced using the CFPE and *E. coli* expression systems were expressed as a mixture of redox states: fully reduced (2dCD4-R2: reduced in D1 and D2), partially reduced (2dCD4-R1: oxidised in D1 and reduced in D2) and fully oxidised (2dCD4-OX: oxidised in D1 and D2) and there is evidence to suggest that the partially

reduced form is the most stable (5). Therefore in the redoximer mixture it might be expected that the less stable 2dCD4-R2 and -OX forms aggregate in solution. However, during complex formation the 2dCD4-R1 form binds to gp120 when the two proteins are mixed, after which the complex is isolated from the unbound 2dCD4-R2 and -OX by a size-exclusion chromatography step prior to measurement by SAS. Since there is evidence suggesting that the 2dCD4-R1 redoximer which uniquely binds to gp120 is the most stable of the three redoximers, then this would explain why 2dCD4-WT in the complex is more stable; as it consists solely of the stable 2dCD4-R1 isoform. Hence the absence of aggregation suggested by the low- $q$  data for the complex. Whereas the 2dCD4-WT contains a mixture of the less stable 2dCD4-R2 and -OX isoforms as well as 2dCD4-R1 explaining the aggregation seen at low- $q$  in the SAS data.

#### **d-2dCD4-WT may exist in a more compact form when bound to gp120 as shown by SANS**

In line with the finding that 2dCD4-WT is more stable when bound to gp120. The SANS contrast variation data suggest that 2dCD4-WT is more compact when bound to gp120 than in isolation. Both the  $R_g$  and  $D_{max}$  for the d-2dCD4-WT/h-gp120 complex in 45.2% D<sub>2</sub>O are smaller than those found for the d-2dCD4-WT in 0 and 45.2% D<sub>2</sub>O alone. At 45.2% D<sub>2</sub>O the h-gp120 is matched out because the SLD of the solvent is equal to that of the gp120 component of the complex. Therefore the scattering observed for the complex in 45.2% D<sub>2</sub>O is solely that of the d-2dCD4-WT. Since the d-2dCD4-WT is a mixture of the 2dCD4-R2, -R1 and -Ox redoximers it is not possible to tell whether this apparent shrinking in the size of the 2dCD4-WT bound to gp120 is as a function of gp120 binding or whether its just as a result of the 2dCD4-R1 isoform being selected by gp120 to form the complex.

The redox isomers in the 2dCD4-WT mixture could possibly have three distinct  $R_g$  and  $D_{max}$  given the results in chapter 5, which suggest that the 2dCD4-R1 is the smallest of the three redoximers. Therefore the  $R_g$  and  $D_{max}$  determined for d-2dCD4-WT could be an average of each of the three isoforms, weighted by their proportion of the d-2dCD4-WT redoximer mixture. To this end the  $R_g$  and  $D_{max}$  observed for gp120-bound d-2dCD4-WT in 45.2% D<sub>2</sub>O could be the size parameters for the 2dCD4-R1 in isolation as well as bound to gp120, if there is no conformational change within CD4 on gp120 binding which has been previously suggested from X-ray crystallography and SAXS data [34, 197]. Conversely these may be the size parameters for 2dCD4-R1 when bound to gp120 due to conformational changes within the protein

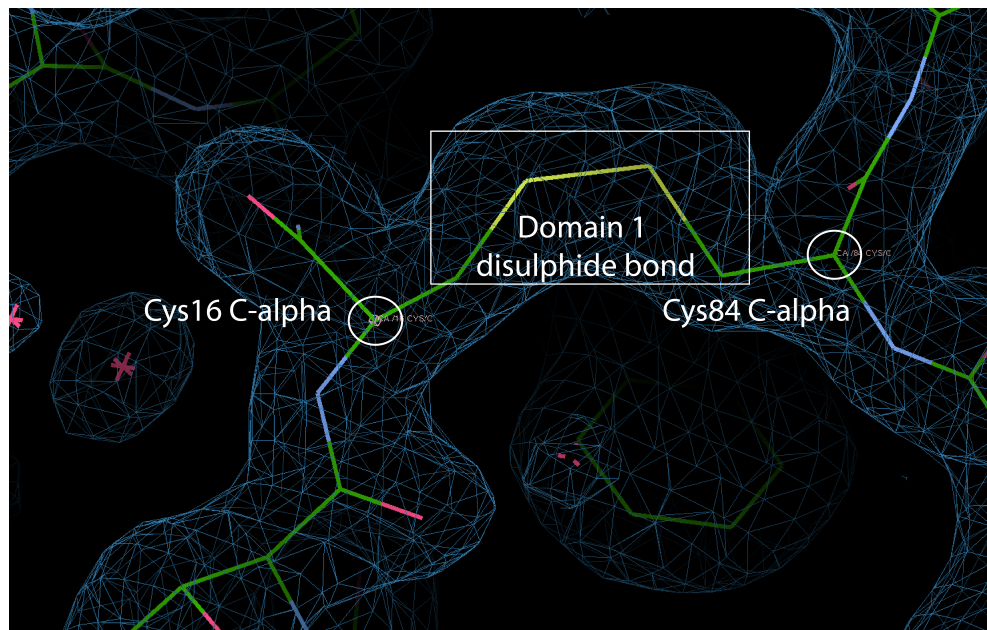
which are distinct from the size parameters from 2dCD4-R1 in isolation. In order to determine whether either of these hypotheses are correct, it would be necessary to isolate the 2dCD4-R1 isoform from 2dCD4-R2 and 2dCD4-Ox or use the Cys to Ala disulphide bond knockout variants to this effect.

### SAS data does not match that of the X-ray crystal structure

Perhaps the most pertinent result from this chapter is the finding that the SAS data collected for gp120-bound and -unbound 2dCD4-WT does not validate the X-ray crystallography data. The CRYSON data (fig. 6.10), which compared the SANS data against PDB entries 3CD4 and 1GC1, suggests that the low- $q$  data containing information on the size of the protein was similar but that the high- $q$  data containing information on the shape is quite different. This could be because of the redox shuffling events at play within CD4 which have not been taken into account during X-ray crystallographic studies. However the 2dCD4-WT construct in PDB entry 3CD4 contains just 182 residues whereas the CFPE expressed d-2dCD4-WT possesses an N-terminal his-tag with cleavage site and therefore the two cannot be precisely compared. Similarly the gp120 used in this thesis has no truncations and was fully glycosylated so similar precautions in analysing the scattering data against the theoretical scattering from the crystal structure are necessary.

Using the X-ray coordinates and reflection data from PDB entry 4H8W [217], which shows the crystal structure of 2dCD4-WT in complex with gp120 and the antibody N5-i5, 2Fo-Fc and Fo-FC electron density maps were calculated using refmac (fig. 6.19). Figure 6.19a which shows the domain 1 structural disulphide bond shows no positive nor negative electron density in the Fo-Fc map around the disulphide bond indicating that the bond is present. Whereas interestingly in figure 6.19b there is negative electron density (at  $\sigma = 3$ ) around the domain 2 allosteric disulphide bond which adds some ambiguity as to whether the disulphide bond is oxidised or reduced. This could be because the disulphide bond is not present in the X-ray data because the isoform of 2dCD4-WT bound to gp120 has shown to be reduced in its domain 2 disulphide bond [63]. Alternatively this could be as a result of ionising radiation from the X-ray beam which causes the disulphide bond to be reduced, especially since this disulphide bond has been shown to be metastable [63, 203] and is therefore more sensitive to ionising radiation.

(A) Domain 1 structural disulphide bond



(B) Domain 2 allosteric disulphide bond

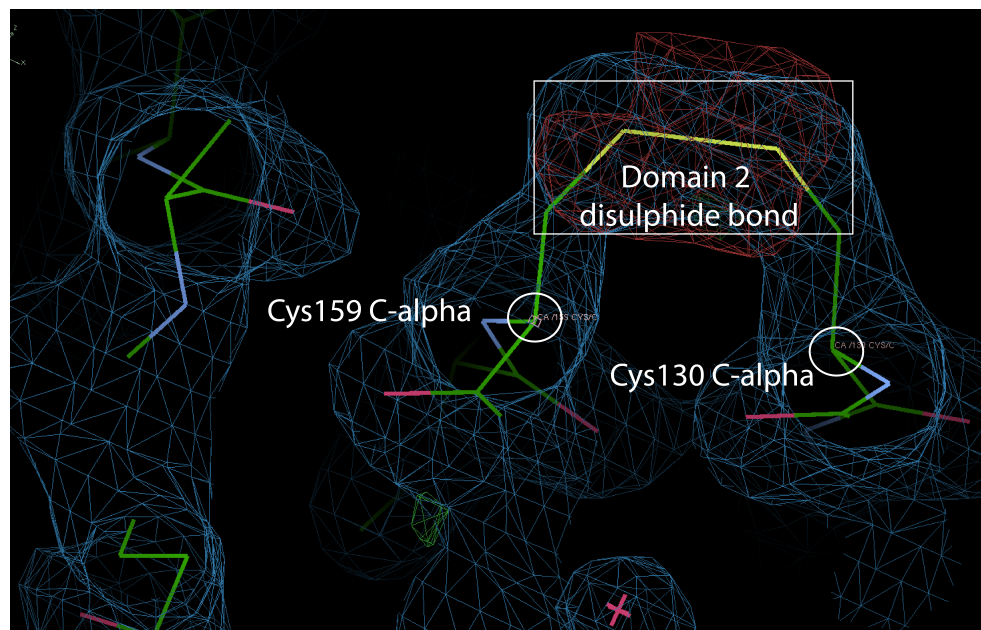


FIGURE 6.19: 2Fo-Fc ( $\sigma = 1$ ) and Fo-Fc ( $\sigma = 3$ ) electron density maps of the disulphide bonds in domains 1 (A) and 2 (B) of 2dCD4-WT when bound to gp120. The C $\alpha$  carbons of the disulphide bonds are circled in white and the disulphide bonds are shown in white boxes. The electron density maps were calculated using the X-ray reflection data and coordinate files of PDB entry 4H8W which is a crystal structure of 2dCD4-WT in complex with gp120 and the antibody N5-i5 [217]. Interestingly there is no negative electron density shown around the domain 1 disulphide bond, whereas there is negative electron density around the domain 2 disulphide bond.

### The relationship between the structure of 2dCD4-WT, its dynamics and gp120 binding

Given the complexity of CD4 binding under certain redox conditions to gp120, the crystal structure may not reveal the subtleties of this dynamic redox process. Owen *et al.* 2018 [218] explored the dynamics of 2dCD4 as a function of domain 1 and domain 2 disulphide bond ablation using protium/deuterium exchange mass-spectrometry. They found regions of high dynamics (corresponding to more than 50% deuteration) at the gp120 binding site in 2dCD4-WT. They also found that ablation of the second domain allosteric disulphide bond decreased the dynamics of the domain whereas ablation of the first domain structural disulphide bond increased the dynamics of the domain which supports the findings in chapter 5 and by Owen *et al.* 2016, that a reduced domain 1 disulphide bond decreases the stability of the protein whereas a reduced domain 2 increases the stability of the protein [203].

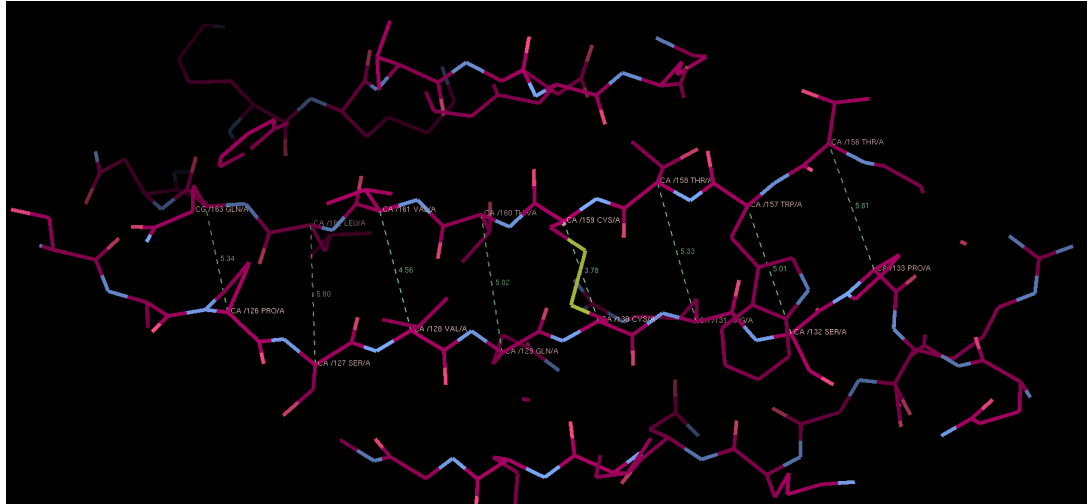
Additionally, it was found that ablation of the second domain disulphide bond had allosteric effects, which increased the dynamics of domain 1, in a similar fashion as if the domain 1 disulphide bond had been reduced. In line with the finding by Cerutti *et al.* 2014 that gp120 binds to an isoform of 2dCD4 in which **either** the domain 1 or domain 2 disulphide bond is reduced (see 5.2) [63], Owen *et al.* 2018 [218] suggest that increased dynamics around the gp120 binding site allowing greater solvent accessibility are necessary for gp120 binding to 2dCD4. The gp120 binding site spans beta-strands C, C' (which contains the key Phe43 residue), C'', and D and the alpha-helix DE (which contains the key Arg59 residue). Since the domain 1 disulphide bond is the stable structural bond, its reduction (shown to destabilise the domain) is unlikely to occur physiologically. Whereas the domain 2 disulphide bond has been shown to be metastable, prone to redox shuffling events and is therefore the allosteric disulphide switch triggering the increased dynamics in domain 1 required for gp120 binding. Completely reduced 2dCD4-WT (2dCD4-R2) cannot bind to gp120 which is explained by loss of tertiary domain 1 structural integrity by reduction of the domain 1 disulphide bond, as discussed in chapter 5.

Reduction of the domain 2 disulphide bond must therefore confer small structural realignments to domain 1 in order for the increase in dynamics of the gp120 binding site to occur. Figure 6.20 and table 6.9 show the C- $\alpha$  distances along  $\beta$ C and  $\beta$ F of domains 2 and domains 4 of CD4 which are the most structurally similar. Domain 2 contains the allosteric disulphide bond of interest formed between beta-strands C and F whereas domain 4 contains a canonical structural



disulphide found in Ig domains between strands B and F. The acquisition of the disulphide bond between the beta-strands of the same sheet in domain 2 has caused the domain to pucker at the disulphide bond, shown by a short C- $\alpha$  distance of 3.78Å, which is the shortest C- $\alpha$  distance along the interface of the two strands.

(A) Domain 2 C- $\alpha$  distances along  $\beta$ C and  $\beta$ F



(B) Domain 4 C- $\alpha$  distances along  $\beta$ C and  $\beta$ F

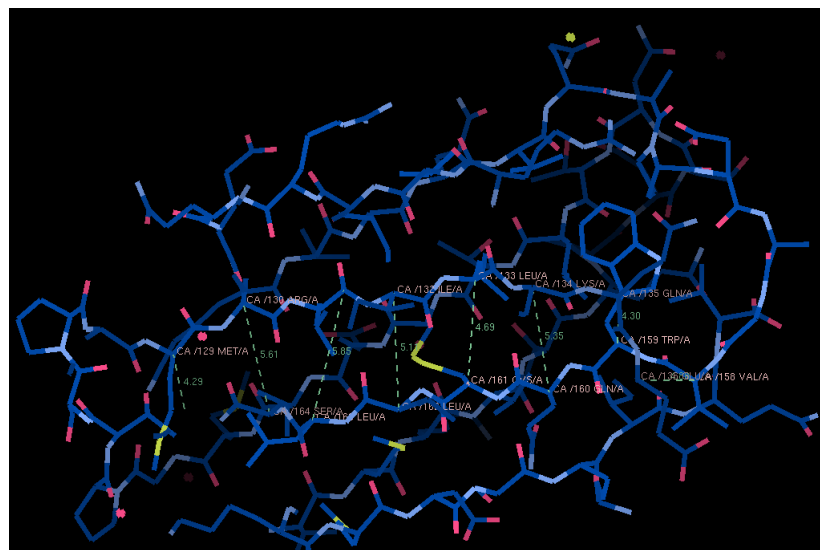


FIGURE 6.20: The C- $\alpha$  distances along the strands  $\beta$ C and  $\beta$ F of domain 2 (A) and domain 4 (B) were measured in WinCoot. Domain 2 (from PDB entry 3CD4) contains a strained disulphide bond formed between  $\beta$ C and  $\beta$ F whereas domain 4 (from PDB entry 1CID) contains an unstrained disulphide bond between  $\beta$ B and  $\beta$ F. The C- $\alpha$  distances are listed in table 6.9 and read from left to right across 6.20a and 6.20b.

Reduction of the second domain allosteric disulphide bond could therefore induce relaxation

Domain 2 C- $\alpha$ residues	Domain 2 C- $\alpha$ distance (Å)	Domain 4 C- $\alpha$ residues	Domain 4 C- $\alpha$ distance (Å)
Pro126-Gln163	5.34		
Ser127-Leu162	5.80		
Val128-Val161	4.56		
Gln129-Thr160	5.02	Met129-165	4.29
<b>Cys130-Cys159</b>	<b>3.78</b>	Arg130-Ser164	5.61
Arg131-Thr158	5.33	131-Leu163	5.85
Ser132-Trp157	5.01	Ile132-Leu162	5.11
Pro133-Thr156	5.81	Leu133-Cys161	4.69
		Lys134-Gln160	5.35
		Gln135-Trp159	4.30

TABLE 6.9: The C- $\alpha$  distances along the strands  $\beta$ C and  $\beta$ F were measured in Win-Coot. The Cys130-Cys159 C- $\alpha$  distance is the shortest where the Trp159 residue in strand C and replaced by a Cys residue to form a disulphide bond of unusual geometry with strand F.

within this domain as the strained bond would no longer be pulling the neighbouring beta-strands together. Structural realignment triggered as a result of this reduction event could then be transferred from domain 2 to domain 1 thereby increasing the dynamics at the gp120 binding site. For example, this could happen through a subtle inward motion of the opposing beta-sheets of domain 2 which could be communicated to domain 1 through the shared domain 1  $\beta$ G domain 2  $\beta$ A strand.

### 6.6.3 Using SAS to study 2dCD4-WT, gp120 and the complex

#### SEC-SAS is a more suitable method for carrying out SAS measurements on 2dCD4-WT

Given the propensity of 2dCD4-WT to aggregate in solution as shown by both SANS and particularly by SAXS, when using the robotic sample changer, and the reduced effect of aggregation in the SEC-SAXS experiment, this suggests that SEC coupled to small-angle scattering measurements are the most suitable way of studying this system. For SAXS this does not pose too much of a problem as SEC columns with small bed volumes such as the S200 increase column used for studying the complex mean that the sample does not suffer from too much dilution with smaller sample volumes (50  $\mu$ L sample in 3 mL column volume). In addition, due to the high-flux of the SAXS beam the sample measurements are short which allows snap-shot 'images' to be taken as the sample elutes from the column. Indeed, this technique was used previously in chapter 5 to study the effect of ablation of the domain 2 disulphide bond on the

structure of 2dCD4-WT, when there was evidence of aggregation in the samples when using the robotic sample changer.

However for SANS *in situ* SEC purification is more problematic especially for a system where 2dCD4-WT cannot initially be easily concentrated to high concentrations. SEC-SANS is a relatively new technique and was first reported by Jordan *et al.* 2016 [204]. The main issue lies in the low neutron beam flux compared to that of X-rays. The sample measuring times for static SANS measurements are in the range of minutes or hours for protein samples since they are weak neutron scatterers. Therefore for SEC-SANS of 2dCD4-WT (a relatively small protein for which initially high concentrations are difficult to achieve) would necessitate long measuring times in the cuvette as the signal would be low. Longer measuring times would negate the usefulness of the *in situ* SEC step as the sample would have time to aggregate and would thus, not be snap-shot of the protein as it elutes from the column.

Although SEC-SANS would be problematic for measuring 2dCD4-WT in isolation or with gp120 matched-out, this would be an excellent answer to the difficulties in isolating gp120 monomers and dimers. By using SEC-SANS the gp120 could be purified straight into the neutron beam by use of the *in situ* SEC step. This would reduce sample loss in preparation of the individual monomeric and dimeric species prior to the experiment and would also allow data collection on monodisperse monomeric and dimeric gp120. Since gp120 is a large protein (compared to 2dCD4-WT) it can be measured by SANS in lower concentrations thus the dilution caused by the SEC step would have less impact on the signal to noise ratio. Additionally, shorter measuring times would not pose as much of a problem for measurement of gp120 compared to 2dCD4-WT.

Finally, SEC-SANS could be used to measure the complex in 0% D<sub>2</sub>O where both components are visible since the complex is large. At the match-out point of gp120, however, the same difficulties would be faced as with measurement of d-2dCD4-WT alone.

### SAS data quality

There are signs in the SANS data which suggest that the gp120-bound and unbound 2dCD4-WT data quality is poor. From the double-logarithmic plot (fig. 6.8a) of the SANS data the signal to noise ratio was low and there was evidence of aggregation at low- $q$ , which was shown

initially by SEC-MALLS-RI data of d-2dCD4-WT in chapter 4. In addition, the shape of the  $P(r)$  functions (fig. 6.8e) do not match which may be because  $P(r)$  analysis was non-trivial due to the noisiness of the data. Finally, the shape of the Kratky curves (fig. 6.8g) are indicative of an unfolded protein. However, the ELISA, pseudo-viral neutralisation assay and 1D-NMR experiments from chapter 4 all showed that the protein was correctly folded. So the shape of the Kratky and  $P(r)$  plots may instead be explained by poor buffer subtraction. This could be because of buffer mismatching which is a common error in SANS due to errors in the  $D_2O$  contrasts and the dialysis often used for changing the buffer.

The SAXS data collected on the gp120-unbound 2dCD4-WT was undoubtedly of poor quality as the double-logarithmic plots (fig. 6.14a) show the presence of multiple species. Many of the curves were left out of the analysis of the SAXS data as they were identified as being significantly different ( $p \leq 0.05$ ) to the other curves by the data comparison function of the Primus programme [213] of the ATSAS suite of software [199]. These curves were different because of radiation damage to which 2dCD4-WT appears sensitive. The reason for this sensitivity may be due to its disulphide bonds. Even though gp120 also possesses disulphide bonds it does not appear to be as sensitive to ionising radiation damage which could be because of its glycosylation which may act as a protectant against ionising radiation considering glycerol and sucrose can be used to protect against ionising radiation damage. Aggregation was also shown to be present in the gp120-bound 2dCD4-WT sample so SEC-SAXS was used as a way of counteracting this.

Regardless, *ab initio* modelling of the SANS data presented models which make reasonable biological sense. However, in order for conformational changes to be observed in 2dCD4-WT as a function of its redox state they would need to be large, dramatic structural realignments within the protein which is not necessarily the case. The differences observed were under 10 Å difference, which is the approximate resolution limit of the technique. The interpretation of the data is therefore speculative and further work using high resolution techniques is necessary in order to confirm or reject the suggested reasoning of the patterns in the data.

#### 6.6.4 The use of *B. choshinensis* and CFPE as recombinant protein expression methods

The SAXS data collected on samples containing d-2dCD4-WT produced using CFPE was generally shown to be aggregated. The samples measured by SAXS after measurement by SANS were thus not freshly prepared, had been kept in D<sub>2</sub>O buffers and exposed to the neutron beam at ambient temperature and then left at 8°C for 24 hours post-SANS prior to being measured by SAXS. So it is unsurprising that the SAXS data was of low quality. In addition, the SEC-SAXS measurement of the CFPE produced d-2dCD4-WT in complex with gp120 shows signs of radiation damage or aggregation. Therefore, it is difficult to say from this data that the CFPE produced protein is suitable for analysis by SAXS from the SAXS data. However, we have already seen in chapter 5 that this sample behaves well by SEC-SAXS when unlabelled and freshly prepared. The *B. choshinensis* expressed protein behaved well in the SEC-SAXS measurement as there was no sign of radiation damage. This may be explained by the fact that the sample was freshly prepared.

#### 6.6.5 Future work

Ultimately, in order to determine whether there really is a difference in the size and conformation of gp120-bound and -unbound 2dCD4-R1 with respect to the 2dCD4-WT redoximer mixture, it would be necessary to fully isolate this isoform from the mixture of redoximers in 2dCD4-WT. Then SEC-SANS could be used to measure d-2dCD4-WT in order to eliminate any aggregation prior to exposure to the neutron beam and to negate the effects of ionising radiation damage by exposure to the X-ray beam. Alternatively, this could be achieved by producing the 2dCD4-D2A variant protein in which the D2 disulphide has been knocked-out by cysteine to alanine substitution, which would also negate the potential damage to the disulphide bond by ionising X-ray radiation for use in SAXS with confidence. Another interesting SAS study would involve segmental labelling of 2dCD4-WT. Alternative segmental labelling of domain 1 and domain 2 would allow the study of how the individual domains change in size as a function of their redox state.

Dynamics have been shown to be at play in this complex redox switch system, whereby reduction of the domain 2 disulphide confers increased dynamics to the gp120 binding site. It

would be interesting to use quasi-elastic neutron scattering look at the protein and solvent dynamics of the system. However this would necessitate very large quantities (100 mg to 1 g) of both hydrogenated and deuterated protein, which are not feasible considering the difficulties obtaining tens of milligrams of sample for SAS experiments.

Unfortunately sample availability proved a significant barrier in measuring all the samples and contrasts necessary to complete a full contrast variation study of the complex. Measurement of the complex in 45.2% and 0% D<sub>2</sub>O was prioritised over measurement of the complex at the contrast match point of d-2dCD4-WT because this thesis work is more concerned with structural realignments within the CD4 counterpart of the complex rather than gp120. In addition, with the lack of information concerning the glycan phase of gp120, further processing of the gp120 data was not possible. In order to fully analyse the 2dCD4-WT-bound and -unbound gp120 samples a full glycoprofile study using mass-spectrometry would need to be carried out. Alternatively gp120 could be deglycosylated and the volume of the sugars could be calculated from the difference between the glycosylated and unglycosylated forms. The unglycosylated gp120 species could also be measured and this compared to the fully glycosylated gp120. Modifications such as those performed by Finzi *et al.* 2010 [215] could also be made to gp120 such as those to prevent the dimeric species from forming. The alteration of the structure or function of gp120 could be controlled for by characterising the modified gp120 using techniques such as ELISA and SEC-MALLS-RI.

Small-angle scattering was the method of choice because SANS studies using contrast variation would allow the study of the envelope structure of d-2dCD4-WT when bound to h-gp120. However, unless the conformational changes on gp120 binding were extreme these were unlikely to be reliably detected by SANS due to the low resolution of the technique. Therefore, it would be preferable to study this system using high resolution techniques such as crystallography. The 2dCD4-WT/gp120 complex has already been studied using X-ray crystallography but there is doubt concerning the second domain disulphide due to the negative density around the second domain disulphide bond, the fact that the system is disulphide bond rich and that disulphide bonds are sensitive to X-rays. It would be interesting in future X-ray crystallography studies to couple raman spectroscopy measurements to shooting crystals of the complex with X-rays to look for changes in the raman spectrum as a result of disulphide bond breaking. Low X-ray dose crystallography could also be used to reduce the risk of ionising radiation

damage.

Alternatively neutron crystallography could be used. The advantage of neutron crystallographic studies of this system would be two-fold (i) neutrons are non-destructive (ii) the protonation (or rather deuteration) state of the thiols could be mapped since neutrons can detect hydrogen atoms to determine the redox state of the domain 2 disulphide bond of 2dCD4-WT when bound to gp120. However, this would require 100% perdeuteration of both 2dCD4-WT and gp120. This would be very challenging as a deuteration strategy for gp120 would need to be developed since this protein was recombinantly expressed in mammalian cells and deuterium is toxic to mammalian cells. In addition, stabilising mutations would need to be made to gp120 and the complex may even necessitate the presence of a stabilising antibody as with many of the current X-ray crystal structures of this system. The other limitation with crystallography is of course the need for crystals and for neutron crystallography these need to be large compared to X-rays.

Another high resolution technique which could be used is NMR. The benefit of NMR is that the complex could be measured in solution thereby negating the need for crystals. However, because the system is quite large (over 100 kDa) triple labelling would be necessary ( $^2H$ ,  $^{13}C$  and  $^{15}N$ ). For 2dCD4-WT this would be relatively easy although expensive to do using the CFPE system as triple labelled instead of deuterium labelled amino acids could be added to the reaction to produce the triple labelled protein. However, for the mammalian expressed gp120 this task would be significantly more complicated considering that deuterium is toxic to mammalian cells as mentioned previously. Instead a triple labelling system would need to be developed which would probably mean that fully glycosylated gp120 would be out of the question.

Finally, cryo-electron microscopy (cryo-EM) could be used but the risks for ionising radiation damage are similar to with X-ray crystallography. The complex with monomeric gp120 is on the small side for cryo-EM but the stable Env trimer mimics the so-called SOSIP trimers [43, 95] (trimeric gp120 and gp41 with stabilising mutations) could be used. There are several cryo-EM structures with 2dCD4-WT in complex with the SOSIP trimers, however the resolution is not quite high-enough to be confident about the presence of the second domain disulphide bond and the grids are blasted with a powerful electron gun so the question of the effect of ionising radiation damage is the same as with the X-ray crystal structure.

## 6.7 Conclusions

Unmodified gp120<sub>ZA</sub>CAP45 is expressed as both a monomer and dimer by HEK293 cells and approximately 37% of its total mass is comprised of a heterogeneous array of glycans. The global match-out point of the gp120 protein and its glycan shield is 45.2% D<sub>2</sub>O. However the chemical composition of the gp120<sub>ZA</sub>CAP45 glycans from HEK293 cell expression was not characterised, therefore detailed analysis of gp120 alone and the gp120 component of the complex was not possible. d-2dCD4-WT produced in the CFPE system has a match-out point of 90.5% D<sub>2</sub>O and may form a more compact structure on gp120 binding due as a result of the ablation of its second domain allosteric disulphide bond as determined by SANS. The SAXS measurement of the samples previously measured by SANS was of poor quality, thus further analyses were not pursued. SEC-SAXS analysis of the 2dCD4-WT analogues produced using the bacterial expression systems yielded good quality data. While the gp120-bound d-2dCD4-WT produced using CFPE suffered from radiation damage, the SANS data indicate that this recombinant protein synthesis technique is suitable for production of deuterated protein for study by SANS. This chapter, therefore, validates the use of *B. choshinensis* and CFPE as methods of recombinant protein expression for use in SAS studies.



## Chapter 7

# Conclusions and future work

### 7.1 Conclusions

CD4 is a cell surface receptor found on cells of the immune system and is the primary receptor for HIV-1. CD4 is comprised of four concatenated immunoglobulin-like domains of which domains 1, 2 and 4 contain disulphide bonds. Domain 1 forms interactions with viral gp120 which forms part of the viral envelope glycoprotein complex and contains a structural disulphide bond. The disulphide bond in domain 2 has been classified as an -RHStaple disulphide bond [73]. This type of disulphide bond is the hallmark configuration for a group of disulphide bonds called the allosteric disulphide bonds, which are believed to affect protein function by redox shuffling of their disulphide bonds [74]. Reduction of the disulphide bond in domain 2 has been shown to be essential for gp120 binding to CD4 [63] and subsequent HIV-1 entry into host cells [112]. CD4 is therefore a redox active protein and is believed to be found as an oxidised monomer in its inactive form [63] and as a domain swapped dimer in its active form which does not bind gp120 [56, 57]. The monomeric redox isoform of CD4 which binds gp120 in which its domain 2 disulphide bond is reduced is functionally and therefore probably structurally unique.

The aim of this work was to use biochemical, biophysical and structural techniques to characterise the impact of the redox state of 2dCD4 on its structure and therefore its ability to bind gp120. 2dCD4, referring to domains 1 and 2 of CD4, was used instead of the full four-domain extracellular portion of CD4 since domain 1 is the minimal gp120 binding domain and domain 2 contains the allosteric bond, also essential to gp120 binding. The intention was to compare the structure and behaviour of 2dCD4-WT to a disulphide knockout variant form of 2dCD4

called 2dCD4-D2A, whose second domain cysteine residues had been substituted for alanine residues. 2dCD4-D2A was to be used in order to understand how ablation of the second domain disulphide bond affects 2dCD4 structure and why this is essential for gp120 binding. This was to be done primarily using contrast variation SANS studies for which the CD4 component of the complex would need to be deuterium labelled.

Chapter 3 describes the investigation and the optimisation of the expression of 2dCD4-WT using three quite different expression systems, prior to testing the expression of 2dCD4-D2A. The first of these was the classical *E. coli* expression system. 2dCD4-WT was expressed by *E. coli* into insoluble inclusion bodies which necessitated denaturation and refolding in order to obtain soluble protein during which much of the protein was lost, significantly reducing yields. The second expression host used was *B. choshinensis* which secretes 2dCD4-WT into the expression medium, however the expression was lengthy, there were large volumes of expression medium containing many non-2dCD4-WT proteins which were also secreted and the ultimate yields were low. *B. choshinensis* was also shown to not grow in minimal media and 2dCD4-WT expression in deuterated Silantes media was poor. Finally, sufficient yields of both unlabelled and labelled 2dCD4-WT were achieved using cell-free protein expression (CFPE). However, while expression of 2dCD4-D2A was comparable to that of 2dCD4-WT using CFPE, the protein was shown to be prone to aggregation due to issues with folding and was therefore not suitable for further characterisation.

After optimising the expression of 2dCD4-WT, the protein was biochemically and biophysically characterised, as described in Chapter 4. 2dCD4-WT from CFPE was shown to be functionally active and correctly folded by enzyme-linked immunosorbant assay, pseudo viral neutralisation assay and 1D NMR experiments. Mass spectrometry analysis showed that deuterium labelled 2dCD4-WT was 73% deuterated at its non-labile hydrogen positions but also indicated that a truncation product with an observed molecular weight of 16.6 kDa was formed. This was confirmed by size-exclusion chromatography coupled to multi-angle laser light scattering and refractive index (SEC-MALLS-RI) data. SEC-MALLS-RI also showed that there was a small amount of aggregation present which could be separated from 2dCD4-WT by the *in situ* SEC step which suggested that SEC coupled to small-angle scattering would be the best way to analyse the sample.

Biophysical characterisation of the fully oxidised (2dCD4-Ox), reduced in domain 2 (2dCD4-R1) and fully-reduced (2dCD4-R2) redox isomers of 2dCD4-WT was performed, as described in Chapter 5. Qualitative analysis of the hydrodynamic volume of the different redox isomers by SEC showed that 2dCD4-R1 has the smallest hydrodynamic volume. This was reasoned to be as a result of relaxation of domain 2 due to reduction of the metastable allosteric disulphide bond [203]. 2dCD4-Ox and 2dCD4-R2 were shown to have larger hydrodynamic volumes than 2dCD4-R1 but the data did not allow determination of whether 2dCD4-R2 was larger than 2dCD4-Ox or *vice versa*.

Attempts to corroborate the SEC data were made using small-angle X-ray scattering (SAXS) experiments as presented in Chapter 5. A DTT titration using the robotic sample changer on BM29 at the ESRF suggested that an increase in DTT concentration caused unfolding or a loss of structure. However, it was difficult to uncouple this as a real effect as a result of disulphide bond reduction from aggregation. 2dCD4-WT with and without 50 mM DTT treatment was hence tested using SEC-SAXS which suggested that the 2dCD4-WT treated with 50 mM DTT (which was therefore fully reduced as shown by its non-reducing SDS-PAGE mobility) was slightly larger than 2dCD4-WT without 50 mM DTT treatment.

Chapter 6 assessed the structure of 2dCD4-WT whilst in complex with gp120, after determination of the match-out point of deuterium labelled 2dCD4-WT (d-2dCD4-WT, 90.5% D<sub>2</sub>O) and unlabelled, fully glycosylated gp120 (45% D<sub>2</sub>O). d-2dCD4-WT was measured by small-angle neutron scattering in its gp120-unbound form and compared to its gp120-bound form with gp120 visible (in 0% D<sub>2</sub>O) and gp120 matched out (in 45% D<sub>2</sub>O). d-2dCD4-WT bound to gp120 (with gp120 matched out) was found to have a more compact structure due to a slight bend between domains 1 and 2 and *ab initio* modelling suggested that domain 2 shrinks slightly in size. Since 2dCD4-WT in isolation is a mixture of the three redox isomers (2dCD4-Ox, -R1 and -R2) it was difficult to uncouple this conformational change on gp120 binding from this being the structure of 2dCD4-R1, since this is the redox isoform binding gp120. This work adds some ambiguity to the current high-resolution crystal data, since the crystallographic data appears to show the presence of at least a partial disulphide bond in the 2dCD4/gp120 complex. The crystal data may therefore not show the full extent of the dynamic structural realignment required for CD4 to bind to gp120.

Finally, SEC-SAXS data was also collected on gp120 in complex with 2dCD4-WT expressed and

refolded from *E. coli* inclusion bodies, *B. choshinensis* secretion and CFPE soluble production which corroborated the published size parameter data, thereby validating *B. choshinensis* and CFPE as novel recombinant expression systems for unlabelled and labelled protein for SAS experiments.

Ultimately this thesis work has shown that: (i) CFPE is a suitable expression system for the production of soluble, functional, match-out labelled and unlabelled recombinant protein suitable for biophysical and structural studies; (ii) reduction of the highly strained second domain disulphide bond of 2dCD4 causes this domain to become smaller so that 2dCD4-R1 (the redox isomer that binds to gp120) has a smaller hydrodynamic volume than 2dCD4-Ox and 2dCD4-R2. This is due to an energetically favourable domain relaxation in domain 2 [203], thanks to the release of the pinch point in the middle of the beta strands at the disulphide bond. Reduction of the disulphide bond in domain 2 is believed to have an allosteric effect on the gp120 binding site in domain 1. The reduction of the disulphide bond in domain 2 is communicated to domain 1 to increase the dynamics at the gp120 binding site to facilitate gp120 binding as shown by Owen *et al.* 2018 [218].

These previously undescribed structural differences as a function of the redox state of the disulphide bonds within 2dCD-WT lay the foundations for future studies to obtain high-resolution data on this system. High-resolution structural data would show on an atomic level the subtle differences in 2dCD4-R1 that cause it to be more compact and potentially how the structural changes as a result of reduction of the second domain disulphide bond are relayed to domain one, increasing its dynamics at the gp120 binding site that in turn allow the CD4 receptor to bind gp120. Understanding the unique structure of 2dCD4-R1 would allow development of anti-CD4 directed immunogens that would prevent HIV-1 binding and therefore entry without necessarily interrupting the normal immune function of CD4.

## 7.2 Future work

In order to concretely determine what the effect of domain 2 disulphide bond ablation is on the structure of 2dCD4 and how this compares to 2dCD4-WT (the redox isomer mixture) and 2dCD4-R2 (fully reduced 2dCD4-WT), further work needs to be carried out in order to establish

an expression protocol for correctly folded 2dCD4-D2A using CFPE. With soluble deuterium labelled 2dCD4-D2A the SANS contrast variation studies of the d-2dCD4/gp120 complex could be repeated and the data compared to that obtained for 2dCD4-WT. Access to even low resolution SAS data obtained 2dCD4-D2A in the gp120-bound and -unbound states would alleviate some of the ambiguity as to whether there is a difference in structure between 2dCD4-WT, 2dCD4-D2A and both bound to gp120.

As mentioned above, high-resolution structural data is needed to decipher the local structural changes around the disulphide bond and allosteric changes that confer increased dynamics to the gp120 binding site in domain 1. High-resolution X-ray crystal structural data already exists for 2dCD4-WT in complex with gp120 but does not fully reflect the dynamic redox shuffling events required for 2dCD4 to bind to gp120. Equally cryo-EM data exists for 2dCD4-WT with the soluble Env trimer but the resolution is too low to determine the redox state of the disulphide bond in domain 2. Use of the 2dCD4-D2A variant protein would allow determination of the structure of 2dCD4-R1 at the point of gp120 binding. Ultimately, high-resolution data collected on gp120-bound and -unbound 2dCD4-D2A would help inform as to how allosteric redox switches can affect protein function depending on their redox state so their implications in disease and infection, such as HIV-1, can be better understood.



# Bibliography

- (1) WHO <http://www.who.int/gho/hiv/en/> **2018**.
- (2) Levy, J. *Microbiological reviews* **1993**, *57*, 183–289.
- (3) <https://www.cdc.gov/hiv/basics/index.html> CDC.
- (4) Lu, K.; Heng, X.; Summers, M. F. *Journal of Molecular Biology* **2011**, *410*, 609–633.
- (5) Splettstoesser, T. Structure of the HIV-1 genome., 2014.
- (6) Briggs, J. A.; Grünewald, K.; Glass, B.; Förster, F.; Kräusslich, H. G.; Fuller, S. D. *Structure* **2006**, *14*, 15–20.
- (7) Briggs, J. A.; Kräusslich, H. G. *Journal of Molecular Biology* **2011**, *410*, 491–500.
- (8) Leitner, T.; Hahn, B.; Foundation, H. M. J.; Kuiken, C.; Foley, B.; Marx, P.; Wolinsky, S.; Lo, C.-c.; Macke, J.; Szinger, J. J.; Thurmond, J.; Yoon, H.; Zhang, M., *HIV Sequence Compendium*, 2008.
- (9) Hallenberger, S.; Bosch, V.; Angliker, H.; Shaw, E.; Klenk, H. D.; Garten, W. *Nature* **1992**, *360*, 358–361.
- (10) Splettstoesser, T. Diagram of the HIV virion., 2014.
- (11) Briggs, J. A. G.; Wilk, T.; Welker, R.; Kra, H.-g.; Fuller, S. D. *The EMBO journal* **2003**, *22*, 1707–1715.
- (12) Zhu, P.; Liu, J.; Bess, J.; Chertova, E.; Lifson, J. D.; Grisé, H.; Ofek, G. A.; Taylor, K. A.; Roux, K. H. *Nature* **2006**, *441*, 847–852.
- (13) Wyatt, R.; Sodroski, J. *Science* **1998**, *280*, 1884–1888.
- (14) Rizzuto, C. D.; Cgacgat, G.; Luc, T.-.; Rizzuto, C. D.; Wyatt, R.; Herna, N.; Sun, Y.; Kwong, P. D.; Hendrickson, W. A.; Sodroski, J. *Science* **1998**, *280*, 1949–1953.
- (15) Wilen, C. B.; Tilton, J. C.; Doms, R. W. *Cold Spring Harbor Perspectives in Medicine* **2012**, *2*, 1–14.

- (16) Deng, H.; Liu, R.; Ellmeier, W.; Choe, S.; Unutmaz, D.; Burkhart, M.; Di Marzio, P.; Marmor, S.; Sutton, R.; Hill, C.; Davis, C.; Peiper, S.; Schall, T.; Littman, D.; Landau, N. *Nature* **1996**, *381*.
- (17) Qin, L.; Kufareva, I.; Holden, L. G.; Wang, C.; Zheng, Y.; Fenalti, G.; Wu, H.; Han, G. W.; Cherezov, V.; Stevens, R. C.; Handel, T. M. *Science* **2015**, *347*, 1117–1122.
- (18) Buzon, V.; Natrajan, G.; Schibli, D.; Campelo, F.; Kozlov, M. M.; Weissenhorn, W. *PLoS Pathogens* **2010**, *6*, 1–7.
- (19) Yu, F.; Lu, L.; Du, L.; Zhu, X.; Debnath, A. K.; Jiang, S. *Viruses* **2013**, *5*, 127–149.
- (20) Ward, A. B.; Wilson, I. a. *Trends in Biochemical Sciences* **2015**, *40*, 101–107.
- (21) Jacobo-Molina, A.; Ding, J.; Nanni, R. G.; Clark, A. D.; Lu, X.; Tantillo, C.; Williams, R. L.; Kamer, G.; Ferris, A. L.; Clark, P. *Proceedings of the National Academy of Sciences* **1993**, *90*, 6320–6324.
- (22) Miller, M. D.; Farnet, C. M.; Bushman, F. D. *Journal of virology* **1997**, *71*, 5382–5390.
- (23) Maertens, G. N.; Hare, S.; Cherepanov, P. *Nature* **2010**, *468*, 326–329.
- (24) Tahirov, T. H.; Babayeva, N. D.; Varzavand, K.; Cooper, J. J.; Sedore, S. C.; Price, D. H. *Nature* **2010**, *465*, 747–751.
- (25) Charnay, N.; Ivanyi-Nagy, R.; Soto-Rifo, R.; Ohlmann, T.; López-Lastra, M.; Darlix, J.-L. *Retrovirology* **2009**, *6*, 74.
- (26) Daugherty, M. D.; Liu, B.; Frankel, A. D. *Nature Structural & Molecular Biology* **2010**, *17*, 1337–1342.
- (27) Hidalgo, L.; Swanson, C. M. *Biochemical Society Transactions* **2017**, *45*, 353–364.
- (28) Hurley, J. H.; Hanson, P. I. *Nature reviews. Molecular cell biology* **2010**, *11*, 556–566.
- (29) Ono, A.; Ablan, S. D.; Lockett, S. J.; Nagashima, K.; Freed, E. O. *Proceedings of the National Academy of Sciences* **2004**, *101*, 14889–14894.
- (30) Von Schwedler, U. K.; Stuchell, M.; Müller, B.; Ward, D. M.; Chung, H. Y.; Morita, E.; Wang, H. E.; Davis, T.; He, G. P.; Cimborra, D. M.; Scott, A.; Kräusslich, H. G.; Kaplan, J.; Morham, S. G.; Sundquist, W. I. *Cell* **2003**, *114*, 701–713.
- (31) De Marco, A.; Müller, B.; Glass, B.; Riches, J. D.; Kräusslich, H. G.; Briggs, J. A. G. *PLoS Pathogens* **2010**, *6*, DOI: [10.1371/journal.ppat.1001215](https://doi.org/10.1371/journal.ppat.1001215).
- (32) Barré-Sinoussi, F.; Ross, A. L.; Delfraissy, J.-F. *Nature reviews. Microbiology* **2013**, *11*, 877–83.



- (33) Allan, J. S.; Coligan, J. E.; Barin, F.; Mclane, M. F.; Sodroski, J. G.; Rosen, C. A.; Haseltine, W. A.; Lee, T. H.; Essex, M. *Science* **1985**, *228*, 1091–1094.
- (34) Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. a. *Nature* **1998**, *393*, 648–659.
- (35) Moore, J. P.; Trkola, A.; Dragic, T. *Current opinion in immunology* **1997**, *9*, 551–62.
- (36) Frost, S. D. W.; Wrin, T.; Smith, D. M.; Pond, S. L. K.; Liu, Y.; Paxinos, E.; Chappay, C.; Galovich, J.; Beauchaine, J.; Petropoulos, C. J.; Little, S. J.; Richman, D. D. *Proceedings of the National Academy of Sciences* **2005**, *102*, 18514–18519.
- (37) Leonard, C. K.; Spellman, M. W.; Riddle, L.; Harris, R. J.; Thomas, J. N.; Gregory, T. J. *J Biol Chem* **1990**, *265*, 10373–10382.
- (38) Mascola, J. R.; Montefiori, D. C. *Nature Medicine* **2003**, *9*, 393–394.
- (39) Van Anken, E.; Sanders, R. W.; Liscaljet, I.; Land, A.; Bontjer, I.; Tillemans, S.; Nabatov, A.; Paxton, W.; Berkhout, B.; Braakman, I. *Molecular biology of the cell* **2008**, *19*, 4298–4309.
- (40) Sanders, R. W.; Derking, R.; Cupo, A.; Julien, J. P.; Yasmeen, A.; de Val, N.; Kim, H. J.; Blattner, C.; de la Peña, A. T.; Korzun, J.; Golabek, M.; de los Reyes, K.; Ketas, T. J.; van Gils, M. J.; King, C. R.; Wilson, I. a.; Ward, A. B.; Klasse, P. J.; Moore, J. P. *PLoS Pathogens* **2013**, *9*, DOI: [10.1371/journal.ppat.1003618](https://doi.org/10.1371/journal.ppat.1003618).
- (41) Do Kwon, Y.; Pancera, M.; Acharya, P.; Georgiev, I. S.; Crooks, E. T.; Gorman, J.; Joyce, M. G.; Guttman, M.; Ma, X.; Narpala, S.; Soto, C.; Terry, D. S.; Yang, Y.; Zhou, T.; Ahlsen, G.; Bailer, R. T.; Chambers, M.; Chuang, G.-Y.; Doria-Rose, N. a.; Druz, A.; Hallen, M. a.; Harned, A.; Kirys, T.; Louder, M. K.; O'Dell, S.; Ofek, G.; Osawa, K.; Prabhakaran, M.; Sastry, M.; Stewart-Jones, G. B. E.; Stuckey, J.; Thomas, P. V.; Tittley, T.; Williams, C.; Zhang, B.; Zhao, H.; Zhou, Z.; Donald, B. R.; Lee, L. K.; Zolla-Pazner, S.; Baxa, U.; Schön, A.; Freire, E.; Shapiro, L.; Lee, K. K.; Arthos, J.; Munro, J. B.; Blanchard, S. C.; Mothes, W.; Binley, J. M.; McDermott, A. B.; Mascola, J. R.; Kwong, P. D. *Nature Structural & Molecular Biology* **2015**, *22*, 522–531.
- (42) Lyumkis, D.; Julien, J. P.; De Val, N.; Cupo, A.; Potter, C. S.; Klasse, P. J.; Burton, D. R.; Sanders, R. W.; Moore, J. P.; Carragher, B.; Wilson, I. A.; Ward, A. B. *Science* **2013**, *342*, 1484–1490.
- (43) Julien, J.-P.; Cupo, A.; Sok, D.; Stanfield, R. L.; Lyumkis, D.; Deller, M. C.; Klasse, P.-J.; Burton, D. R.; Sanders, R. W.; Moore, J. P.; Ward, A. B.; Wilson, I. a. *Science* **2013**, *342*, 1477–1483.

- (44) Pancera, M.; Zhou, T.; Druz, A.; Georgiev, I. S.; Soto, C.; Gorman, J.; Huang, J.; Acharya, P.; Chuang, G.-y.; Ofek, G.; Stewart-jones, G. B. E.; Stuckey, J.; Bailer, R. T.; Joyce, M. G.; Louder, M. K.; Tumba, N.; Yang, Y.; Zhang, B.; Cohen, M. S.; Haynes, B. F.; Mascola, J. R.; Morris, L.; Munro, J. B.; Blanchard, S. C.; Mothes, W.; Connors, M.; Kwong, P. D. *Nature* **2014**, *514*, 455–461.
- (45) Ozorowski, G.; Pallesen, J.; De Val, N.; Lyumkis, D.; Cottrell, C. A.; Torres, J. L.; Copps, J.; Stanfield, R. L.; Cupo, A.; Pugach, P.; Moore, J. P.; Wilson, I. A.; Ward, A. B. *Nature* **2017**, *547*, 360–361.
- (46) Wang, H.; Cohen, A. A.; Galimidi, R. P.; Gristick, H. B.; Jensen, G. J.; Bjorkman, P. J. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, E7151–E7158.
- (47) Bartesaghi, A.; Merk, A.; Borgnia, M. J.; Milne, J. L. S.; Subramaniam, S. *Nature Structural & Molecular Biology* **2013**, *20*, 1352–1357.
- (48) Dalglish, a. G.; Beverley, P. C.; Clapham, P. R.; Crawford, D. H.; Greaves, M. F.; Weiss, R. a. *Nature* **1984**, *312*, 763–767.
- (49) Lee, B.; Sharron, M.; Montaner, L. J.; Weissman, D.; Doms, R. W. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 5215–5220.
- (50) König, R. *Current Opinion in Immunology* **2002**, *14*, 75–83.
- (51) Li, Q.-J.; Dinner, A. R.; Qi, S.; Irvine, D. J.; Huppa, J. B.; Davis, M. M.; Chakraborty, A. K. *Nature immunology* **2004**, *5*, 791–799.
- (52) De Aós, I.; Metzger, M. H.; Exley, M.; Dahl, C. E.; Misra, S.; Zheng, D.; Varticovski, L.; Terhorst, C.; Sancho, J. *The Journal of biological chemistry* **1997**, *272*, 25310–8.
- (53) Love, P. E.; Hayes, S. M. ITAM-mediated signaling by the T-cell antigen receptor., 2010.
- (54) Wu, H.; Kwong, P. D.; Hendrickson, W. a. *Nature* **1997**, *387*, 527–530.
- (55) Lamarre, D.; Capon, D. J.; Karp, D. R.; Gregory, T.; Long, E. O.; Sékaly, R. P. *The EMBO journal* **1989**, *8*, 3271–3277.
- (56) Moldovan, M.-C.; Yachou, A.; Lévesque, K.; Wu, H.; Hendrickson, W. a.; Cohen, E. a.; Sékaly, R.-P. *Journal of immunology (Baltimore, Md. : 1950)* **2002**, *169*, 6261–6268.
- (57) Moldovan, M.-C.; Sabbagh, L.; Breton, G.; Sékaly, R.-P.; Krummel, M. F. *Journal of immunology (Baltimore, Md. : 1950)* **2006**, *176*, 5438–5445.
- (58) Maekawa, A.; Schmidt, B.; Fazekas de St Groth, B.; Sanejouand, Y.-H.; Hogg, P. J. *Journal of immunology (Baltimore, Md. : 1950)* **2006**, *176*, 6873–6878.

- (59) Sakihama, T.; Smolyar, a.; Reinherz, E. L. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 6444–6448.
- (60) Bourgeois, R.; Mercier, J.; Paquette-Brooks, I.; Cohen, E. a. *Retrovirology* **2006**, *3*, 31.
- (61) Fournier, M.; Peyrou, M.; Bourgoïn, L.; Maeder, C.; Tchou, I.; Foti, M. *Molecular Immunology* **2010**, *47*, 2594–2603.
- (62) Sanejouand, Y. H. *Proteins: Structure, Function and Genetics* **2004**, *57*, 205–212.
- (63) Cerutti, N.; Killick, M.; Jugnarain, V.; Papathanasopoulos, M.; Capovilla, A. *Journal of Biological Chemistry* **2014**, *289*, 10455–10465.
- (64) Maddon, P. J.; Littman, D. R.; Godfrey, M.; Maddon, D. E.; Chess, L.; Axel, R. *Cell* **1985**, *42*, 93–104.
- (65) Clark, S. J.; Jefferies, W. A.; Barclay, A. N.; Gagnon, J.; Williams, A. F. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84*, 1649–1653.
- (66) Ryu, S.-E.; Kwong, P.; Truneh, A.; Porter, T.; Arthos, J.; Rosenberg, M.; Dai, X.; Xuong, N.-h.; Axel, R.; Sweet, R.; Hendrickson, W. A. *Nature* **1990**, *348*, 419–426.
- (67) Wang, J. H.; Yan, Y. W.; Garrett, T. P.; Liu, J. H.; Rodgers, D. W.; Garlick, R. L.; Tarr, G. E.; Husain, Y.; Reinherz, E. L.; Harrison, S. C. *Nature* **1990**, *348*, 411–418.
- (68) Brady, R. L.; Dodson, E. J.; Dodson, G. G.; Lange, G.; Davis, S. J.; Williams, a. F.; Barclay, a. N. *Science (New York, N.Y.)* **1993**, *260*, 979–983.
- (69) Williams, A.; Barclay, A. *Annual Review of Immunology* **1988**, *6*, 381–405.
- (70) Bork, P.; Holm, L.; Sander, C. The immunoglobulin fold: Structural classification, sequence patterns and common core., 1994.
- (71) Garrett, T. P. J.; Wang, J.; Yan, Y.; Liu, J.; Harrison, S. C. Refinement and Analysis of the Structure of the First 2 Domains of Human CD4.pdf., 1993.
- (72) Guttman, M.; Kahn, M.; Garcia, N. K.; Hu, S.-L.; Lee, K. K. *Journal of Virology* **2012**, *86*, 8750–8764.
- (73) Schmidt, B.; Ho, L.; Hogg, P. J. *Biochemistry* **2006**, *45*, 7429–7433.
- (74) Wouters, M. A.; Lau, K. K.; Hogg, P. J. *BioEssays* **2004**, *26*, 73–79.
- (75) Karlsson Hedestam, G. B.; Fouchier, R. a. M.; Phogat, S.; Burton, D. R.; Sodroski, J.; Wyatt, R. T. *Nature reviews. Microbiology* **2008**, *6*, 143–155.
- (76) Kwong, P. D.; Doyle, M. L.; Casper, D. J.; Cicala, C.; Leavitt, S. A.; Majeed, S.; Steenbeke, T. D.; Venturi, M.; Chaiken, I.; Fung, M.; Katinger, H.; Parren, P. W.; Robinson, J.; Van

- Ryk, D.; Wang, L.; Burton, D. R.; Freire, E.; Wyatt, R.; Sodroski, J.; Hendrickson, W. A.; Arthos, J. *Nature* **2002**, *420*, 678–682.
- (77) Labrijn, A. F.; Poignard, P.; Raja, A.; Zwick, M. B.; Delgado, K.; Franti, M.; Binley, J.; Vivona, V.; Grundner, C.; Huang, C.-C.; Venturi, M.; Petropoulos, C. J.; Wrin, T.; Dimitrov, D. S.; Robinson, J.; Kwong, P. D.; Wyatt, R. T.; Sodroski, J.; Burton, D. R. *Journal of Virology* **2003**, *77*, 10557–10565.
- (78) Binley, J. M.; Ban, Y. E. A.; Crooks, E. T.; Eggink, D.; Osawa, K.; Schief, W. R.; Sanders, R. W. *Journal of Virology* **2010**, *84*, 5637–5655.
- (79) Frey, G.; Peng, H.; Rits-Volloch, S.; Morelli, M.; Cheng, Y.; Chen, B. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 3739–44.
- (80) Poignard, P.; Moulard, M.; Golez, E.; Vivona, V.; Franti, M.; Venturini, S.; Wang, M.; Parren, P. W. H. I.; Burton, D. R. *Journal of virology* **2003**, *77*, 353–65.
- (81) Moore, P. L.; Crooks, E. T.; Porter, L.; Zhu, P.; Cayanan, C. S.; Grise, H.; Corcoran, P.; Zwick, M. B.; Franti, M.; Morris, L.; Roux, K. H.; Burton, D. R.; Binley, J. M. *Journal of virology* **2006**, *80*, 2515–28.
- (82) Klein, J. S.; Gnanapragasam, P. N. P.; Galimidi, R. P.; Foglesong, C. P.; West, A. P.; Bjorkman, P. J. *Proceedings of the National Academy of Sciences* **2009**, *106*, 7385–7390.
- (83) Stamatatos, L.; Morris, L.; Burton, D. R.; Mascola, J. R. *Nature medicine* **2009**, *15*, 866–870.
- (84) Kwong, P. D.; Mascola, J. R.; Nabel, G. J. *Nature Reviews Immunology* **2013**, *13*, 693–701.
- (85) Mouquet, H. *Trends in Immunology* **2014**, *35*, 549–561.
- (86) Simek, M. D.; Rida, W.; Priddy, F. H.; Pung, P.; Carrow, E.; Laufer, D. S.; Lehrman, J. K.; Boaz, M.; Tarragona-Fiol, T.; Miiro, G.; Birungi, J.; Pozniak, A.; McPhee, D. A.; Manigart, O.; Karita, E.; Inwoley, A.; Jaoko, W.; DeHovitz, J.; Bekker, L.-G.; Pitisuttithum, P.; Paris, R.; Walker, L. M.; Poignard, P.; Wrin, T.; Fast, P. E.; Burton, D. R.; Koff, W. C. *Journal of Virology* **2009**, *83*, 7337–7348.
- (87) Burton, D.; Pyati, J.; Koduri, R.; Sharp, S.; Thornton, G.; Parren, P.; Sawyer, L.; Hendry, R.; Dunlop, N.; Nara, P.; Et, A. *Science* **1994**, *266*, 1024–1027.
- (88) Wu, X.; Yang, Z. Y.; Li, Y.; Hogerkorp, C. M.; Schief, W. R.; Seaman, M. S.; Zhou, T.; Schmidt, S. D.; Wu, L.; Xu, L.; Longo, N. S.; McKee, K.; O'Dell, S.; Louder, M. K.; Wycuff, D. L.; Feng, Y.; Nason, M.; Doria-Rose, N.; Connors, M.; Kwong, P. D.; Roederer, M.; Wyatt, R. T.; Nabel, G. J.; Mascola, J. R. *Science* **2010**, *329*, 856–861.

- (89) Zhou, T.; Georgiev, I.; Wu, X.; Yang, Z. Y.; Dai, K.; Finzi, A.; Kwon, Y. D.; Scheid, J. F.; Shi, W.; Xu, L.; Yang, Y.; Zhu, J.; Nussenzweig, M. C.; Sodroski, J.; Shapiro, L.; Nabel, G. J.; Mascola, J. R.; Kwong, P. D. *Science* **2010**, *329*, 811–817.
- (90) Burton, D. R.; Walker, L. M.; Phogat, S. K.; Chan-Hui, P. Y.; Wagner, D.; Phung, P.; Goss, J. L.; Wrin, T.; Simek, M. D.; Fling, S.; Mitcham, J. L.; Lehrman, J. K.; Priddy, F. H.; Olsen, O. A.; Frey, S. M.; Hammond, P. W.; Protocol, G.; Kaminsky, S.; Zamb, T.; Moyle, M.; Koff, W. C.; Poignard, P. *Science* **2009**, *326*, 285–289.
- (91) Walker, L. M.; Huber, M.; Doores, K. J.; Falkowska, E.; Pejchal, R.; Julien, J. P.; Wang, S. K.; Ramos, A.; Chan-Hui, P. Y.; Moyle, M.; Mitcham, J. L.; Hammond, P. W.; Olsen, O. A.; Phung, P.; Fling, S.; Wong, C. H.; Phogat, S.; Wrin, T.; Simek, M. D.; Koff, W. C.; Wilson, I. A.; Burton, D. R.; Poignard, P. *Nature* **2011**, *477*, 466–470.
- (92) Trkola, A.; Pomales, A. B.; Yuan, H.; Korber, B.; Maddon, P. J.; Allaway, G. P.; Katinger, H.; Barbas, C. F.; Burton, D. R.; Ho, D. D. *Journal of virology* **1995**, *69*, 6609–17.
- (93) Muster, T.; Steindl, F.; Purtscher, M.; Trkola, A.; Klima, A.; Himmler, G.; Rucker, F.; Katinger, H. *Journal of virology* **1993**, *67*, 6642–7.
- (94) Huang, J.; Ofek, G.; Laub, L.; Louder, M. K.; Doria-Rose, N. A.; Longo, N. S.; Imamichi, H.; Bailer, R. T.; Chakrabarti, B.; Sharma, S. K.; Alam, S. M.; Wang, T.; Yang, Y.; Zhang, B.; Migueles, S. A.; Wyatt, R.; Haynes, B. F.; Kwong, P. D.; Mascola, J. R.; Connors, M. *Nature* **2012**, *491*, 406–412.
- (95) Julien, J.-P.; Lee, J. H.; Cupo, A.; Murin, C. D.; Derking, R.; Hoffenberg, S.; Caulfield, M. J.; King, C. R.; Marozsan, A. J.; Klasse, P. J.; Sanders, R. W.; Moore, J. P.; Wilson, I. a.; Ward, A. B. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 4351–6.
- (96) Moore, J. P.; Sattentau, Q. J.; Klasse, P. J.; Burkly, L. C. *J. Virol.* **1992**, *66*, 4784–4793.
- (97) Song, R.; Franco, D.; Kao, C.-Y.; Yu, F.; Huang, Y.; Ho, D. D. *Journal of Virology* **2010**, *84*, 6935–6942.
- (98) Freeman, M. M.; Seaman, M. S.; Rits-Volloch, S.; Hong, X.; Kao, C. Y.; Ho, D. D.; Chen, B. *Structure* **2010**, *18*, 1632–1641.
- (99) Iacob, S. A.; Iacob, D. G. *Frontiers in Microbiology* **2017**, *8*, 1–8.
- (100) Matthias, L. J.; Hogg, P. J. *Antioxidants & redox signaling* **2003**, *5*, 133–138.
- (101) Hogg, P. *Disulfide Bond Analysis.*, 2015.
- (102) Hogg, P. J. *Trends in Biochemical Sciences* **2003**, *28*, 210–214.

- (103) Hogg, P. J. *Journal of Thrombosis and Haemostasis* **2009**, *7*, 13–16.
- (104) Bechtel, T. J.; Weerapana, E. From structure to redox: The diverse functional roles of disulfides and implications in disease., 2017.
- (105) Donoghue, N.; Yam, P. T.; Jiang, X. M.; Hogg, P. J. *Protein science : a publication of the Protein Society* **2000**, *9*, 2436–2445.
- (106) Metcalfe, C.; Cresswell, P.; Ciaccia, L.; Thomas, B.; Barclay, a. N. *Open Biology* **2011**, *1*, 110010–110010.
- (107) Cook, K. M.; McNeil, H. P.; Hogg, P. J. *Journal of Biological Chemistry* **2013**, *288*, 34920–34929.
- (108) Chen, V. M.; Ahamed, J.; Versteeg, H. H.; Berndt, M. C.; Ruf, W.; Hogg, P. J. *Biochemistry* **2006**, *45*, 12020–12028.
- (109) Butera, D.; Cook, K.; Chiu, J. *Blood* **2014**, *123*, 2000–2007.
- (110) Groitl, B.; Jakob, U. *Biochimica et Biophysica Acta - Proteins and Proteomics* **2014**, *1844*, 1335–1343.
- (111) Lawrence, D. a.; Song, R.; Weber, P. *Journal of leukocyte biology* **1996**, *60*, 611–618.
- (112) Matthias, L. J.; Yam, P. T. W.; Jiang, X.-M.; Vandegraaff, N.; Li, P.; Pournourios, P.; Donoghue, N.; Hogg, P. J. *Nature immunology* **2002**, *3*, 727–732.
- (113) Matthias, L. J.; Azimi, I.; Tabrett, C. a.; Hogg, P. J. *Journal of Biological Chemistry* **2010**, *285*, 40793–40799.
- (114) Choi, H.; Aboulfatova, K.; Pownall, H. J.; Cook, R.; Dong, J. F. *Journal of Biological Chemistry* **2007**, *282*, 35604–35611.
- (115) Barbouche, R.; Miquelis, R.; Jones, I. M.; Fenouillet, E. *Journal of Biological Chemistry* **2003**, *278*, 3131–3136.
- (116) Ryser, H. J.; Levy, E. M.; Mandel, R.; DiSciullo, G. J. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 4559–4563.
- (117) Gallina, A.; Hanley, T. M.; Mandel, R.; Trahey, M.; Broder, C. C.; Viglianti, G. a.; Ryser, H. J. P. *Journal of Biological Chemistry* **2002**, *277*, 50579–50588.
- (118) Ashish; Juncadella, I. J.; Garg, R.; Boone, C. D.; Anguita, J.; Krueger, J. K. *Journal of Biological Chemistry* **2008**, *283*, 2761–2772.
- (119) Ryser, H. J. P.; Flückiger, R. *Drug Discovery Today* **2005**, *10*, 1085–1094.

- (120) Kwong, P. D.; Ryu, S. E.; Hendrickson, W. a.; Axel, R.; Sweet, R. M.; Folena-Wasserman, G.; Hensley, P.; Sweet, R. W. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87*, 6423–6427.
- (121) Moolla, N.; Killick, M.; Papathanasopoulos, M.; Capovilla, A. *Biochimica et Biophysica Acta - General Subjects* **2016**, *1860*, 1854–1863.
- (122) Lynch, G. W.; Turville, S.; Carter, B.; Sloane, A. J.; Chan, A.; Muljadi, N.; Li, S.; Low, L.; Armati, P.; Raison, R.; Zoellner, H.; Williamson, P.; Cunningham, A.; Church, W. B. *Immunology and Cell Biology* **2006**, *84*, 154–165.
- (123) Fenouillet, E.; Barbouche, R.; Courageot, J.; Miquelis, R. *The Journal of infectious diseases* **2001**, *183*, 744–752.
- (124) Auwerx, J.; Isacson, O.; Söderlund, J.; Balzarini, J.; Johansson, M.; Lundberg, M. *International Journal of Biochemistry and Cell Biology* **2009**, *41*, 1269–1275.
- (125) Azimi, I.; Matthias, L. J.; Center, R. J.; Wong, J. W. H.; Hogg, P. J. *Journal of Biological Chemistry* **2010**, *285*, 40072–40080.
- (126) Reiser, K.; Franois, K. O.; Schols, D.; Bergman, T.; Jörnvall, H.; Balzarini, J.; Karlsson, A.; Lundberg, M. *International Journal of Biochemistry and Cell Biology* **2012**, *44*, 556–562.
- (127) Brown, T., *Gene cloning & DNA analysis: an introduction*, 6th; Wiley-Blackwell: Chichester, 2010.
- (128) Qiagen *Product manual* **2012**.
- (129) Promega **2010**, 1–12.
- (130) Okamoto, A.; Kosugi, A.; Koizumi, Y.; Yanagida, F.; Udaka, S. *Bioscience, Biotechnology, and Biochemistry* **1997**, *61*, 202–203.
- (131) Klebe, R. J.; Harriss, J. V.; Sharp, Z. D.; Douglas, M. G. *Gene* **1983**, *25*, 333–341.
- (132) Li, J.; Han, W.; Yu, Y. In *Protein Engineering - Technology and Application*, 2013, pp 33–61.
- (133) Held, D.; Kilz, P. *Monitoring Polymerization Reactions: From Fundamentals to Applications* **2014**, *i*, 171–199.
- (134) Irvine, G. B. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]* **2001**, *Chapter 5*, Unit 5.5.
- (135) Beer, A. *Ann. der Phys. und Chemie* **1852**, *86*, 78–88.
- (136) Pace, C.; Schmid, F. In *Protein Structure: A Practical Approach*, Creighton, T., Ed.; IRL Press: Oxford, UK, 1997, pp 253–259.
- (137) Margolis, J.; Kenrick, K. G. *Nature* **1969**, *221*, 1056–1057.

- (138) Laemmli, U. K. *Nature* **1970**, *227*, 680–685.
- (139) Reynolds, J. A.; Tanford, C. *Proceedings of the National Academy of Sciences* **1970**, *66*, 1002–1007.
- (140) De St. Groth, S.; Webster, R.; Datyner, A. *Biochimica et Biophysica Acta* **1963**, *71*, 377–391.
- (141) Burnette, W. N. *Analytical Biochemistry* **1981**, *112*, 195–203.
- (142) Van Weemen, B. K.; Schuurs, A. H. *FEBS Letters* **1971**, *15*, 232–236.
- (143) Engvall, E.; Perlmann, P. *Immunochemistry* **1971**, *8*, 871–874.
- (144) Lequin, R. M. *Clinical Chemistry* **2005**, *51*, 2415–2418.
- (145) Mascola, J. R.; D'Souza, P.; Gilbert, P.; Hahn, B. H.; Haigwood, N. L.; Morris, L.; Petropoulos, C. J.; Polonis, V. R.; Sarzotti, M.; Montefiori, D. C. *Journal of Virology* **2005**, *79*, 10103–10107.
- (146) Montefiori, D. C. In *HIV protocols*; Humana Press: 2009, pp 395–405.
- (147) Edman, P. *Acta Chemica Scandinavica* **1950**, *4*, 283–293.
- (148) Reim, D.; Speicher, D. *Current Protocols in Protein Science* **1997**, DOI: [10.1002/0471140864.ps1110s57](https://doi.org/10.1002/0471140864.ps1110s57).
- (149) Edman, P.; Begg, G. *European Journal of Biochemistry* **1967**, *1*, 80–91.
- (150) Zaccai, N.; Serdyuk, I.; Zaccai, G., *Methods in Molecular Biophysics: Structure Dynamics, Function for Biology and Medicine*, 2nd; Cambridge University Press: Cambridge, 2017.
- (151) Willoughby, R. C.; Browner, R. F. *Analytical Chemistry* **1984**, *56*, 2626–2631.
- (152) Yamashita, M.; Fenn, J. B. *Journal of Physical Chemistry* **1984**, *88*, 4671–4675.
- (153) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. *Rapid Communications in Mass Spectrometry* **1988**, *2*, 151–153.
- (154) Podzimek, S. *Journal of Applied Polymer Science* **1994**, *54*, 91–103.
- (155) Wyatt, P. J. *Analytica Chimica Acta* **1993**, *272*, 1–40.
- (156) Zimm, B. H. *The Journal of Chemical Physics* **1948**, *16*, 1099–1116.
- (157) Pauli, W. *Zeitschrift für Physik* **1925**, *31*, 765–783.
- (158) Gerlach, W.; Stern, O. *Annalen der Physik* **1924**, *379*, 673–699.
- (159) Arnold, J. T.; Packard, M. E. *The Journal of Chemical Physics* **1951**, *19*, 1608–1609.
- (160) Bloch, F.; Hansen, W. W.; Packard, M. *Physical Review* **1946**, *70*, 474–485.
- (161) Overhauser, A. W. *Physical Review* **1953**, *92*, 411–415.
- (162) Schanda, P.; Forge, V.; Brutscher, B. *Magnetic Resonance in Chemistry* **2006**, *44*, 177–184.



- (163) Svergun, D.; Koch, M.; Timmins, P.; May, R., *Small Angle X-ray and Neutron Scattering from Solutions of Biological Macromolecules*, 1st ed.; Oxford University Press: Oxford, 2013.
- (164) Williams, C. In *Neutron and synchrotron radiation for condensed matter studies: Volume 1*, 1993; Chapter 10, pp 235–245.
- (165) Pernot, P.; Round, A.; Barrett, R.; De Maria Antolinos, A.; Gobbo, A.; Gordon, E.; Huet, J.; Kieffer, J.; Lentini, M.; Mattenet, M.; Morawe, C.; Mueller-Dieckmann, C.; Ohlsson, S.; Schmid, W.; Surr, J.; Theveneau, P.; Zerrad, L.; McSweeney, S. *Journal of synchrotron radiation* **2013**, *20*, 660–4.
- (166) Pernot, P.; Theveneau, P.; Giraud, T.; Fernandes, N. R.; Nurizzo, D.; Spruce, D.; Surr, J.; McSweeney, S.; Round, A.; Felisaz, F.; Foedinger, L.; Gobbo, A.; Huet, J.; Villard, C.; Cipriani, F. *Journal of Physics: Conference Series* **2010**, *247*, 1–8.
- (167) Round, A.; Felisaz, F.; Fodinger, L.; Gobbo, A.; Huet, J.; Villard, C.; Blanchet, C. E.; Pernot, P.; McSweeney, S.; Roessle, M.; Svergun, D. I.; Cipriani, F. *Acta Crystallographica Section D: Biological Crystallography* **2015**, *71*, 67–75.
- (168) David, G.; Pérez, J. *Journal of Applied Crystallography* **2009**, *42*, 892–900.
- (169) Round, A.; Brown, E.; Marcellin, R.; Kapp, U.; Westfall, C. S.; Jez, J. M.; Zubieta, C. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *69*, 2072–2080.
- (170) <https://www.ill.eu/users/instruments/instruments-list/d22/characteristics/>.
- (171) Haertlein, M.; Moulin, M.; Devos, J. M.; Laux, V.; Dunne, O.; Trevor Forsyth, V., *Biomolecular Deuteration for Neutron Structural Biology and Dynamics*, 1st ed.; Elsevier Inc.: 2016; Vol. 566, pp 113–157.
- (172) Dunne, O.; Weidenhaupt, M.; Callow, P.; Martel, A.; Moulin, M.; Perkins, S. J.; Haertlein, M.; Forsyth, V. T. *European Biophysics Journal* **2017**, *46*, 425–432.
- (173) Guinier, A. *Annales de Physique* **1939**, *11*, 161–237.
- (174) Perkins, S. *European Journal of Biochemistry* **1986**, *157*, 169–180.
- (175) Guinier, A.; Fournet, G. *Journal of Polymer Science* **1956**, *19*, 594–594.
- (176) Glatter, O. *Journal of Applied Crystallography* **1977**, *10*, 415–421.
- (177) Kratky, O.; Porod, G. *Recueil des Travaux Chimiques des Pays Bas* **1949**, *68*, 1106–1122.
- (178) Debye, P. J. *Phys. Colloid. Chem* **1947**, *51*, 18–32.
- (179) <https://www-ssrl.slac.stanford.edu/~saxs/analysis/assessment.htm>.
- (180) Garlick, R. L.; Kirschner, R. J.; Eckenrode, F. M.; Tarpley, W. G.; Tomich, C. S. *AIDS research and human retroviruses* **1990**, *6*, 465–479.

- (181) Udaka, S.; Yamagata, H. *Antonie van Leeuwenhoek* **1993**, *64*, 137–43.
- (182) Tanaka, R.; Kosugi, K.; Mizukami, M.; Ishibashi, M.; Tokunaga, H.; Tokunaga, M. *Protein Expression and Purification* **2004**, *37*, 385–391.
- (183) Tanaka, H.; Baba, T. *Cellular and molecular life sciences : CMLS* **2005**, *62*, 344–54.
- (184) Tanio, M.; Tanaka, T.; Kohno, T. *Analytical Biochemistry* **2008**, *373*, 164–166.
- (185) Kim, T. W.; Kim, D. M.; Choi, C. Y. *Journal of Biotechnology* **2006**, *124*, 373–380.
- (186) Carlson, E. D.; Gan, R.; Hodgman, E. C.; Jewet, M. C. **2014**, *30*, 1185–1194.
- (187) Katzen, F.; Chang, G.; Kudlicki, W. *Trends in Biotechnology* **2005**, *23*, 150–156.
- (188) Zawada, J. F.; Yin, G.; Steiner, A. R.; Yang, J.; Naresh, A.; Roy, S. M.; Gold, D. S.; Heinsohn, H. G.; Murray, C. J. *Biotechnology and Bioengineering* **2011**, *108*, 1570–1578.
- (189) Yin, G.; Swartz, J. R. *Biotechnology and Bioengineering* **2004**, *86*, 188–195.
- (190) Staunton, D.; Schlinkert, R.; Zanetti, G.; Colebrook, S. A.; Campbell, I. D. *Magnetic Resonance in Chemistry* **2006**, *44*, 2–9.
- (191) Matsuda, T.; Koshiha, S.; Tochio, N.; Seki, E.; Iwasaki, N.; Yabuki, T.; Inoue, M.; Yokoyama, S.; Kigawa, T. *Journal of Biomolecular NMR* **2007**, *37*, 225–229.
- (192) Takahashi, H.; Shimada, I. *Journal of Biomolecular NMR* **2010**, *46*, 3–10.
- (193) Kang, S. H.; Kim, D. M.; Kim, H. J.; Jun, S. Y.; Lee, K. Y. *Biotechnology Progress* **2005**, *21*, 1412–1419.
- (194) Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinel, T. *Molecular & Cellular Proteomics* **2006**, *5*, 2336–2349.
- (195) Xiao, Q.; Zhang, F.; Nacev, B. A.; Liu, J. O.; Pei, D. *Biochemistry* **2010**, *49*, 5588–5599.
- (196) Gagoski, D.; Polinkovsky, M. E.; Mureev, S.; Kunert, A.; Johnston, W.; Gambin, Y.; Alexandrov, K. *Biotechnology and Bioengineering* **2016**, *113*, 292–300.
- (197) Ashish; Garg, R.; Anguita, J.; Krueger, J. K. *Biophysical journal* **2006**, *91*, L69–71.
- (198) Wingfield, P. T.; Palmer, I.; Liang, S. In *Current Protocols in Protein Science*; 1; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1995; Vol. 00, pp 6.5.1–6.5.27.
- (199) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I. *Journal of Applied Crystallography* **2017**, *50*, 1212–1225.
- (200) Bioisis: welcome - index.
- (201) Franke, D.; Svergun, D. I. *Journal of Applied Crystallography* **2009**, *42*, 342–346.
- (202) Svergun, D. I. *Journal of Applied Crystallography* **1992**, *25*, 495–503.

- (203) Owen, G. R.; Channell, J. A.; Forsyth, V. T.; Haertlein, M.; Mitchell, E. P.; Capovilla, A.; Papathanasopoulos, M.; Cerutti, N. M. *Biochemistry* **2016**, *55*, 2227–2237.
- (204) Jordan, A.; Jacques, M.; Merrick, C.; Devos, J.; Forsyth, V. T.; Porcar, L.; Martel, A. *Journal of Applied Crystallography* **2016**, *49*, 2015–2020.
- (205) Panico, M.; Bouché, L.; Binet, D.; O'Connor, M. J.; Rahman, D.; Pang, P. C.; Canis, K.; North, S. J.; Desrosiers, R. C.; Chertova, E.; Keele, B. F.; Bess, J. W.; Lifson, J. D.; Haslam, S. M.; Dell, A.; Morris, H. R. *Scientific Reports* **2016**, *6*, 1–17.
- (206) Goh, J. B.; Ng, S. K. *Critical Reviews in Biotechnology* **2017**, *0*, 1–17.
- (207) Croset, A.; Delafosse, L.; Gaudry, J. P.; Arod, C.; Glez, L.; Losberger, C.; Begue, D.; Krstanovic, A.; Robert, F.; Vilbois, F.; Chevalet, L.; Antonsson, B. *Journal of Biotechnology* **2012**, *161*, 336–348.
- (208) Raska, M.; Czernekova, L.; Moldoveanu, Z.; Zachova, K.; Elliott, M. C.; Novak, Z.; Hall, S.; Hoelscher, M.; Maboko, L.; Brown, R.; Smith, P. D.; Mestecky, J.; Novak, J. *AIDS Research and Therapy* **2014**, *11*, 1–16.
- (209) Whitten, A. E.; Cai, S.; Trehwella, J. *Journal of Applied Crystallography* **2008**, *41*, 222–226.
- (210) Svergun, D. I.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccai, G. *Proceedings of the National Academy of Sciences* **1998**, *95*, 2267–2272.
- (211) Volkov, V. V.; Svergun, D. I. In *Journal of Applied Crystallography*, 2003; Vol. 36, pp 860–864.
- (212) Konarev, P. V.; Petoukhov, M. V.; Svergun, D. I. *Journal of Applied Crystallography* **2016**, *49*, 953–960.
- (213) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H.; Svergun, D. I. *Journal of Applied Crystallography* **2003**, *36*, 1277–1282.
- (214) Chamberlain, T. W.; Biskupek, J.; Skowron, S. T.; Bayliss, P. A.; Bichoutskaia, E.; Kaiser, U.; Khlobystov, A. N. *Small* **2015**, *11*, 622–629.
- (215) Finzi, A.; Pacheco, B.; Zeng, X.; Kwon, Y. D.; Kwong, P. D.; Sodroski, J. *Journal of Virological Methods* **2010**, *168*, 155–161.
- (216) Cao, L.; Diedrich, J. K.; Kulp, D. W.; Pauthner, M.; He, L.; Park, S. K. R.; Sok, D.; Su, C. Y.; Delahunty, C. M.; Menis, S.; Andrabi, R.; Guenaga, J.; Georgeson, E.; Kubitz, M.; Adachi, Y.; Burton, D. R.; Schief, W. R.; Yates, J. R.; Paulson, J. C. *Nature Communications* **2017**, *8*, 1–13.

- 
- (217) Acharya, P.; Tolbert, W. D.; Gohain, N.; Wu, X.; Yu, L.; Liu, T.; Huang, W.; Huang, C.-C.; Kwon, Y. D.; Louder, R. K.; Luongo, T. S.; McLellan, J. S.; Pancera, M.; Yang, Y.; Zhang, B.; Flinko, R.; Foulke, J. S.; Sajadi, M. M.; Kamin-Lewis, R.; Robinson, J. E.; Martin, L.; Kwong, P. D.; Guan, Y.; DeVico, A. L.; Lewis, G. K.; Pazgier, M. *Journal of virology* **2014**, *88*, 12895–906.
- (218) Owen, G. R.; Le, D.; Stoychev, S.; Cerutti, N. M.; Papathanasopoulos, M. *Biochemical and Biophysical Research Communications* **2018**, *497*, 811–817.

## Appendix A

# Plasmids and cDNA sequences

### A.1 2dCD4 clones

A sequence alignment of the cDNA for 2dCD4-WT expressed in *E. coli*, *B. choshinensis* and CFPE and the cDNA for the 2dCD4-D2A variant expressed in CFPE is shown here. The respective vector maps for each expression system are also shown.

E. coli-2dCD4-WT	-----MKKVVLGKKGDTVELTCTASQKRSIQFHWKNSNQIK	36
CFPE-C-2dCD4-WT	-----MAQMKKVVVLGKKGDTVELTCTASQKRSIQFHWKNSNQIK	39
CFPE-2dCD4-D2A	-----MAQMKKVVVLGKKGDTVELTCTASQKRSIQFHWKNSNQIK	39
B. choshinensis-2dCD4-WT	HHHHHPMS-DYDIPPTENLYFQGAKKVVVLGKKGDTVELTCTASQKRSIQFHWKNSNQIK	59
CFPE-N-2dCD4-WT	----KPYDGFHHHHHENLYFQ-SKKVVLGKKGDTVELTCTASQKRSIQFHWKNSNQIK	54
	*****	
E. coli-2dCD4-WT	ILGNQGSFLTKGPSKLNDRADSRRLWDQGNFPLIINKLKIEDSDTYItextbfCEVEDQ	96
CFPE-C-2dCD4-WT	ILGNQGSFLTKGPSKLNDRADSRRLWDQGNFPLIINKLKIEDSDTYI-----CEVEDQ	93
CFPE-2dCD4-D2A	ILGNQGSFLTKGPSKLNDRADSRRLWDQGNFPLIINKLKIEDSDTYItextbfCEVEDQ	99
B. choshinensis-2dCD4-WT	ILGNQGSFLTKGPSKLNDRADSRRLWDQGNFPLIINKLKIEDSDTYI-----CEVEDQ	113
CFPE-N-2dCD4-WT	ILGNQGSFLTKGPSKLNDRADSRRLWDQGNFPLIINKLKIEDSDTYI-----CEVEDQ	108
	*****	
E. coli-2dCD4-WT	KEEVQLLVFGLTANS DTHLLQGQSLTTLTLESPPGSSPSVQCRRSPRGKNIQGGKTLVSQSL	156
CFPE-C-2dCD4-WT	KEEVQLLVFGLTANS DTHLLQGQSLTTLTLESPPGSSPSVQCRRSPRGKNIQGGKTLVSQSL	153
CFPE-2dCD4-D2A	KEEVQLLVFGLTANS DTHLLQGQSLTTLTLESPPGSSPSVQARS PRGKNIQGGKTLVSQSL	159
B. choshinensis-2dCD4-WT	KEEVQLLVFGLTANS DTHLLQGQSLTTLTLESPPGSSPSVQCRRSPRGKNIQGGKTLVSQSL	173
CFPE-N-2dCD4-WT	KEEVQLLVFGLTANS DTHLLQGQSLTTLTLESPPGSSPSVQCRRSPRGKNIQGGKTLVSQSL	168
	*****	
E. coli-2dCD4-WT	ELQDSGTWTCTVLQNQKVEFKIDIVVLA FQKLEHHHHHH-----	196
CFPE-C-2dCD4-WT	ELQDSGTWTCTVLQNQKVEFKIDIVVLAENLYFQSLERAPGGGSHHHHHH	204
CFPE-2dCD4-D2A	ELQDSGTWTATVLQNQKVEFKIDIVVLAENLYFQSLERAPGGGSHHHHHH	210
B. choshinensis-2dCD4-WT	ELQDSGTWTCTVLQNQKVEFKIDIVVLA FQKLE-----	207
CFPE-N-2dCD4-WT	ELQDSGTWTCTVLQNQKVEFKIDIVVLA M-----	198
	*****	

FIGURE A.1: Sequence alignment of the cDNA for the 2dCD4-WT protein which was expressed in: *E. coli*, *B. choshinensis* and by CFPE with an N-terminal and C-terminal His-tag and the cDNA for 2dCD4-D2A variant protein. Sequence alignment was done using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

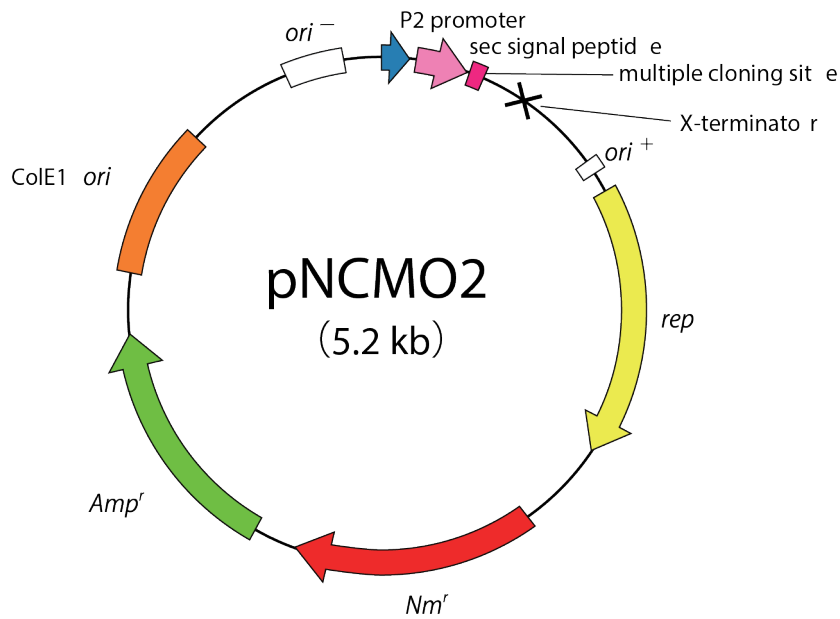


FIGURE A.2: Vector map for the pNCMO2 plasmid used for expression of N-terminal His-tagged 2dCD4-WT in *B. choshinensis*. The cDNA was inserted between the Nco1 and Xho1 restriction sites.

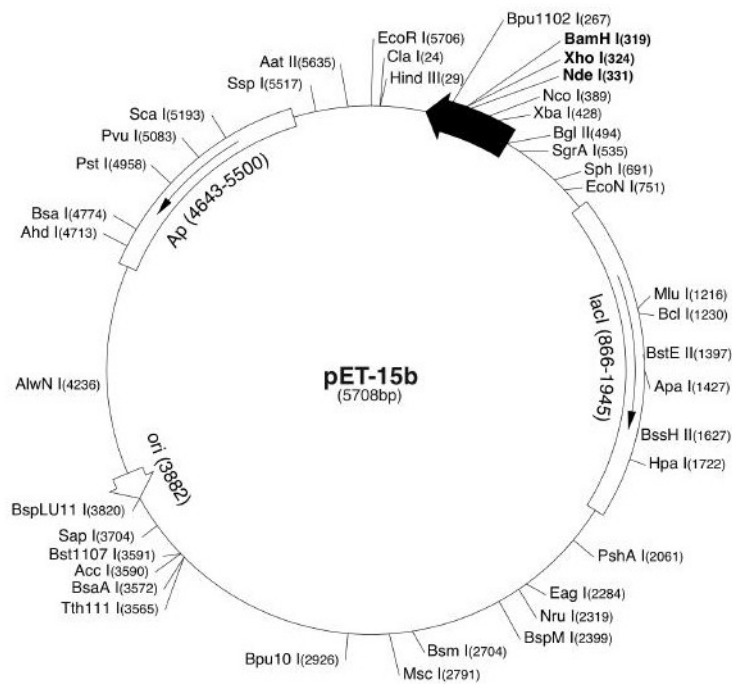


FIGURE A.3: Vector map for the pET15b plasmid used for expression of C-terminal His-tagged protiated 2dCD4-WT in *E. coli*. The cDNA was inserted between the Xho1 and Nde1 restriction sites.

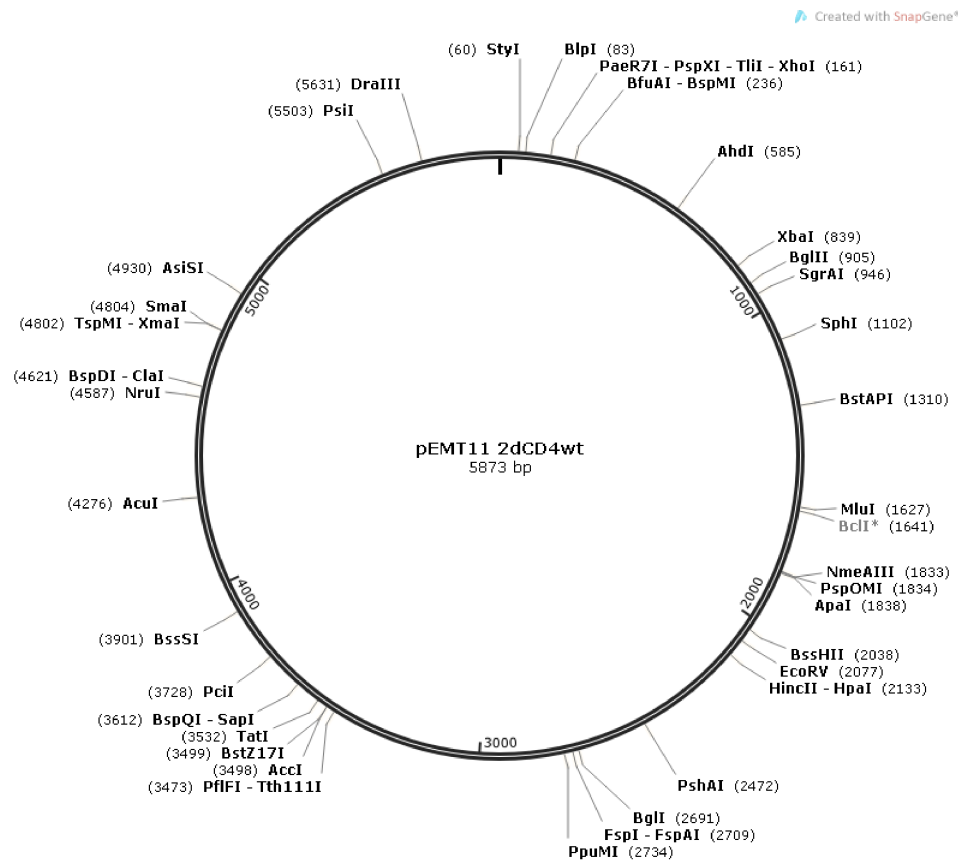


FIGURE A.4: Vector map for the pEM11 plasmid used for expression of C-terminal His-tagged deuterated 2dCD4-WT in *E. coli*. The cDNA was inserted between the Nco1 and Xho1 restriction sites.

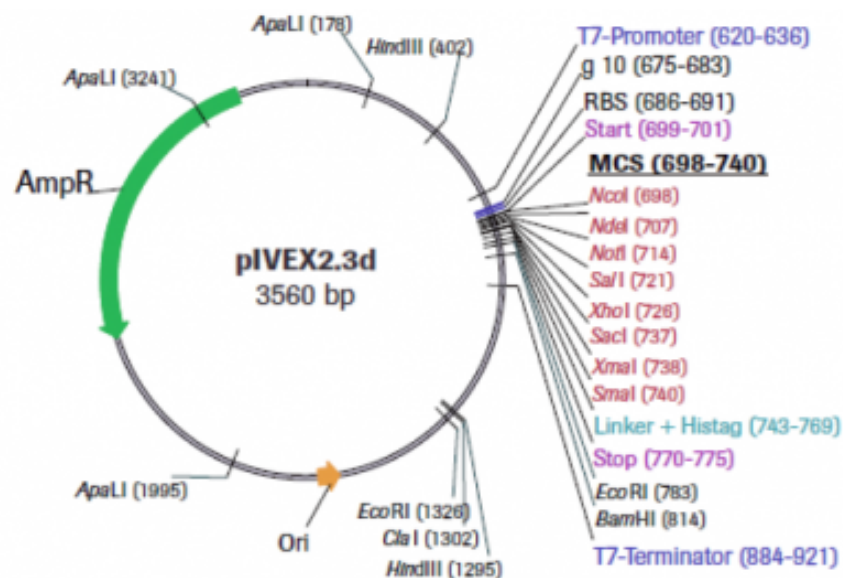


FIGURE A.5: Vector map for the piVEX2.3d plasmid used for expression of N-terminal and C-terminal His-tagged 2dCD4-WT and the 2dCD4-D2A variant protein. The cDNA was inserted between the Nde1 and Xho1 restriction sites.

## A.2 gp120 clones

A sequence alignment of the cDNA for gp120 BAL and gp120<sub>ZA</sub>CAP45 is shown here. The respective vector maps are also shown.

```

gp120ZACAP45      MPMGSLQPLATLYLLGMLVASVLAV-MGNLWVTVYYGVPVWKEAKATLFCASDARAYEKE 59
gp120BAL          --MGSLQPLATLYLLGMLVASCLGNAEKLWVTVYYGVPVWKEATTLFCASDRKAYDTE 58
                  ***** * . :***** :***** :*:.*
gp120ZACAP45      VHNVWATHACVPTDPNPQEIYLGNVTFENFMWKNMVDQMHEDIISLWDQSLKPCVKLTP 119
gp120BAL          VHNVWATHACVPTDPNPQVELKNVTFENFMWKNMVEQMHEDIISLWDQSLKPCVKLTP 118
                  *****: * *****:*.*****:*****
gp120ZACAP45      LCVTLRCTNATING-----SLTEEVKNCSEFNITTELRDQKQAYALFYRPD 165
gp120BAL          LCVTLNCTDLRNATNGNDTNTSSSRGMVGGGEMKNCSEFNITTNIRGKVKQKEYALFYKLD 178
                  *****.*: * *****:*. * * * * * * * * * *
gp120ZACAP45      VVPLNKNPSGNSSEYILINCNTSTITQACPQVDFPIPIHYCAPAGYAILKCNNKTFNG 225
gp120BAL          IAPIDNNS---NNRYRLISCNTSVITQACPQVDFPIPIHYCAPAGFAILKCKDKKFNNG 234
                  :.*::** .*. * * * * * * * * * * * * * * * * * * * * * * * * * * * *
gp120ZACAP45      TGPCNNVSTVQCTHGIKPVVSTQLLNGSLAEEDIIKSENLNNTIIVHLNKSVEIV 285
gp120BAL          KGPCTNVSTVQCTHGRPVVSTQLLNGSLAEDEVVIRSANFADNAKVIIVQLNESVEIN 294
                  .*.* *****:*****:*****:.*:* *::* * * * * * * * * * * * * *
gp120ZACAP45      CRRPNNNTRKSIIRIGPQAFYATNDIIGDIRQAHCNINNSTWNRILEQIKKKLREHFLNR 345
gp120BAL          CTRPNNNTRKSIHIGPRAFYTTEIIGDIRQAHCNLSRAKWNNTLNKIVIKLREQFGNK 354
                  * *****:*****:***:*.:*****:..*. * * * * * * * * * * * *
gp120ZACAP45      TIEFESPSGGDLEVTHSFNCGGEFFYCNTRLFKWSS-----NVTNDTITIPCRIK 397
gp120BAL          TIVFKHSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWNVTEESNNTVENNTITLPCRIL 414
                  ** * : * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
gp120ZACAP45      QFINMWQGVGRAMYAPPIEGNITCNSSITGLLLTRDGGKTDNDTEIFRPGGGNMKDNWR 457
gp120BAL          QIINMWQEVGRAMYAPPIRGQIRCSSNITGLLLTRDGGPED-NKTEVFRPGGGMDRDNWR 473
                  *:* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
gp120ZACAP45      NELYKQVVEIKPLGVAPTEARRRVVEREKR----- 488
gp120BAL          SELYKQVVKIEPLGVAPTKAKRRVVREREKRHHHHHH 510
                  .*****:*.*****:*.*****:*****

```

FIGURE A.6: Sequence alignment of the cDNA for the gp120 BAL and gp120<sub>ZA</sub>CAP45 proteins. Sequence alignment was done using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).



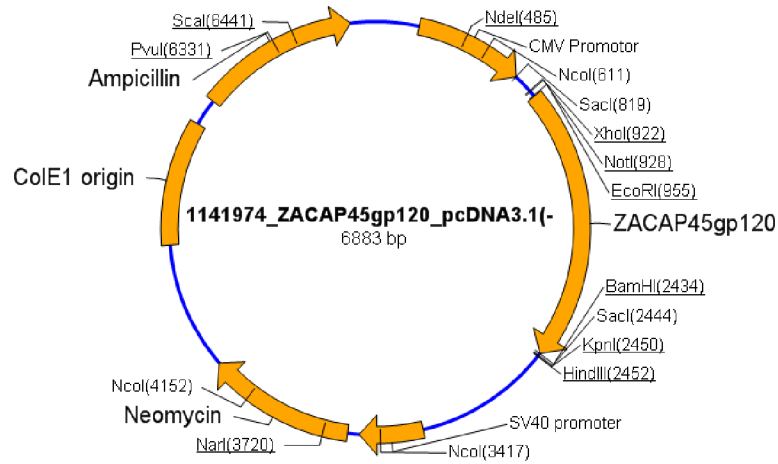


FIGURE A.7: Vector map for the pcDNA 3.1(-)\_A012 plasmid used for expression of untagged gp120<sub>ZACAP45</sub> in HEK293 cells. The cDNA was inserted between the EcoR1 and BamH1 restriction sites.

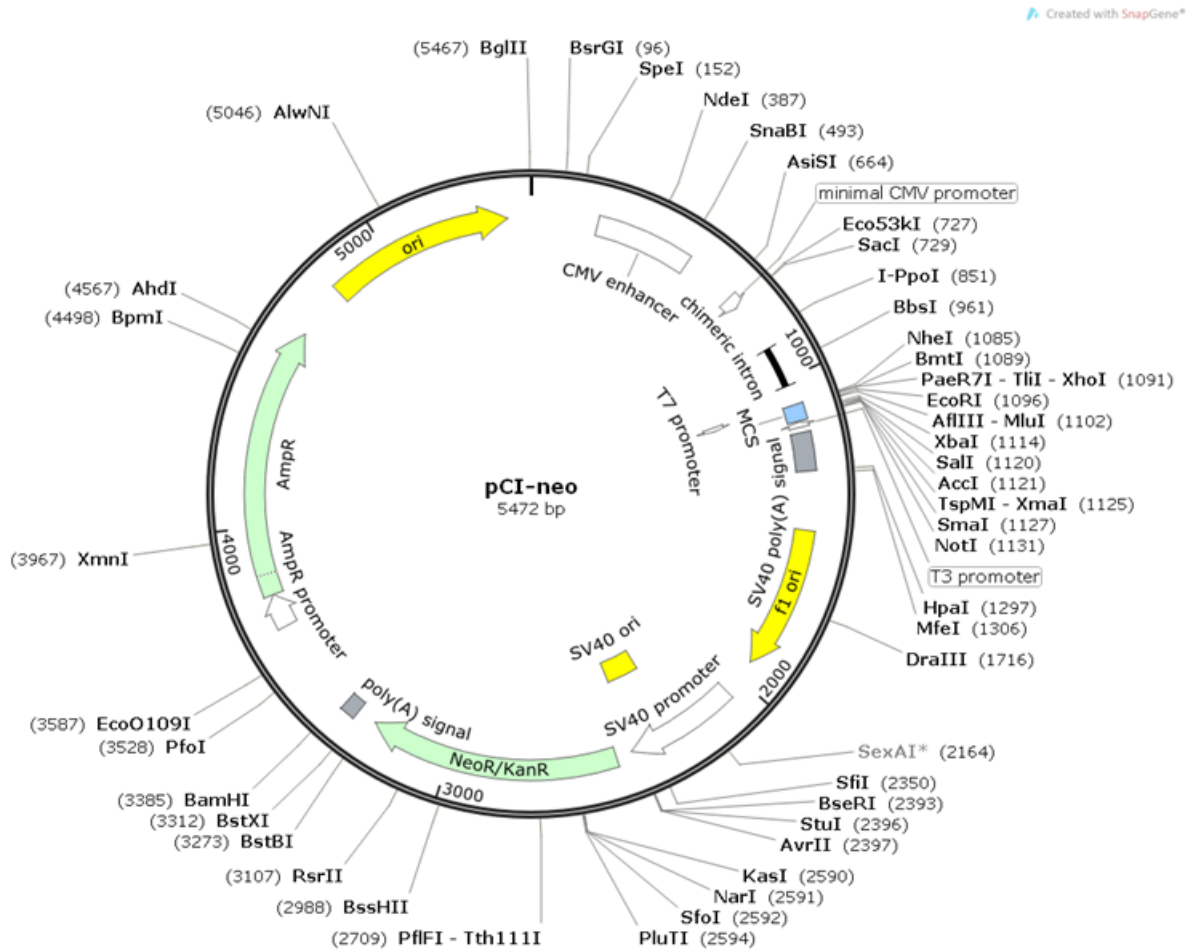


FIGURE A.8: Vector map for the pCIneo plasmid used for expression of C-terminal His-tagged gp120<sub>BAL</sub> in HEK293 cells. The cDNA was inserted between the Xho1 and EcoR1 restriction sites.



## Appendix B

# Buffer recipes

### B.1 Bacterial growth media

#### B.1.1 Liquid growth media

- 1 L LB: 25 g of pre-mixed LB powder (Sigma, 10 g bacto-typtone, 10 g NaCl, 5 g yeast extract) was dissolved in ddH<sub>2</sub>O to a final volume of 1 L.
- 1 L 2SY: 40 g bacto-soytone (BD Bacto™), 5 g yeast extract (BD Bacto™) were dissolved in ddH<sub>2</sub>O to a final volume of 900 mL and sterilised by autoclaving as described below. 20 g glucose and 150 mg CaCl<sub>2</sub>·2H<sub>2</sub>O were dissolved in ddH<sub>2</sub>O to a final volume of 100 mL and filter sterilised using a 0.22 µm vacuum filter (Corning) and added to the sterile soytone and yeast solution.

Liquid growth media were sterilised by autoclaving at 121°C, 15 psi for 30 minutes.

#### B.1.2 Enfors minimal medium

- 1 L H-Enfors: 6.86 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.56 g KH<sub>2</sub>PO<sub>4</sub>, 6.48 g Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, 0.49 g (NH<sub>4</sub>)–H–citrate and 5 g glycerol were dissolved in ddH<sub>2</sub>O to a final volume of 998 mL and sterilised by autoclaving as described below. 1 mL sterile 1 M MgSO<sub>4</sub> and 1 mL sterile metal salts solution were added to the sterile media under the biological fume hood.
- Metal salt solution: 0.23 g L<sup>-1</sup> MgSO<sub>4</sub>, 0.50 g L<sup>-1</sup> CaCl<sub>2</sub>·2H<sub>2</sub>O, 16.7 g L<sup>-1</sup> FeCl<sub>3</sub>·6H<sub>2</sub>O, 0.18 g L<sup>-1</sup> ZnSO<sub>4</sub>·7H<sub>2</sub>O, 0.16 g L<sup>-1</sup> CuSO<sub>4</sub>·5H<sub>2</sub>O, 0.15 g L<sup>-1</sup> MnSO<sub>4</sub>·4H<sub>2</sub>O, 0.18 g L<sup>-1</sup> CoCl<sub>2</sub>·6H<sub>2</sub>O and 20.1 g L<sup>-1</sup> Na–EDTA.

- 1 L 85% D-Enfors: the components as listed above were dissolved in 85% D<sub>2</sub>O. The medium was filter sterilised rather than by autoclaving.
- 4% NaOD: 10 mL NaOD (10 M stock) was diluted with 90 mL of D<sub>2</sub>O and filter sterilised.
- Feed for fermenter: 30 g glycerol was diluted in 85% D-Enfors and sterilised by filtration.

### B.1.3 Solid growth media

- 500 mL LB-agar: 12.5 g pre-mixed LB powder (Sigma, 5 g bacto-typtone, 5 g NaCl, 2.5 g yeast extract) and 7 g agar (Sigma) were added to a final volume of 1 L with ddH<sub>2</sub>O.
- 500 mL 2SY-agar: 20 g bacto-soytone (BD Bacto™), 2.5 g yeast extract (BD Bacto™) and 7.5 g agar were dissolved in ddH<sub>2</sub>O to a final volume of 450 mL and sterilised by autoclaving as described below. 10 g glucose and 75 mg CaCl<sub>2</sub>·2H<sub>2</sub>O were dissolved in dH<sub>2</sub>O to a final volume of 100 mL and filter sterilised using a 0.22 µm vacuum filter (Corning) and added to the sterile soytone and yeast solution.

Solid growth media were sterilised by autoclaving at 121°C, 15 psi for 30 minutes. Before the liquid cooled down it was poured under a biological fume hood into sterile plastic petri dishes (25 mL per petri dish) and 1 X antibiotic was added. The media was allowed to solidify before storing at 4-8°C.

## B.2 Antibiotic stock solutions

- Ampicillin: 100 mg mL<sup>-1</sup> in ddH<sub>2</sub>O and stored at -20°C
- Kanamycin: 30 mg mL<sup>-1</sup> in ddH<sub>2</sub>O and stored at -20°C
- G-418: 50 mg mL<sup>-1</sup> in 1 X PBS (sterile) and stored at 4°C
- Neomycin: 10 mg mL<sup>-1</sup> in ddH<sub>2</sub>O and stored at -20°C

## B.3 Purification buffers

### B.3.1 *E. coli* purification buffers

- Wash buffer 1: 1 X PBS, 20 mM imidazole, 8 M urea, 2 mM  $\beta$ -mercaptoethanol, 0.5 M NaCl and 50 mM glycine
- Wash buffer 2: 1 X PBS, 50 mM imidazole, 8 M urea, 2 mM  $\beta$ -mercaptoethanol, 0.5 M NaCl and 50 mM glycine
- Elution buffer: 1 X PBS, 500 mM imidazole, 8 M urea, 2 mM  $\beta$ -mercaptoethanol, 0.5 M NaCl and 50 mM glycine
- Refolding buffer A: 1 X PBS, 50 mM glycine, 10% w/v sucrose, 1 mM EDTA, 1 mM GSH, 0.1 mM GSSG, 4 M urea, 15.6 mM NaOH (pH 9.6)
- Refolding buffer B: 1 X PBS, 15 mM Na<sub>2</sub>CO<sub>3</sub>, 35 mM NaHCO<sub>3</sub>, 10% w/v sucrose, 1 mM EDTA, 0.1 mM GSH, 0.01 mM GSSG (pH 9.6)
- SEC buffer: 1 X PBS, 10% w/v sucrose

1000 X PBS contains 80 g NaCl, 2 KCl, 26.8 g Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O and 2.4 g KH<sub>2</sub>PO<sub>4</sub>. The pH is adjusted to pH 7.4 using HCl and the volume is made up to 1 L with ddH<sub>2</sub>O.

### B.3.2 *B. choshinensis* purification buffers

- Loading buffer: 100 mM Tris-HCl pH 8.0, 300 mM NaCl and 5 mM imidazole
- Wash buffer 1: 100 mM Tris-HCl pH 8.0, 1 M NaCl and 20 mM imidazole
- Wash buffer 2: 100 mM Tris-HCl pH 8.0, 300 mM NaCl and 40 mM imidazole
- Elution buffer: 100 mM Tris-HCl pH 8.0, 300 mM NaCl and 300 mM imidazole
- SEC buffer: 100 mM Tris-HCl pH 9.0, 300 mM NaCl, 5% w/v sucrose

### B.3.3 CFPE purification buffers

- Loading buffer: 20mM Tris-HCl pH 7.5, 300 mM NaCl and 5 mM imidazole

- Wash buffer 1: 20 mM Tris-HCl pH 7.5, 1 M NaCl and 20 mM imidazole
- Wash buffer 2: 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 40 mM imidazole
- Elution buffer: 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 300 mM imidazole
- SEC buffer: 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% w/v sucrose

*B. choshinensis* and CFPE IMAC purifications were done using an ÅKTA prime plus system (GE Healthcare). SEC purifications were done using either: ÅKTA purifier (GE Healthcare) or BioLogic DuoFlow™ or NGC™ (both BioRad) fast protein liquid chromatography systems.

## B.4 SDS-PAGE

12% Tris-Tricine gels were cast using the components outlined in table B.1. 10 or 15 well combs were added to the top of the gel and they were left to polymerise. The gels were stored moist at 4-8° before use. SDS-PAGE was run using a Mini-Protean® tetra vertical electrophoresis cell connected to a PowerPac™ universal power supply (both BioRad). Cathode buffer was poured into the middle of the chamber and anode buffer around the outside of the chamber (see below). Samples were mixed with 5 X reducing loading buffer and boiled for 5 minutes before loading onto the gel. The gel was run at 150 V until the gel front reached the bottom of the gel.

Components	12% Resolving gel (mL)	4% Stacking gel (mL)
30/0.8% Bis/ Acrylamide	12	1
3 M Tris-HCl pH 8.45	10	3
10% SDS	0.3	0.1
100% glycerol	3	-
10% ammonium persulphate	0.3	0.1
TEMED	0.03	0.01
Add ddH <sub>2</sub> O to a final volume	30 mL	12 mL

TABLE B.1: Ingredients for the 12% resolving gel and 4% stacking gel to make 12% Tris-Tricine SDS-PAGE.

- Cathode buffer: 121.1 g Tris base, 179.2 g tricine and 10 g SDS were dissolved in ddH<sub>2</sub>O to a total volume of 1 L.
- 10 X anode buffer: 242.g Tris base was dissolved in 500 mL ddH<sub>2</sub>O. The pH was adjusted to 8.9 with concentrated HCl and diluted to a final volume of 1 L with ddH<sub>2</sub>O.

After an SDS-PAGE run was complete the gel could be either: (i) stained with Coomassie blue or (ii) transferred onto a nitrocellulose membrane for western blotting.

#### **B.4.1 Coomassie blue staining**

For Coomassie blue staining the gel was incubated with agitation for 30 minutes in Coomassie blue stain. The stain was removed and the gel was destained by incubation with destain and tissue to soak up the excess Coomassie blue until the protein bands became sufficiently clear against the background stain.

- 1 X Coomassie blue stain: 2.5 g Coomassie blue G-250 was dissolved in 520 mL 96% ethanol and 100 mL 100% acetic acid and diluted to a total volume of 1 L with ddH<sub>2</sub>O
- 1 X destain: 313 mL 96% ethanol and 200 mL 100% acetic acid were diluted to a total volume of 4 L with ddH<sub>2</sub>O.

#### **B.4.2 Western blotting**

For western blotting the proteins were transferred from the SDS-PAGE to a Trans-Blot®mini nitrocellulose membrane using a Trans-blot®Turbo™ transfer system (both BioRad). After transfer, the nitrocellulose membrane was blocked for 1 hour with 5% w/v blocking solution. The blocking solution was removed and the membrane was incubated with the primary antibody solution for 1 hour. The primary antibody solution was removed and the membrane was washed three times for 5 minutes each time in PBS-T. The membrane was incubated with the secondary antibody solution for an hour then washed as described previously. The membrane was then developed with 5 mL Western Blue®stabilised substrate for ALP (Promega) for 10 minutes. The membrane was washed with ddH<sub>2</sub>O before a picture was taken.

- Blocking solution: 5% w/v milk in PBS-T (1% v/v PSB with 0.05% w/v Tween)
- Primary antibody solution: 1:3 000 primary antibody (mouse anti-His, Sigma) in blocking solution
- Secondary antibody solution 1:7 500 secondary antibody (goat anti-mouse ALP conjugated, Sigma) in blocking solution





## Appendix C

# Mass spectrometry analysis of 2dCD4-WT analogues

### C.1 Mass-spectra of 2dCD4-WT analogues

The liquid chromatography, electro-spray ionisation, time-of-flight mass spectra (LC-ESI-TOF MS) and the deconvoluted spectra for the 2dCD4-WT analogues expressed using *E. coli*, *B.choshinensis* and cell-free protein expression (CFPE) are presented here. This appendix corresponds to the LC-ESI-TOF mass spectrometry results summarised in chapter 4 section 4.4.4.

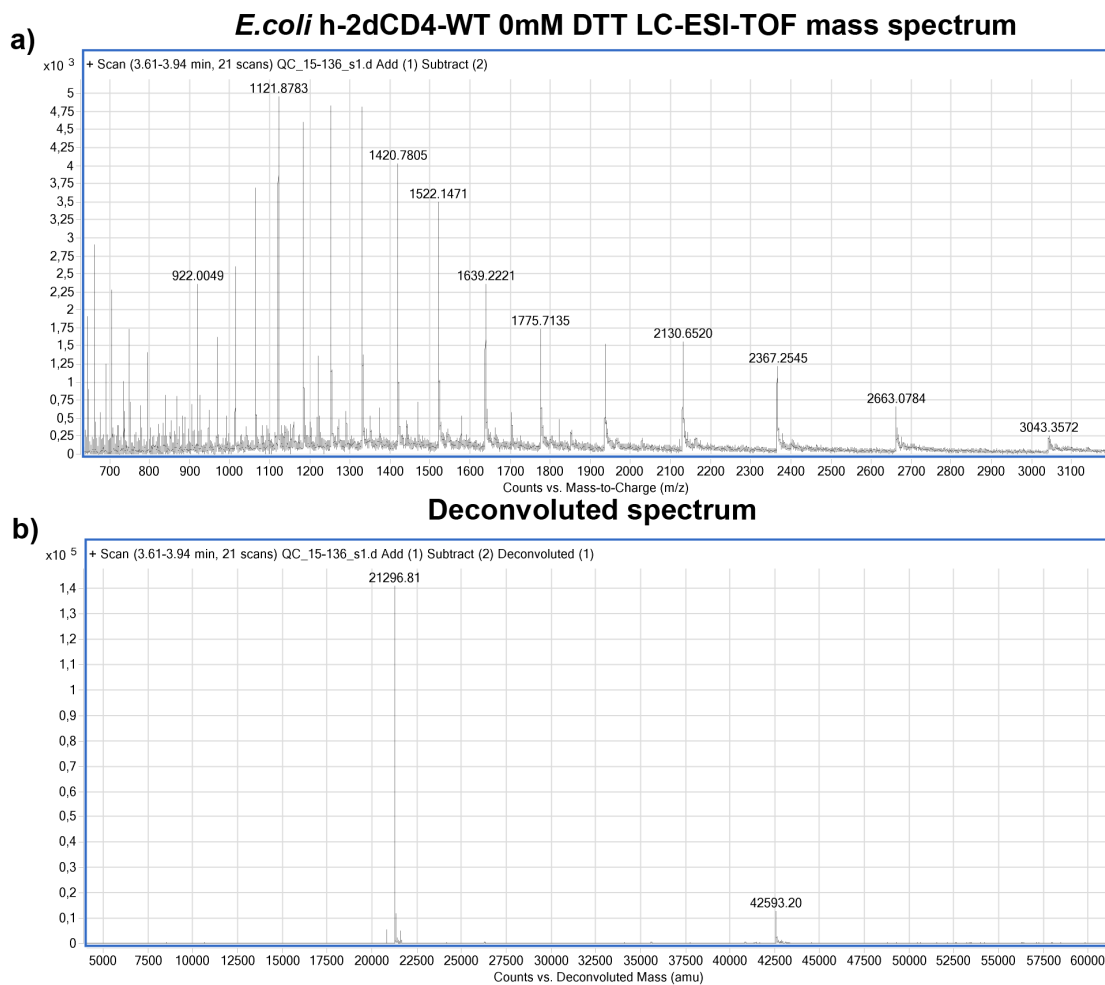


FIGURE C.1: (a) Liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum of *E. coli* expressed h-2dCD4-WT in 0 mM DTT. (b) Deconvoluted liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum of *E. coli* expressed h-2dCD4-WT in 0mM DTT. A primary species at 21 297 Da corresponding to an oxidised monomer and a secondary species corresponding to an oxidised dimer at 42 593 Da were observed.

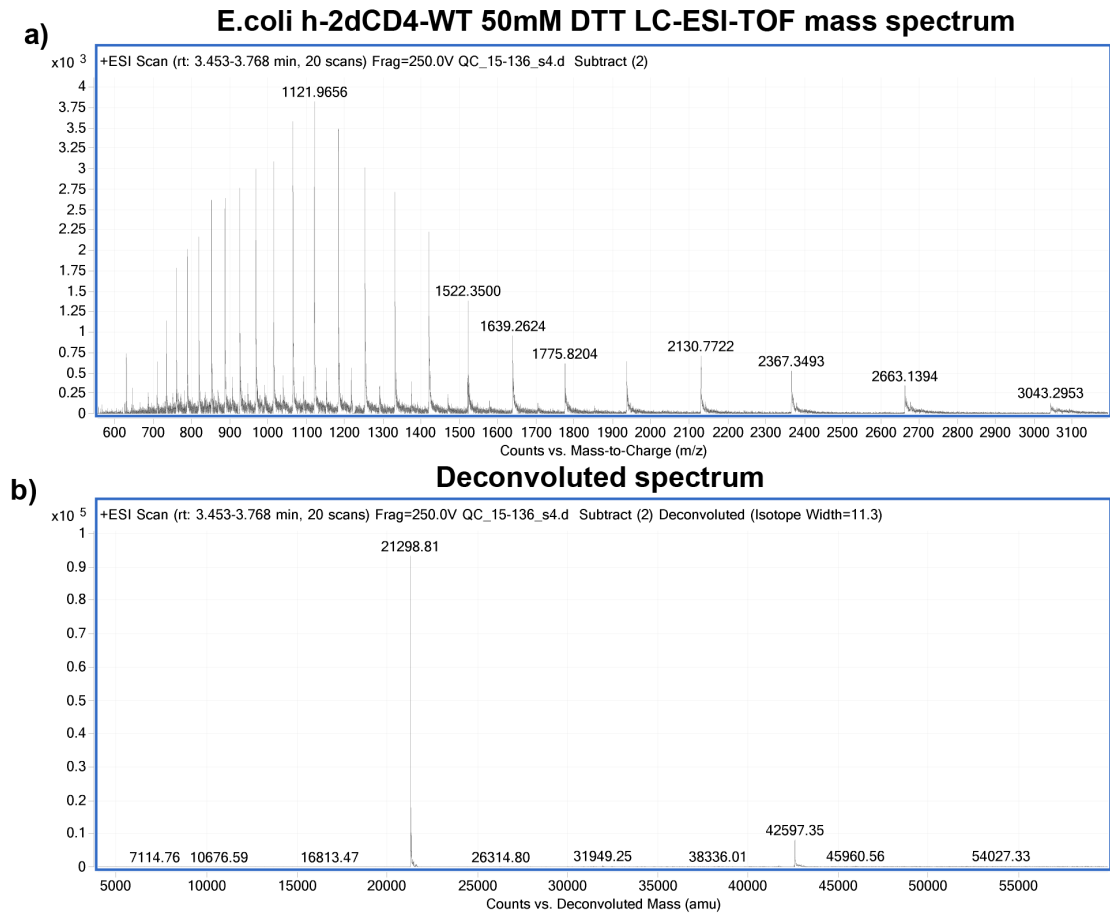


FIGURE C.2: (a) Liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum (LC-ESI-TOF MS) of *E. coli* expressed h-2dCD4-WT in 50 mM DTT. (b) Deconvoluted LC-ESI-TOF MS. A primary species at 21 299 Da corresponding to a reduced monomer and a secondary species corresponding to a reduced dimer at 42 597 Da were observed.

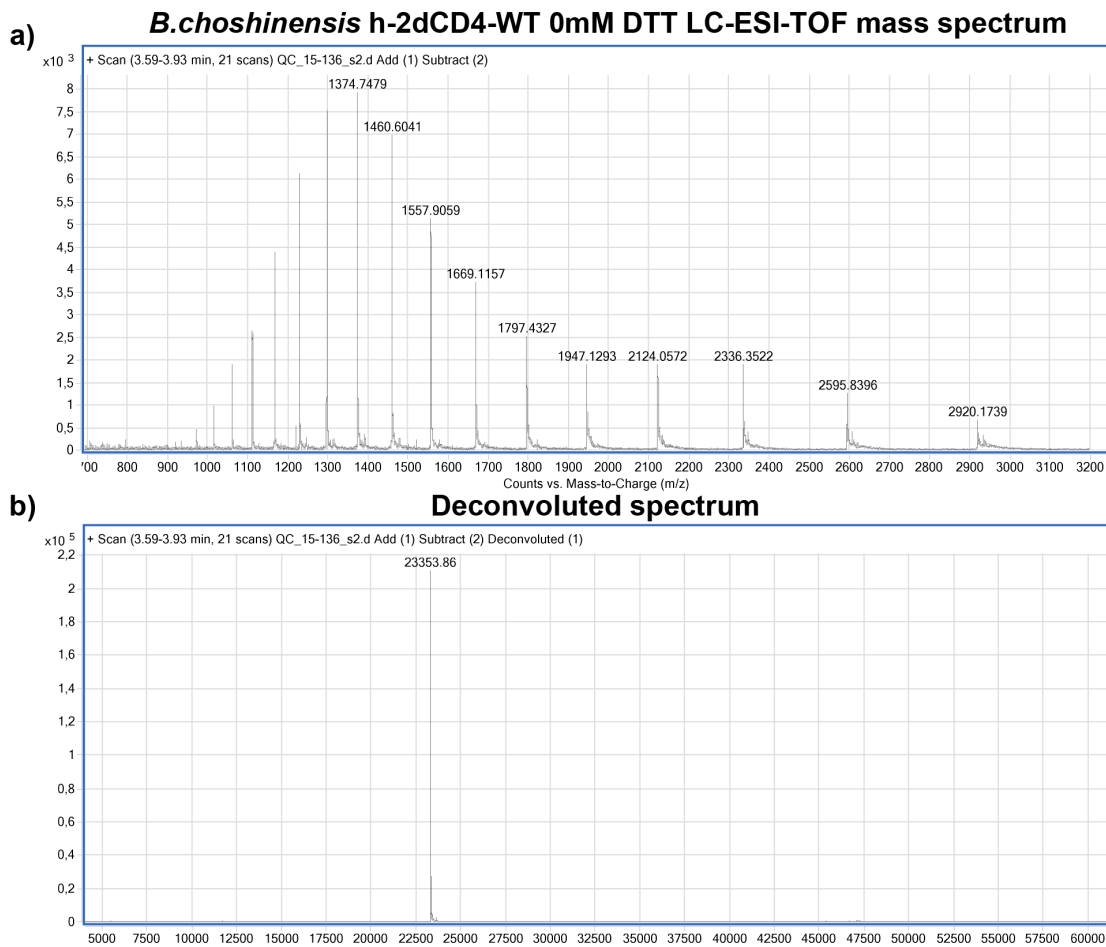


FIGURE C.3: (a) Liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum (LC-ESI-TOF MS) of *B. choshinensis* expressed h-2dCD4-WT in 0 mM DTT. (b) Deconvoluted LC-ESI-TOF MS. A single species at 23 354 corresponding to an oxidised monomer was observed.

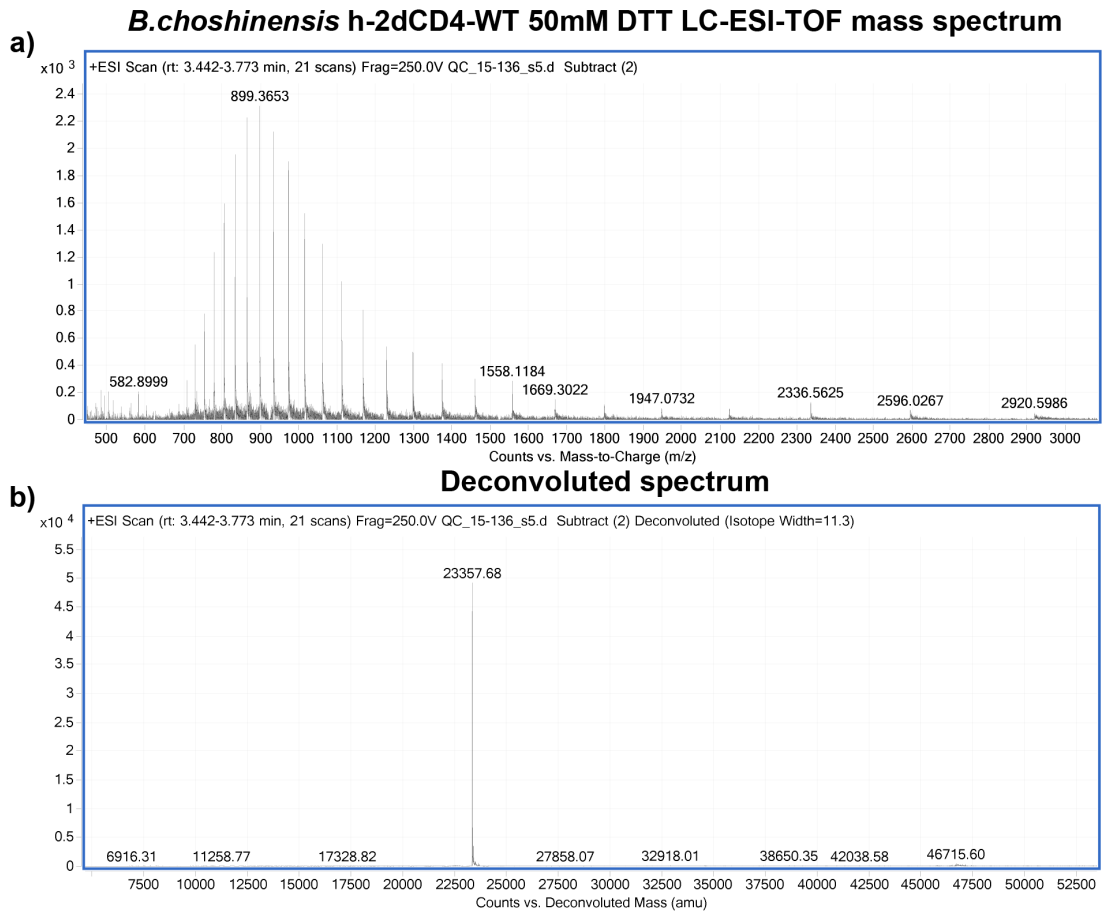


FIGURE C.4: (a) Liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum (LC-ESI-TOF MS) of *B. choshinensis* expressed h-2dCD4-WT in 50 mM DTT. (b) Deconvoluted LC-ESI-TOF MS. A single species at 23 358 Da corresponding to a reduced monomer was observed.

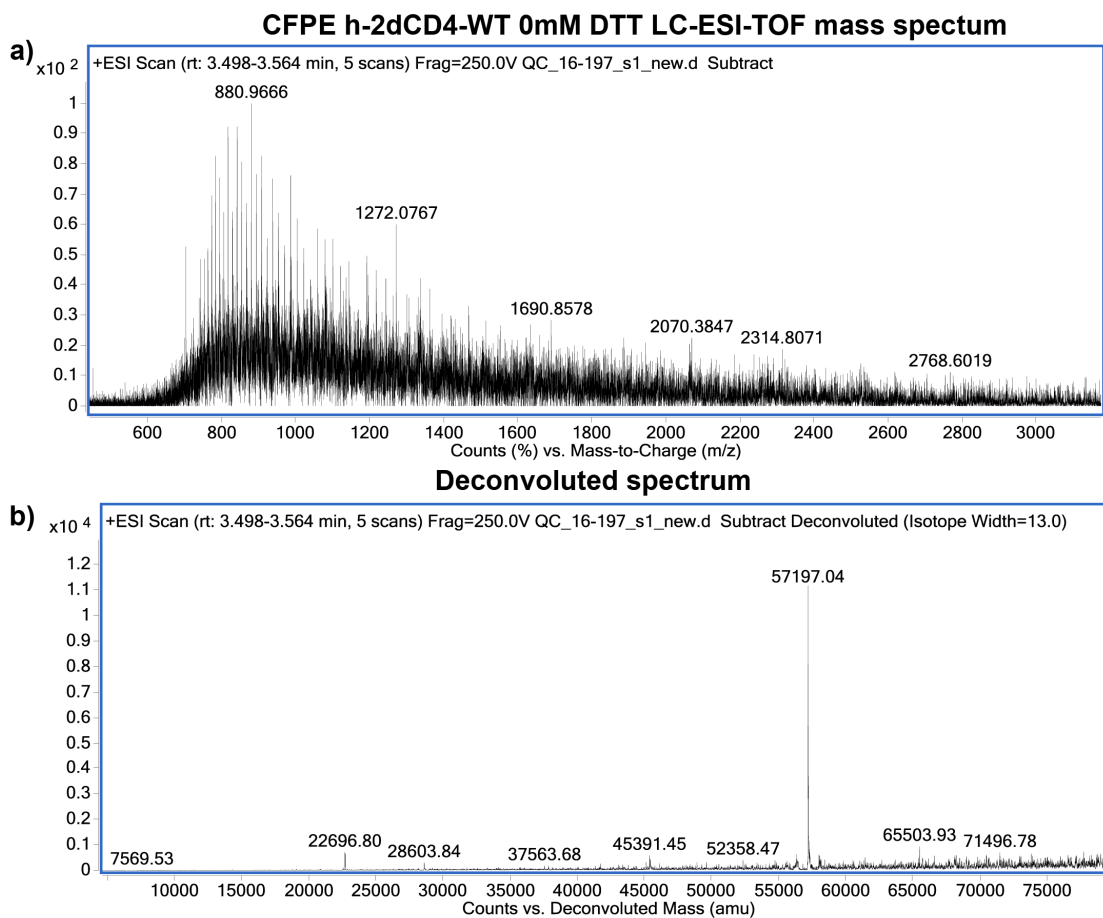


FIGURE C.5: (a) Liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum (LC-ESI-TOF MS) of h-2dCD4-WT produced using cell-free protein expression in 0 mM DTT. The spectrum is very noisy. (b) Deconvoluted LC-ESI-TOF MS. A primary species at 57 197 Da which does not correspond to an h-2dCD4-WT dimer or trimer and a secondary species at 22 697 Da corresponding to an oxidised h-2dCD4-WT monomer were observed.

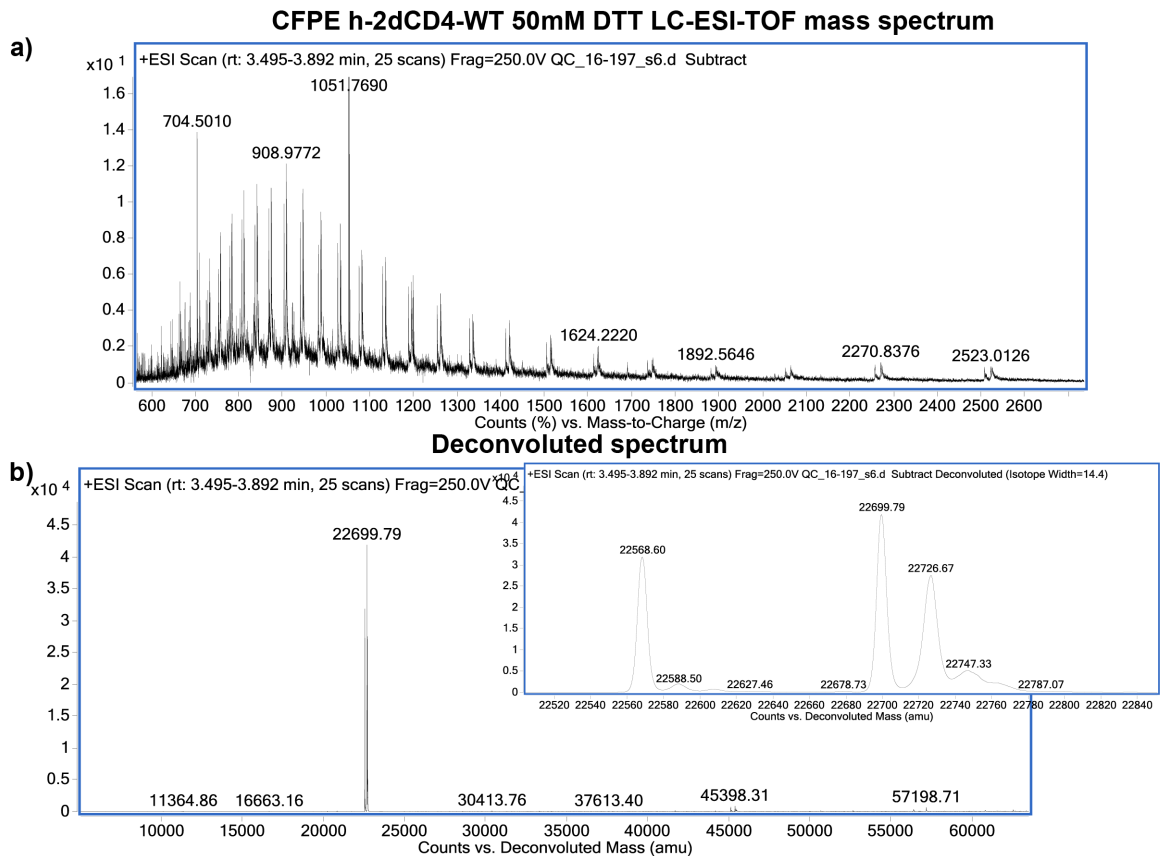


FIGURE C.6: (a) Liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum (LC-ESI-TOF MS) of h-2dCD4-WT produced using cell-free protein expression in 50 mM DTT. A repeating pattern of multiple peaks can be observed. (b) Deconvoluted LC-ESI-TOF MS and a zoomed-in view of the main peak (insert). A primary species at 22 700 Da corresponding to a reduced h-2dCD4-WT monomer is observed. A slightly lower and slightly higher molecular mass species are observed at 22 567 and 22 727 Da, respectively.

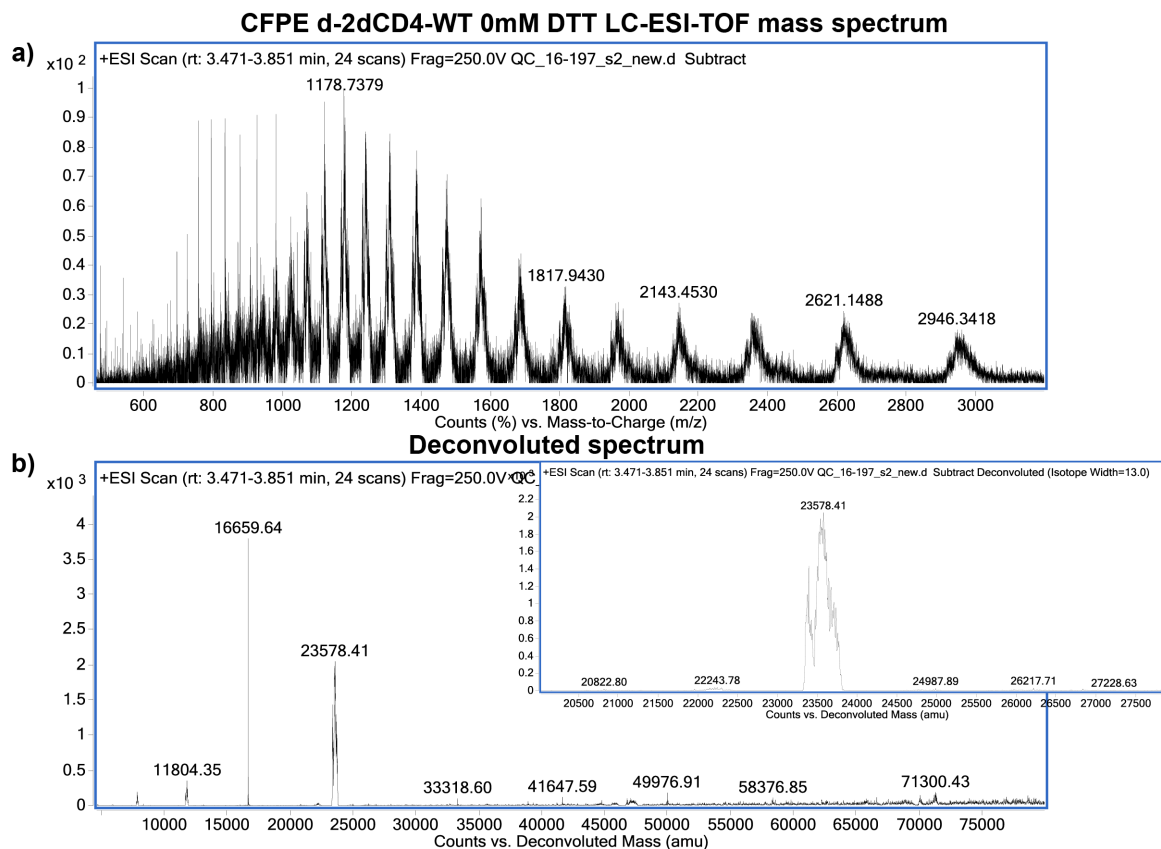


FIGURE C.7: (a) Liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum (LC-ESI-TOF MS) of d-2dCD4-WT produced using cell-free protein expression in 0 mM DTT. A repeating pattern of multiple, broad peaks can be observed. (b) Deconvoluted LC-ESI-TOF MS with a zoomed-in view (insert). A primary species at 23 578 Da which corresponds to an oxidised, partially deuterated 2dCD4-WT monomer and a secondary species at 16 660 Da corresponding to a CFPE truncation product were observed. The main 23 578 Da peak appears broad with many sub-peaks.



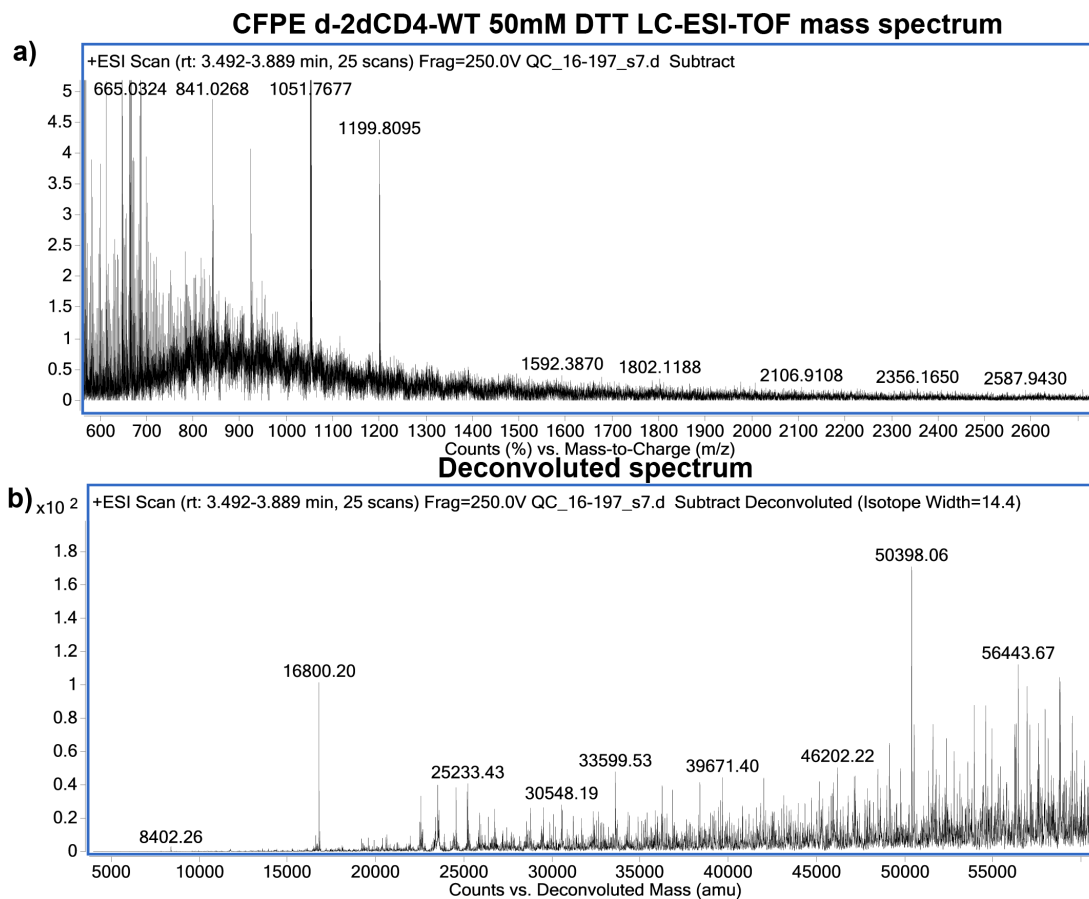


FIGURE C.8: (a) Liquid chromatography, electro-spray ionisation, time-of-flight mass spectrum (LC-ESI-TOF MS) of d-2dCD4-WT produced using cell-free protein expression in 50 mM DTT. The spectrum is noisy with irregular peaks. (b) Deconvoluted LC-ESI-TOF MS. The data analysis was non-trivial as many peaks of differing molecular weight were observed in the deconvoluted spectrum.

## C.2 Percentage deuteration determination of 2dCD4-WT

The percentage deuteration achieved using deuterated amino acids in the cell-free protein expression reaction was calculated from the difference between the hydrogenated and deuterated molecular mass determined from mass-spectrometry.

## For Hydrogenated Protein

AA	Number of AA	Hydrogenated AA MW	Hydro. AA MW x Nb.
Ala	6	71.0788	426.4728
Arg	6	156.1876	937.1256
Asn	10	114.1039	1141.039
Asp	10	115.0886	1150.886
Cys	4	103.1448	412.5792
Glu	11	129.1155	1420.2705
Gln	17	128.1308	2178.2236
Gly	16	57.052	912.832
His	8	137.1412	1097.1296
Ile	10	113.1595	1131.595
Leu	23	113.1595	2602.6685
Lys	18	128.1742	2307.1356
Met	2	131.1986	262.3972
Phe	6	147.1766	883.0596
Pro	7	97.1167	679.8169
Ser	20	87.0782	1741.564
Thr	13	101.1051	1314.3663
Trp	3	186.2133	558.6399
Tyr	2	163.176	326.352
Val	12	99	1188
end effect	1	18	18
Total	204		

H mass 22690.1533 Da

(A)

## For Deuterated Protein (with only non exchangeable D)

AA	non exchangeable D*	Deuterated MW AA (non-exchangeable D only)	Not fully deut AA MW x Nb.
Ala	4	75.0788	450.4728
Arg	7	163.1876	979.1256
Asn	3	117.1039	1171.039
Asp	3	118.0886	1180.886
Cys	3	106.1448	424.5792
Glu	5	134.1155	1475.2705
Gln	5	133.1308	2263.2236
Gly	2	59.052	944.832
His	5	142.1412	1137.1296
Ile	10	123.1595	1231.595
Leu	10	123.1595	2832.6685
Lys	9	137.1742	2469.1356
Met	8	139.1986	278.3972
Phe	8	155.1766	931.0596
Pro	7	104.1167	728.8169
Ser	3	90.0782	1801.564
Thr	5	106.1051	1379.3663
Trp	8	194.2133	582.6399
Tyr	7	170.176	340.352
Val	8	107	1284
end effect	0	18	18
Total	120		

D Mass 23904.1533 Da  
(including non-exchangeable D only)

\*pH effects not accounted for

For Info:  
Nb. Non Exch. D  
Nb. Exch. H/D  
Total H/D

1214  
399  
1613

(B)

## For Fully Deuterated Protein

AA	exch. H/D*	Perdeuterated AA MW	Hxl (perdeuterated AA xnb)
Ala	1	76.0788	456.4728
Arg	5	168.1876	1009.1256
Asn	3	120.1039	1201.039
Asp	2	120.0886	1200.886
Cys	2	108.1448	432.5792
Glu	2	136.1155	1497.2705
Gln	3	136.1308	2314.2236
Gly	1	60.052	960.832
His	2	144.1412	1153.1296
Ile	1	124.1595	1241.595
Leu	1	124.1595	2855.6685
Lys	3	140.1742	2523.1356
Met	1	140.1986	280.3972
Phe	1	156.1766	937.0596
Pro	2	106.1167	742.8169
Ser	2	92.0782	1841.564
Thr	2	108.1051	1405.3663
Trp	2	196.2133	588.6399
Tyr	2	172.176	344.352
Val	1	108	1296
end effect	3	20	20
Total	42		

Fully Deuterated Mass

24302.1533 Da

Experimental Mass (Mass spec. realised in H<sub>2</sub>O)

23578

% Deuteration

73.13399506

(C)

FIGURE C.9: The percentage deuterium labelling using the cell-free protein expression method was calculated using the following spread sheets. (a) The hydrogenated mass (b) non-labile H deuterated mass and (c) full deuterated mass were calculated to be 22 690, 23 904 and 24 032 Da, respectively. (b) The number of non-labile (1 214) and labile H (399) were calculated and (c) from the experimental mass determined by mass-spectrometry in H<sub>2</sub>O (23 578 Da), the percentage deuteration was calculated to be 73%.

## Appendix D

# List of activities

### Publications

- Owen, G.R., Channell, J.A., Forsyth, V.T., Haertlein, M., Mitchell, E.P., Capovilla, A., Papatheanasopoulos, M. and Cerutti, N.M. (2016) 'Human CD4 metastability is a function of the allosteric disulfide bond in domain 2', *Biochemistry* 55, pp. 2227-2237.

### Conferences with oral contribution

- **24<sup>th</sup> Congress and General Assembly of the International Union of Crystallography, August 2017 - Hyderabad, India:** Structural studies of dynamic CD4 changes relevant to HIV-1 infection.
- **FASEB: Functional Disulfide Bonds in Health and Disease, July 2016 - Steamboat Springs, USA:** Studying the allosteric disulphide bond in domain 2 of human CD4 relevant to HIV-1 infection.
- **European Conference on Neutron Scattering, September 2015 - Zaragosa, Spain:** SANS studies of the host-pathogen interaction underlying HIV infection: towards rational anti-viral therapy.

### Awards

- **UK Neutron and Muon Users' Meeting, June 2016 - Warwick, UK:** Best poster prize.

**Courses and other conferences**

- **HERCULES European School on neutrons and synchrotron radiation of science, 2016 - Grenoble, France:** poster contribution
- **Crystallography tutorial, 2016 - Grenoble France**
- **Introduction to advanced and emerging biophysical methods for integrative biology - 2, 2016 - Grenoble, France**
- **All you need is neutrons 2, 2016 - Grenoble, France:** oral contribution
- **Introduction to advanced and emerging biophysical methods for integrative biology - 1, 2015 - Grenoble, France**
- **All you need is neutrons 1, 2015 - Grenoble, France:** oral contribution
- **HIV R4P conference, 2014 - Cape Town, South Africa**

**Student days and user meetings**

- **NMUM, 2016 & 2017 - Warwick, UK:** poster contribution
- **PSB student day, 2015-2017 - Grenoble, France:** poster and oral clip contributions
- **ILL student day, 2015-2017 - Grenoble, France:** poster and oral clip contributions
- **ESRF users' meeting 2015 & 2016 - Grenoble, France:** poster and oral clip contributions.

## Human CD4 Metastability Is a Function of the Allosteric Disulfide Bond in Domain 2

Gavin R. Owen,<sup>\*,†</sup> Jennifer A. Channell,<sup>‡,§,||</sup> V. Trevor Forsyth,<sup>‡,§</sup> Michael Haertlein,<sup>§</sup> Edward P. Mitchell,<sup>‡,||</sup> Alexio Capovilla,<sup>†</sup> Maria Papatathanasopoulos,<sup>†</sup> and Nichole M. Cerutti<sup>†</sup>

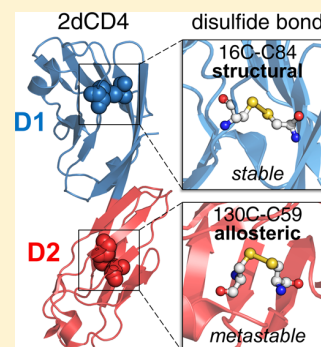
<sup>†</sup>HIV Pathogenesis Research Unit, Department of Molecular Medicine and Haematology, Faculty of Health Sciences, University of the Witwatersrand, 7 York Road, Parktown, 2193, Johannesburg, South Africa

<sup>‡</sup>Faculty of Natural Sciences, Keele University, Keele, Staffordshire ST5 5BG, United Kingdom

<sup>§</sup>Life Sciences Group, Institut Laue-Langevin, 71 Avenue des Martyrs, 38042, Grenoble, France

<sup>||</sup>European Synchrotron Radiation Facility, 71 Avenue des Martyrs, 38042, Grenoble, France

**ABSTRACT:** CD4 is expressed on the surface of specific leukocytes where it plays a key role in the activation of immunostimulatory T-cells and acts as a primary receptor for HIV-1 entry. CD4 has four ecto-domains (D1–D4) of which D1, D2, and D4 contain disulfide bonds. Although disulfide bonds commonly serve structural or catalytic functions, a rare class of disulfide bonds possessing unusually high dihedral strain energy and a relative ease of reduction can impact protein function by shuffling their redox state. D2 of CD4 possesses one such “allosteric” disulfide. While it is becoming accepted that redox exchange of the D2 allosteric disulfide plays an essential role in regulating CD4 activity, the biophysical consequences of its reduction remain incompletely understood. By analyzing the hydrodynamic volume, secondary structure, and thermal stability of the reduced and nonreduced forms of the single D1 and D2 domains, as well as the various redox isomers of two domain CD4, we have shown that ablation of the allosteric disulfide bond in domain 2 results in both a favorable structural collapse and an increase in the stability of CD4. Conversely, ablating the structural disulfide of D1 results in destabilizing structural rearrangements in CD4. These findings expand our understanding of the mechanisms by which oxidoreduction of the D2 allosteric disulfide regulates CD4 function; they reveal the intrinsic disulfide-dependent metastability of D2 and illustrate that redox shuffling of the allosteric disulfide results in previously undescribed conformational changes in CD4 that are likely important for its interaction with its protein partners.



The cluster of differentiation 4 receptor (CD4) is a 55 kDa type I integral membrane glycoprotein which plays a substantial role during adaptive immune responses. In T-cells, CD4 mediates maturation,<sup>1</sup> stabilizes T-cell receptor interactions with peptide-major histocompatibility complex class II (MHCII) complexes on antigen presenting cells, and mediates intracellular T-cell signaling.<sup>2</sup> In addition to T-cells, CD4 has also been shown to be expressed on monocytes, monocyte-derived macrophages, Langerhans cells, B lymphocytes, dendritic cells, eosinophils, megakaryocytes, and mast cells.<sup>3–8</sup> The role of CD4 during immune function can also be diverted by the human immunodeficiency virus type 1 (HIV-1) envelope glycoprotein gp120 as the viral receptor utilizes CD4 as the primary receptor as well as the coreceptor CCR5/CXCR4 during the entry process.<sup>9–11</sup> Functional binding sites for cognate ligands MHCII and gp120 have been mapped to the first two domains of CD4 (2dCD4), the subject of this study.

Each human CD4 monomer comprises a linear repeat of four concatenated N-terminal immunoglobulin-like ecto-domains (D1–D4, residues 1–371) linked to the cell membrane via the C-terminal transmembrane domain and a short cytoplasmic tail (residues 372–433).<sup>12,13</sup> The immunoglobulin domains

comprise two opposing  $\beta$ -sheets with 7–9  $\beta$ -strands each, and all but domain 3 have a single distinctive disulfide bond (Cys16–Cys84, Cys130–Cys159, and Cys303–Cys345 in D1, D2, and D4, respectively). All three disulfides, positioned in a hydrophobic core, have a closed configuration with low solvent accessibility.<sup>14–16</sup> The disulfide bonds in D1 and D4 function primarily to stabilize the protein by linking opposing sheets and have molecular configurations typical of conventional, structure-stabilizing disulfide bonds. In contrast, D2 is atypical, comprising ~80 amino acids rather than ~100 found in common Ig domains, and its disulfide bond joins neighboring strands within the same  $\beta$ -sheet rather than neighboring strands in opposite  $\beta$ -sheets. This rare spatial arrangement and unfavorable disulfide geometry, an “–RHStaple” disulfide configuration, increases the angular strain on the covalent linkage. The high dihedral strain energy in these disulfide bonds is caused by the torsional energy of the linkage in addition to the energy of deformation in the  $\beta$ -sheet that they occur. These

**Received:** February 18, 2016

**Revised:** March 16, 2016

**Published:** March 23, 2016

unstable disulfides can thus introduce metastable areas within the folded protein in which they exist, often decreasing the stability of the native protein as a whole.<sup>17</sup> These factors have been shown to influence the redox potential of the disulfide, increasing the ease of their reduction<sup>17,18</sup> and providing them with functional roles: they are able to regulate protein function in a nonenzymatic manner by triggering conformational changes when they switch between their bonded and nonbonded states.<sup>19</sup> The allosteric nature of the D2 disulfide<sup>19–21</sup> thus renders it susceptible to redox-mediated rearrangements at the cell surface<sup>22,23</sup> and enables the formation of functionally distinct monomeric, dimeric, and oligomeric forms of CD4.<sup>24–28</sup> A model for the presentation of CD4 on the cell surface has been proposed, wherein the CD4 receptor may be present as either an oxidized monomer, a reduced monomer, or a disulfide-linked dimer, depending on the redox state of the D2 allosteric disulfide.<sup>29</sup> The disulfide-linked dimeric CD4 is the preferred form for binding MHCII expressed on the surface of antigen-presenting cells such as macrophages,<sup>30</sup> and the reduced monomeric form is the preferred form for binding the envelope glycoprotein of HIV-1 during viral entry.<sup>29</sup>

By performing molecular dynamics simulations on several hundred disulfide bonded proteins, including specific bioinformatic analysis of the disulfides of CD4, Zhou et al. revealed that allosteric disulfide configurations such as the –RHStaple carry a significantly larger degree of internal tensile prestress than other disulfide bonds as a result of the direct stretching of the sulfur–sulfur bond.<sup>31</sup> As mechanical stretching of S–S bonds is recognized to accelerate their cleavage, the authors proposed that the internal mechanical prestress carried by allosteric disulfides destabilize them, thereby enhancing their reactivity and susceptibility to reduction. The substantial prestress discovered in the Cys130–Cys159 allosteric disulfide of CD4 has thus been suggested as a major factor in the elevated redox activity of the bond observed in the context of HIV entry and MHCII engagement.<sup>31</sup>

Despite these recent insights into functional implications of disulfide bond rearrangements in CD4, the biophysical consequences thereof are poorly understood. The crystal structures of CD4 as defined in the unliganded form<sup>12,13</sup> and in complex with gp120,<sup>32</sup> MHCII,<sup>2</sup> and MHCII–TCR are readily superimposable (RMS of 0.98 Å), suggesting that no major structural reordering of CD4 is required for binding gp120, and that CD4 acts as a rigid docking site for its cognate receptors. Although the Cys residues in the crystal structures of unliganded CD4 as well as CD4 in complex with these receptors have been assigned as oxidized (based on atomic proximity calculations), these assignments are unable to reflect the dynamic redox events that are now known to represent an important component of the biological function of CD4. Indeed, several authors have suggested that significant structural rearrangements occur in specific amino acid residues in the first two domains of CD4 post-gp120 binding by using CD4 specific mAbs,<sup>33</sup> molecular simulation studies of entropic changes in D1,<sup>34</sup> as well as normal-mode analysis of 2dCD4.<sup>35</sup> In addition, upon refinement of the initial 2dCD4 structure, Ryu et al. demonstrated flexibility (although limited) in the region of CD4 involved in gp120 binding as well as a detectable motion between D1 and D2.<sup>36</sup>

In this study, we describe the biophysical analysis of the redox isomers of 2dCD4 and single domain CD4 (1dCD4-D1/D2) using reducing and nonreducing SDS-PAGE, size-

exclusion chromatography, circular dichroism spectroscopy, and thermal unfolding. These studies are part of ongoing efforts to gain further insight into the consequences of oxidoreduction of each disulfide bond, and the molecular basis for receptor specificity of the distinct CD4 redox isomers. Our results suggest that while the highly ordered  $\beta$ -sheeted secondary structures typical of the Ig-family of proteins are able to form in 1dCD4-D1 when its resident structural disulfide bond is intact, the  $\beta$ -sheeted secondary structure observed for 1dCD4-D2 is less ordered and highly dependent on the redox state of its allosteric disulfide bond. We also illustrate the metastable nature of the allosteric disulfide bond of D2 and the effect that its reduction has on the structure and stability of the distinct CD4 isomers, demonstrating previously undescribed allosteric bond-dependent biophysical changes to CD4 that are likely to be important for its interaction with HIV-1 gp120 and MHCII.

## ■ MATERIALS AND METHODS

**Expression Plasmids.** DNA cassettes encoding the first two amino-terminal domains of wild-type (WT) human CD4 (2dCD4-WT, residues 1–183) and the Cys/Ala variants C16A/C84A and C130A/C159A were generated and subcloned into pET15b vectors as described previously.<sup>37,38</sup> In addition, two single domain CD4 cassettes were synthesized by GENEART (Regensburg, Germany) following codon optimization for expression in *Escherichia coli* and subsequently subcloned into pET15b vectors as described by Cerutti et al.<sup>37</sup> domain 1 (1dCD4-D1, residues 1–98 of CD4) containing L5I, A55V, I76P, L96I, and F98L substitutions, which have been documented to maintain secondary structural content, increase protein solubility, affinity for HIV-1 gp120 and neutralize pseudoviruses to a similar degree;<sup>39</sup> and domain 2 (1dCD4-D2, residues 99–178 of CD4, wild-type sequence).

**CD4 Expression, Purification, and Refolding.** The recombinant wild-type and Cys/Ala mutants of 2dCD4, as well as the 1dCD4-D1 and 1dCD4-D2 variants were expressed and purified by standard denaturing metal-chelate affinity chromatography and refolded using oxidative refolding protocols according to methods described previously.<sup>37</sup>

**Hydrodynamic Volume Analyses of CD4 Variants by Nonreducing/Reducing SDS-PAGE and Size-exclusion Chromatography.** The unfolded hydrodynamic volumes of single domain CD4 variants were analyzed by reducing and nonreducing SDS-PAGE using precast 4–12% gradient NuPAGE gels (Invitrogen). Prior to gel loading, the samples were treated with lithium dodecyl sulfate (LDS) loading buffer with or without DTT (50 mM) and incubated at room temperature for 10 min. Iodoacetamide was then added to a final concentration of 50 mM to alkylate and block any thiol groups present (i.e., reduced cysteines). Gels were stained with Coomassie Blue R-250. The native hydrodynamic volumes of the 2dCD4-WT redox isomers were analyzed by size-exclusion chromatography (SEC). A HiLoad 16/600 Superdex 75 preparative grade (GE Healthcare) size-exclusion column was first equilibrated with PBS and then 2 mL 2dCD4-WT at 1.4 mg/mL was loaded onto the column and resolved at a flow rate of 0.25 mL/min.

### Far-UV Circular Dichroism Analysis of CD4 Variants.

Far-UV circular dichroism (CD) spectra were recorded using a Jasco J-1500 spectropolarimeter and represent an average of five accumulations. Measurements were performed at 20 and 90 °C using 10  $\mu$ M of each CD4 variant in PBS, pH 7.4. CD measurements of the fully reduced forms of 2dCD4, 1dCD4-

## Biochemistry

## Article

D1, and 1dCD4-D2 proteins were obtained by first incubating each variant in PBS, pH 7.4 with 100 mM DTT for 1 h at RT. This was followed by the addition of 200 mM iodoacetamide for 1 h at RT to block exposed thiols before dialyzing into fresh PBS, pH 7.0, after which measurements were recorded. All spectra were corrected for solvent and normalized by calculating the mean residue ellipticity,  $[\theta]$  (degrees square centimeters per decimole) using the equation:

$$[\theta] = \frac{100\theta}{Cnl}$$

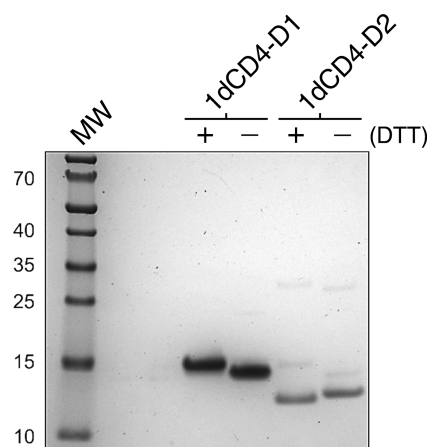
where  $\theta$  is the measured ellipticity (millidegrees),  $C$  is the protein concentration (millimolar),  $n$  is the number of residues in the protein, and  $l$  is the path length (centimeters). In addition, predicted spectra for 2dCD4 and 1dCD4 variants were obtained using the DichroCalc spectra prediction algorithm.<sup>40</sup> Amino acid crystal structure coordinates for each domain were extracted from PDB ID 1CDH and submitted for analysis.

**Temperature-Induced Unfolding Studies.** Temperature-induced denaturation of proteins were monitored by measuring the amount of  $\beta$ -sheet secondary structure present using ellipticity at 215 nm, employing a Jasco J-1500 spectropolarimeter over a temperature range of 20–90 °C. Thermal unfolding profiles were recorded using 10  $\mu$ M of each CD4 variant in PBS, pH 7.4. Fully reduced forms of 2dCD4, 1dCD4-D1, and 1dCD4-D2 were obtained as described above. The temperature was increased at a rate of 1 °C/min, which was controlled by a Jasco PTC-510 Peltier-type temperature control system. The temperature unfolding profiles were performed in duplicate, averaged, and normalized by converting the data to mean residue ellipticity. As irreversibility of unfolding precluded the calculation of thermodynamic parameters, lines of best fit were generated using SigmaPlot v11.0 (SPSS Science, Chicago, Illinois USA) to describe the thermal-unfolding profiles and for qualitative analysis of the data.

## RESULTS

**Disulfide Reduction in D1 and D2 of CD4 Increases and Decreases their Unfolded Hydrodynamic Volume Respectively.** Under nonreducing conditions, native wild-type 2dCD4 has been found to exist in solution in three distinct oxidation states in equilibrium. We have previously reported the assignment of the oxidation states of purified 2dCD4 isomers resolved by nonreducing SDS-PAGE: 2dCD4<sup>R2</sup> (where both the D1 and D2 disulfide bonds are fully reduced), 2dCD4<sup>R1</sup> (where the D2 disulfide is reduced), and 2dCD4<sup>Ox</sup> (where both D1 and D2 disulfides are oxidized) (see Figure 1 in Cerutti et al.<sup>38</sup>). Similarly, reducing and nonreducing SDS-PAGE analysis was performed on 1dCD4-D1 and 1dCD4-D2 (Figure 1).

By far the most common effect of reduction of disulfide bond containing proteins analyzed by SDS-PAGE is to decrease relative electrophoretic mobility.<sup>41</sup> This is essentially attributable to an increase in hydrodynamic volume caused by the relaxation and extension of disulfide-linked secondary structural elements. Consistent with this, reduction of 1dCD4-D1 resulted in a decrease in relative electrophoretic mobility and a related increase in apparent molecular mass of approximately 4 kDa (Figure 1, lane 1), a finding which can be attributed to the extension of previously disulfide-linked secondary structural elements.<sup>41</sup> Remarkably and in contrast, the effect of reduction



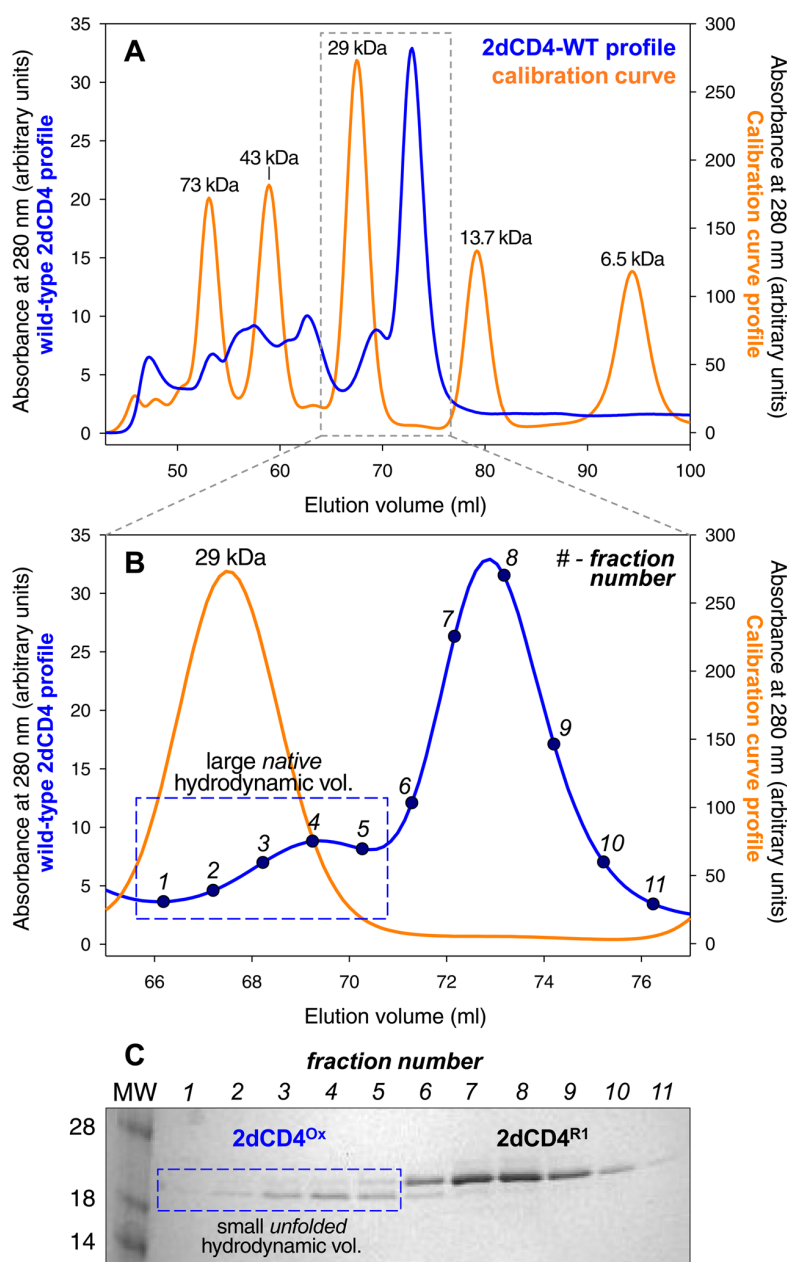
**Figure 1.** Changes to the unfolded hydrodynamic volume of domain 1 and domain 2 of CD4 upon reduction of their disulfide bonds. Purified recombinant 1dCD4-D1 and 1dCD4-D2 were subjected to both reducing and nonreducing denaturing PAGE by treating each with LDS loading buffer with or without DTT, which were then blocked with iodoacetamide and resolved by SDS-PAGE. Gels were stained with Coomassie Blue R250. Molecular mass, in kDa, is shown on the left.

on 1dCD4-D2 was to reduce apparent molecular mass by approximately 2 kDa (Figure 1, lane 3), suggesting that disulfide bond reduction in D2 decreases the hydrodynamic volume of this protein, including in its LDS unfolded state. A possible implication of this is that reduction of the allosteric D2 disulfide in CD4 may be associated with a release of the high energy contained in the disulfide bond, and an inward motion (or “collapse”) of the ordered,  $\beta$ -sheet structure. It is speculated that this leads to a more stable, hydrodynamically compact and energy-minimized structural intermediate that possesses secondary structure elements resistant—at least partially—to LDS-induced denaturation.

Thus, we reason that the increase in the unfolded hydrodynamic volume of 1dCD4-D1 upon treatment with DTT observed during denaturing PAGE (Figure 1) is reflective of a less compact, more extended state of the domain upon unfolding allowed by the reduction of the 16C–C84 and is consistent with the expected effect of reduction of a conventional structural disulfide. In contrast, reduction of the D2 disulfide leads to a more compacted, LDS-resistant structure with lower hydrodynamic volume, suggesting stabilization and possible proximation of secondary structural elements in 1dCD4-D2.

**Reduction of the D2 Disulfide Decreases the Native Hydrodynamic Volume of 2dCD4, as Revealed by Size-exclusion Chromatography.** To test the hypothesis that reduction of the D2 disulfide of CD4 results in the collapse of the secondary structure of the domain, we evaluated the differences in the native hydrodynamic volumes of the three known wild-type 2dCD4 redox isomers by attempting to resolve the distinct forms under nonreducing, nondenaturing conditions by size-exclusion chromatography (SEC). Indeed, the native hydrodynamic volumes of these isoforms are sufficiently distinct to enable isolation by SEC (Figure 2).

Nonreducing, denaturing SDS-PAGE analysis of the fractions collected following SEC separation of the redox isomers (Figure 2C) revealed that the smaller peak eluting at  $\pm 69$  mL (Figure 2B, blue box) corresponds to the 2dCD4 isoform



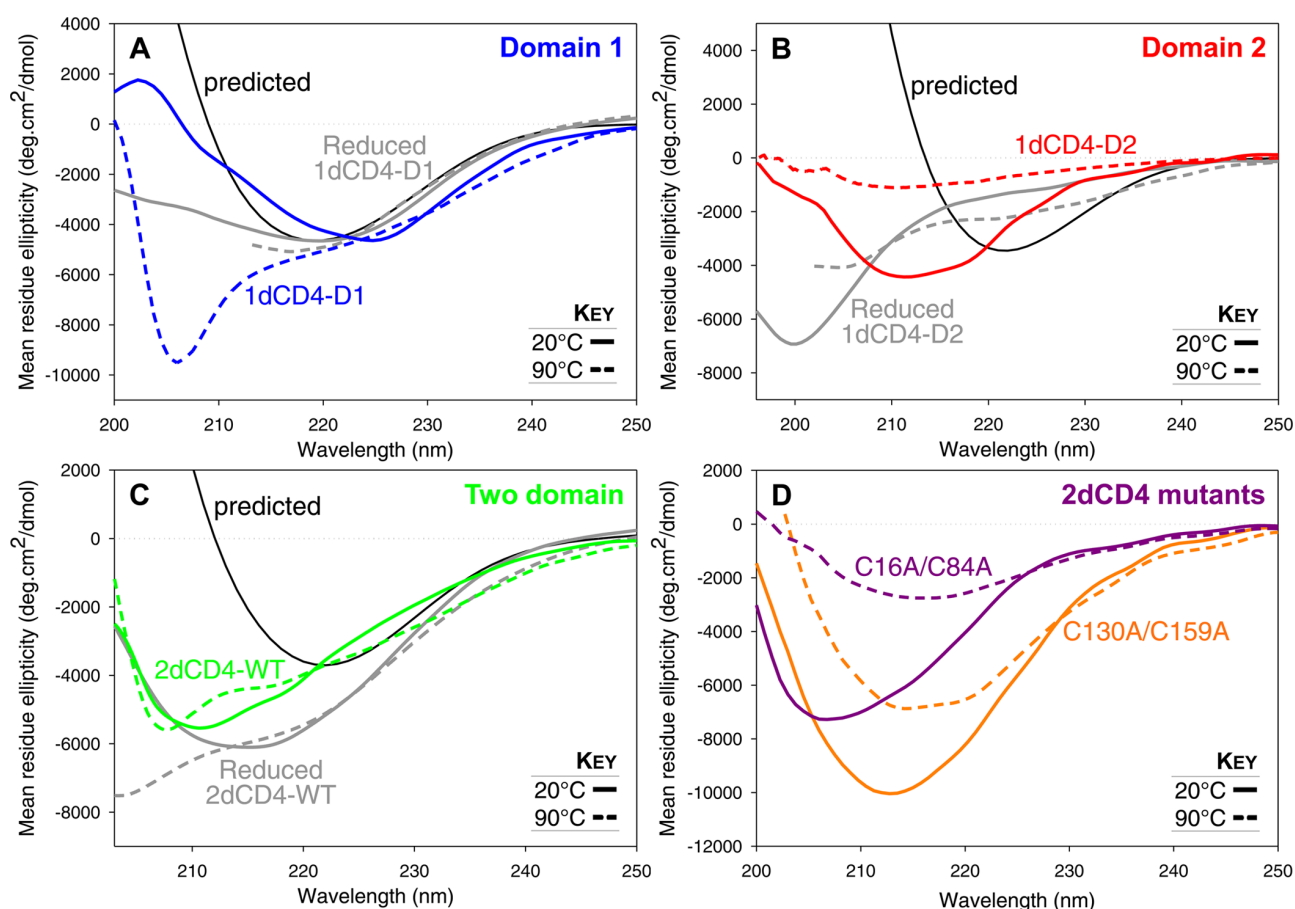
**Figure 2.** Fully oxidized 2dCD4-WT redox isomer has a larger native hydrodynamic volume than its partially reduced form. The full chromatogram for the separation of the redox isomers of 2dCD4-WT (blue) by a Superdex 75 size-exclusion column (A), and the portion of the chromatogram containing the 2dCD4-WT isomers of interest (B) were superimposed over the chromatogram of the column calibration (orange). 12% Bis-Tris SDS-PAGE analysis of the eluted fractions (C) indicated that the peak at 69 mL (blue box in B) corresponds to the fully oxidized species (2dCD4<sup>Ox</sup>) that resolves at  $\pm 22$  kDa under denaturing conditions. This species possesses a larger native hydrodynamic volume than the protein in the SEC peak at 73 mL, a protein that under denaturing conditions corresponds to the partially reduced 2dCD4 species (2dCD4<sup>R1</sup>) that resolves at  $\pm 25$  kDa possessing a relatively greater unfolded hydrodynamic volume. The species which resolves at 69 mL/22 kDa falls within the range of the 29–13.7 kDa calibration proteins and therefore does not represent a wild-type 2dCD4 dimer.

which resolves electrophoretically at 22 kDa (fully oxidized form) under nonreducing denaturing conditions, whereas the larger peak eluting at  $\pm 73$  mL corresponds to a species of 2dCD4 which resolves at 25 kDa (partially reduced form lacking the domain 2 disulfide). The resolution position and the corresponding implied hydrodynamic volume of these 2dCD4 species are therefore different in their native form as observed by SEC compared to their resolution by SDS-PAGE in which the proteins are denatured. These data suggest that the natively

folded fully oxidized 2dCD4 species has a larger hydrodynamic volume than the native structure of the partially reduced 2dCD4, whereas the inverse is true for their denatured forms.

A decrease in the hydrodynamic volume of a denatured protein when subjected to nonreducing SDS-PAGE commonly indicates the introduction of a disulfide bond because covalent linkages prevent the full extension of the polypeptide chain when it unfolds, as observed for the redox isomers of 2dCD4.<sup>38</sup> Size-exclusion chromatography on the other hand can provide





**Figure 3.** Ablation of the individual disulfide bonds in the single and two domain variants of CD4 alters their secondary structure. Far-UV circular dichroism spectra of the nonreduced and fully reduced forms of the domain 1 (A, blue), domain 2 (B, red), and two domain (C, green) variants of CD4, at both 20 °C (solid lines) and 90 °C (dashed lines) reveal the effect that both ablation of their disulfide bonds and thermal unfolding have on their secondary structure. Predicted far-UV CD spectra for these CD4 variants are shown as black lines. The impact that the separate reduction of the individual disulfide bonds of domain 1 and domain 2 have on the secondary structure of full-length 2dCD4 was indicated by the far-UV CD spectra of mutants lacking the ability to form the domain 1 (C16A/C84A, purple) and domain 2 (C130A/C159A, orange) disulfide bonds (D).

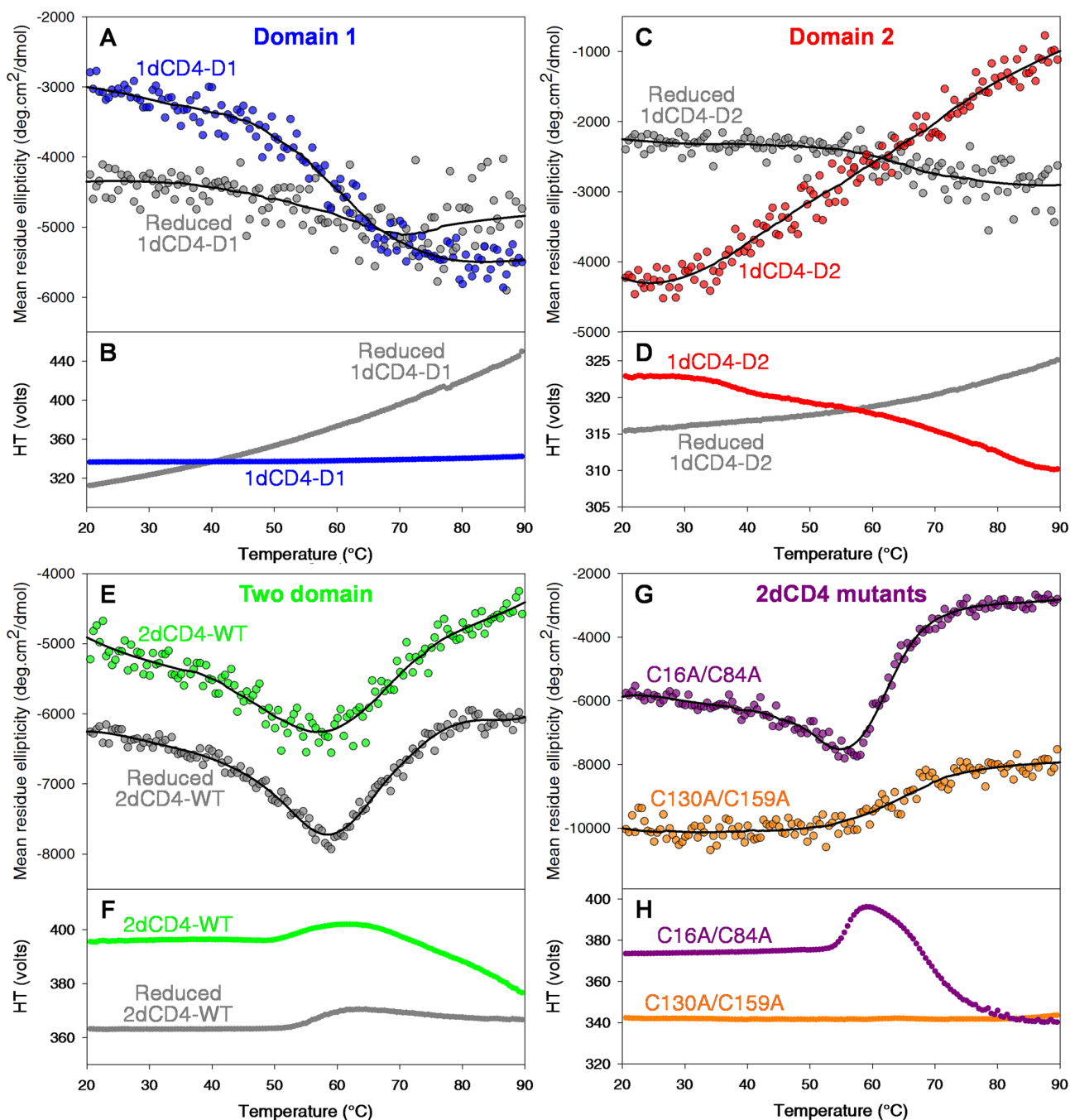
information on the changes to the hydrodynamic volume of a protein in its native, properly folded form, and can thus reveal the structural response of a protein when it experiences bond breakage such as disulfide reduction. With respect to 2dCD4, SEC reveals that the addition of a disulfide bond to domain 2 appears to distinctly increase the native hydrodynamic volume of the two domain protein.

The observation that the partially reduced form of natively folded 2dCD4 possesses a smaller hydrodynamic volume compared to its fully oxidized native counterpart, reinforces the proposition that reduction of the D2 disulfide causes structural rearrangements in D2 characterized by the relaxation or possible inward collapse of some secondary structure elements.

**Disulfide Reduction Affects the Secondary Structure of 2dCD4 and its Individual Domains, 1dCD4-D1 and 1dCD4-D2.** In order to gain more insight into the structure of each domain and their individual contributions to defining the secondary structure content of 2dCD4, as well as to begin to understand the unusual effect that reduction of the D2 disulfide has on the hydrodynamic volume of the protein, the secondary structures of both the individual CD4 domains (1dCD4-D1 and 1dCD4-D2) as well as 2dCD4 were characterized by CD

spectroscopy under both native, nonreducing conditions and after the reduction and blocking of their disulfide bonds (Figure 3A,B).

The CD spectrum of D1 displays typical highly ordered  $\beta$ -sheet fold topology, with an ellipticity minimum at approximately 223 nm and positive ellipticity below 206 nm (Figure 3A). Interestingly, this spectrum has characteristics similar to the spectrum predicted by the DichroCalc CD spectra prediction algorithm<sup>40</sup> for D1 extracted from PDB ID 1CDH (coordinates for amino acids 1–98). In contrast, 1dCD4-D2 generated a CD spectrum more consistent with  $\beta$ -stranding or short irregular  $\beta$ -sheeted structures,<sup>42</sup> with a significant decrease in the position of the ellipticity minimum which occurs between 210–213 nm (Figure 3B). Moreover, the predicted and experimental ellipticity minima for D2 (220 and 212 nm respectively) are substantially different, indicating that the structure observed in PDB ID 1CDH reflects only a single specific isomer and suggests that in solution, this molecule adopts a different or more dynamic structure. It is also important to note that the experimental spectrum of 1dCD4-D2 differs significantly from that of proteins that are classified as natively unfolded, which display considerably lower negative ellipticities at 200 nm.<sup>43,44</sup> These data are consistent with a



**Figure 4.** Thermal stability of the single and two domain variants of CD4 is modified by the redox state of their individual disulfide bonds. Temperature-induced unfolding of the domain 1 (A, blue), domain 2 (C, red), and two domain (E, green) variants of CD4, in both their nonreduced and reduced forms, was monitored by measuring ellipticity at 215 nm to indicate the effect that ablation of their disulfide bonds has on the stability of the proteins. The unfolding data are the average of two replicates, and lines of best fit are presented for clarity. Turbidity (dynode voltage) plots (B, D, and F) illustrate aggregation of the variants followed by precipitation of the aggregates during their thermal denaturation, indicated by the increase and reduction in voltage, respectively. Thermal denaturation and turbidity curves of the 2dCD4 mutants that are unable to form the domain 1 and domain 2 disulfide bonds (C16A/C84A and C130A/C159A, respectively) provide insight into the effect that reduction of the individual disulfide bonds in each domain has on the stability of full-length 2dCD4 (G and H).

model in which the ordered and inflexible D1 architecture conferred by its structural disulfide bond is contrasted with the metastability of D2 and its allosteric disulfide bond.

The CD spectrum of 2dCD4 under native conditions shows that it possesses an extensive  $\beta$ -sheeted structure, and while the ellipticity minimum at approximately 211 nm is indicative of

the presence of more irregular  $\beta$ -strands, it is not to the extent of individual 1dCD4-D2 (Figure 3C). The CD spectrum of the reduced state of 2dCD4 exhibits both a red shift (minimum at  $\pm 215$  nm) and a decrease in the magnitude of the negative ellipticity compared to 2dCD4-WT under native conditions (minimum at 211 nm) (Figure 3C).<sup>45</sup>

Reduction of the 130C–C159 disulfide bond in D2 results in a drastic change in the nature of its CD spectra, including large decreases in the wavelength and ellipticity of its minima (from approximately 212 to 200 nm) (Figure 3B). Collapse of the secondary structure elements of D2 of CD4 upon reduction of the allosteric intrastrand 130C–C159 disulfide likely results in the drastic reorientation and relative positioning of the  $\beta$ -strands that make up the domain.

Reduction of both disulfide bonds in 2dCD4 may result in similar conversions in D2, but these changes may be mitigated by the presence of D1, or by likely detrimental structural modifications that take place in the first domain upon reduction of its structural disulfide bond. These structural alterations are apparent when comparing the CD spectra of the reduced form of D1 with that of its native redox form, where reduction causes a blue shift in its ellipticity minima and a substantially shallower ascending arm below 220 nm (Figure 3A).

**Ablating the Disulfides of D1, D2, and 2dCD4 Impacts the Thermal Stability of the Proteins.** To measure the relative stabilities of D1 and D2, and determine the contribution that each domain makes to the overall stability of 2dCD4, we carried out temperature-induced denaturation of each variant while monitoring their secondary structure content using ellipticity measurements at 215 nm. This technique was also used to determine the effect that the redox state of each disulfide bond has on the stability of the domains in which they exist.

During temperature-induced denaturation, the CD4 variants showed complex and markedly distinct multistate thermal unfolding transitions as revealed by the line of best fit qualitatively describing each data set (Figure 4). Consequently, single representative  $T_M$  values cannot be accurately assigned to the profiles. Thermal unfolding was also shown to be irreversible as a result of the aggregation and/or precipitation revealed by the respective increase and decrease in dynode voltage (indicative of solution turbidity<sup>46</sup>) (Figure 4B,D,F,H). Nonetheless, thermal stability can be inferred from the temperature at which each CD4 protein begins to unfold ( $T_U$ ; the temperature at which the initial change in the secondary structure is observed). These points are fully distinct and can thus be used as clear indicators of relative protein stability.

Thermal denaturation of 1dCD4-D2 in its native redox state revealed that the individual domain is particularly unstable and begins to unfold and immediately precipitate at a low temperature of 33 °C (Figure 4C and Table 1). The finding

**Table 1. Temperature at Which Each CD4 Variant Begins to Unfold ( $T_U$ ) during Thermal Denaturation Monitored by Measuring the Ellipticity at 215 nm<sup>a</sup>**

CD4 variant		$T_U$ (°C)	precipitation at 90 °C
two domain	2dCD4-WT	40	yes
	2dCD4-WT, reduced	40	yes
	2dCD4-C16A/C84A	41	yes
	2dCD4-C130A/C159A	53	no
single domain	1dCD4-D1	47	no
	1dCD4-D1, reduced	39	yes
	1dCD4-D2	33	yes
	1dCD4-D2, reduced	55	no

<sup>a</sup>The presence or absence of visible precipitation at the completion of unfolding at 90 °C is indicated.

that D2 with an intact 130C–C159 disulfide is thermally unstable is supported by the gradual, linear, and non-cooperative decrease in the negative ellipticity at 215 nm, as well as a concomitant and progressive decrease in the dynode voltage from 33 °C (Figure 4D). These data indicate that the introduction of high temperatures causes excessive precipitation that steadily displaces D2 from solution, decreasing the concentration of soluble protein contributing to signal.

Remarkably, when the 130C–C159 disulfide is reduced, however, 1dCD4-D2 becomes significantly more recalcitrant to heat-induced unfolding. Reduced D2 begins to unfold at approximately 55 °C (22 °C higher than native D2), with little to no visible precipitation of the protein even at the final temperature of 90 °C (Figure 4C and Table 1). The steady increase in the dynode voltage of reduced D2 observed during unfolding does however reveal that the protein undergoes continual yet slight aggregation during thermal denaturation, but it does not experience the distinct precipitation shown by nonreduced 1dCD4-D2 (Figure 4D).

Unfolding of D1 of CD4 in its native redox state (possessing a fully oxidized 16C–C84 disulfide bond) begins at approximately 47 °C and proceeds non-cooperatively (Figure 4A and Table 1). The minor increase in the dynode voltage over the entire temperature range indicates that only a small degree of aggregation occurs as the protein unfolds (Figure 4B), consistent with the lack of visible precipitation in the cuvette at 90 °C.

Interestingly, thermal denaturation of D1 and exposure to temperatures as high as 90 °C does not destroy all the secondary structure present under native redox conditions. The secondary structure is considerably different from the structure at 20 °C as evidenced by the fact that the unfolded spectrum lacks a characteristic minimum at 223 nm. However, at this wavelength there is a distinct shoulder to a striking trough at 206 nm that has large negative ellipticity (Figure 3A), indicating that a substantial degree of secondary structure still remains in 1dCD4-D1 even when exposed to disruptive high temperatures. This spectrum is not a product of light scattering concealing the true CD spectrum, as the dynode voltage remains below 600 mV above 204 nm. The alteration to the structure of D1 induced by high temperatures is, however, irreversible upon cooling (data not shown).

Because of the similarity in the spectra of fully reduced 1dCD4-D1 at both 20 and 90 °C (at least at the wavelengths at which the thermal unfolding data is reliable with an appropriate signal-to-noise ratio) (Figure 3A), it is difficult to monitor its thermal unfolding transition accurately. Nonetheless, it is clear that the folded baseline of the curve only remains unchanged until  $\pm 39$  °C at the maximum until it begins to unfold (Figure 4A and Table 1). Furthermore, the noticeable change in the ellipticity at  $\pm 70$  °C and subsequent gradual decrease in the negative ellipticity is similar to that exhibited during the precipitation of native D2 as the temperature is raised. This is consistent with the precipitated reduced 1dCD4-D1 protein observed in the cuvette at 90 °C.

It is a possibility that 1dCD4-D1 may already be at least partly unfolded at 20 °C when its disulfide bond is reduced, suggesting that there may not be a significant change in the structure and stability of the reduced state of D1 upon temperature induced structural rearrangement or when shifting between two (or more) forms of several possible unfolded states. Nonetheless, as native D1 begins to unfold at the higher temperature of 47 °C and does not precipitate upon complete

unfolding, it is clear that reduction of the 16C–C84 disulfide bond of 1dCD4-D1 is destabilizing and induces substantial changes in the structure of this region of 2dCD4 when it is in the absence of D2.

The nature of thermal unfolding of 2dCD4 under native redox conditions initially resembles that of native D1, yet it begins to unfold non-cooperatively at a comparatively lower temperature of approximately 40 °C (Figure 4E and Table 1). Moreover, notable aggregation of the protein begins at  $\pm 47$  °C, followed by robust precipitation at  $\pm 58$  °C, as revealed by the respective increase and decrease in dynode voltage at those temperatures (Figure 4F), as well as the sharp change in the ellipticity at 215 nm. The gradual decrease in the negative ellipticity above 58 °C indicates that aggregation and precipitation of native 2dCD4 is occurring in a manner not unlike that described above for native D2 and reduced D1, which concurs with the observable precipitation present in the cuvette at 90 °C.

Despite this, however, the population of denatured yet soluble 2dCD4 that has not precipitated by 90 °C does not unfold to form a complete random coil, and retains a high degree of secondary structure similar to that of D1. The heat unfolded spectrum of the remaining soluble 2dCD4 also has a distinct trough at 206 nm with a large shoulder extending to higher wavelengths (Figure 3C). The accuracy of this spectrum is reiterated by its corresponding dynode voltage, which again measures below 600 mV at wavelengths above 205 nm (data not shown).

The thermal unfolding of a fully reduced two domain CD4 in which both 16C–C84 and 130C–C159 of D1 and D2 respectively are ablated is very similar to that of native 2dCD4. Reducing both disulfide bonds together appears not to affect the stability of the protein as the temperature at which it starts to unfold stays at  $\pm 40$  °C (Figure 4E and Table 1). The fully reduced protein then unfolds in a manner comparable to the pathway followed by native 2dCD4, where aggregation occurs followed by extensive precipitation at  $\pm 58$  °C (Figure 4E,F).

Thus, in the context of 2dCD4 under native (nonreducing) conditions, D1 and its intact structural disulfide bond confer stability on the metastable D2 with its destabilizing allosteric bond. Conversely, under reducing conditions, stabilization of D2 upon reduction of 130C–C159 and the subsequent structural collapse of the  $\beta$ -sheets that this induces, confers enough stability to full-length reduced 2dCD4 that the unfavorable changes to the structure and stability of D1 caused by the reduction of its 16C–C84 disulfide bond are mitigated.

**Disulfide-deficient Mutants Reveal the Contribution of the Individual Disulfides in D1 and D2 to the Structure and Stability of 2dCD4.** In order to evaluate the contribution of the individual disulfide bonds to the structure and stability of full-length 2dCD4, we generated a panel of identical 2dCD4 proteins containing double Cys/Ala substitutions at the disulfide bond residues in each domain. By ablation of each disulfide bond separately, these mutant proteins can therefore simulate the various 2dCD4-WT redox states.

Analyzing the secondary structure of the nondisulfide bonded D1 mutant, 2dCD4-C16A/C84A, by CD spectroscopy produced a spectrum with an ellipticity minimum at 206 nm (Figure 3D). The minimum has both greater negative ellipticity and is distinctly blue-shifted when compared to the minimum exhibited by 2dCD4-WT at 211 nm (Figure 3C). These

characteristics are indicative of a protein that possesses  $\beta$ -sheeted structures that are more irregular and/or shorter than those present in 2dCD4-WT.

In contrast, the CD spectra of 2dCD4-C130A/C159A, lacking the ability to form the D2 allosteric disulfide bond, displays a slightly red-shifted minimum (at 213 nm) with increased negative ellipticity that is close to double that of 2dCD4-WT (Figure 3D). This is indicative of a higher secondary structure content for 2dCD4-C130A/C159A relative to both wild-type and C16A/C84A mutant 2dCD4, as well as to fully reduced 2dCD4 (Figure 3C).

2dCD4-C16A/C84A thermally unfolds in a manner similar to wild-type and reduced 2dCD4; the mutant begins to unfold at a temperature of  $\pm 41$  °C (Figure 4G and Table 1) and begins to precipitate out of solution at  $\pm 58$  °C (Figure 4H). However, it precipitates with much higher cooperativity compared to 2dCD4-WT.

Thermal unfolding of 2dCD4-C130A/C159A on the other hand begins at  $\pm 53$  °C, a temperature that is higher than the unfolding of both wild-type and fully reduced 2dCD4, as well as the 2dCD4-C16A/C84A domain 1 mutant (Figure 4G and Table 1). In addition, no significant aggregation occurred across the entire temperature range of denaturation as indicated by the particularly stable dynode voltage (Figure 4H) and the lack of precipitation in the cuvette upon completion of unfolding. As the decrease in the negative ellipticity during the unfolding transition of this mutant cannot be attributed to aggregation and precipitation, this protein must unfold via a pathway that is distinct from the other 2dCD4 variants, whose unfolding transitions begin with an increase in negative ellipticity while exhibiting decreases in signal only as a result of precipitation. Furthermore, at the completion of thermal unfolding at 90 °C, the D2 disulfide mutant still exhibits a well-defined CD spectrum with an ellipticity minimum of just above  $-7000$  deg $\cdot$ cm $^2$ /dmol indicating the presence of considerable secondary structure in the protein at high temperatures.

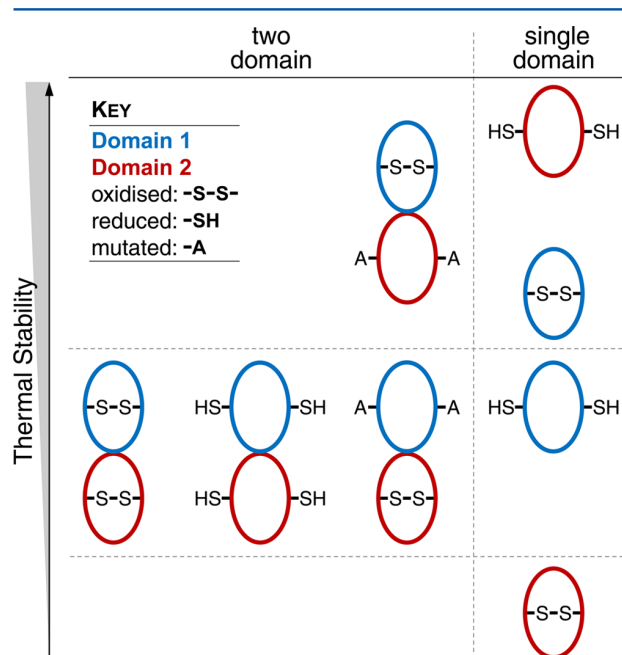
## DISCUSSION

While it is now generally accepted that reduction of one or more of the disulfides of CD4 is important for potentiating the molecular changes required for effective engagement of CD4 with TCR-MHCII complexes and HIV-1 gp120,<sup>47,48</sup> the biophysical and structural consequences of CD4 disulfide reduction that underlie its ligand-binding specificities are not known. The role that the atypical D2 allosteric disulfide bond in particular plays in the structure and stability of CD4 has also until now remained relatively unclear. It had been previously suggested that the allosteric D2 disulfide is not an essential part of the fold since the homologous D4 domain of CD4 has a conventional intersheet disulfide bridge and the D1/D2 and D3/D4 domains are structurally comparable.<sup>36</sup> Herein, however, it has been demonstrated that the D2 disulfide does indeed influence the character of the first two domains of CD4 as its redox state has a significant impact on the secondary structural elements and the stability of the protein. In line with this, this paper describes biophysical data on the secondary structure and stability of several redox variants of one domain and two-domain CD4 (2dCD4) as part of ongoing efforts to gain further insights into the consequences and biological importance of CD4 D2 oxidoreduction.

**Allosteric Disulfide Bond Dependent Changes in the Structure and Stability of CD4.** We examined the effect that reduction or ablation of each disulfide bond in both single

domain and two domain CD4 variants has on the thermal stability and secondary structure of the proteins by measuring their thermal unfolding using circular dichroism.

The least stable variant of the single domain CD4 proteins was 1dCD4-D2 with its putative allosteric disulfide bond intact. However, this protein experienced significant restructuring and a large increase in thermal stability (the greatest change in any CD4 variant) upon reduction of its disulfide bond (described schematically in Figure 5). This effect is consistent with a



**Figure 5.** Effect that ablation of the individual disulfide bonds has on the thermal stability of both the two domain and single domain variants of CD4. The diagram illustrates the relative thermal stabilities of the partially or fully reduced forms of full-length 2dCD4-WT, domain 1 (blue), and domain 2 (red). Ablating the allosteric disulfide bond in domain 2 alone appears to increase the relative stabilities of both the D2 single domain and two domain variants of CD4.

reordering of secondary structural elements that is likely to be induced by torsional energy released from the strained D2 disulfide bond. That these changes result in domain compaction and stabilization is supported additionally by the decrease in hydrodynamic volume suggested by reducing and nonreducing SDS-PAGE analysis. The structural reconfiguration and increase in the stability of D2 thus occurs to such an extent that certain structural elements are apparently able to resist full LDS-induced denaturation in their more energetically favorable form. In contrast, reduction of the disulfide bond in 1dCD4-D1 (which exhibits the highest stability of any nonreduced one or two domain CD4 variant) results in decreased stability and a corresponding increase in the hydrodynamic volume, a typical biophysical consequence of structural disulfide bond ablation. Taken together, these data are consistent with a model in which the highly ordered, inflexible D1 architecture is contrasted by the metastability of D2, biophysical characteristics that are shared with their respective putative structural and allosteric disulfide bonds.

Several attempts to generate a homogeneously pure form of CD4 with oxidized disulfides in both domains—an objective that will greatly facilitate a more detailed biophysical character-

ization of the contributions of each disulfide in D1 and D2 to global structural dynamics of CD4—were unsuccessful. This endeavor is further challenged by the diverse biophysical consequences of reduction of the disulfides in each domain. Nevertheless, considered alongside knowledge on the impact of disulfide bond reduction on each domain in isolation, our analysis of the comparative changes in hydrodynamic volume, secondary structure, and thermal stability of reduced and oxidized forms of wild-type 2-domain CD4 (2dCD4) and a range of 2dCD4 Cys/Ala variants provided further support to the notion that reduction of the allosteric disulfide bond results in substantial reordering and stabilization of D2. Moreover, these studies have generated two additional insights that warrant further investigation: that an intact D1 disulfide is required to keep the torsionally strained D2 disulfide intact in the native state, and that local structural perturbations induced by its reduction are propagated beyond the second domain.

Upon identifying the significant prestress that exists within the D2 disulfide, Zhou et al. questioned whether a structural change would occur in CD4 upon ablation of the allosteric bond by modeling the protein with the D2 pair of cysteines in a reduced state.<sup>31</sup> While the simulations they performed did not indicate any observable conformational changes in CD4 following the reduction of its prestressed allosteric disulfide, the authors acknowledged that the redox-dependent regulation of the receptor may entail larger structural changes that take place at time-scales that extend past their nanosecond simulations. Our current findings can thus be seen as experimental confirmation of the theoretical prediction that, at longer time scales, the allosteric disulfide of D2 induces conformational changes in the domain when shuffling between its redox states.

**Functional Implications of the Metastability of the D2 Allosteric Disulfide.** Despite these insights, and others that have demonstrated how the redox state of the D2 disulfide defines the ligand-binding specificity of CD4,<sup>30,38,47</sup> the biological importance of CD4 domain 2 metastability remains unclear. It has been proposed that the oxidized and partially reduced forms of CD4 exist in equilibrium on the cell surface, and secretion of thioredoxin by activated T-cells facilitates the reduction of the D2 allosteric disulfide to trigger necessary conformational changes in CD4.<sup>47,48</sup> Although the exact structural changes that take place in CD4 are not fully understood, the data presented here demonstrate that fundamental, stabilizing structural changes occur in D2 upon reduction of its disulfide bond. The metastability of D2 may “prime” the receptor for rapid structural changes induced by cell surface associated oxidoreductases such as thioredoxin, a putative physiological reductant of CD4.<sup>38,49,50</sup> In addition to important oxidation reactions in which they participate, allosteric disulfides that decrease protein stability or rigidity may confer functionally important flexibility on some part of the folded protein. Internal conformational flexibility has long been understood to be an important contributing factor to the activity of a protein,<sup>51</sup> and the instability imparted by the allosteric disulfide bond of CD4 may thus play an important role in providing the flexibility required for fine and dynamic regulation of CD4 interactions.

**Conclusions.** Numerous proteins have been found whose functions are mediated by the redox state of an allosteric disulfide bond, including botulinum neurotoxins, von Willebrand factor, plasminogen, and Src kinase to name only a few examples (reviewed by Cook and Hogg<sup>52</sup>). A recent study has

discovered that in addition to CD4, many varied redox-regulated disulfide bond-containing proteins exist on the surface of T-cells.<sup>53</sup> The allosteric disulfides of these proteins may differ in the manner in which they are cleaved (by either oxidoreductases or by thiol/disulfide exchange), but the conformations of the disulfides and the secondary structures that they link share many common features. Moreover, the exact mechanisms by which the reduction of many of the known allosteric disulfide bonds induces functional changes in their proteins are poorly understood. The insight gained into the redox-dependent biophysical changes of CD4 may thus act as a model to aid in the functional description of other uncharacterized allosteric-bond containing proteins and help explain how that function is mediated. It would be of significance to test experimentally whether domain proximation following reduction of prestressed –RHStaple allosteric disulfide bonds is a common regulatory mechanism in proteins, particularly given the implication of these disulfides in human disease.<sup>52</sup>

In summary, the data presented here describe the metastable nature of the D2 disulfide bond, a characteristic that is now widely understood to be important to regulation of CD4 function by assisting in the presentation of its different isomers at the cell surface. This study thus begins to elucidate the influence that the redox state of the disulfide bonds have on the structure and stability of the different and distinct CD4 isomers. We demonstrate that reduction of the D2 disulfide results in previously undescribed changes to the conformation and stability of the first two domains of CD4, changes that can be hypothesized to be important for the interaction of full length CD4 with HIV gp120 and MHCII. Current efforts in our laboratory are focused on obtaining more detailed descriptions of the changes to the structure and conformational dynamics that redox shuffling of the metastable disulfide in D2 is likely to effect, using X-ray crystallography and hydrogen–deuterium exchange mass spectrometry. Biochemical knowledge of this nature will not only provide the information required to better understand the specificities of CD4 ligand binding but a more general appreciation of the importance of the structure and stability of allosteric disulfides in other biological systems. A full understanding of the disulfide-shuffling induced changes in CD4 will also be of value in the rational design of novel anti-HIV pharmaceuticals.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [gavinrayowen@gmail.com](mailto:gavinrayowen@gmail.com). Tel: +27 11 717 2164.

### Funding

Financial support was provided by the South African Medical Research Council through a SHIP (Strategic Health Innovation Partnerships) grant as well as by the South African National Research Foundation through the Thuthuka Programme and a PROTEA grant.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank H. W. Dirr for generously allowing the use of the Jasco J-1500 spectropolarimeter.

## ABBREVIATIONS

CD, circular dichroism; CD4, cluster of differentiation 4; D1–D4, domain 1–4 of CD4; 2dCD4, two domain CD4 variant comprising both domain 1 and domain 2; 1dCD4-D1, single domain CD4 variant comprising only domain 1; 1dCD4-D2, single domain CD4 variant comprising only domain 2; WT, wild-type; MHCII, major histocompatibility complex class II; HIV-1, the human immunodeficiency virus type 1; LDS, lithium dodecyl sulfate; SEC, size-exclusion chromatography

## REFERENCES

- (1) Trobridge, P. A., Forbush, K. A., and Levin, S. D. (2001) Positive and negative selection of thymocytes depends on Lck interaction with the CD4 and CD8 coreceptors. *J. Immunol.* 166, 809–818.
- (2) Wang, J. H., Meijers, R., Xiong, Y., Liu, J. H., Sakihama, T., Zhang, R., Joachimiak, A., and Reinherz, E. L. (2001) Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10799–10804.
- (3) Basch, R. S., Kouri, Y. H., and Karpatkin, S. (1990) Expression of CD4 by human megakaryocytes. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8085–8089.
- (4) Cutrona, G., Leanza, N., Ulivi, M., Majolini, M. B., Taborelli, G., Zupo, S., Baldari, C. T., Roncella, S., and Ferrarini, M. (1999) Apoptosis induced by crosslinking of CD4 on activated human B cells. *Cell. Immunol.* 193, 80–89.
- (5) Lee, B., Sharron, M., Montaner, L. J., Weissman, D., and Doms, R. W. (1999) Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5215–5220.
- (6) Li, Y., Li, L., Wadley, R., Reddel, S. W., Qi, J. C., Archis, C., Collins, A., Clark, E., Cooley, M., Kouts, S., Naif, H. M., Alali, M., Cunningham, A., Wong, G. W., Stevens, R. L., and Krilis, S. A. (2001) Mast cells/basophils in the peripheral blood of allergic individuals who are HIV-1 susceptible due to their surface expression of CD4 and the chemokine receptors CCR3, CCR5, and CXCR4. *Blood* 97, 3484–3490.
- (7) Lucey, D. R., Dorsky, D. I., Nicholson-Weller, A., and Weller, P. F. (1989) Human eosinophils express CD4 protein and bind human immunodeficiency virus 1 gp120. *J. Exp. Med.* 169, 327–332.
- (8) Wood, G. S., Warner, N. L., and Warnke, R. A. (1983) Anti-Leu-3/T4 antibodies react with cells of monocyte/macrophage and Langerhans lineage. *J. Immunol.* 131, 212–216.
- (9) Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 312, 763–767.
- (10) Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C., and Montagnier, L. (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 312, 767–768.
- (11) Robey, E., and Axel, R. (1990) CD4: collaborator in immune recognition and HIV infection. *Cell* 60, 697–700.
- (12) Ryu, S. E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X. P., Xuong, N. H., Axel, R., Sweet, R. W., et al. (1990) Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* 348, 419–426.
- (13) Wang, J. H., Yan, Y. W., Garrett, T. P., Liu, J. H., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. (1990) Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* 348, 411–418.
- (14) Brady, R. L., Dodson, E. J., Dodson, G. G., Lange, G., Davis, S. J., Williams, A. F., and Barclay, A. N. (1993) Crystal structure of domains 3 and 4 of rat CD4: relation to the NH2-terminal domains. *Science* 260, 979–983.

- (15) Garrett, T. P., Wang, J., Yan, Y., Liu, J., and Harrison, S. C. (1993) Refinement and analysis of the structure of the first two domains of human CD4. *J. Mol. Biol.* 234, 763–778.
- (16) Wu, H., Kwong, P. D., and Hendrickson, W. A. (1997) Dimeric association and segmental variability in the structure of human CD4. *Nature* 387, 527–530.
- (17) Wouters, M. A., Fan, S. W., and Haworth, N. L. (2010) Disulfides as redox switches: from molecular mechanisms to functional significance. *Antioxid. Redox Signaling* 12, 53–91.
- (18) Roos, G., Fonseca Guerra, C., and Bickelhaupt, F. M. (2015) How the disulfide conformation determines the disulfide/thiol redox potential. *J. Biomol. Struct. Dyn.* 33, 93–103.
- (19) Schmidt, B., Ho, L., and Hogg, P. J. (2006) Allosteric disulfide bonds. *Biochemistry* 45, 7429–7433.
- (20) Schmidt, B., and Hogg, P. J. (2007) Search for allosteric disulfide bonds in NMR structures. *BMC Struct. Biol.* 7, 49.
- (21) Wouters, M. A., Lau, K. K., and Hogg, P. J. (2004) Cross-strand disulphides in cell entry proteins: poised to act. *BioEssays* 26, 73–79.
- (22) Matthias, L. J., Yam, P. T., Jiang, X. M., Vandegraaff, N., Li, P., Pombourios, P., Donoghue, N., and Hogg, P. J. (2002) Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nat. Immunol.* 3, 727–732.
- (23) Tagaya, Y., Maeda, Y., Mitsui, A., Kondo, N., Matsui, H., Hamuro, J., Brown, N., Arai, K., Yokota, T., Wakasugi, H., et al. (1989) ATL-derived factor (ADF), an IL-2 receptor/Tac inducer homologous to thioredoxin; possible involvement of dithiol-reduction in the IL-2 receptor induction. *EMBO J.* 8, 757–764.
- (24) Lynch, G. W., Slaytor, E. K., Elliott, F. D., Saurajen, A., Turville, S. G., Sloane, A. J., Cameron, P. U., Cunningham, A. L., and Halliday, G. M. (2003) CD4 is expressed by epidermal Langerhans' cells predominantly as covalent dimers. *Exp. Dermatol.* 12, 700–711.
- (25) Lynch, G. W., Sloane, A. J., Raso, V., Lai, A., and Cunningham, A. L. (1999) Direct evidence for native CD4 oligomers in lymphoid and monocytoid cells. *Eur. J. Immunol.* 29, 2590–2602.
- (26) Sakihama, T., Smolyar, A., and Reinherz, E. L. (1995) Oligomerization of CD4 is required for stable binding to class II major histocompatibility complex proteins but not for interaction with human immunodeficiency virus gp120. *Proc. Natl. Acad. Sci. U. S. A.* 92, 6444–6448.
- (27) Konig, R., Shen, X., and Germain, R. N. (1995) Involvement of both major histocompatibility complex class II alpha and beta chains in CD4 function indicates a role for ordered oligomerization in T cell activation. *J. Exp. Med.* 182, 779–787.
- (28) Weber, S., and Karjalainen, K. (1993) Mouse CD4 binds MHC class II with extremely low affinity. *Int. Immunol.* 5, 695–698.
- (29) Matthias, L. J., Azimi, I., Tabrett, C. A., and Hogg, P. J. (2010) Reduced monomeric CD4 is the preferred receptor for HIV. *J. Biol. Chem.* 285, 40793–40799.
- (30) Maekawa, A., Schmidt, B., Fazekas de St Groth, B., Sanejouand, Y. H., and Hogg, P. J. (2006) Evidence for a domain-swapped CD4 dimer as the coreceptor for binding to class II MHC. *J. Immunol.* 176, 6873–6878.
- (31) Zhou, B., Baldus, I. B., Li, W., Edwards, S. A., and Grater, F. (2014) Identification of allosteric disulfides from prestress analysis. *Biophys. J.* 107, 672–681.
- (32) Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393, 648–659.
- (33) Denisova, G., Zwickel, J., and Gershoni, J. M. (1995) Binding of HIV-1 gp120 to an anti-V3 loop antibody reveals novel antigen-induced epitopes. *FASEB J.* 9, 127–132.
- (34) Hsu, S. T., Peter, C., van Gunsteren, W. F., and Bonvin, A. M. (2005) Entropy calculation of HIV-1 Env gp120, its receptor CD4, and their complex: an analysis of configurational entropy changes upon complexation. *Biophys. J.* 88, 15–24.
- (35) Sanejouand, Y. H. (1996) Normal-mode analysis suggests important flexibility between the two N-terminal domains of CD4 and supports the hypothesis of a conformational change in CD4 upon HIV binding. *Protein Eng., Des. Sel.* 9, 671–677.
- (36) Ryu, S. E., Truneh, A., Sweet, R. W., and Hendrickson, W. A. (1994) Structures of an HIV and MHC binding fragment from human CD4 as refined in two crystal lattices. *Structure* 2, 59–74.
- (37) Cerutti, N., Mendelow, B. V., Napier, G. B., Papathanasopoulos, M. A., Killick, M., Khati, M., Stevens, W., and Capovilla, A. (2010) Stabilization of HIV-1 gp120-CD4 receptor complex through targeted interchain disulfide exchange. *J. Biol. Chem.* 285, 25743–25752.
- (38) Cerutti, N., Killick, M., Jugnarain, V., Papathanasopoulos, M., and Capovilla, A. (2014) Disulfide reduction in CD4 domain 1 or 2 is essential for interaction with HIV glycoprotein 120 (gp120), which impairs thioredoxin-driven CD4 dimerization. *J. Biol. Chem.* 289, 10455–10465.
- (39) Chen, W., Feng, Y., Gong, R., Zhu, Z., Wang, Y., Zhao, Q., and Dimitrov, D. S. (2011) Engineered single human CD4 domains as potent HIV-1 inhibitors and components of vaccine immunogens. *J. Virol.* 85, 9395–9405.
- (40) Bulheller, B. M., and Hirst, J. D. (2009) DichroCalc—circular and linear dichroism online. *Bioinformatics* 25, 539–540.
- (41) Wingfield, P. T., Palmer, I., and Liang, S. M. (1995) Folding and purification of insoluble (inclusion body) proteins from *Escherichia coli*. *Curr. Protoc. Protein Sci.* 6, 6.5.1–6.5.27.
- (42) Wu, J., Yang, J. T., and Wu, C. S. (1992) Beta-II conformation of all-beta proteins can be distinguished from unordered form by circular dichroism. *Anal. Biochem.* 200, 359–364.
- (43) Denning, D. P., Uversky, V., Patel, S. S., Fink, A. L., and Rexach, M. (2002) The *Saccharomyces cerevisiae* nucleoporin Nup2p is a natively unfolded protein. *J. Biol. Chem.* 277, 33447–33455.
- (44) Sanchez-Puig, N., Veprintsev, D. B., and Fersht, A. R. (2005) Human full-length Securin is a natively unfolded protein. *Protein Sci.* 14, 1410–1418.
- (45) Manning, M. C., and Woody, R. W. (1987) Theoretical determination of the CD of proteins containing closely packed antiparallel beta-sheets. *Biopolymers* 26, 1731–1752.
- (46) Benjwal, S., Verma, S., Röhm, K. H., and Gursky, O. (2006) Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Sci.* 15, 635–639.
- (47) Matthias, L. J., Yam, P. T. W., Jiang, X.-M., Vandegraaff, N., Li, P., Pombourios, P., Donoghue, N., and Hogg, P. J. (2002) Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1. *Nat. Immunol.* 3, 727–732.
- (48) Matthias, L. J., Azimi, I., Tabrett, C. A., and Hogg, P. J. (2010) Reduced Monomeric CD4 Is the Preferred Receptor for HIV. *J. Biol. Chem.* 285, 40793–40799.
- (49) Markovic, I., Stantchev, T. S., Fields, K. H., Tiffany, L. J., Tomic, M., Weiss, C. D., Broder, C. C., Strebel, K., and Clouse, K. A. (2004) Thiol/disulfide exchange is a prerequisite for CXCR4-tropic HIV-1 envelope-mediated T-cell fusion during viral entry. *Blood* 103, 1586–1594.
- (50) Ou, W., and Silver, J. (2006) Role of protein disulfide isomerase and other thiol-reactive proteins in HIV-1 envelope protein-mediated fusion. *Virology* 350, 406–417.
- (51) Bennett, W. S., Huber, R., and Engel, J. (1984) Structural and functional aspects of domain motions in proteins. *Crit. Rev. Biochem.* 15, 291–384.
- (52) Cook, K. M., and Hogg, P. J. (2013) Post-translational control of protein function by disulfide bond cleavage. *Antioxid. Redox Signaling* 18, 1987–2015.
- (53) Metcalfe, C., Cresswell, P., Ciaccia, L., Thomas, B., and Barclay, A. N. (2011) Labile disulfide bonds are common at the leucocyte cell surface. *Open Biol.* 1, 110010.