



University  
of Glasgow

Vijayakrishnan, Swetha (2009) *The architectural complexity of the human PDC core assembly*.  
PhD thesis.

<http://theses.gla.ac.uk/573/>

Copyright and moral rights for this thesis are retained by the author

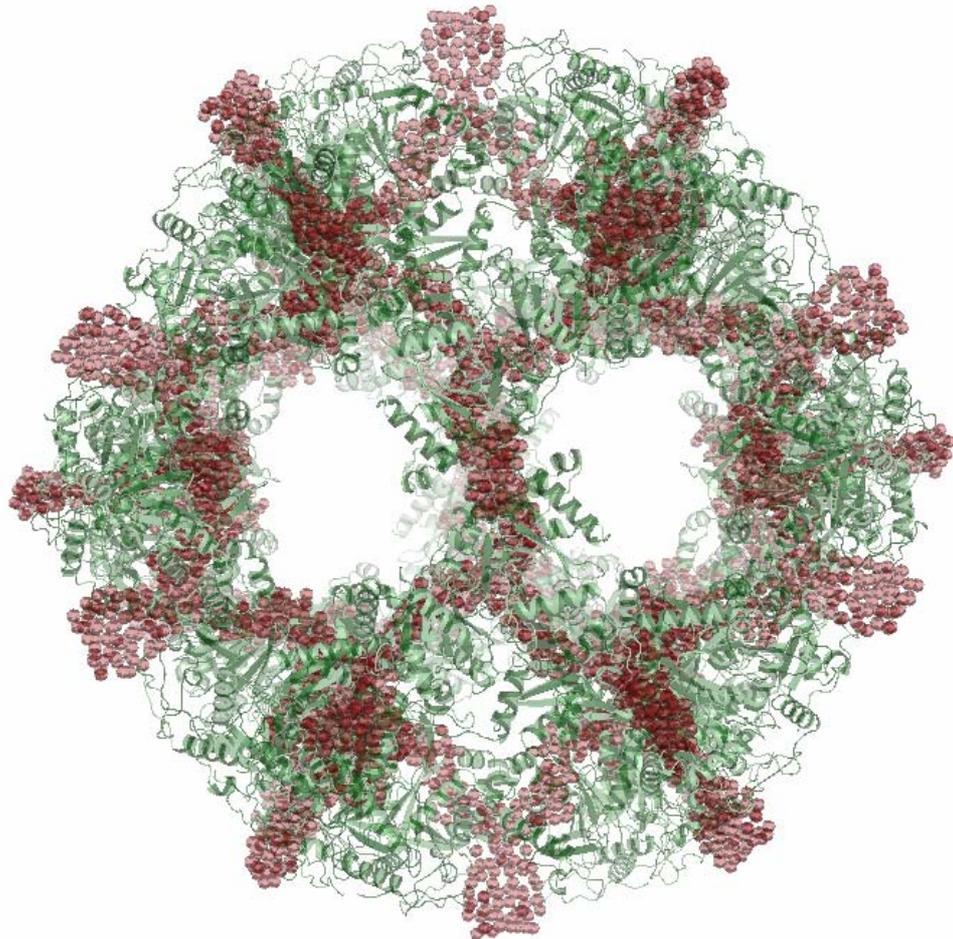
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

# The architectural complexity of the human PDC core assembly



Swetha Vijayakrishnan

*A thesis submitted for the Doctor of Philosophy*

Department of Molecular & Cellular Biology, IBLS  
University of Glasgow, February 2009

*The important thing is not to stop questioning. Curiosity has its own reason for existing.*

Albert Einstein

*The important thing in science is not so much to obtain new facts, but as to discover new ways of thinking about them.*

William Lawrence Bragg

*The doubter is a true man of science; he doubts only himself and his interpretations, but he believes in science.*

Claude Bernard

## Abstract

The mammalian pyruvate dehydrogenase complex (PDC) is a key multi-enzyme assembly linking the glycolytic pathway to the TCA cycle via the specific conversion of pyruvate to acetyl CoA and, as such, is responsible for the maintenance of glucose homeostasis in humans. PDC comprises a central pentagonal dodecahedral core of 60 dihydrolipoamide acetyltransferase (E2) and 12 E3 binding protein (E3BP) subunits. Presently, two conflicting models of PDC (E2+E3BP) core organisation exist: the ‘addition’ (60+12) and ‘substitution’ (48+12) models. In addition to its catalytic role, the multi-domain E2/E3BP core provides the structural framework to which 30 pyruvate decarboxylase (E1) heterotetramers and 6-12 dihydrolipoamide dehydrogenase (E3) homodimers are proposed to bind at maximal occupancy. The formation of specific E2:E1 and E3BP:E3 subcomplexes are characteristic of eukaryotic PDCs and are critical for normal complex function. Despite the availability of limited structural data, the exact subunit organisation and mechanism of operation of the mammalian E2/E3BP core remains unknown.

This thesis describes the large-scale purification of tagged, recombinant human PDC cores, full-length rE2 and rE2/E3BP, truncated E2/E3BP, peripheral rE3 enzyme as well as native E2/E3BP core (bE2/E3BP) purified from bovine heart. The ability to purify large amounts of pure protein has enabled the characterisation of the individual cores as well as the E2/E3BP:E3 complex using a variety of biochemical and biophysical techniques.

Full-length rE2/E3BP, rE2, bE2/E3BP, truncated E2/E3BP (tLi19/tLi30) and rE2/E3BP:E3 were analysed in solution by analytical ultracentrifugation (AUC). While AUC of the cores supported the substitution model of core organisation, the stoichiometry of interaction was determined to be 2:1 (rE2/E3BP:E3). This was further complemented by gel filtration chromatography (GFC) and small angle neutron scattering (SANS), implying the possible existence of a network of E3 ‘cross-bridges’ linking pairs of E3BP molecules across the surface of the E2 core assembly. Low resolution solution structures obtained for rE2/E3BP, bE2/E3BP and tLi19/tLi30 by small angle x-ray scattering (SAXS) and SANS revealed the presence of icosahedral cores with open pentagonal faces favouring

the substitution model of core organisation. These solution structures also indicated high structural similarity between the recombinant and native cores, as well as with the crystal structure obtained previously for the truncated bacterial E2 core. In addition, homology modelling and superimpositions of high- and low-resolution structures of the core revealed conservation of the overall pentagonal dodecahedral morphology despite evolutionary diversity. Evidence for the substitution model of core organisation was further substantiated by negative stain EM of the recombinant and bovine E2/E3BP cores.

SANS stoichiometry data indicated the binding of 10 E3 dimers per E2/E3BP core. Although this could correspond to approximately 1:1 stoichiometry between E2/E3BP:E3, subsequent radiolabelling studies suggested possible variation in core subunit composition between the native and recombinant E2/E3BP cores. Therefore, as opposed to the 48E2+12E3BP substitution model based on AUC and SAXS studies with the recombinant E2/E3BP core, rE2/E3BP cores produced in this study indicated a higher level of incorporation of E3BPs with a maximum core composition of 40E2+20E3BP. On the basis of this new finding we have proposed the 'variable E3BP substitution model', wherein the number of E3BPs within the core can range from 0 to a maximum of 20, thus resulting in variable populations of E2/E3BP cores. Despite this core variability, the highly controlled regulatory mechanisms *in vivo* may bias the core composition towards an average of 48E2+12E3BP. However, as the over-expression of the recombinant E2/E3BP core in our study is not as tightly regulated as *in vivo*, higher number of E3BPs (>12) is observed to be integrated into the core. This new level of architectural complexity and variable subunit composition in mammalian PDC core organisation is likely to have important implications for the catalytic mechanism, overall complex efficiency and tissue-specific regulation by the intrinsic PDC kinases (PDKs) in normal and disease states.

The E2 cores of the PDC family are known to be highly flexible, exhibiting inherent size variability reflective of the 'breathing' of the core. Integration of E3BP into the E2 core assembly would then be expected to have significant consequences for the structural assembly, affecting the 'breathing' and in turn the function and regulation of the complex. Unfolding studies to assess core stability via circular dichroism (CD) and tryptophan fluorescence revealed lower stability of the rE2/E3BP core as compared to cores composed exclusively of rE2

subunits, thus implying the contribution of E3BP towards core destabilisation. In addition, crosslinking studies indicated weak dimerisation of rE3BP, which may be a key factor promoting core destabilisation. The lower stability of the E2/E3BP core may be of benefit in mammals where sophisticated fine tuning is required to obtain cores with optimal catalytic and regulatory efficiencies.

SAXS solution structures of E2/E3BP cores obtained were unable to locate the exact positions of E3BP within the core. However, SANS in combination with contrast matching of selectively deuterated components as well as cryo-EM, EM tomography and single molecule studies could be used in future for determination of the exact locations of E3BP, and validating the importance of E2/E3BP core organisation and subunit composition for overall PDC function and regulation.

# Table of Contents

<b>Abstract .....</b>	<b>iii</b>
<b>Table of Contents.....</b>	<b>vi</b>
<b>List of Tables.....</b>	<b>xiv</b>
<b>List of Figures .....</b>	<b>xv</b>
<b>Acknowledgements .....</b>	<b>xviii</b>
<b>Author’s Declaration.....</b>	<b>xxi</b>
<b>Abbreviations .....</b>	<b>xxii</b>
<b>Chapter 1 Introduction to PDC .....</b>	<b>25</b>
1.1 Multi-enzyme complexes .....	25
1.2 Reaction mechanism of PDC .....	26
1.3 Insights into PDC architecture .....	27
1.3.1 Organisation of the E2/E3BP core assembly .....	27
1.3.2 The structure of PDC components.....	28
1.3.2.1 Pyruvate decarboxylase (E1) .....	28
1.3.2.2 Dihydrolipoyl acetyltransferase (E2) .....	30
1.3.2.2.1 The lipoyl domain (LD) .....	30
1.3.2.2.2 The subunit binding domain (SBD)...	32
1.3.2.2.3 The C-terminal domain (CTD) .....	33
1.3.2.2.4 The linker regions .....	33
1.3.2.3 Dihydrolipoamide dehydrogenase (E3) .....	34
1.3.2.4 E3 binding protein (E3BP) .....	36
1.3.3 Subunit composition and association of PDC .....	37
1.4 PDC regulation .....	40
1.5 PDC defects.....	43
1.5.1 Genetic defects .....	43
1.5.2 Alzheimer’s disease.....	45
1.5.3 Primary biliary cirrhosis (PBC) .....	46

1.5.4	Diabetes .....	47
1.6	Project aims .....	48

## **Chapter 2 Introduction to biophysical techniques ..... 50**

2.1	Introduction.....	50
2.2	Analytical ultracentrifugation (AUC) .....	50
2.2.1	Sedimentation velocity (SV).....	52
2.2.1.1	Theory .....	52
2.2.1.1.1	The mechanical picture of SV.....	53
2.2.1.1.2	The thermodynamic picture of SV ...	56
2.2.1.2	SV data analysis .....	61
2.2.2	Sedimentation equilibrium (SE).....	62
2.2.2.1	Theory.....	63
2.2.2.1.1	The mechanical picture of SE .....	63
2.2.2.1.2	The thermodynamical picture of SE.	65
2.2.2.2	SE data analysis .....	67
2.3	Small angle scattering .....	67
2.3.1	X-rays .....	69
2.3.1.1	Theory of SAXS .....	69
2.3.1.2	Data treatment .....	72
2.3.1.3	<i>Ab initio</i> modelling.....	76
2.3.1.4	Instrumentation.....	78
2.3.2	Neutrons .....	81
2.3.2.1	Small angle neutron scattering (SANS) .....	81
2.3.2.2	Contrast variation .....	82
2.3.2.3	Instrumentation.....	84
2.3.2.4	Data treatment .....	85
2.4	Circular dichroism (CD) .....	86
2.5	Tryptophan fluorescence .....	88
2.6	Negative stain electron microscopy (EM) .....	90

<b>Chapter 3</b>	<b>Materials and Methods .....</b>	<b>92</b>
3.1	Bacterial strains.....	92
3.1.1	Chemicals and standard materials.....	92
3.1.2	Preparation of competent cells.....	92
3.1.3	Bacterial transformation .....	94
3.1.4	Bacterial media.....	94
3.1.5	Oligonucleotide primers.....	94
3.1.6	Ethanol precipitation of DNA .....	95
3.2	Molecular Biology .....	95
3.2.1	Plasmid preparation.....	95
3.2.2	Agarose gel electrophoresis .....	95
3.2.3	DNA extraction from agarose gels .....	96
3.2.4	Polymerase chain reaction.....	96
3.2.5	Restriction digestion and plasmid ligation .....	97
3.3	Protein methods .....	97
3.3.1	Over-expression and solubility .....	97
3.3.2	Protein purification.....	98
3.3.2.1	Cell lysis.....	98
3.3.2.2	Metal chelate affinity chromatography .....	98
3.3.2.3	High-capacity HQ anion exchange chromatography .....	99
3.3.2.4	Gel filtration chromatography.....	100
3.3.2.5	PDC purification from bovine heart.....	100
3.3.2.6	Sucrose gradient centrifugation .....	101
3.3.3	Dialysis.....	102
3.3.4	Protein concentration .....	102
3.3.5	Polyacrylamide gel electrophoresis (PAGE).....	102
3.3.5.1	Sodium dodecyl sulphate PAGE (SDS-PAGE) ...	102
3.3.5.2	Native PAGE.....	103
3.3.6	Glutaraldehyde crosslinking .....	104
3.3.7	Western blotting .....	104
3.3.7.1	His-tag antibody .....	104
3.3.7.2	Other antibodies .....	105
3.3.8	Determination of protein concentration .....	105

	3.3.8.1	Protein absorbance .....	105
	3.3.8.2	Bradford assay .....	106
	3.3.8.3	Biuret method .....	107
	3.3.	Determination of Radioactivity .....	107
	3.3.9.1	[ <sup>14</sup> C] N-ethyl maleimide (NEM) labelling .....	107
	3.3.9.2	Scintillation counting.....	108
	3.3.9.3	Fluorography.....	108
3.4		Biophysical methods.....	108
	3.4.1	Buffer density and viscosity calculations.....	108
	3.4.2	Sedimentation velocity (SV) .....	109
	3.4.3	Sedimentation equilibrium (SE).....	110
	3.4.4	Small angle x-ray scattering (SAXS) .....	111
	3.4.5	Small angle neutron scattering (SANS) .....	112
	3.4.6	Negative staining electron microscopy (EM).....	113
	3.4.7	Circular dichroism (CD) .....	113
	3.4.8	Tryptophan fluorescence .....	114
3.5		Computational methods .....	114
	3.5.1	Sequence alignments .....	114
	3.5.2	<i>Ab initio</i> modelling of SAXS and SANS data .....	114
	3.5.3	Homology modelling .....	115
	3.5.4	Hydrodynamic modelling.....	115
	3.5.5	Calculation of scattering curves from crystal structures .....	116
	3.5.6	Superimposition of <i>ab initio</i> and crystal structure models ...	117
	3.5.7	Mathematical modelling of core subunit organisation .....	117
3.6		Nomenclature for stoichiometry experiments used in this study .....	119

## **Chapter 4 Cloning, over-expression and purification . 120**

4.1		Introduction.....	120
4.2		Materials and methods .....	122
	4.2.1	Cloning of C-terminal constructs: tE2 and tE3BP .....	122
	4.2.2	Protein over-expression .....	125
	4.2.3	Protein purification .....	126
4.3		Results .....	126
	4.3.1	Cloning of truncated clones tE2 and tE3BP .....	126

4.3.1.1	Cloning of tE2: tLi19 and tLi27 .....	128
4.3.1.2	Cloning of tLi30 .....	129
4.3.2	Protein over-expression and purification .....	130
4.3.2.1	Full-length recombinant E2/E3BP core (rE2/E3BP) .....	130
4.3.2.2	Full-length recombinant E2 core (rE2) .....	132
4.3.2.3	Recombinant E3 (rE3 and dE3) .....	134
4.3.2.4	Truncated recombinant E2/E3BP core (tE2/tE3BP) .....	135
4.3.2.5	Truncated recombinant E2 core (tE2).....	136
4.3.2.6	Full-length recombinant E3BP-28b.....	137
4.3.2.7	Purification of bovine E2/E3BP core (bE2/E3BP) ... .....	139
4.4	Discussion.....	140

## Chapter 5 Characterisation of the full-length cores:

<b>E2/E3BP and E2 .....</b>	<b>142</b>	
5.1	Introduction.....	142
5.2	Materials and methods .....	146
5.2.1	Protein purification .....	146
5.2.2	Sedimentation velocity.....	146
5.2.3	Sedimentation equilibrium .....	147
5.2.4	Small angle x-ray scattering.....	148
5.2.5	Small angle neutron scattering .....	148
5.2.6	<i>Ab initio</i> modelling of SAXS and SANS data .....	149
5.2.7	Hydrodynamic modelling.....	150
5.2.8	Circular dichroism.....	150
5.2.9	Tryptophan fluorescence .....	151
5.2.10	Negative stain electron microscopy.....	152
5.3	Results and modelling.....	152
5.3.1	Recombinant E2/E3BP core (rE2/E3BP) .....	152
5.3.1.1	Shape determination of rE2/E3BP core .....	152
5.3.1.1.1	SV AUC.....	152

	5.3.1.1.2	SE AUC.....	154
5.3.1.2		Solution structure of rE2/E3BP core .....	157
	5.3.1.2.1	SAXS, SANS and ab initio modelling..	
		.....	157
	5.3.1.2.2	Hydrodynamic modelling .....	164
5.3.1.3		Stability of rE2/E3BP core.....	164
5.3.1.4		Towards a high-resolution structure .....	168
	5.3.1.4.1	Negative stain EM .....	168
	5.3.1.4.2	Preliminary crystallisation trials	168
5.3.2		Bovine E2/E3BP core (bE2/E3BP).....	169
	5.3.2.1	Shape determination of bE2/E3BP .....	169
		5.3.2.1.1 SV AUC.....	169
		5.3.2.1.2 SE AUC .....	170
	5.3.2.2	<i>Ab initio</i> solution structure of bE2/E3BP....	171
	5.3.2.3	Electron microscopy of bE2/E3BP core .....	174
5.3.3		Recombinant E2 core (rE2) .....	177
	5.3.3.1	Shape determination of rE2 .....	177
		5.3.3.1.1 SV AUC .....	177
		5.3.3.1.2 SE AUC .....	178
		5.3.3.1.3 Negative stain EM .....	181
	5.3.3.2	Solution structure of rE2 core.....	182
	5.3.3.3	Stability of the rE2 core.....	183
5.4		Discussion.....	186

## Chapter 6 Characterisation of the truncated C-terminal

	<b>PDC core: tE2/tE3BP .....</b>	<b>191</b>
6.1	Introduction.....	191
6.2	Material and methods .....	193
	6.2.1 Protein purification.....	193
	6.2.2 Sedimentation velocity.....	193
	6.2.3 Sedimentation equilibrium .....	194
	6.2.4 SANS and <i>ab initio</i> modelling .....	194
	6.2.5 Homology and hydrodynamic modelling .....	194

6.2.6	Negative stain electron microscopy.....	195
6.2.7	CD and tryptophan fluorescence .....	195
6.3	Results and modelling.....	196
6.3.1	Solution shape determination of tE2/tE3BP .....	196
6.3.1.1	Sedimentation velocity analysis.....	196
6.3.2	Molecular weight determination of tE2/tE3BP .....	198
6.3.2.1	Sedimentation equilibrium analysis .....	198
6.3.3	Solution structure and modelling of tE2/tE3BP .....	200
6.3.3.1	SANS and <i>ab initio</i> modelling.....	200
6.3.3.2	Hydrodynamic modelling .....	201
6.3.3.3	Homology modelling .....	203
6.3.4	Conserved C-terminal domains: superimposition of x-ray crystal and SAXS solution cores .....	203
6.3.5	Stability of the tE2/tE3BP core .....	205
6.3.6	High-resolution structure of tE2/tE3BP: the goal.....	209
6.3.6.1	Negative stain EM of tE2/tE3BP .....	209
6.3.6.2	Crystallisation of truncated E2/E3BP core ...	209
6.4	Discussion .....	211

## **Chapter 7 Binding stoichiometry of E2/E3BP:E3 and core variation – an investigation .....214**

7.1	Introduction .....	214
7.2	Binding stoichiometry of rE2/E3BP:E3.....	216
7.2.1	Materials and methods .....	216
7.2.1.1	Protein purification .....	216
7.2.1.2	Sedimentation velocity .....	216
7.2.1.3	Gel filtration .....	217
7.2.1.4	SANS contrast variation.....	217
7.2.2	Results and analysis.....	217
7.2.2.1	E2/E3BP:E3 binding indicates a 2:1 stoichiometry.....	217
7.2.2.1.1	Gel filtration profile of E2/E3BP:E3. .....	217

	7.2.2.1.2	SDS-PAGE of E2/E3BP:E3 .....	219
	7.2.2.1.3	SV analysis of E2/E3BP:E3 .....	220
	7.2.2.2	10E3s bind to the rE2/E3BP core .....	220
7.3		Subunit composition of the rE2/E3BP core .....	225
	7.3.1	Materials and methods .....	225
	7.3.1.1	Protein purification .....	225
	7.3.1.2	[ <sup>14</sup> C]-NEM radiolabelling and counting .....	225
	7.3.1.3	Fluorography .....	226
	7.3.1.4	Mathematical modelling .....	227
	7.3.2	Results and analysis.....	227
	7.3.2.1	Distinctive subunit organisations: rE2/E3BP and bE2/E3BP cores .....	227
	7.3.2.1.1	Variable composition of rE2+rE3BP based on SANS .....	227
	7.3.2.1.2	[ <sup>14</sup> C]-NEM radiolabelling .....	228
	7.3.2.1.3	Fluorography.....	229
	7.3.2.2	E2/E3BP core composition: investigation by mathematical modelling.....	232
7.4		Discussion .....	234
<b>Chapter 8 Conclusions.....</b>			<b>240</b>
<b>Appendix.....</b>			<b>245</b>
<b>Bibliography .....</b>			<b>246</b>

## List of Tables

2.1	X-ray and neutron scattering lengths of some elements.....	82
3.1	List of common chemicals used and their sources .....	93
3.2	Plasmids used in this study .....	96
3.3	Antibiotics and their final concentrations in over-expression studies .....	98
3.4	SDS-PAGE gel and buffer recipes used in this study. ....	103
3.5	Extinction coefficients and molecular mass of proteins used in this study.....	106
3.6	Calculated densities and viscosities of buffers used in AUC.....	109
4.1	PCR cycle reactions .....	123
4.2	<i>E. coli</i> bacterial strains and conditions for protein over-expression.. .....	125
4.3	Purification buffers used in this project.....	127
5.1	Hydrodynamic parameters for rE2/E3BP derived from SV data...	154
5.2	Variable core organisation of rE2/E3BP and its effect on molecular weight .....	157
5.3	Hydrodynamic parameters of rE2 derived from SV analysis .....	179
5.4	Variable core organisation of rE2 and its effect on molecular weight .....	181
6.1	Hydrodynamic parameters of tLi19/tLi30 derived from SV analysis .. .....	198
7.1	$R_g$ of SANS rE2/E3BP:dE3 mixtures .....	223
7.2	Variable substitution core models of the rE2/E3BP core .....	228
7.3	Core stoichiometries and their theoretical [ $^{14}\text{C}$ ] radiolabelled E2:E3BP ratios .....	229
7.4	Determination of subunit composition of bovine and recombinant E2/E3BP cores via [ $^{14}\text{C}$ ]-NEM radiolabelling.....	230
7.5	Various E2/E3BP core compositions and number of trimers.....	233

# List of Figures

1.1	Cellular metabolism of 2-oxo acid dehydrogenase complexes.....	26
1.2	Multi-step reactions catalysed by PDC.....	26
1.3	E2 core of PDC.....	28
1.4	Structure of the human pyruvate decarboxylase.....	29
1.5	Modular structure of E2 and E3BP.....	31
1.6	Structure of the LDs of <i>B. stearrowthermophilus</i> and human E2.....	31
1.7	Structure of the E2-SBD of <i>B. stearrowthermophilus</i> .....	32
1.8	Crystal structure of human E3.....	35
1.9	Models of PDC core organisation.....	38
1.10	Association of E1 and E3 with E2-SBD in <i>B. stearrowthermophilus</i> ....	38
1.11	'Cross-bridge' model of human PDC.....	39
1.12	Short-term regulation of PDC activity.....	42
2.1	Optical system of the XL-A analytical ultracentrifuge.....	51
2.2	Sedimentation velocity AUC.....	53
2.3	Forces on a particle under the influence of a gravitational field ..	54
2.4	Schematic diagram of SE analytical ultracentrifugation.....	63
2.5	General features of a solution scattering curve.....	68
2.6	Schematic representation of SAXS.....	70
2.7	Scattering intensities and $p(r)$ functions of various geometrical bodies.....	75
2.8	Instrumentation layout and experimental hutch of the X33 beamline at the EMBL/DESY synchrotron.....	80
2.9	Scattering length densities of biomolecules in solvents of various $D_2O/H_2O$ ratios.....	83
2.10	Instrumentation layout of D22 at the ILL nuclear reactor source ..	85
2.11	Origin of the CD effect.....	86
2.12	Far-UV CD of various secondary structural components.....	87
2.13	Intrinsic tryptophan fluorescence.....	89
2.14	Schematic diagram of a transmission electron microscope (TEM)..... .....	91
4.1	Primer sequences for the truncated CTD constructs of rE2 and rE3BP.....	123
4.2	Cloning of truncated E2 constructs - tLi19 and tLi27.....	128

4.3	Cloning of the truncated E3BP tLi30 construct .....	129
4.4	Over-expression and solubility of rE2/E3BP at various temperatures .....	131
4.5	Purification of rE2/E3BP .....	132
4.6	Purification of rE2 .....	133
4.7	Purification of rE3 .....	134
4.8	Purification of tLi19/tLi30 core .....	136
4.9	Purification of tLi19-14b .....	137
4.10	Purification of rE3BP-28b .....	138
4.11	Purification of bE2/E3BP .....	139
5.1	Sequence alignment of the domains of eukaryotic E2 and E3BP .	145
5.2	Sedimentation velocity analysis of rE2/E3BP core .....	153
5.3	Sedimentation equilibrium analysis of rE2/E3BP .....	155
5.4	SDS-PAGE of post-SE rE2/E3BP core samples .....	156
5.5	Small angle x-ray scattering of rE2/E3BP .....	158
5.6	Small angle neutron scattering of rE2/E3BP .....	159
5.7	<i>Ab initio</i> models of the rE2/E3BP core .....	161
5.8	Superimposition of 10 <i>ab initio</i> GASBOR models of rE2/E3BP core ... .....	163
5.9	AtoB generated model of rE2/E3BP core .....	164
5.10	CD spectra of GdmCl-induced denaturation of rE2/E3BP .....	166
5.11	GdmCl-induced change in the fluorescence emission spectrum of rE2/E3BP .....	167
5.12	Negative stain EM image of rE2/E3BP .....	169
5.13	Sedimentation velocity analysis of bE2/E3BP core.....	170
5.14	Sedimentation equilibrium analysis of bE2/E3BP .....	172
5.15	Small angle x-ray scattering of bE2/E3BP core .....	173
5.16	<i>Ab initio</i> models of bE2/E3BP core .....	174
5.17	Superimposition of 10 <i>ab initio</i> GASBOR models of bE2/E3BP ....	176
5.18	Electron micrograph of bE2/E3BP .....	177
5.19	Sedimentation velocity analysis of rE2.....	178
5.20	Sedimentation equilibrium analysis of rE2 .....	180
5.21	SDS-PAGE of post SE samples of rE2 .....	181
5.22	Negative stain image of rE2.....	182
5.23	Small angle scattering curves of rE2 .....	183

5.24	CD analysis of GdmCl-induced denaturation of rE2 .....	184
5.25	GdmCl-induced change in the fluorescence emission spectrum of rE2 .....	185
6.1	Sedimentation velocity analysis of tLi19/tLi30.....	197
6.2	Sedimentation equilibrium analysis of tLi19/tLi30 .....	199
6.3	Small angle neutron scattering of the tLi19/tLi30 core.....	201
6.4	Solution structure of the tLi19/tLi30 core .....	202
6.5	AtoB generated model of tLi19/tLi30 .....	202
6.6	Homology model of the tLi19/tLi30 core .....	203
6.7	Superimposition of the crystal structure of tE2 and the <i>ab initio</i> model of tLi19/tLi30.....	204
6.8	GdmCl-induced denaturation of tLi19/tLi30 .....	206
6.9	GdmCl-induced fluorescence emission spectra of tLi19/tLi30....	208
6.10	Negative stain EM of tLi19/tLi30 .....	210
6.11	Crystallisation of tLi19/tLi30 .....	210
7.1	Gel filtration analysis of rE2/E3BP:rE3 binding stoichiometry....	218
7.2	SDS-PAGE analysis of E2/E3BP:E3 stoichiometry .....	219
7.3	Sedimentation velocity analysis of E2/E3BP:E3 stoichiometry....	220
7.4	Stoichiometry studies of dE3:rE2/E3BP using SANS.....	222
7.5	Small angle neutron scattering of deuterated E3 .....	224
7.6	Size distribution $p(r)$ analysis of dE3:rE2/E3BP at 10:1 saturation	225
7.7	Incorporation of [ <sup>14</sup> C]-NEM into bovine heart PDC .....	231
7.8	Incorporation of [ <sup>14</sup> C]-NEM into rE2/E3BP core.....	231
7.9	Chemical crosslinking of full-length rE3BP .....	233
7.10	Mathematical modelling of E2/E3BP core organisation.....	235
7.11	Various core models and their number of final states .....	236
7.12	Cross-bridge formation in human PDC.....	237

# Acknowledgements

There have been numerous people providing support and encouragement throughout my four years of chaos and madness.

Firstly, a big thank you to my supervisors Prof. Gordon Lindsay and Dr. Olwyn Byron for their endless patience, continuing support, enthusiasm, inspiration and sense of humour throughout the course of this study. I am grateful to Gordon for his patience in answering all my questions, especially the silly ones and inspiring me to be a good scientist. I have thoroughly enjoyed all our discussions and working with you, Gordon. I thank Olwyn for her inspiration, interesting tips and ideas (including the non-scientific ones) and cheerful persona that has lifted my spirits during bad patches. Thanks to Olwyn for also patiently explaining several computer programs and all those nice times at the synchrotron trips and AUC meetings during the past three years. It has been an absolute pleasure working with you, Olwyn.

I would also like to thank Dr. Sharon Kelly, University of Glasgow for carrying out the CD and fluorescence experiments and for advice on the interpretation of data. I would also like to thank Profs David Gilbert and Prof. Peter Kropholler, University of Glasgow for all their help in trying to mathematically model PDC core organisation. I am also grateful to Dr. David Bhella, University of Glasgow for all his expertise and help in carrying out EM experiments. I would also like to thank Prof. Carol Robinson and Dr. Elisabetta Boeri Erba, University of Cambridge for all their help in carrying out mass spectrometry experiments of the PDC core. Special thanks to Dr. Mattia Rocco, Istituto Nazionale per la Ricerca sul Cancro, Genoa, Italy for his invaluable help and advice on hydrodynamic modelling.

In the course of my three years study at the University of Glasgow, I am indebted to several people who have helped me in various ways to accomplish this thesis. Happily they are too numerous to individually acknowledge, but I thank them all. A special big thank you to Judith Harriman for helping me with all technical problems associated with thesis formatting. Thank you to Mads Gabrielsen for all his help in setting up protein crystallisations. The *ab initio*

models of PDC core in Chapters 5 and 6 could not have been generated without the unfailing and expert assistance of Dr. Graeme Stewart, University of Glasgow. Thanks Graeme for all the technical assistance and making the simulations on the complex computer cluster look so easy! Special thanks to Dr. Martin Boocock for sharing all his secrets and protocols of genetics, especially of running DNA acrylamide gels. Thanks Martin, for your inspiration, sense of humour and for being a good friend over the course of my stay in Glasgow. I have enjoyed all our discussions especially the intellectual ones!

All scattering data carried out in this project would not have been possible without the unfailing assistance of several people. Thanks to Dr. Gnter Grossmann at the SRS Daresbury Laboratory, UK for his invaluable help and advice, to Dr. Dimitri Svergun and Dr. Manfred Rssle from the EMBL outstation at the Deutsches Elektronen Synchrotron in Hamburg, Germany for all their help in data collection and initial analysis and Dr. Phil Callow from the Institut Laue Langevin in Grenoble, France for all his help on SANS data collection, initial treatment and invaluable advice. Thanks a lot for all those insightful scientific ideas and discussions, Phil.

Special thanks to all members of L232 and L233, past and present who made the lab a great place to work in and life very enjoyable. In particular, I am greatly indebted to Alison, Donna, Hiba, Zhenbo, Mischa, Geeta, Mridu, Lindsay and Adrienne for revealing all their technical secrets in biology and putting up with my never-ending line of questions. Thanks guys for all your support and friendship over the past four years!

I would also like to thank my friend and mentor back in India, Dr. Shekhar Mande for his continuing support and understanding over all these years. His initial guidance and undying belief in the challenging aspects of biology inspired me to pursue a PhD in biophysics. I am ever grateful to him for that. It has been a pleasure knowing you, Shekhar. I am also greatly indebted to Prof. William Cushley and Dr. Darren Monckton for all their assistance and guidance relating to the PhD program over the past four years.

Special thanks to all my friends who have provided joyful emotional support whilst I have been studying. They are too many to name but I thank them all. A

big thank you to Mridu and Sandeep for all the visits to the restaurants, all the laughs and the good times we spent together. I would also like to thank my friend Goutham for all his emotional support, understanding, friendship (over all these years) and interesting overseas phone conversations that have brightened many a rainy morning. My special thanks to Frances Nicoll for making me feel at home far away from home. I have thoroughly enjoyed all our interesting conversations alongside the gallons of tea, Frances.

I would like to thank my best friend and husband, Sreenu for being who he is, loving and understanding. His continuing emotional support, friendship, love and inspiration have helped me through all the happy and difficult times over the past several years. Thank you for cheering me up during the bad patches and for always being there, Sreenu. Successful completion of this PhD would not have been possible without your constant encouragement, enthusiasm and inspiration. Thank you for assisting me with all computer related glitches as well. You are a life saviour!

And lastly I thank my parents, my brother, Shyam and my sister, Shilpa who have been my pillar of support and strength all my life. Their constant encouragement, enthusiasm, love and understanding have got me through many rough times for which I am always indebted. Thank you for always believing in me.

## **Author's Declaration**

I hereby declare that the work presented in this thesis is my own, except where otherwise cited or acknowledged. No part of this thesis has been presented for any other degree.

Swetha Vijayakrishnan

February 2009

## Abbreviations

A <sub>260</sub>	Absorbance at 260 nm
A <sub>280</sub>	Absorbance at 280 nm
aa	Amino acid
APS	Ammonium persulphate
Amp	Ampicillin
AUC	Analytical ultracentrifugation
HQ	Anion exchange chromatography
Å	Angstrom
bp	Base pairs
BCODC	Branched-chain 2-oxoacid dehydrogenase complex
BSA	Bovine serum albumin
Cam	Chloramphenicol
CD	Circular dichroism
CoA	Coenzyme A
CV	Column volume
Cryo-EM	Cryo-electron microscopy
CTD	C-terminal domain
Cys	Cysteine
Da	Daltons
E2	Dihydrolipoamide acetyltransferase
E3	Dihydrolipoamide dehydrogenase
<i>K<sub>d</sub></i>	Dissociation constant
DTT	Dithiothreitol
E3BP	E3 binding protein
EM	Electron microscopy
V <sub>e</sub>	Elution volume

PEB	2 mM EDTA, 0.01% NaN <sub>3</sub> , 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5
PEBS100	2 mM EDTA, 0.01% NaN <sub>3</sub> , 100 mM NaCl, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5
EDTA	Ethylenediaminetetra-acetate
FAD	Flavin adenine dinucleotide
GFC	Gel filtration chromatography
$\Delta G$	Gibbs free energy
GdmCl	Guanadinium chloride
HBM	Hydrodynamic bead model
IPTG	Isopropyl- $\beta$ -D-thiogalactose
ITC	Isothermal titration calorimetry
Kan	Kanamycin
Kb	Kilo base
kDa	Kilo Dalton
LD	Lipoyl domain
LB	Luria-Bertani broth
MS	Mass spectrometry
$D_{max}$	Maximum particle diameter
MDa	Mega Dalton
MC	Metal chelate chromatography
min	Minutes
MWCO	Molecular weight cutoff
nm	Nano metre
NAD <sup>+</sup>	Nicotinamide adenine nucleotide
TEMED	NNN'N'-tetramethylethylenediamine
NMR	Nuclear magnetic resonance
ORF	Open reading frame
OD	Optical density

OGDC	2-oxoglutarate dehydrogenase complex
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PBC	Primary biliary cirrhosis
E1	Pyruvate decarboxylase
PDC	Pyruvate dehydrogenase complex
PDK	Pyruvate dehydrogenase kinase
PDP	Pyruvate dehydrogenase phosphatase
[ <sup>14</sup> C]-NEM	Radiolabelled [ <sup>14</sup> C]-N Ethyl maleimide
$R_g$	Radius of gyration
rpm	Revolutions per minute
rmsd	Root mean square deviation
SE	Sedimentation equilibrium
SV	Sedimentation velocity
SANS	Small angle neutron scattering
SAXS	Small angle x-ray scattering
SDS	Sodium dodecyl sulphate
SBD	Subunit binding domain
S	Svedberg units ( $10^{-13}$ s)
ThDP	Thiamine diphosphate
TCA	Tricarboxylic cycle
UV	Ultraviolet
R	Universal gas constant
$V_o$	Void volume
v/v	volume to volume
w/v	weight to volume

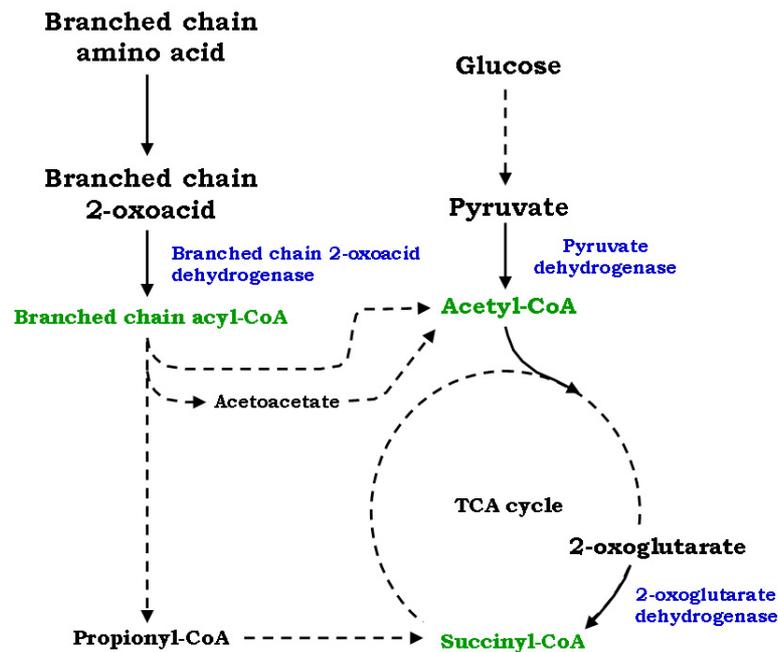
# Chapter 1

## Introduction to PDC

### 1.1 Multi-enzyme complexes

Many enzymes are thought to associate into large multienzyme multifunctional complexes. The evolution of these macromolecular assemblies allows for simultaneous substrate channelling and active site coupling of multi-step reactions, thus decreasing the extent and time of diffusion, promoting the stabilisation of intermediates and increasing the overall efficiency of the reaction cycle.

One such family of stable molecular machines is the mitochondrial 2-oxoacid dehydrogenase complexes, high  $M_r$  (4-10 million) multi-enzyme assemblies that serve as models for the study of protein-protein interactions, enzyme cooperativity and active site coupling. Principal members include the pyruvate dehydrogenase (PDC), 2-oxoglutarate dehydrogenase (OGDC) and branched-chain 2-oxoacid dehydrogenase complexes (BCODC). All three of these massive assemblies are located in the mitochondrial matrix compartment, associated with the inner membrane (Perham, 1991). In the case of PDC and BCO DC, they are composed of three enzymes E1, E2 and E3 and regulated by unique mitochondrial protein kinases and protein phosphatases. The lipoamide cofactor assists all three complexes in their catalytic function, via its 'swinging arm' mechanism visiting all the active sites during the multi-step reaction cycle (Perham, 1991; 2000). All three complexes have strategic roles in primary energy metabolism. PDC links glycolysis with the tricarboxylic acid (TCA) cycle, catalysing the irreversible decarboxylation of pyruvate to acetyl-CoA (Fig. 1.1). It serves as the primary regulator of glucose homeostasis in mammals (Sugden and Holness, 2003). Similarly OGDC is involved in the conversion of substrate 2-oxoglutarate to succinyl-CoA (Sheu and Blass, 1999), while BCO DC catalyses the catabolism of the branched chain amino acids valine, leucine and isoleucine (Patel and Harris, 1995).



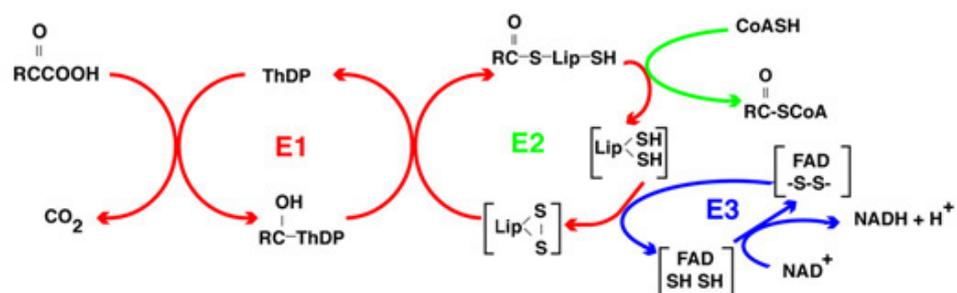
**Figure 1.1 Cellular metabolism of 2-oxo acid dehydrogenase complexes**

The three mitochondrial 2-oxoacid dehydrogenase multienzyme complexes (PDC, OGDC, BCODC) linking important pathways in energy metabolism are shown. The enzymes are represented in blue and their conversion products in green. The image was taken from Smolle (2005).

This chapter will provide insights into the reaction mechanism, structure, macromolecular organisation, regulatory role and diseases associated with eukaryotic PDC.

## 1.2 Reaction mechanism of PDC

As indicated, PDC catalyses the decarboxylation of pyruvate, resulting in the formation of acetyl-CoA via three distinct steps as shown in Fig. 1.2.



**Figure 1.2 Multi-step reactions catalysed by PDC**

Taken from Milne, et al (2002).

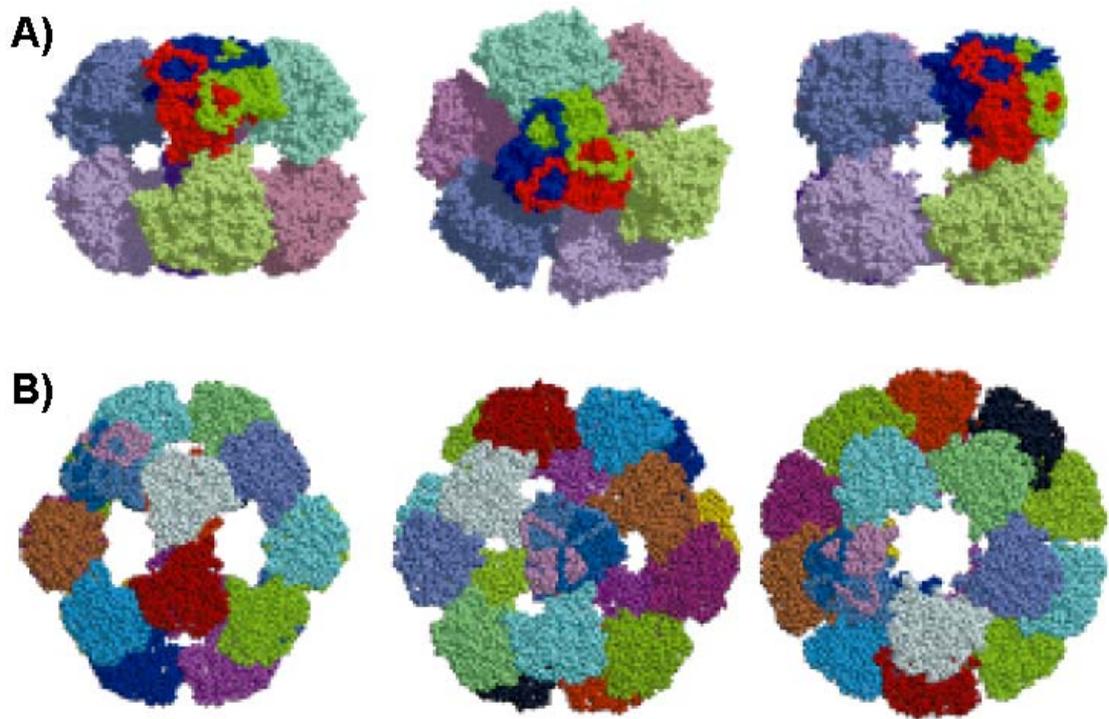
Firstly, the thiamine diphosphate (ThDP)-dependent enzyme pyruvate decarboxylase (E1) catalyses the decarboxylation of pyruvate, as well as the subsequent transfer of the resulting acetyl group from ThDP onto the lipoamide moiety of dihydrolipoamide acetyltransferase (E2) in a reductive acetylation step. E2 then proceeds to catalyse acetyl group transfer in the presence of CoASH to form acetyl-CoA. In the absence of CoASH, acetyl groups remain bound to the lipoamide cofactor. In the final step, the dihydrolipoamide moiety of E2 is reoxidised by dihydrolipoamide dehydrogenase (E3), transferring electrons to the isoalloxazine ring of FAD and subsequently onto the electron acceptor  $\text{NAD}^+$ , regenerating the lipoyl moiety to undergo another reaction cycle (Reed, 2001). The overall reaction is summarised thus



## 1.3 Insights into PDC architecture

### 1.3.1 Organisation of the E2/E3BP core assembly

PDC comprises multiple copies of three distinct enzymes E1, E2 and E3. In addition, eukaryotic PDC contains another unique subunit termed E3 binding protein (E3BP), or protein X (De Marcucci and Lindsay, 1985; Jilka, et al., 1986). The central core of PDC is made up of E2 (and E3BP in eukaryotes), to which all other constituent enzymes bind forming either a 24meric (in Gram-negative bacteria like *Escherichia coli*) or 60meric (in eukaryotes and some Gram-positive bacteria) core with octahedral (432) or icosahedral (532) symmetry, respectively (Fig. 1.3) (Oliver and Reed, 1982; Wagenknecht, et al., 1990). Thus the PDC core has the morphology of an octahedron in Gram-negative bacteria, while in eukaryotes and Gram-positive bacteria, it assembles as a pentagonal dodecahedron or icosahedron. In addition to its role in catalysis, E2 provides the structural and organisational framework upon which the assembly and function of the entire complex is dependent. The E2 core interacts with E1 and E3 across the edges and faces, respectively (Patel and Roche, 1990; Perham, 2000) and in eukaryotes with E3BP that specifically mediates E3 binding to the complex (De Marcucci and Lindsay, 1985; Jilka, et al., 1986).



**Figure 1.3 E2 core of PDC**

(A) The octahedral core of *Azotobacter vinelandii* PDC is shown with views along the two-, three- and four-fold axes of symmetry (from left to right). (B) Similarly, the icosahedral core of *Bacillus stearothermophilus* PDC is depicted with views along the two-, three- and five-fold axes of symmetry (left to right). In both cases, the three different subunits of a trimer are in different colours, readily seen on the three-fold views. In addition, every trimeric unit (8 in *A. vinelandii* and 20 in *B. stearothermophilus*) is represented in a different colour. The image was taken from Perham (2000).

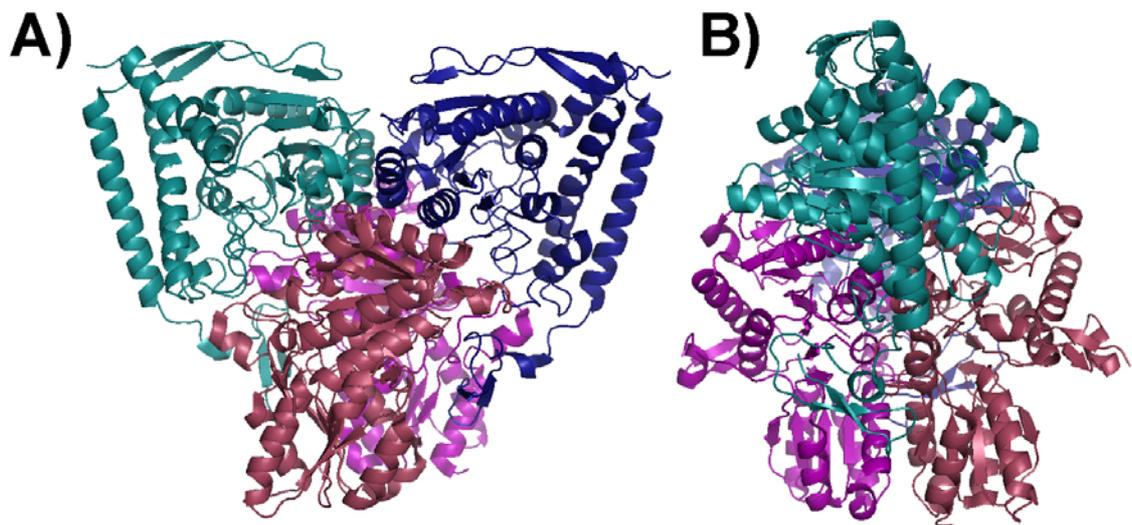
## 1.3.2 The structure of PDC components

### 1.3.2.1 Pyruvate decarboxylase (E1)

E1 is a ThDP-dependent decarboxylase that catalyses the first rate-limiting irreversible step in the catalytic cycle (Berg, et al., 1998; Cate, et al., 1980; Danson, et al., 1978). The initial decarboxylation of pyruvate is succeeded by the transfer of the ThDP-bound acetyl group onto the E2-lipoamide via a reductive acetylation reaction (Kern, et al., 1997; Pan and Jordan, 1998). E1-PDC forms an  $\alpha_2$  homodimer in Gram negative bacteria and an  $\alpha_2\beta_2$  heterotetramer in Gram positive bacteria and eukaryotes. Two active sites are located in E1, both requiring the ThDP cofactor and  $Mg^{2+}$  ions for activity. A number of E1 crystal structures have been determined in recent years including E1-PDC from *E. coli* (Arjunan, et al., 2002) and humans (Ciszak, et al., 2001;

2003). Comparison of these structures with other ThDP-dependent enzymes like transketolase (Fiedler, et al., 2002) indicates high similarity in structure and function (Hawkins, et al., 1989; Muller, et al., 1993; Schellenberger, 1998).

E1  $\alpha$  and  $\beta$  subunits are related by a crystallographic two-fold symmetry axis and arranged in a tetrahedral manner (Fig. 1.4). The active sites that bind to ThDP are located in a 20 Å funnel-shaped tunnel at the  $\alpha$ - $\beta$  subunit interface. They are readily accessible to the lysine residue of the lipoyl moiety during the catalytic reaction.



**Figure 1.4 Structure of the human pyruvate decarboxylase**

(A) Front and (B) side views of the crystal structure of the human E1 heterotetramer (PDB ID 1NI4) are shown. The two  $\alpha$  subunits (cyan and dark blue) and  $\beta$  subunits (dark pink and purple) are indicated. The figure was generated with Pymol (Delano Scientific, USA).

Hydrophobic contacts maintain the stability of the interaction between the  $\alpha$  and  $\beta$  subunits of E1. It is hypothesized that the catalytic action of E1 is induced by three critical residues (Glu59, Gly136, Val138) via movement of a pair of helices having the G $\Phi$ XXG motif, where  $\Phi$  is usually a  $\beta$ -branched chain amino acid (Ciszak et al., 2003). Biochemical, kinetic and spectral studies have shown that E1 carries out a flip-flop enzymatic action: while one active site is involved in the pyruvate decarboxylation reaction, the other is simultaneously carrying out the reductive acetylation of E2-lipoamide (Khailova and Korochkina, 1985; Khailova, et al., 1990; Kovina and Kochetov, 1998; Sergienko and Jordan, 2002; Sergienko, et al., 2000; Yi, et al., 1996). It is also proposed from the recent

crystal structure (Ciszak, et al., 2003) that both the E1  $\alpha\beta$  heterodimers undergo concerted motion resulting in a “push and pull” conformational change in the active sites. However, recently Frank and colleagues (2004) showed a novel type of active site communication in E1 mediated by the shuttling of a single proton in a 20 Å tunnel lined with acidic residues connecting the two active sites. Thus it seems that communication between the two active sites in E1 is crucial for their coordinated action during the catalytic cycle.

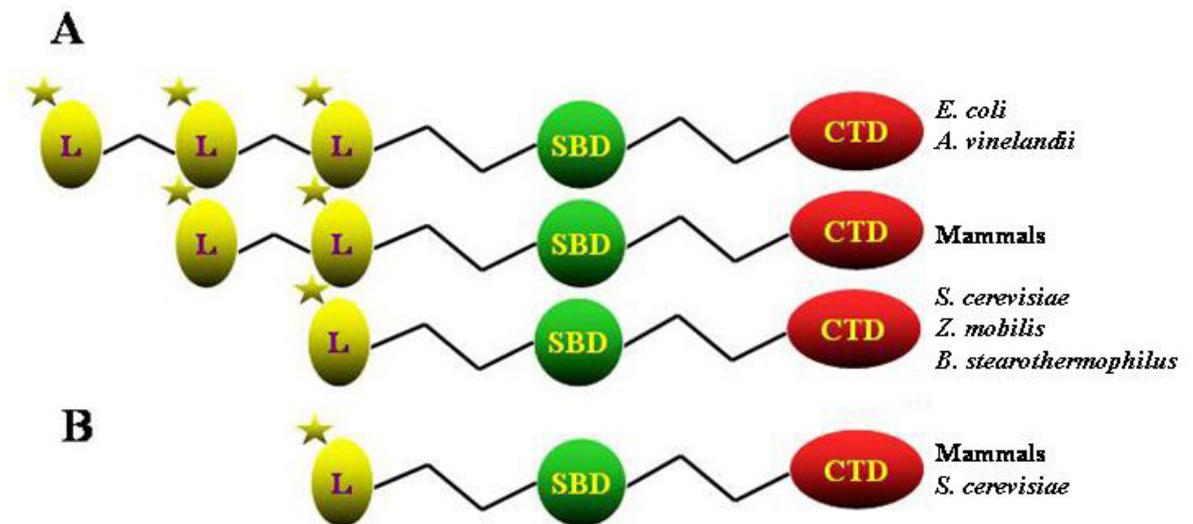
### 1.3.2.2 Dihydrolipoyl acetyltransferase (E2)

In addition to the globular structure, E2 (and E3BP) has a similar multi domain modular structure (Fig. 1.5) (Borges, et al., 1990). The N-terminal domain contains one to three 80 residue lipoyl domains (LD), the exact number varying between species (Perham, 1991; Reed and Hackert, 1990); a peripheral subunit binding domain (SBD) responsible for association with E1 and/or E3 (Hipps, et al., 1994; Lessard and Perham, 1995; Packman and Perham, 1986; Westphal, et al., 1995); and finally a C-terminal domain (CTD) (about 250 residues) responsible for the catalytic action (acetyltransferase activity) and self-association (Perham, 1991; Reed and Hackert, 1990). The domains are interconnected by long, highly flexible, extended linker regions of about 30 residues in length, thus conferring a high degree of conformational flexibility to the N terminal lipoyl ‘swinging arm’, permitting it to visit all three active sites in turn during catalysis.

#### 1.3.2.2.1 *The lipoyl domain (LD)*

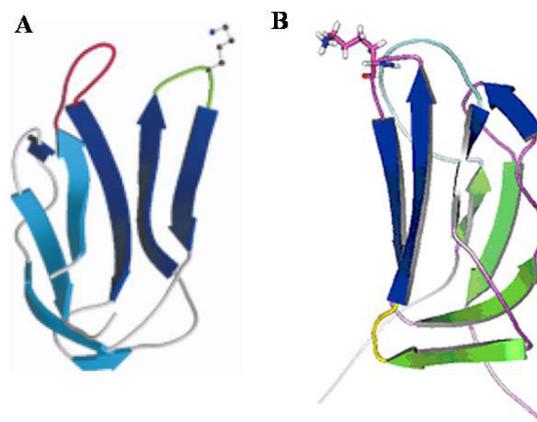
The solution structures of E2-LD from *B. stearothermophilus* PDC (Dardel, et al., 1993), *E. coli* PDC and OGDC (Green, et al., 1995; Ricaud, et al., 1996), *A. vinelandii* PDC and OGDC (Berg, et al., 1994; 1995; 1996; 1997) and human PDC (Howard, et al., 1998) have been solved using NMR spectroscopy. All structures are very similar despite low sequence identity (25% between E2-LD of *A. vinelandii* PDC and OGDC) (Berg, et al., 1997). In addition, no significant differences have been observed between the lipoylated and unlipoylated forms (Berg, et al., 1994; Dardel, et al., 1991). The LD forms a flattened  $\beta$ -barrel with

a symmetric 2-fold axis and a conserved DKA motif at the tip of a type I  $\beta$ -turn (Fig. 1.6).



**Figure 1.5 Modular structure of E2 and E3BP**

Schematic representation of the domain structure of E2 (**A**) and E3BP (**B**). The number of lipoyl domains (L) varies from one to three for different species. The lipoylation sites are indicated by a star. Subunit binding (SBD) and C-terminal domains (CTD) are also shown.



**Figure 1.6 Structure of the LDs of *B. stearrowthermophilus* and human E2**

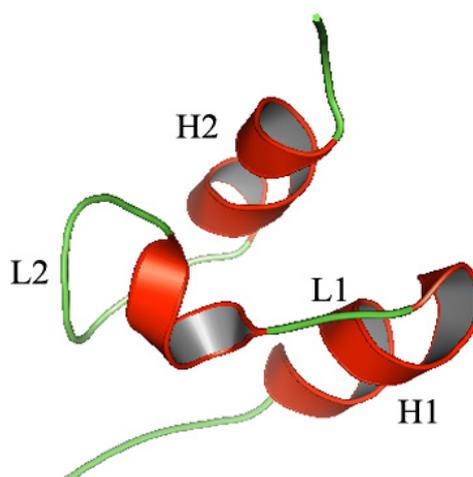
NMR structures of the (**A**) *B. stearrowthermophilus* (PDB ID 1LAB) and (**B**) non-lipoylated inner human E2 (PDB ID 1FYC) LD are shown. The two  $\beta$ -sheets of *B. stearrowthermophilus* are depicted in light blue and blue, while the human sheets of LD are shown in green and blue. The lysine residue required for lipoylation is drawn in stick representation. Surface loops involved in E1 binding are coloured cyan and yellow in the human LD domain, while it is depicted in red in the LD of *B. stearrowthermophilus*. Images were drawn with Pymol (Delano Scientific, USA).

The position of the lysine (K) residue at the tip of the turn is crucial for lipoylation; movement of K by a single residue towards the N- or C-terminus

abolishes lipoylation (Wallis and Perham, 1994). The aspartate (D) and alanine (A) of the conserved motif are necessary for the acetylation reaction but are not required for lipoylation (Jones, et al., 2000a; Jones, et al., 2000b; Wallis and Perham, 1994). Despite structural similarities between all the LDs, interactions are very specific with their cognate E1 enzymes, mediated by the variable surface loop connecting strands 1 and 2. Deletion of this loop in *B. stearothermophilus* results in the complete elimination of reductive acetylation and non recognition by E1 (Wallis, et al., 1996).

#### 1.3.2.2.2 The subunit binding domain (SBD)

One of the smallest functional domains known to date, the SBD (aa 35-50) is required for interaction with the E1 and/or E3 enzymes. In *B. stearothermophilus*, the E2-SBD is capable of binding both E1 and E3, while in mammalian PDC it specifically interacts only with E1, whereas E3 binds to the related E3BP-SBD. The SBD structures of *B. stearothermophilus* on its own (Fig. 1.7) (Kalia, et al., 1993) as well as in association with E1 (Frank, et al., 2005) or E3 (Mande, et al., 1996); and the human E3BP-SBD complexed with E3 have been determined (Ciszak, et al., 2006). The *B. stearothermophilus* and human SBD structures superimpose quite well and consist of two short, parallel  $\alpha$ -helices connected by a very short  $3^{10}$ -helix and two loops (Fig. 1.7).



**Figure 1.7 Structure of the E2-SBD of *B. stearothermophilus***

Cartoon representation of the solution structure of E2-SBD of *B. stearothermophilus* (PDB ID 2PDD). The two  $\alpha$ -helices (H1, H2) and the two loops (L1, L2) are indicated. Image was drawn with Pymol (Delano Scientific, USA).

### 1.3.2.2.3 *The C-terminal domain (CTD)*

The CTD of E2 forms the central core of PDC by its association into trimers located at each vertex. That the trimer is the morphological building block is indicative of its structural importance in the assembly of the complex. The octahedral core of *A. vinelandii* PDC (Mattevi, et al., 1993b; 1992a) consists of 8 trimers, while the icosahedral cores of *B. stearrowthermophilus* (Izard, et al., 1999; Milne, et al., 2002), yeast (Stoops, et al., 1992; 1997) and bovine PDC (Behal, et al., 1994; Zhou, et al., 2001b) are formed by the association of 20 trimers. These trimers are loosely stacked, and their dynamic movements facilitate the ‘breathing’ motion of the core (Kong, et al., 2003; Zhou, et al., 2001a). Interestingly, the inter-trimer contacts are much weaker than the intra-trimer ones (Izard, et al., 1999), providing a flexible yet well connected core, characteristics that may be crucial for optimal PDC function. High resolution cryoelectron microscopy (cryo-EM) reconstructions of human (Yu, et al., 2008), bovine (Zhou, et al., 2001b) and *S. cerevisiae* (Stoops, et al., 1992) E2 cores, as well as crystal structures of the E2 cores of *A. vinelandii* (Mattevi, et al., 1992a) and *B. stearrowthermophilus* (Izard, et al., 1999) have been determined. The CTDs possess the conserved sequence motif DHRXXDG, with the histidine and aspartate necessary for catalysis (Radford, et al., 1987). In addition, the active site is located within a long channel (30 Å) that runs across the trimer on the CTD subunit interface. Interestingly, the two substrates that take part in the transacetylation reaction enter the E2 active site from opposite directions: CoASH enters from the inside (Izard, et al., 1999) while the acetylated LD enters from the outside (Mattevi, et al., 1993c; 1993b; 1992a). Furthermore, cryo-EM and x-ray crystal structures of the E2 core indicate a hollow and highly solvated particle (Fig. 1.3), with only about 20% of the total volume being occupied by protein residues (Stoops, et al., 1992). This hollow core may be inherently flexible, a property that may be crucial to mediate substrate channelling and active site coupling.

### 1.3.2.2.4 *The linker regions*

The LDs, SBD and CTD of E2 (and E3BP) are interconnected by sequences of 30-50 amino acids, rich in alanine, proline and other charged amino acids, and are commonly referred to as ‘linker regions’. The abundance of alanine and proline

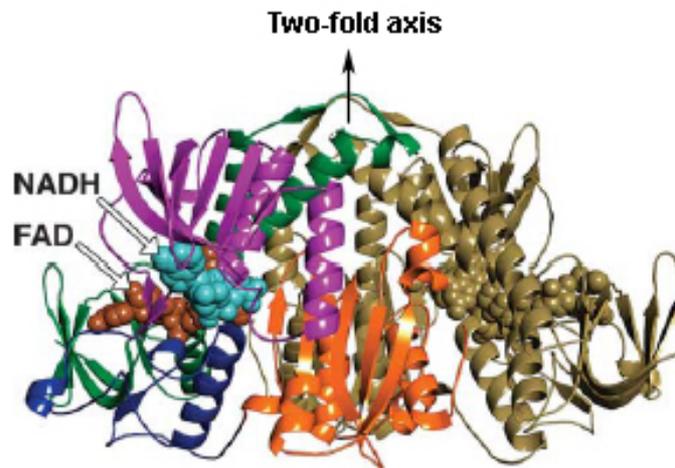
give rise to a solvent exposed, rigid and extended structure (Green, et al., 1992; Radford, et al., 1989a; Texter, et al., 1988). The linker regions possess an inherent flexibility, thereby allowing the LDs to visit all three active sites within the complex (Green, et al., 1992; Perham, et al., 1981). Mutation studies have also shown that the overall length of the linker regions can be manipulated to a considerable degree without affecting the activity of PDC (Miles, et al., 1988).

### 1.3.2.3 Dihydrolipoamide dehydrogenase (E3)

E3 is a flavoprotein belonging to a family of FAD-dependent disulphide oxidoreductases that catalyse the transfer of electrons between pyridine nucleotides and disulphides. The E3 component is the same in all the 2-oxoacid dehydrogenase complexes, although exceptions are found in prokaryotes where various E3 enzymes are encoded for the different complexes (Burns, et al., 1989; Carothers, et al., 1989). In PDC, E3 catalyses the oxidation of the dihydrolipoamide group attached to one of the LDs of E2 via a two-step ping-pong mechanism. In the first step, electrons are transferred from the reduced lipoamide group via a reactive disulphide bridge onto the FAD, forming a stable intermediate complex. Subsequently, electrons from the intermediate complex are transferred to the electron acceptor  $\text{NAD}^+$  (Wilkinson and Williams, 1981; Williams, 1965; 1992).

Three-dimensional structures of E3 from several different sources namely, *A. vinelandii* (Mattevi, et al., 1991; Schierbeek, et al., 1989), *Pseudomonas putida* (Mattevi, et al., 1992b), *Pseudomonas fluorescens* (Mattevi, et al., 1993a), *B. stearothermophilus* (Mande, et al., 1996), *S. cerevisiae* (Toyoda, et al., 1998b) and most recently from humans (Fig.1.8) (Brautigam, et al., 2006) have been solved by x-ray crystallography. Despite a low sequence identity (about 40%), the tertiary structures of prokaryotic and eukaryotic E3 are very similar. E3 forms a dimer of identical subunits, each comprising four domains: the FAD domain (aa 1-150), the NAD domain (aa 151-208), the central domain (aa 281-350) and the interface domain (aa 351-474). Each homodimer requires the non-covalent binding of one FAD molecule per monomer for correct folding and self-association (Lindsay, et al., 2000). The monomers are related by a non-crystallographic two-fold axis, while the dimers are related via a

crystallographic two-fold axis. The FAD and NAD domains (as the names suggest) are responsible for binding FAD and NAD molecules. The interface domains are involved in dimerisation, while the central domain is pivotal for binding to E2 or E3BP. The two active sites on each homodimer are located separately at the subunit interface in an 11 Å deep cleft near the FAD cofactors.



**Figure 1.8 Crystal structure of human E3**

The structure of human E3 homodimer crystallised in the presence of NADH and FAD (PDB ID 1ZMC) is depicted. The domains of one monomer are coloured: FAD-binding domain (green); NAD-binding domain (purple); central domain (blue) and the interface domain (orange). The other monomer is coloured tan. Atomic representation of the bound FAD (red) and NADH (cyan) molecules for the first monomer and in tan for the second monomer are shown. The quasi-symmetric two-fold axis is indicated. The image was taken from Brautigam, et al (2006).

All important active site residues as indicated by mutagenesis studies and structural comparison with other disulphide reductases are conserved within all E3 enzymes (Claiborn, et al., 1994). The pair of cysteines involved in disulphide bridge formation and the binding of FAD and NAD, along with key histidine and glutamine residues (His 452 and Glu 457 in human E3), play a crucial role in the efficient functioning of the E3 enzyme. Mutations of these residues in humans and other organisms significantly reduce or completely abolish E3 activity (Berry, et al., 1989; Deonarain, et al., 1989; Kim and Patel, 1992; Liu, et al., 1995; Williams, et al., 1989), confirming the importance and requirement for their high degree of conservation. The lipoamide enters the active site via a pocket on the *cis* side of the flavin plane, while the NAD associates with a pocket on the reverse side of the FAD (Mattevi, et al., 1993b).

### 1.3.2.4 E3 binding protein (E3BP)

E3BP is an integral component of PDC present only in eukaryotes. It was initially thought to be an E2 proteolytic product in bovine PDC preparations. However, elegant proteolysis and immunological studies provided the first evidence that this protein (initially named protein X) contributed to the function of PDC (De Marcucci and Lindsay, 1985; De Marcucci, et al., 1986; Hodgson, et al., 1986; Jilka, et al., 1986; Neagle, et al., 1989). The cloning and sequencing of the *Pdx1* gene encoding E3BP in *S. cerevisiae* (Behal, et al., 1989) and human (Harris, et al., 1997), indicated a high sequence similarity between proteins E2 and E3BP, with the only significant difference being the presence of one LD in E3BP in comparison to two LDs in E2 (Fig. 1.5). The N-terminal amino acid sequences of human and yeast E3BP LDs and SBDs are 46% and 38% identical to their cognate E2 proteins, respectively. Similarly, the amino acid sequence of human E3BP CTD is 50% identical to that of E2, while the yeast E3BP is substantially different (Harris, et al., 1997). The active site motif DHRXXDG in E2 harbouring the histidine crucial for acetyltransferase activity is replaced by serine in human E3BP, while it is completely absent from yeast (Harris, et al., 1997) and therefore human E3BP does not possess acetyltransferase activity.

Proteolytic studies carried out in order to ascertain the function of E3BP confirmed its primary role in high-affinity binding to E3 (Lawson, et al., 1991a; Neagle and Lindsay, 1991; Rahmatullah, et al., 1989a; Rahmatullah, et al., 1989b). These observations were confirmed by subsequent *in vitro* binding studies (McCartney, et al., 1997). Furthermore, some studies had indicated a possible role of E3BP-LD in the reductive acetylation reaction (De Marcucci, et al., 1986; Jilka, et al., 1986; Rahmatullah and Roche, 1987). Deletion of E3BP-LD resulted in the loss of PDC activity (Gopalakrishnan, et al., 1989; Lawson, et al., 1991b; Powers-Greenwood, et al., 1989; Rahmatullah, et al., 1990), but this was later shown to be primarily due to reduced E3 binding (Sanderson, et al., 1996a). However, other functions carried out by E3BP in the catalytic cycle remain undetermined. As PDC is a pivotal enzyme in the cellular metabolic pathway, it is possible that the evolution of E3BP in eukaryotes serves to increase the efficiency of the assembly by way of its specialised binding to E3. However, it is

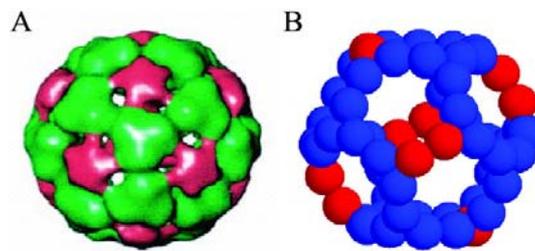
possible that there are additional structural and enzymatic roles of E3BP that are yet to be elucidated.

### **1.3.3 Subunit composition and association of PDC**

The fundamental metabolic and clinical importance of PDC has prompted the detailed examination of its structure, as it serves as a paradigm for multi-enzyme complexes in general. The entire PDC is made up of multiple copies of enzymes E1, E2, E3 and E3BP (in eukaryotes). In Gram-negative bacteria like *E. coli* PDC is thought to comprise 24 E2s forming the inner octahedral core, which in turn is non-covalently associated with 12 E1 tetramers along the edges and 6 E3 dimers across the core faces (Reed, 2001). Subsequent studies with yeast (Maeng, et al., 1994; 1996) and bovine (Sanderson, et al., 1996b) PDC established the presence of 12 E3BP molecules per 60meric core, thought to be associated with the 12 icosahedral faces. This was confirmed by cryo-EM reconstructions of *S. cerevisiae* PDC that located 12 copies of E3BP anchored inside the pentagonal openings of the E2 core along the vertices of the basic trimeric unit (Fig. 1.9A), with one E3 dimer bound per E3BP monomer (Stoops, et al., 1997). This resulted in an E2+E3BP core composition of 60+12 and was referred to as the 'addition model'. In contrast, a recent study based on analytical ultracentrifugation (AUC) and small angle x-ray scattering (SAXS) has suggested that the inner core domain of human E3BP, which has a significantly longer CTD than its yeast counterpart, is integrated into the 60-meric core of human PDC to give an E2+E3BP composition of 48+12 (E2+E3BP) subunit stoichiometry (Hiromasa, et al., 2004). This organisation of the mammalian core leads to a new 'substitution model', where 12 E2 molecules are replaced by 12 E3BP instead of being added to the 60-meric core (Fig. 1.9B).

Although the dissimilarity in sequence of yeast E3BP compared to its cognate E2 and its human counterpart has been thought to result in the difference in their core organisations, further work is required to confirm this hypothesis.

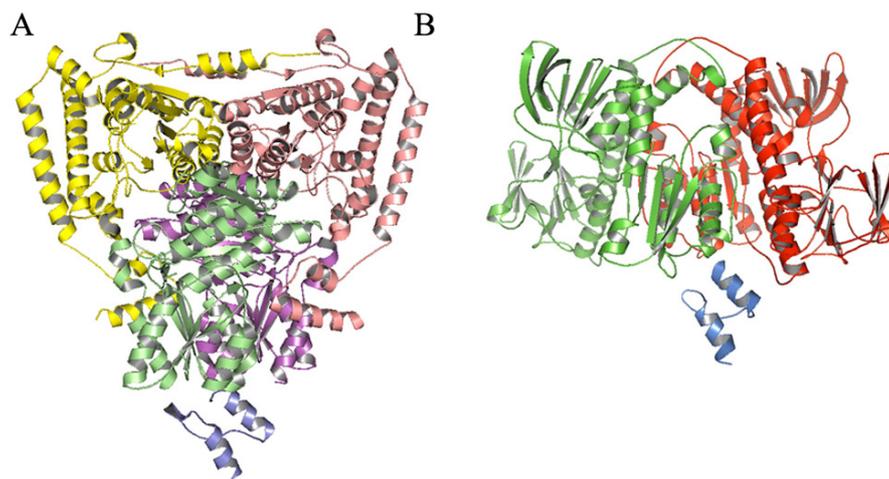
In eukaryotes the stability and overall macromolecular organisation of the complex is governed by specific and tight protein-protein interactions between E2-SBD and E1, and between E3BP-SBD and E3 respectively (Fig. 1.10). However,



**Figure 1.9 Models of PDC core organisation**

The **(A)** addition (yeast) and **(B)** substitution (human) models of subunit composition of the PDC core comprising E2 and E3BP are shown. In the addition model 12 E3BPs (pink) bind in the 12 pentagonal faces of the 60-meric E2 (green), while in the substitution model 12 E3BPs (red) are thought to replace 12 E2s (blue), resulting in core compositions of 60E2+12E3BP and 48E2+12E3BP, respectively. Images were taken from Reed (2001) and Hiromasa, et al (2004).

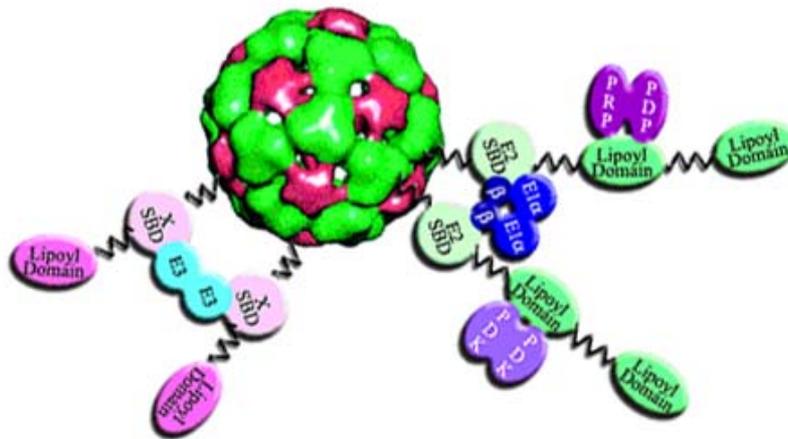
in *B. stearothersophilus*, owing to the absence of E3BP, E1 and E3 compete for binding sites on E2, and the interaction of E2 with either one prevents binding of the other (Fig. 1.10) (Mande, et al., 1996). Biochemical and structural data provide evidence of a 1:1 stoichiometry of E1 and E3 binding with the SBD of E2 in *B. stearothersophilus* PDC (Frank, et al., 2005; Hipps, et al., 1994; Jung, et al., 2002a; 2002b; Lessard, et al., 1998; Mande, et al., 1996). The E2-SBD binding sites on E1 and E3 are located across and close to the two-fold axis of symmetry, respectively (Frank, et al., 2005; Mande, et al., 1996). Occupation of both binding sites brings about steric clashes in one of the loop regions.



**Figure 1.10 Association of E1 and E3 with E2-SBD in *B. stearothersophilus***

Cartoon representations of the complexes formed by *B. stearothersophilus* **(A)** E1 ( $\alpha$ , yellow/pink;  $\beta$ , green/purple) with E2-SBD (blue) (PDB ID 1W85) and **(B)** E3 (green/red) with E2-SBD (blue) (PDB ID 1EBD). Diagrams were taken from Frank, et al. (2005) and Mande, et al. (1996).

However, the exact stoichiometry of association of E1 and E3 with the SBDs of E2 and E3BP in mammalian PDC still remains unclear. While the recent crystal structure of the E3BP-SBD:E3 complex indicates a 1:1 stoichiometry (Brautigam, et al., 2006; Cizak, et al., 2006) similar to the *B. stearothermophilus* PDC, some recent solution structural studies suggested a 2:1 binding stoichiometry (Smolle, et al., 2006). Such a stoichiometry may result in the formation of ‘cross-bridges’ that would provide a well connected and robust networked surface (Fig. 1.11), and may in turn facilitate enhanced movement and transfer of intermediates of the catalytic reactions, thus increasing the efficiency of substrate channelling and active site coupling.



**Figure 1.11** ‘Cross-bridge’ model of human PDC

Schematic representation of the formation of ‘cross-bridges’ spanning the core surface as a result of the 2:1 stoichiometries between E2:E1 and E3BP:E3 in human PDC. The CTD icosahedral core (from the cryo-EM reconstruction) extending outwards with the SBDs of E2 (green) associating with tetrameric E1 (blue), and E3BP (pink) associating with dimeric E3 (cyan) in a 2:1 stoichiometry is indicated for a small subset of the core proteins. PDC kinase (PDK, violet) and phosphatase (PDP, magenta) interact with the inner E2 LD. The image was taken from Smolle, et al. (2006).

Very interestingly, PDC-deficient patients who completely lack E3BP still possess 10-20% PDC activity as compared to controls (Ling, et al., 1998). This has been attributed to the residual low affinity binding of E2 with E3 even in the absence of E3BP (Richards, 1999), thus providing clear evidence of the importance of E3BP in the activity of PDC.

In this context, depending on the model of subunit organisation and binding stoichiometry, there are several possibilities for the actual macromolecular

organisation of mammalian PDC. Based on the addition model, PDC would comprise 60E2+12E3BP associated with 60 or 30 E1 and 12 or 6 E3 at maximal occupancy based on 1:1 or 2:1 stoichiometries of binding. Similarly, the substitution model would then comprise 48E2+12E3BP with respective binding to 60 or 30 E1 and 12 or 6 E3 subject to binding stoichiometries being 1:1 or 2:1. However further work is required to confirm and elucidate the precise molecular organisation and subunit composition of the mammalian PDC.

## 1.4 PDC regulation

As PDC is located at a significant point in the cellular metabolic pathway channelling intermediates into the TCA cycle, it controls a vital regulatory step in carbohydrate utilisation. PDC regulation is of crucial importance in balancing metabolism towards glucose or fatty acids and ketone bodies (Randle, 1986). Thus, it is down-regulated during starvation or diabetes in order to preserve carbohydrate reserves for various tissues that are dependent primarily upon glucose for their energy metabolism, while it is up-regulated after a full meal to maximise energy production and fatty acid biosynthesis. Control of PDC plays an important role in dictating the fuel used by various tissues of the body in different nutritional and hormonal states. Its regulatory mechanisms are sensitive and yet efficient enough to accommodate the complexities of tissue-specific metabolic requirements. PDC is regulated by two major mechanisms: the first being through metabolite inhibition or the end-product inhibition, while the second involves covalent modification of the complex by means of a phosphorylation/dephosphorylation mechanism.

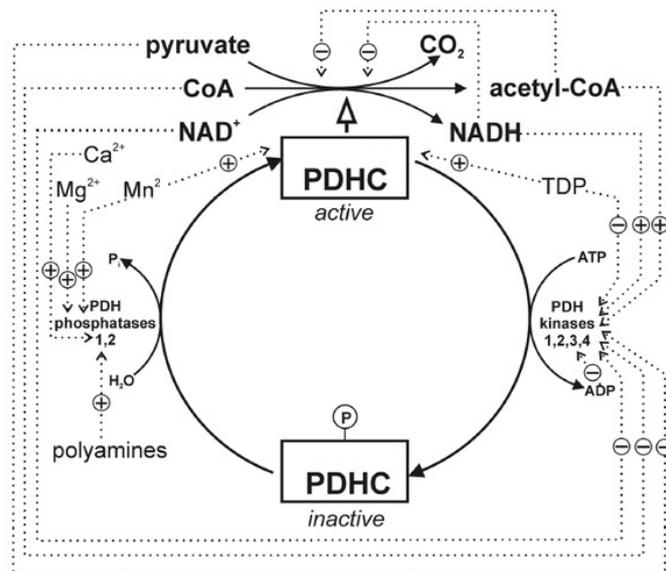
Mediation of the activity state of PDC via phosphorylation/dephosphorylation and the factors associated with it are described as 'acute or short term' regulation (Fig. 1.12). This control is exercised by a specific PDC protein kinase (PDK; EC 2.7.1.99) tightly bound to the complex and a specific loosely bound PDC phosphatase (PDP; EC 3.1.3.43). Both PDK and PDP interact with the rate-limiting E1 component of PDC (Harris, et al., 2002), phosphorylating any one of the three different serine residues (Ser264, Ser271, Ser203) of human E1 $\alpha$ , resulting in enzyme inactivation (Bao, et al., 2004a; Kolobova, et al., 2001). The

dephosphorylation by PDP in the presence of  $Mg^{2+}$  and  $Ca^{2+}$  ions results in PDC reactivation (Linn, et al., 1969; Wieland, 1983).

Four PDK isoforms (1-4) are expressed in various mammalian tissues. PDK1 and 4 are highly expressed in the heart, liver and muscle, while PDK3 is prevalent in kidney, brain and testes and PDK2 is present in most tissues (Bowker-Kinley, et al., 1998; Gudi, et al., 1995; Sugden and Holness, 2003). These isoforms can form both homo- and heterodimers (Boulatnikov and Popov, 2003) and display different kinetic properties and specific activities (Bowker-Kinley, et al., 1998; Gudi, et al., 1995; Popov, et al., 1994; Rowles, et al., 1996). Although these kinases phosphorylate serine residues, interestingly they are markedly different from other serine/threonine protein kinases and do not exhibit any sequence similarity. However, they resemble the histidine protein kinases in their sequence (Popov, et al., 1992) and three-dimensional structures (Machius, et al., 2001; Popov, et al., 1993; Steussy, et al., 2001), despite differences in their catalytic mechanisms (Harris, et al., 2002; Steussy, et al., 2001; Tuganova, et al., 2001). PDK1 phosphorylates all three Ser residues on E1 $\alpha$ , while the other three isoforms can act only on sites 1 and 2. Moreover, all four kinases phosphorylate the three serine sites with different efficiencies (Korotchkina and Patel, 2001). Phosphorylation at a single site is sufficient for complete inactivation and is independent of the status of the two remaining sites (Kolobova, et al., 2001; Korotchkina and Patel, 1995; Korotchkina and Patel, 2001; Patel and Korotchkina, 2001; Yeaman, et al., 1978). PDKs1-3 interact with the E2 inner LD while PDK4 interacts mostly with the E3BP-LD (Roche, et al., 2003). The interaction with the LD brings about a conformational change in the PDK, thus causing an increase in its activity. Furthermore, NADH is thought to stimulate kinase activity by changing the reduced/acetylated state of the E2-LDs (Bao, et al., 2004a; Bao, et al., 2004b; Patel and Korotchkina, 2001; Roche, et al., 2003). Only 1-2 molecules of PDK are present per PDC complex (Yeaman, 1989) and this limited number is sufficient to completely inactivate PDC. Previous binding experiments have also suggested the formation of transient cross-links between adjacent pairs of E2, that consequently result in a 'hand-over-hand' mechanism of PDK migration on the surface of the complex (Liu, et al., 1995).

Dephosphorylation reactivates PDC, and is catalysed by two PDP heterodimeric isoforms (PDP1 and 2) each comprising a catalytic and regulatory subunit. Analogous to PDK, PDP expression is tissue-specific. PDP1 is expressed in muscle while PDP2 is predominantly present in liver and adipose tissue (Huang, et al., 1998). PDP1 the dominant isoform, requires  $Mg^{2+}$  and is stimulated by micromolar concentrations of  $Ca^{2+}$  binding to E2-LD (Damuni and Reed, 1987; Denton, et al., 1972; Linn, et al., 1969; Pettit, et al., 1972; Reed, et al., 1996); this interaction is further enhanced by the presence of lipoamide (Chen, et al., 1996; Pettit, et al., 1972).

Pyruvate,  $NADH/NAD^+$  and acetyl-CoA/CoA ratios act as key effectors in feedback inhibition of PDC by influencing the oxidation reduction reaction and acetylation state of the lipoamide cofactors inducing 'short term' regulation of the complex. While pyruvate inhibits PDK and acts as a positive inhibitor,  $NADH$  and acetyl-CoA activate PDK and thereby act as negative effectors. This 'short-term' feedback control mechanism is tightly regulated under normal body requirements as is shown schematically in Fig.1.12.



**Figure 1.12 Short-term regulation of PDC activity**

The various activators and inhibitors of PDC (PDHC) and its regulatory enzymes are shown. Adapted from Strumilo (2005). While thiamine diphosphate (TDP) acts as a key allosteric modulator, inducing conformational and functional changes in PDC, the divalent ions  $Mg^{2+}$ ,  $Ca^{2+}$  and  $Mn^{2+}$  significantly increase the effect of TDP towards this modulation.

Long-term regulation of PDC depends on hormonal or nutritional changes and involves the participation of transcriptional and translational mechanisms (Patel and Harris, 1995; Patel and Korotchkina, 2003). A major determinant of PDC activity under various physiological conditions is the amount of PDK. For example, starvation increases PDK2 expression in the liver and kidney only at the level of transcription, while dominant regulation of PDK4 expression in these tissues is at the level of translation (Harris, et al., 2001). In addition, PDK2 and PDK4 expression increases in many tissues during starvation, diabetes and hyperthyroidism. While insulin exposure in diabetic rats suppresses PDK2 expression, it is less effective in decreasing PDK4 expression (in the presence of high levels of fatty acids), suggesting that PDK2 activity may account for short-term inhibition of PDC, while PDK4 activity may correspond to long-term inhibition of the complex, leading to several disease states (Harris, et al., 2001).

## **1.5 PDC defects**

### **1.5.1 Genetic defects**

PDC is involved in several pathological conditions. Aerobic glucose oxidation is an essential requirement for the brain, which is heavily dependent on PDC for energy production. Deficiency of PDC owing to defects in one or more of its constituent enzymes results in accumulation of lactate and causes primary lactic acidosis in infants and young children. The presentation of PDC deficiency is metabolic and neurological, as the muscle and brain tissues appear to be the primary affected sites. While the metabolic form manifests as lactic acidosis, the neurological manifestations include hypotonia, weakness, ataxia, spasticity, cerebellar degeneration, seizure, lethargy, blindness, mental retardation, Leigh syndrome and even death (Brown, et al., 1989a; 1994; De Vivo, 1998; Di Mauro and De Vivo, 1996; Robinson, 1995; Robinson, et al., 1987). However, mild cases where patients survive into adulthood are constantly being found, thus expanding the clinical spectrum of PDC deficiency disease (Brown, et al., 2002b; Head, et al., 2005; Seyda, et al., 2000). Current treatments include prescription of a carbohydrate-restricted high-fat ketogenic diet (Wexler, et al., 1997; Wijburg, et al., 1992), supplementation with thiamine/vitamin B1 or cofactors like lipoic acid or ThDP (Byrd, et al., 1989; Naito, et al., 1998; Naito, et al.,

2002), and the administration of the PDK inhibitor dichloroacetate (DCA) that reactivates PDC (Morten, et al., 1999). However, all these treatments have met with only moderate success.

In spite of the genetic and biochemical complexity, the great majority of cases of PDC deficiency are due to mutations in the gene encoding the E1 $\alpha$  subunit (Robinson, 1995). More than 80 mutations have been identified so far, which have been classified into three categories: missense, nonsense and insertion/deletion mutations (Fouque, et al., 1998; Lissens, et al., 2000). The severity of the disease depends on the nature of the mutation. *PDHA1* is located on the short arm of the X chromosome (Brown, et al., 1989b; Dahl, et al., 1992; Szabo, et al., 1990). Owing to X chromosome inactivation of the normal E1 $\alpha$  gene, the expression pattern is mosaic in women. In contrast, males are hemizygous and likely to be more severely affected by disruptive mutations (Lissens, et al., 2000). The *PDHA2* gene, exclusively expressed in the testes and encoding for E1 $\alpha$ , has also been identified, and no mutations have been observed to date (Dahl, 1995; 1990). Only two missense mutations in the *PDHB* gene encoding E1 $\beta$  have been identified so far (Brown, et al., 2004).

While the majority of PDC deficient patients harbour mutations in E1 $\alpha$ , a few have mutations in the E2, E3BP or E3 proteins of PDC. Only two mutations have been reported in E2 to date: a missense mutation of phenylalanine (mutated to leucine) in the E2 active site, and the removal of a glutamic acid residue from the outer LD owing to a 3-bp 'in frame' deletion within the *DLAT* gene (Head, et al., 2005). However, a third E2 mutation of a valine (Val455) residue in the CTD, close to the active site has been recently investigated in our laboratory (Singh, 2008). E3BP deficiency is an extremely rare inborn error of metabolism resulting from the premature termination of the *PDX1* gene product encoding E3BP, with only 12 patients being identified so far (Aral, et al., 1997; Brown, et al., 2002b; De Meirleir, et al., 1998; Dey, et al., 2002; 2003; Geoffroy, et al., 1996; Ling, et al., 1998; Marsac, et al., 1993; Ramadan, et al., 2004). Premature termination of E3BP can be caused by deletions in the mitochondrial import sequence (Ling, et al., 1998) or coding regions (Aral, et al., 1997) or due to exon skipping (Aral, et al., 1997). Interestingly, patients with E3BP deficiency show significant PDC residual activity (10-20 % of controls) (Marsac, et al., 1993), suggesting that

mammalian E2 has retained a residual capacity to bind E3 even in the complete absence of E3BP. This was further confirmed by *in vitro* binding studies (McCartney, et al., 1997). The residual activity is thought to be the major factor responsible for prolonged survival, a feature exhibited by most E3BP deficient patients. Some cases of E3BP deficiency exhibit a milder form of the disease (Head, et al., 2005).

E3 deficiency accounts for a number of well characterized cases with most mutations present in the FAD, central and interface domains, affecting FAD binding, catalysis and dimerisation (Odievre, et al., 2005). While over 20 cases having been identified, the majority of them are found among Ashkenazi Jews, exhibiting a single substitution of Gly194 with cysteine in the NAD domain (Hong, et al., 2003). As E3 is common to PDC, OGDC and BCODC, its impairment results in deleterious consequences with high levels of plasma and urine lactate being produced (Hong, et al., 1997; Shany, et al., 1999).

### **1.5.2 Alzheimer's disease**

Alzheimer's disease (AD) is the most common age related neurodegenerative disease. The main pathway for glucose oxidation is the TCA cycle, which is initiated by acetyl-CoA generated by PDC. Therefore PDC activity is critical for providing energy to the brain. Decreased activity of metabolic enzymes like PDC, OGDC and isocitrate dehydrogenase have been associated with the development of AD (Bubber, et al., 2005; Hoyer, 2004; Sheu, et al., 1994). Diminished metabolism always accompanies clinical AD. However, reduction in glucose metabolism can precede overt clinical symptoms by decades (Reiman, et al., 1996; Small, et al., 1995). Slight reductions in brain metabolism impair judgement, memory, orientation and other brain functions within seconds. Severe and prolonged impairments of brain glucose oxidation lead to more severe brain damage including permanent dementia (Plum and Posner, 1980). However, treatment with glucose and insulin improves memory in AD patients transiently (Craft, et al., 2000). Recent experiments suggest that abnormal glucose metabolism is critical to the pathophysiology of AD (Blass and Gibson, 1999). Even slight reductions in brain metabolism reduce brain function, resulting in an increased production of amyloid beta peptide and

hyperphosphorylation of tau considered to be the hallmarks of AD (Gabuzda, et al., 1994; Planel, et al., 2004). The molecular mechanism behind the decrease of PDC activity in AD is still unclear but oxidative stress is suggested to be one of the possibilities (Blass and Gibson, 1999; Sims, et al., 2000), where the brain is unable to handle the increased production of reactive oxygen species (Shi, et al., 2005). Defects in PDC may also impair the production of acetylcholine and other neurotransmitters derived from the TCA cycle including glutamate and GABA (Blass and Gibson, 1991; Gibson and Blass, 1976; Klivenyi, et al., 2004; Shoffner, 1997).

### **1.5.3 Primary biliary cirrhosis (PBC)**

Primary biliary cirrhosis (PBC) is a chronic progressive cholestatic autoimmune liver disease of unknown etiology. It is viewed both as a model and paradox for autoimmune diseases (Gershwin and Mackay, 1991). PBC is primarily a disease of middle-aged women, with most cases occurring between 40-60 years (Iwayama, et al., 1992; Mackay, et al., 2000). Histopathologically, PBC is characterized by chronic progressive destruction of the biliary epithelial cells lining the small intrahepatic bile ducts with portal inflammation, leading initially to fibrosis and later to cirrhosis (Nakanuma and Ohta, 1979; Nakanuma, et al., 1995). The diagnostic determinant of PBC is the presence of a high level of anti-mitochondrial antibodies (AMA), long before clinical signs or symptoms appear (Gershwin, et al., 1998).

A major advancement towards defining the pathogenesis of PBC has been the precise identification of the mitochondrial proteins that are targeted by the anti-mitochondrial antibodies. A puzzling feature of PBC is that the autoimmune attack is predominantly organ specific (in the liver), but the mitochondrial autoantigen is not tissue specific. Although the AMA targets the E2 component of PDC, OGDC and BCODC, E3BP and in some cases E1 $\alpha$ -PDC, its primary reactivity (> 95%) is directed against E2-PDC (Coppel, et al., 1988; Van de Water, et al., 1988a; Yeaman, et al., 1988). In PDC, immunogenicity is brought about by the E2 LDs, with the inner LD being the predominant epitope and the outer LD being a relatively weak one (Fussey, et al., 1990; Surh, et al., 1990; Van de Water, et al., 1988b). Several studies have shown the importance of bound lipoic acid in

AMA recognition (Fussey, et al., 1990; Leung, et al., 1990; Quinn, et al., 1993). An additional target of some AMA is the E1 $\alpha$  subunit of PDC, with its autoepitope located at the phosphorylation and thiamine pyrophosphate binding site (Iwayama, et al., 1991). Cross-reactivities of sera have also been detected for E2 from OGDC and BCODC (Fussey, et al., 1988; 1991), as well as E3BP-PDC (Surh, et al., 1989; Yeaman, et al., 1988) and both E1-PDC subunits (Fussey, et al., 1989). Additionally, aberrant expression of E2 and E3BP antigens on the membrane of biliary epithelial cells has been observed in PBC affected individuals (Joplin and Gershwin, 1997; Joplin, et al., 1997). The break-down of tolerance and mechanisms of tissue injury and disease caused by these AMA in PBC is unclear and remain to be defined.

### **1.5.4 Diabetes**

Diabetes is a consequence of the imbalance between glucose and lipid metabolism. PDC regulates fuel use and synthesis. In the well-fed state, liver PDC is active promoting the synthesis of fatty acids (FA), while during starvation, its activity is significantly reduced to conserve the 3-carbon compounds for gluconeogenesis. However, when there is an excessive production of glucose in the liver via gluconeogenesis, there is a highly significant increase in the blood glucose levels (in excess of 10 mM) (Boden, 2001; Randle, et al., 1994). Increasing the activity of PDC in the peripheral tissues is an attractive method for excessive glucose disposal. Studies in hepatoma cells showed elevated expression of PDK2 and PDK4 (upregulated by glucocorticoids and the peroxisome proliferator-activated receptor  $\alpha$  ligand (PPAR $\alpha$ ) WY-14,643) during starvation and diabetes, which can be down-regulated by insulin to a certain extent (Huang, et al., 2002). While insulin reduces the glucocorticoid up-regulated expression of PDK2 and PDK4, it is not effective in reducing the effects of the PPAR $\alpha$  ligand WY-14,643, that acts as a mimic of FA. The diminished effect of insulin, characteristic of diabetes (Type II) is partly responsible for the failure of the liver to reduce gluconeogenesis, thus leading to hyperglycemia. Hence, diabetes is referred to as a condition of 'starvation amidst plenty', where the PDKs remain in the starvation mode in spite of the supplementation of nutrients by insulin. As PDK2 and PDK4 modulate the energy flux in the liver, they are attractive therapeutic drug targets for diabetes, as inhibition would

increase glucose uptake in both liver and peripheral tissues in the case of PDK2, and in peripheral tissues only in the case of PDK4 (Mayers, et al., 2005). Some PDK2 inhibitors developed recently lowered glucose levels and increased PDC activity in obese Zucker rats which are used as model animals for type II diabetes (Mayers, et al., 2003; Morrell, et al., 2003). The development of a sound and efficient PDK inhibitor for all PDK1-4 isoforms poses a challenging problem in the search for a therapy for type-II diabetes, as these PDKs possess different properties and structures.

## 1.6 Project aims

An abundance of structural, biochemical and biophysical information has been gathered over the years for the various components/enzymes of prokaryotic (*B. stearothermophilus* and *A. vinelandii*) and eukaryotic (*S. cerevisiae*, bovine and human) PDCs and serves as a useful background for this research project. Although several of the recently solved mammalian structures: namely human E1 (Ciszak, et al., 2001; 2003), inner E2 LD (Howard, et al., 1998), E3 (Brautigam, et al., 2005) and E3 complexed with the E3BP-SBD (Brautigam, et al., 2006; Ciszak, et al., 2006) provide great insights into the binding and functioning of the individual enzymes, little is known of the mammalian PDC core that serves as the essential framework for the association and efficient functioning of the complex. In the light of the importance of PDC in cellular metabolism, elucidating the stoichiometry of association of the various components and the macromolecular organisation of the core is imperative for understanding its regulatory and metabolic functions in detail and the basis of its impairment in disease states. Previous microcalorimetry experiments (Brown, 2002a), AUC, SAXS and isothermal titration calorimetric (ITC) experiments (Brown, 2002a; Smolle, et al., 2006) have indicated a 2:1 binding stoichiometry between recombinant proteins E3BP:E3 and E2:E1 of human PDC. This implies the formation of so called 'cross-bridges' spanning the core surface, forming a well connected network of subunits that could potentially facilitate faster movement of the catalytic intermediates and regulatory molecules PDK and PDP. However these data were obtained from experiments conducted on monomeric E2 and E3BP constructs that possessed only the SBD or LD domains, and not the CTD. Therefore, it is of crucial significance and importance to test these binding

stoichiometries for fully-assembled core (using full-length E2 and E3BP) that may pose greater geometric constraints, and test the feasibility of a 2:1 (E3BP:E3 and E2:E1) binding stoichiometry in a more physiologically relevant system.

Another important question to be answered in the field of human PDC research is what is the precise molecular organisation of the core that provides the basic structural and mechanistic framework for its underpinning optimal functioning of the entire complex? Two E2+E3BP models have been proposed: a 60+12 additive model based on cryo-EM analysis and a 48+12 substitution model based on AUC and SAXS studies.

This project focussed on the characterisation and determination of the solution structures of recombinant human E2 and E2/E3BP cores (Chapter 5) and CTD truncated E2/E3BP core (Chapter 6) by means of biophysical techniques including AUC, SAXS, SANS and negative stain EM. The first ever solution structures of the full and truncated E2/E3BP cores were obtained using SAXS and SANS in combination with molecular modelling approaches. This project also includes investigation and confirmation of the stoichiometry of association (1:1 or 2:1) of recombinant human E2/E3BP core:E3 by a range of techniques, including AUC, GFC and SANS (Chapter 7). In addition, validation of the exact subunit organisation of the E2/E3BP core was carried out using various approaches such as SANS, [ $^{14}\text{C}$ ] radiolabelling and mathematical modelling (Chapter 7). Finally, investigation of core stability (E2 and the full and truncated E2/E3BP cores) using circular dichroism (CD) and tryptophan fluorescence was also conducted and the data are presented in the relevant chapters.

The design of new CTD truncated constructs involving standard molecular cloning techniques is described in Chapter 4. All biophysical methods (Chapter 2) require high yields of pure protein. This entailed large scale purification employing standard chromatographic techniques as described in Chapter 4.

## Chapter 2

# Introduction to biophysical techniques

## 2.1 Introduction

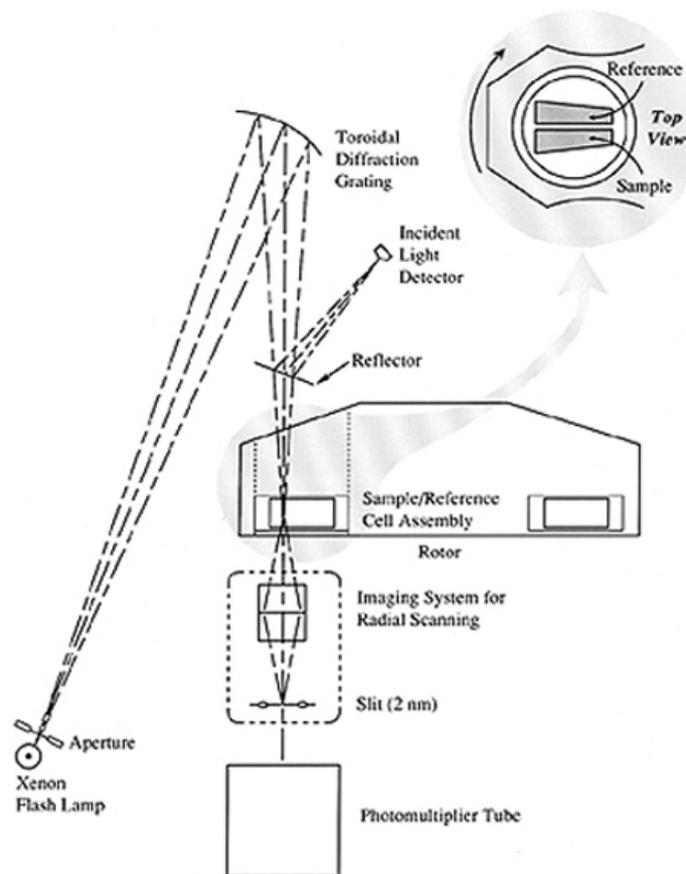
This chapter summarises the biophysical techniques employed in this detailed investigation of subunit and structural variations in the human E2/E3BP core of PDC: analytical ultracentrifugation (AUC), small angle x-ray/neutron scattering (SAXS/SANS), circular dichroism (CD), tryptophan fluorescence and negative stain electron microscopy (EM).

## 2.2 Analytical ultracentrifugation (AUC)

Constructed by Theodor Svedberg in the 1920s, the first ultracentrifuge was extensively used in the field of polymer research during the 1940s and 1950s, but was superseded by simpler techniques like chromatography and gel electrophoresis in the 1970s. However, the development of modern, commercial computer-based AUCs in the 1990s by Beckman Instruments (Palo Alto, USA) in the form of the Optima XL-A and XL-I triggered a new era in solution interaction analysis.

Ultracentrifugation involves the spinning of macromolecules at very high speeds, typically 3,000 to 60,000 rpm, resulting in the movement of particles through the solvent in the direction of the centrifugal force and sedimentation as a function of time, dependent on the particle's mass, size and shape. Sedimentation results in depletion of the solute from the meniscus with the formation of a distinct solute/solvent boundary. The movement of this boundary can be observed by monitoring the solute concentration using absorbance (Fig. 2.1) (Giebeler, 1992; Hanlon, et al., 1962; Schachman, et al., 1962), interference (Laue, 1994; Schachman, 1959; Yphantis, et al., 1994), schlieren (Svedberg and Pedersen, 1940), or fluorescence (Schmidt and Riesner, 1992)

optics. Although both absorbance and interference optics were used for the AUC experiments in this study, only interference data are presented in this thesis.



**Figure 2.1 Optical system of the XL-A analytical ultracentrifuge**

Schematic representation of the absorbance optical system in the XL-A analytical ultracentrifuge. Taken from Ralston (1993).

AUC has several advantages over complementary biophysical methods. As molecules are characterised in solution without any isotope labelling, chemical modification or standards for comparison, this more closely matches the ionic and hydrated environment within a cell than most other biophysical techniques. In addition to being non-destructive, rapid and simple, the various optical systems used for data acquisition and the different sample cell path lengths permit analysis of the solution behaviour of any macromolecule over a large range of sample concentrations. Moreover, by adjusting the rotor speed, AUC can be used to analyse macromolecules over a large molecular weight range (Lebowitz, et al., 2002).

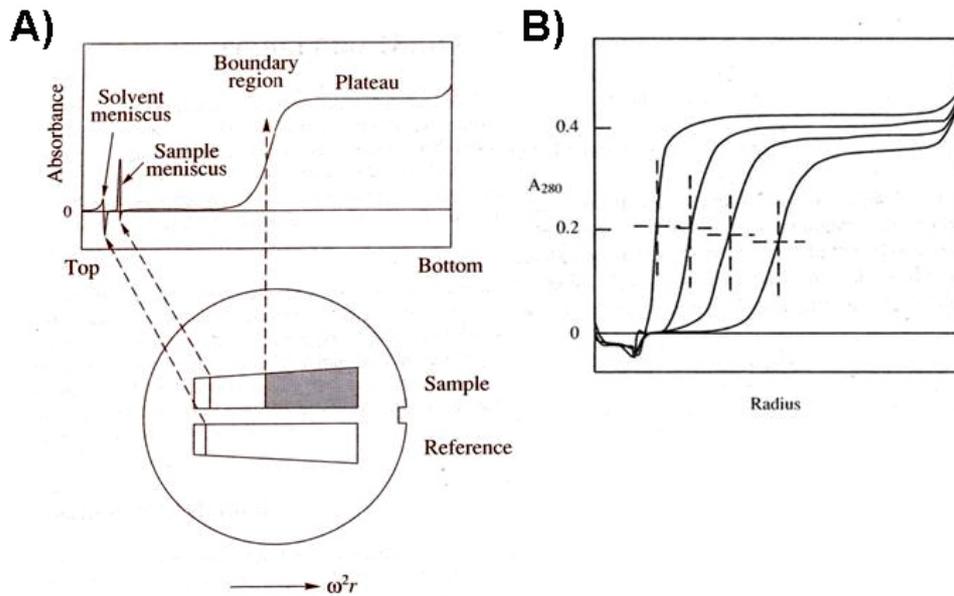
AUC provides two complementary views of macromolecule behaviour in solution, namely sedimentation velocity (SV) and sedimentation equilibrium (SE). SV provides first-principle hydrodynamic information about the size, shape and heterogeneity of the molecule. On the other hand first-principle thermodynamic information on the molar mass, association constants, stoichiometries of interaction and non-ideality are obtained from SE data. In this thesis, AUC was used to determine the homogeneity, sedimentation coefficient, apparent molecular weight and stoichiometry of constituent enzymes of PDC, in particular the oligomeric E2/E3BP core.

### **2.2.1 Sedimentation velocity (SV)**

In a SV study the solute and solvent are loaded into separate channels of a two-channel centrepiece (Fig. 2.2) and subjected to centrifugation at high rotor speeds (typically 20,000-60,000 rpm), with the exact rotor speed chosen dependent on the size of the macromolecule. At the beginning of the experiment the macromolecule is distributed homogeneously throughout the radial range recorded. However, during the run the high centrifugal force depletes most particles from the meniscus, forming a boundary that moves with time until all the particles form a pellet towards the bottom of the cell (Fig. 2.2A). Macromolecules are separated as the result of their different rates of migration through the centrifugal field, and the concentration distribution across the cell at various times during sedimentation is monitored by a set of absorbance/interference boundary profiles (Fig. 2.2B). The rate of movement of this boundary gives the sedimentation coefficient 's'. This parameter is dependent both on the mass of the sedimenting species and the frictional coefficient, which in turn is a measure of its effective size, shape and hydrodynamic hydration.

#### **2.2.1.1 Theory**

The theory of sedimentation velocity can be derived in two ways, (1) mechanical and (2) thermodynamical (Tanford, 1961). The latter involves a rigorous treatment of components in a sedimenting system even in extreme conditions of non-ideality and polydispersity. An outline of the theory is summarized below.



**Figure 2.2 Sedimentation velocity AUC**

(A) Schematic diagram of a double-sector centrifuge cell. The sample and reference solvent are placed in different sectors. (B) The movement of the sedimenting boundary detected via interference or absorbance optics is indicated. Taken from [www-bioc.rice.edu/bios576/AU/AU\\_Page.html](http://www-bioc.rice.edu/bios576/AU/AU_Page.html)

### 2.2.1.1.1 The mechanical picture of SV

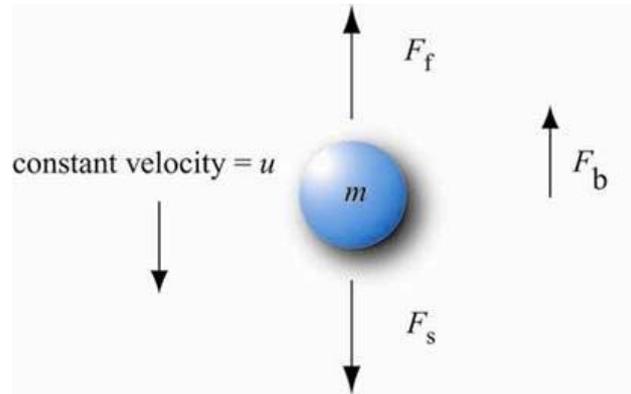
The theory given by Tanford (1961) is summarised here. A particle suspended in solvent and subjected to a gravitational field experiences several forces as shown in Fig. 2.3.

The sedimenting or gravitational force,  $F_s$ , is proportional to the mass ( $m$ , g) of the particle and the acceleration. For a particle subjected to a rotation, this acceleration is determined by the distance from the axis of rotation,  $r$  (cm), and the square of the angular velocity,  $\omega$  (rad s<sup>-1</sup>).

The gravitational force on the particle is then

$$F_s = m_h \omega^2 r \quad (2.1)$$

where  $m_h$  is the hydrated mass of the particle (g) having a corresponding hydrated volume.



**Figure 2.3 Forces on a particle under the influence of a gravitational field**

The various forces experienced by a solvated particle of mass,  $m$  under the influence of a gravitational field. The centrifugal, buoyant and frictional forces are denoted as  $F_s$ ,  $F_b$  and  $F_f$  respectively. Taken from [www-bioc.rice.edu/bios576/AU/AU\\_Page.html](http://www-bioc.rice.edu/bios576/AU/AU_Page.html)

The hydrated mass can in turn be defined as

$$m_h = \frac{M}{N_A} (1 + \delta_1) \quad (2.2)$$

where  $\delta_1$  is the hydration (or solvation) of the particle (g of solvent per g of solute),  $M$  is the molar mass of the particle (g/mol) and  $N_A$  is the Avogadro's number ( $6.023 \times 10^{23} \text{ mol}^{-1}$ ). Substituting equation 2.2 into 2.1 gives

$$F_s = \frac{M}{N_A} (1 + \delta_1) \omega^2 r \quad (2.3)$$

From Archimedes principle, the buoyant force,  $F_b$ , is equal to the weight of the fluid displaced and expressed as,

$$F_b = -m_0 \omega^2 r \quad (2.4)$$

The mass of the solvent displaced by the particle ( $m_0$ ) in the case of infinite dilution can be defined in terms of  $u_h$  and  $m_0$  with,

$$u_h = \frac{M}{N_A} (\bar{v} + \delta_1 v_1^0) \quad (2.5)$$

$$m_o = \rho_o u_h = \frac{M}{N_A} (\bar{v} \rho_o + \delta_1 \rho_o v_1^o) \quad (2.6)$$

where  $u_h$  is the velocity of the particle,  $v_1^o$  and  $\rho_o$  are the specific volume and density of the solvent, respectively.

But  $v_1^o = 1/\rho_o$ , therefore

$$m_o = \frac{M}{N_A} (\bar{v} \rho_o + \delta_1) \quad (2.7)$$

Now the total force acting on the particle is given by

$$F_t = F_s + F_b \quad (2.8)$$

Substituting equations 2.3, 2.4 and 2.7 into equation 2.8 gives

$$F_t = \frac{M}{N_A} r \omega^2 (1 - \bar{v} \rho_o) \quad (2.9)$$

This resultant force  $F_t$  opposes the frictional drag on the particle that moves through the viscous liquid with a constant velocity such that

$$F_t = F_f, \quad F_f = uf \quad (2.10)$$

where  $f$  is the frictional coefficient of the sedimenting particle

Combining equations 2.9 and 2.10 gives

$$uf = \frac{M}{N_A} r \omega^2 (1 - \bar{v} \rho_o)$$

The sedimentation coefficient is then defined as

$$s = \frac{u}{r \omega^2} = \frac{M}{N_A} \frac{(1 - \bar{v} \rho_o)}{f} \quad (2.11)$$

Simplistically, at a given radial position,  $r$

$$s = \frac{ur}{r\omega^2} = \frac{1}{r\omega^2} \frac{dr}{dt} \quad (2.12)$$

However, the consideration of diffusion poses limitations to equation 2.12, but this is overcome by the thermodynamic treatment of sedimentation.

### 2.2.1.1.2 *The thermodynamic picture of SV*

The thermodynamic treatment of SV data as derived by Tanford (1961) is presented here. The high velocities employed during SV result in the establishment of equilibrium with the formation of a steep radial concentration gradient across the solution. At this equilibrium, the total chemical potential ( $\mu_t$ ) within the solution will be constant

$$\mu_t = \mu_c + E_p \quad (2.13)$$

$$\frac{\partial \mu_t}{\partial r} = \frac{\partial \mu_c}{\partial r} + \frac{\partial E_p}{\partial r} \quad (2.14)$$

Here  $\mu_c$  is the chemical potential and  $E_p$  is the mechanical potential energy. At mechanical equilibrium, the mechanical potential energy is exactly equal to the kinetic energy,  $E_k$ :

$$\partial E_k = \int F \partial r$$

$$\frac{\partial E_p}{\partial r} = \frac{\partial E_k}{\partial r} = -F = -Mr\omega^2 \quad (2.15)$$

Substituting equation 2.15 into equation 2.14 yields

$$\frac{\partial \mu_t}{\partial r} = \frac{\partial \mu_c}{\partial r} - Mr\omega^2 \quad (2.16)$$

This can be expressed as a thermodynamic flow equation

$$J = -L \frac{\partial \mu_t}{\partial r} = -L \left[ \frac{\partial \mu_c}{\partial r} - Mr\omega^2 \right] \quad (2.17)$$

Here  $J$  is the cross sectional flow ( $\text{mol cm}^{-2} \text{s}^{-1}$ ) and  $L$  is a coefficient that is a function of temperature, pressure and composition. Equation 2.17 can be reduced to the form (van Holde, 1971)

$$J = L \left[ \omega^2 r M (1 - \bar{v}\rho) - \frac{RT}{C} \frac{\partial C}{\partial r} \right] \quad (2.18)$$

where  $L = C/N_A f$ ,  $R$  is the gas constant,  $T$  is the temperature and  $C$  is the solute concentration. This gives the Lamm equation (Fujita, 1975)

$$J = \frac{M(1 - \bar{v}\rho)}{N_A f} \omega^2 r C - \frac{RT}{N_A f} \frac{\partial C}{\partial r} \quad (2.19)$$

$$J = s\omega^2 r C - D \frac{\partial C}{\partial r} \quad (2.20)$$

From the above equation the sedimentation ( $s$ ) and diffusion ( $D$ ) coefficients are obtained and defined as

$$s = \frac{M(1 - \bar{v}\rho)}{N_A f}, \quad D = \frac{RT}{N_A f} \quad (2.21)$$

It is clearly seen from equation 2.20 that sedimentation opposes diffusion, and the coefficients of sedimentation and diffusion are related by the equation

$$s = \frac{M(1 - \bar{v}\rho)}{N_A f} = \frac{MD(1 - \bar{v}\rho)}{RT} \quad (2.22)$$

The sedimentation and diffusion coefficients,  $s$  and  $D$  are reported in Svedberg units,  $S$  ( $10^{-13} \text{s}$ ) and  $\text{cm}^2 \text{s}^{-1}$ , respectively.

Protein molecules in solution carry with them a volume of bound water owing to the polar interaction of surface residues with water, and also due to the entrapment of water within clefts/cavities present in the molecule. This is

known as hydration ( $\delta$ ) and, as it depends on the surface composition and structure of the protein, it is difficult to quantify theoretically (Squire and Himmel, 1979) or measure experimentally (Kuntz and Kauzmann, 1974) with a high degree of accuracy. Although the hydration of proteins has a negligible effect on some experimentally measured parameters such as the radius of gyration,  $R_g$  (as measured by small angle scattering), it significantly affects others such as the sedimentation coefficient,  $s$ .

The frictional coefficient of an anhydrous sphere ( $f_0$ ) is given by Stokes' Law (Tanford, 1961)

$$f_0 = 6\pi\eta R_0 \quad (2.23)$$

where  $\eta$  is the viscosity of the solvent (Poise) and  $R_0$  is the anhydrous radius of the particle (cm)

$$R_0 = \left( \frac{3M\bar{v}}{4\pi N_A} \right)^{1/3} \quad (2.24)$$

The ratio of  $f_0$  to the experimentally measured frictional coefficient  $f$  can be considered to be the product of two terms (Squire and Himmel, 1979; Teller, 1973)

$$\frac{f}{f_0} = \frac{f}{f_h} \frac{f_h}{f_0} \quad (2.25)$$

The first term  $f/f_h$  represents the deviation of  $f$  from  $f_h$  owing to hydrodynamic hydration, while the second term  $f_h/f_0$  reflects the deviation from sphericity of the shape of the molecule and is known as the Perrin ratio,  $P$  (Perrin, 1936).

The maximum value of hydration  $\delta_{max}$  can be determined by setting the Perrin function,  $P$ , to its minimum possible value ( $P = 1$  for a sphere), and calculating the deviation of  $f/f_0$  due to hydration.

From equations 2.23 to 2.25

$$f = Pf_0 \left( \frac{V_h N_A}{\bar{v} M} \right)^{1/3} \quad (2.26)$$

The hydrated volume  $V_h$  and the anhydrous frictional coefficient  $f_0$  are described by the equations

$$V_h = \frac{M v_s}{N_A} = \frac{M}{N_A} (\bar{v} + \delta v^0) \quad (2.27)$$

$$f_0 = 6\pi\eta R_0 = 6\pi\eta \left( \frac{3M\bar{v}}{4\pi N_A} \right)^{1/3} \quad (2.28)$$

The Svedberg equation in terms of the frictional coefficient of a molecule can be written as (van Holde, 1971)

$$s^0 = \frac{M(1 - \bar{v}\rho)}{N_A f} \quad (2.29)$$

Substituting equations 2.26-2.28 into equation 2.29, setting  $P = 1$  and rearranging for  $\delta$  gives

$$\delta_{max} = \frac{1}{v^0} \left[ \frac{4\pi N_A}{3M} \left( \frac{M(1 - \bar{v}\rho)}{6\pi\eta N_A s^0} \right)^3 - \bar{v} \right] \quad (2.30)$$

This  $\delta_{max}$  represents the maximum hydration that can be calculated for a molecule given,  $M$ ,  $\bar{v}$  and  $s^0$ . Typically this value ranges from 0.15-0.50 g/g although most proteins are thought to have a hydration of about 0.3-0.4 g/g (García de la Torre, 2001).

Based on the Svedberg equation, the anhydrous ( $s_0$ ) and hydrated ( $s_\delta$ ) sedimentation coefficients are defined as (van Holde, 1971)

$$s_0 = \frac{M(1 - \bar{v}\rho)}{N_A f_h}, \quad s_\delta = \frac{M(1 - \bar{v}\rho)}{N_A f} \quad (2.31)$$

Rearranging equation 2.31 for  $s_\delta$  gives (Tanford, 1961)

$$s_\delta = s_0 \frac{f_h}{f} = s_0 \left( \frac{M\bar{v}}{V_h N_A} \right)^{1/3} \quad (2.32)$$

Substituting for  $V_h$  from equation 2.27 into 2.32, gives the relationship between the hydrated ( $s_\delta$ ) and the anhydrous ( $s_0$ ) sedimentation coefficients.

$$s_\delta = s_0 \left( \frac{\bar{v}}{\bar{v} + \delta_1 v_1^o} \right)^{1/3} \quad (2.33)$$

Combining the Svedberg (van Holde, 1971) and Stokes' equations (Tanford, 1961) and substituting for all constants with values for water at 20°C, the sedimentation coefficient  $s_{sphere}$  of an anhydrous spherical particle in terms of  $M$  (g/mol),  $\bar{v}$  (ml/g) and  $\rho$  (g/ml) can be obtained.

$$s_{sphere} = 0.012 \frac{M^{2/3} (1 - \bar{v}\rho)}{\bar{v}^{1/3}} \quad (2.34)$$

Spherical proteins have minimal surface area exposed to solvent and consequently a low frictional coefficient  $f_0$ . However, any asymmetry in particle shape increases the interaction of protein residues on the surface with the solvent, causing the value of  $s$  to decrease as a consequence.

To compare sedimentation coefficients obtained under different experimental conditions, the value of  $s$  obtained experimentally is corrected to standard conditions (in water at 20°C) using the equation (van Holde, 1971)

$$s_{20,w} = s_{T,b} \left( \frac{\eta_{T,b}}{\eta_{20,w}} \right) \left( \frac{(1 - \bar{v}\rho)_{20,w}}{(1 - \bar{v}\rho)_{T,b}} \right) \quad (2.35)$$

where the subscripts  $T$  and  $b$  refer to the temperature and buffer conditions used in the experiment, and  $20,w$  indicates standard conditions in water.

### 2.2.1.2 SV data analysis

Although several approaches have been developed for the analysis of SV data e.g. SEDANAL (Stafford, 1992; 1994; Stafford and Sherwood, 2004), UltraScan (Demeler, 2004; 2005) LAMM (Behlke and Ristau, 1997), only methods used in this thesis will be described, namely the sedimentation coefficient distribution and finite element analyses as implemented in the program SEDFIT (Schuck, 2000; Schuck, et al., 2002).

The macromolecular concentration distribution  $\chi$  as a function of time  $t$  and radial position  $r$  under the influence of sedimentation and diffusion processes in a sector-shaped sample cell is described by the Lamm equation (Lamm, 1929),

$$\frac{\partial \chi(r,t)}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left[ rD \frac{\partial \chi(r,t)}{\partial r} - s\omega^2 r^2 \chi(r,t) \right] \quad (2.36)$$

In SEDFIT the SV profiles are directly modelled with numerical finite element solutions of the Lamm equation. SEDFIT has advantages over other methods in terms of achieving a higher resolution and being able to model SV profiles even when no clear visible boundary is observed in the actual data due to the effects of diffusion (Lebowitz, et al., 2002; Schuck, 2000). Moreover, it is also capable of modelling a wide range of species, from salts (typically  $s < 0.1$  S) to large viral capsids ( $s > 1000$  S). SEDFIT also contains a comprehensive set of tools to adjust the analysis for the special noise structure of interference optical data (Schuck and Demeler, 1999). Several models are available to the user to best fit the data. Most commonly used ones include the ‘non-interacting species model’ that can be used to fit a maximum of 4 species, and the ‘interacting species model’ that is used for fitting self-association reactions such as monomer-dimer association etc.

The differential sedimentation coefficient distribution,  $c(s)$  developed by Schuck (2000) requires no prior knowledge of the number of species present in the sample or their interaction(s). The  $c(s)$  deconvolutes diffusion effects by direct boundary modelling of the SV profiles as distributions of Lamm equation solutions.  $c(s)$  is defined as

$$a(r,t) = \int c(s) \chi(s, D(s), r, t) ds + \varepsilon \quad (2.37)$$

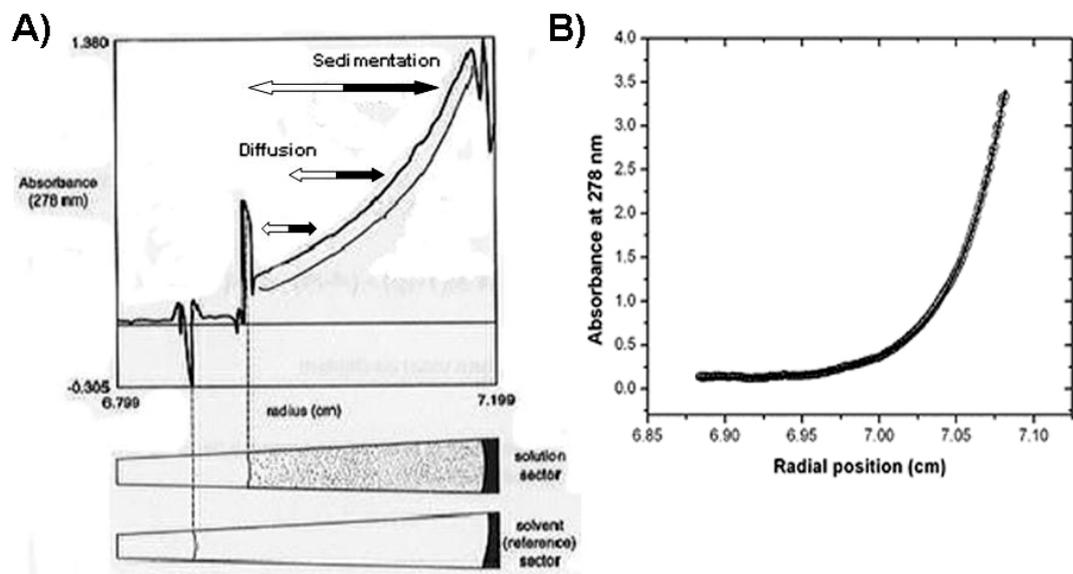
where  $a(r,t)$  represents the observed sedimentation data,  $c(s)$  is the concentration of species with sedimentation coefficients between  $s$  and  $s+ds$ ,  $\chi(s, D, r, t)$  is the solution of the Lamm equation for a single species and  $\varepsilon$  is the noise component. In simple terms, the species present in an SV system have specific sedimentation ( $s$ ) and diffusion ( $D$ ) coefficients, and therefore contribute to the radial- and time-dependent absorbance/interference  $a(r,t)$  that is proportional to the Lamm equation solution  $\chi(s, D, r, t)$  and the loading concentration. However, this approach assumes an average frictional ratio  $f/f_0$  for all species in the sample, and that none of the species interact with one another on the time-scale of the experiment. The  $c(s)$  distribution can also be used to analyse mixtures of interacting proteins: however, the peaks in the  $c(s)$  profile may then correspond to interconverted species rather than any of the individual sedimenting macromolecules. This can be overcome by the analysis of various concentrations of the same protein mixture. The area under the peaks in the  $c(s)$  sedimentation coefficient distributions can be integrated to yield weight-average sedimentation coefficients.

All SV analysis carried out in this thesis involved the use of the  $c(s)$  distribution as the first step to determine the number of species present in the experimental sample, and to obtain initial estimates of their sedimentation coefficients and loading concentrations. All  $c(s)$  analysis resulted in less than four species and the data were subsequently modelled using finite element analysis, via the non-interacting species model to determine true sedimentation coefficients for all species. The sedimentation coefficients were then corrected for the effect of concentration by extrapolation to infinite dilution, to yield the value of  $s$  independent of concentration,  $s_{20,w}^0$ .

## 2.2.2 Sedimentation equilibrium (SE)

In SE experiments, macromolecules are subjected to a lower centrifugal field (i.e. low rotor speeds). As the particles in the sample begin to sediment towards the bottom of the cell forming a concentration gradient, the process of diffusion opposes the force of sedimentation (Fig. 2.4A). After an appropriate period of

time, these two opposing forces reach thermodynamic equilibrium resulting in no net movement of molecules within the concentration gradient. At this stage the distribution of each species in the sample within this concentration gradient is described by an exponential as shown in Fig. 2.4B. As a consequence higher molecular weight species will be located at the bottom of the cell, while the low molecular weight species will dominate at the top. Parameters such as the molecular weight and association/dissociation constants can be determined from a typical SE run. As with SV, equations pertaining to SE can be derived from either a mechanical or thermodynamical standpoint.



**Figure 2.4 Schematic diagram of SE analytical ultracentrifugation**

(A) Schematic representation of the forces involved during an SE run. The increased sedimentation of the solute (black arrow) with radial distance is balanced by the reverse flow of diffusion (open arrow) that increases with the concentration gradient. (B) Characteristic SE profile  $a(r)$  as a function of radial position  $r$ . (A) was taken from [http://www-bioc.rice.edu/bios576/AU/AU\\_Page.html](http://www-bioc.rice.edu/bios576/AU/AU_Page.html)

## 2.2.2.1 Theory

### 2.2.2.1.1 The mechanical picture of SE

In SE, a state of equilibrium is attained by the balance of the opposing forces of sedimentation and diffusion. In this state, the concentration gradient of a homogeneous single ideal macromolecule in a two-component system is given by (Svedberg and Pedersen, 1940; Tanford, 1961; van Holde, 1971)

$$\frac{d \ln(c)}{dr^2} = \frac{M(1 - \bar{v}\rho)\omega^2}{2RT} \quad (2.38)$$

where  $c$  is the macromolecular concentration ( $M$ ) and  $R$  is the gas constant ( $8.314 \times 10^7 \text{ erg K}^{-1} \text{ mol}^{-1}$ ). Ideality here refers to particles in solution that do not interact with each other. Therefore, for the ideal single species plotting a graph of  $\ln(c)$  versus  $r^2$ , results in a straight line with the slope being  $M(1 - \bar{v}\rho)\omega^2 / 2RT$ . The molecular mass can then be easily obtained from this slope knowing the partial specific volume  $\bar{v}$  of the macromolecule in question.

However, complications arise when the above plot is not a straight line. A downward curving plot, reflecting the decrease in molecular weight with increasing concentration, indicates nonideality and concentration dependence of the molecular weight. On the other hand, an upward curvature is an indication that the sample is polydisperse, either owing to impurity or sample aggregation. In this case,  $d \ln(c) / dr^2$  yields an average molecular weight  $M_w$  assuming that all the species in the sample have the same partial specific volume  $\bar{v}$ . If polydispersity is a reflection of a self-associating system, SE runs carried out at different rotor speeds will result in the overlap of  $M_w$  versus  $c$  plots, and consequently an average molecular weight independent of concentration can be determined with high accuracy.

When the experimental sample contains more than one species, the sedimentation profile  $a(r)$  is given by the sum of the exponentials recorded for each species and takes the form (Lebowitz, et al., 2002)

$$a(r) = \sum_{i=1}^n c_{i,0} \varepsilon_i d \exp \left[ \frac{M_i (1 - \bar{v}_i \rho) \omega^2}{2RT} (r^2 - r_0^2) \right] + \delta \quad (2.39)$$

Here,  $n$  represents the number of species,  $c_{i,0}$  and  $\varepsilon_i$  are the molar concentration and extinction coefficient of species  $i$  at the reference position  $r_0$ ,  $d$  denotes the optical path length used during the experiment (12 or 3 mm) and  $\delta$  represents the baseline offset that compensates for all non-sedimenting material as well as small imperfections in the cell assembly and data acquisition. The extinction

coefficient in equation 2.39 is replaced by the specific signal increment (2.75 fringes per mg/ml) when interference optics are used (Lebowitz, et al., 2002).

The buoyant molar mass corresponds to the resulting mass of the molecule when displaced by the solvent and is given by  $M_b = M(1 - \bar{v}\rho)$ . The steepness of the concentration gradient in SE experiments is highly dependent on the square of the rotor speed and this buoyant molar mass of the macromolecule. Typically, SE experiments are carried out at 2-4 different rotor speeds and 7-10 sample concentrations in order to accurately determine the molecular weight of the experimental system.

### 2.2.2.1.2 *The thermodynamical picture of SE*

Although the kinetic derivation for SE is sufficient in most cases, real biological systems often exhibit significant nonideality. Hence, the thermodynamic approach provides a more natural description of a system at sedimentation equilibrium. In addition, from the measurement of concentration distribution, valuable information on the thermodynamic properties of the system can be obtained. The theory developed by Goldberg (1953) is summarized below.

The Gibbs-Duhem equation of total potential for a system in equilibrium is given by

$$V dp^{(x)} = \sum_{i=0}^r n_i^{(x)} d\mu_i^{(x)} \quad (2.40)$$

Here  $n_i^{(x)}$  is the number of moles of solute  $i$  in phase  $x$  that has a volume  $V$ . The change in pressure in the vicinity of this phase and the chemical potential of component  $i$  in phase  $x$  are denoted by  $dp^{(x)}$  and  $\mu_i^{(x)}$ , respectively. On solving equation 2.40, a general equation for equilibrium in a centrifugal field is obtained and takes the form

$$(1 - \bar{v}\rho) dE_p = \left( \frac{\partial \mu}{\partial c} \right) dc \quad (2.41)$$

With  $dE_p = mr\omega^2 dr$

$$M(1 - \bar{v}\rho)r\omega^2 = \frac{\partial\mu}{\partial c} \frac{dc}{dr} \quad (2.42)$$

However, for dilute solutions  $\mu$  can be expressed as a power series such that

$$\mu_1 = \mu_1^0 - RTV^0 c \left( \frac{1}{M} + Bc + Cc^2 + \dots \right) \quad (2.43)$$

where subscript 1 refers to solvent,  $V^0$  is the molar volume of the solute,  $c$  its concentration,  $B$  and  $C$  denote the second and third virial coefficients, respectively. The concentration  $c$  (M) becomes negligible beyond the first order and therefore  $\partial\mu/\partial c$  can be expressed in terms of a sedimentation virial coefficient (Tanford, 1961)

$$\frac{\partial\mu}{\partial c} = RT \left( \frac{1}{c} + BM \right) \quad (2.44)$$

Thus equation 2.42 becomes

$$\frac{1}{r} \frac{dc}{dr} = \frac{M(1 - \bar{v}\rho)\omega^2}{RT(1 + BMc)} \quad (2.45)$$

An apparent molecular weight  $M^{app}$  can be defined as

$$M^{app} = \frac{M}{1 + BMc} \quad (2.46)$$

so that equation 2.45 takes the form

$$\frac{1}{r} \frac{dc}{dr} = \frac{M^{app}(1 - \bar{v}\rho)\omega^2}{RT} \quad (2.47)$$

This is the fundamental expression for SE which clearly indicates that from the knowledge of the radial dependence of concentration of a macromolecule in equilibrium, its molecular weight can be determined.

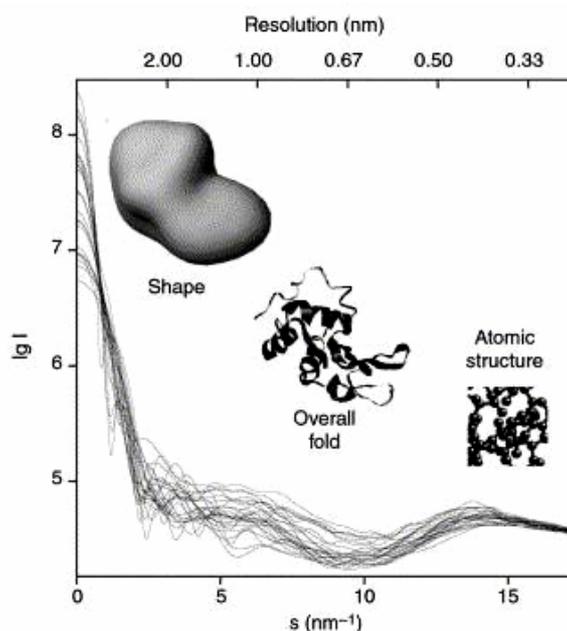
### 2.2.2.2 SE data analysis

The weight-average molecular mass  $M_w$  was determined by fitting the equation 2.38 to the data via programs MicroCal ORIGIN (part of the Beckman XL-A/XL-I software suite) and SEDPHAT (Schuck, 2003; Vistica, et al., 2004). MicroCal ORIGIN was also used to determine estimates of the second virial coefficient,  $B$ , to improve fitting of the experimental SE data. However, inclusion of  $B$  seemed to worsen the fits and hence all further SE analysis was carried out using only SEDPHAT.

Global analysis of SE data using SEDPHAT was performed by simultaneously fitting the data sets from all cells at each rotor speed. Global fitting improves consistency by fitting with the same baseline and ensures higher reliability of the obtained results. In addition, cells were also individually analysed to enable determination of the concentration-independent molecular weight,  $M_w^0$ , from the plot of the inverse of the apparent molecular weight,  $M_{app}$  versus protein concentration,  $c$ . Fitting models available in SEDPHAT that were used in this thesis include the non-interacting discrete species model (that fits with up to four species), and the monomer  $\rightarrow$  n-mer self-association models.

## 2.3 Small angle scattering

Small angle scattering (SAS) probes the structure of native biological macromolecules in solution at low (10-20 Å) resolution. The solution sample is irradiated by a highly collimated beam of x-rays (SAXS) or neutrons (SANS) and the scattering intensity  $I$  is recorded by the detector. For dilute solutions (concentration of nM) the particles are randomly oriented and give rise to isotropic intensity that is dependent only on the scattering angle ( $2\theta$ ) between the incident and scattered beam. SAS here is based on elastic scattering and the wavelength of radiation  $\lambda$  remains unchanged. The SAS data provide direct determination of parameters such as molecular mass ( $M$ ), radius of gyration ( $R_g$ ), hydrated volume ( $V$ ) and maximum diameter ( $D_{max}$ ). The intensity of scattering,  $I(s)$  is measured as a function of the scattering angle  $2\theta$  or momentum transfer  $s$  ( $s = 4\pi \sin\theta/\lambda$ ) as shown in Fig. 2.5.



**Figure 2.5** General features of a solution scattering curve

X-ray scattering patterns of proteins with different folds and molecular masses are shown. The axes,  $s$  and  $\lg I$  denote the scattering angle and logarithm of scattering intensity, respectively. The low angle ( $2\text{--}3\text{ nm}^{-1}$ ) region determines the particle shape, while higher angles ( $> 3\text{ nm}^{-1}$ ) yield information on the tertiary and quaternary structure of the macromolecules. Taken from Svergun and Koch (2002).

At small angles, values of  $R_g$  and the forward scattering intensity at zero angle  $I(0)$  are obtained. At higher angles, further information on the overall fold and tertiary structure can be obtained.

During the past decade the development of *ab initio* methods for three-dimensional modelling of SAS data coupled with the advent of high brilliance x-ray synchrotrons and high-flux nuclear reactors have paved the way for the present renaissance of SAS in structural biology. In addition, the measurement of structural data in solution over a wide range of molecular sizes reflects the advantage of SAS over complementary techniques like electron microscopy.

The theoretical aspects, data treatment procedures and *ab initio* modelling of SAS data obtained using x-rays and neutrons are discussed below.

## 2.3.1 X-rays

Small angle x-ray scattering (SAXS) is used to study the structure of materials that can vary in size from ten to several thousand Å, such as proteins and viral capsids (Heller, et al., 2004; Lee, et al., 2004). This technique is not limited to biological macromolecules and has been used widely to characterise synthetic polymers, nanoparticles, etc. SAXS exploits coherent scattering in which electrons oscillating under the influence of the electric field of the x-ray beam act as secondary sources, emitting x-rays with the same wavelength as the incident beam, but 180° out of phase. The typical wavelength of x-rays used for SAXS is 1.5 Å. SAXS has the advantage of probing structures of highly flexible proteins (in solution) that are very hard to crystallise. The following sections will focus only on aspects relevant to the work described in this thesis, namely the solution scattering of monodisperse systems of identical, non-interacting particles.

### 2.3.1.1 Theory of SAXS

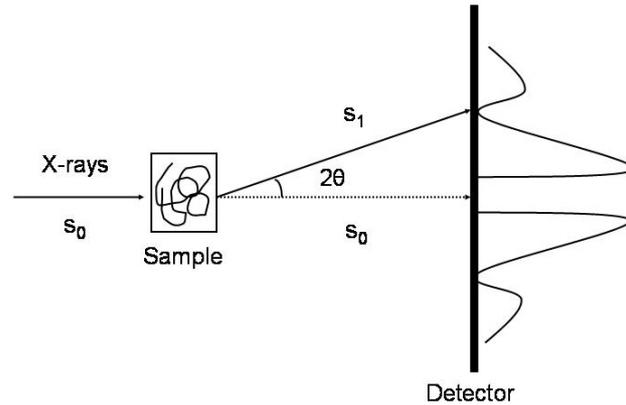
Only the electrons in the sample scatter upon interaction with the x-rays. In an ideal monodisperse system all particles have the same chemical and shape composition and do not interact with each other. As a result upon scattering, all particles have the same non-directional (isotropic) scattering amplitude. Furthermore, the recorded scattering intensity is the sum of the intensities of all individual particles. The uniform particle distribution in solution and isotropic scattering leads to the spherical averaging of the single particle scattering and a one-dimensional scattering curve with low resolution.

If  $\rho(r)$  is the electron density of the sample at a point  $r$  and  $s_0$  is the incident wave vector ( $|s_0| = 1/\lambda$ , where  $\lambda$  is the wavelength of the x-rays), then at distances much greater than the size of the sample the amplitude of the scattered radiation is (Vachette and Svergun, 2000)

$$F(s) = \int_{V_r} \rho(r) e^{-2i\pi rs} dV_r \quad (2.48)$$

where  $F(s)$  is the Fourier transform of the electron density distribution  $\rho(r)$ ,  $s$  denotes the scattering vector and  $V_r$  is the volume of the sample. The resultant scattering vector  $s$  is the difference between the scattered ( $s_1$ ) and incident ( $s_0$ ) wave vectors ( $s = s_1 - s_0$ ) (Fig. 2.6), and is

$$s = |s| = \frac{2 \sin \vartheta}{\lambda} \cong \frac{2\vartheta}{\lambda} \quad (2.49)$$



**Figure 2.6 Schematic representation of SAXS**

Vectorial and schematic representation of an x-ray scattering experiment is shown. Incident x-rays ( $s_0$ ) on the sample in solution results in scattering due to the electrons present in the sample at an angle  $2\theta$  with a resultant scattering vector of  $s = s_1 - s_0$ . The one dimensional scattering waveform as recorded by the detector is shown. Taken from Vachette and Svergun (2000)

The scattered intensity  $I(s)$  is defined as,

$$I(s) = F(s) \cdot F^*(s) = \int_{V_r} \int_{V_{r'}} \rho(r) \rho(r') e^{-2i\pi(r-r')s} dV_r dV_{r'} \quad (2.50)$$

where  $F^*(s)$  is the complex conjugate of  $F(s)$ , and  $r$  and  $r'$  denote radial vectors inside the particle. The integrals are over the particle volumes  $V_r$  and  $V_{r'}$  and  $\rho(r)$  is the electron density of the sample.

Scattering of particles in solution originates from the contrast of electron density between the particle and the homogenous solvent (having electron density  $\rho_0$ ) (Vachette and Svergun, 2000), hence equation 2.48 takes the form

$$F(s) = \int_V \Delta\rho(r) e^{-2i\pi r \cdot s} dV_r, \quad \text{where } \Delta\rho(r) = \rho(r) - \rho_0 \quad (2.51)$$

Owing to Brownian motion only a spherical average intensity of the sample is obtained from SAXS; the isotropic scattering intensity  $I(s)$  expressed in terms of the distance distribution function  $p(r)$  is (Vachette and Svergun, 2000)

$$I(s) = 4\pi \int_0^\infty p(r) \frac{\sin 2\pi r s}{2\pi r s} dr \quad (2.52)$$

The  $p(r)$  distance distribution function is

$$p(r) = \frac{1}{\pi} \int_0^\infty r s I(s) \sin 2\pi r s ds \quad (2.53)$$

For the ideal monodisperse sample,  $p(r)$  is the histogram of distances between all pairs of points (volume elements) in the sample. The scattering intensity at zero angle  $I(0)$  can be derived from equation 2.50

$$I(0) = \iint_V \Delta\rho(r) \rho(r') dV_r dV_{r'} = \Delta m^2 = m^2 - m_0^2 \quad (2.54)$$

where  $m$  represents the total number of electrons in the particle and  $m_0$  denotes the number of electrons of the solvent displaced by the particle.

The scattered intensity can be expanded in powers of  $s^2$  to give

$$I(s) = I(0) \left[ 1 - \frac{4\pi}{3} R_g^2 s^2 + k s^4 + \dots \right] \quad (2.55)$$

However, close to the origin (at low angles), the expansion can be restricted to the first order term (Guinier and Fournet, 1955)

$$I(s) \cong I(0) \exp\left(-\frac{4\pi^2}{3} R_g^2 s^2\right) \quad (2.56)$$

where  $R_g$  is the radius of gyration ,

$$R_g^2 = \frac{\int_{V_r} \Delta\rho(r) r^2 dV_r}{\int_{V_r} \Delta\rho(r) dV_r} \quad (2.57)$$

Thus at very low scattering angles, the scattering pattern can be approximated by a Gaussian distribution, the width of which is proportional to the square of the radius of gyration,  $R_g$  of the particle (as defined by equation 2.56). This is known as the Guinier approximation (Guinier and Fournet, 1955).  $R_g$  is a measure of the (non)sphericity of the particle and is experimentally determined from the plot of  $\ln I(s)$  versus  $s^2$ , also commonly referred to as the Guinier plot. It can also be determined from the slope using linear regression, and the intensity at zero angle,  $I(0)$  is obtained by extrapolation to zero scattering angle. The Guinier region which represents the scattering angle range over which the Guinier approximation is valid, is usually defined as  $0.2 \leq sR_g \leq 1.3$  although values may differ for asymmetric particles (Perkins, 1988). As the Guinier approximation holds true only for ideal, monodisperse particles, sample polydispersity due to aggregate formation, will greatly hamper accurate and reliable determination of  $R_g$ . To avoid this problem, scattering is usually recorded at several low sample concentrations.

### 2.3.1.2 Data treatment

Ideal single particle scattering intensities  $I(s)$  cannot be measured directly in SAXS. Instead, a discrete set of intensities  $I_{exp}(s_i)$  at  $i$  angles ( $i = 1, 2, \dots, N$ ) within the angular range  $s_{min} \leq s \leq s_{max}$  is recorded (Vachette and Svergun, 2000). However, this data set contains statistical errors and smearing effects owing to beam divergence, polychromaticity and detector resolution (Feigin and Svergun, 1987). Therefore, the main task of the data processing is in restoring the ideal scattering intensity  $I(s)$  from the experimental set  $I_{exp}(s_i)$ . For monodisperse systems,  $I(s)$  is related to the particle distance distribution function  $p(r)$  by the Fourier transform

$$I(s) = 4\pi \int_0^{D_{max}} p(r) \frac{\sin 2\pi rs}{2\pi rs} dr \quad (2.58)$$

where  $D_{max}$  is the maximum particle dimension (Vachette and Svergun, 2000). The function  $p(r)$  has the same information content as  $I(s)$ , and data processing can be done ‘indirectly’ by restoring  $p(r)$ . This ‘indirect transform approach’ first used by Glatter (1977) introduces an important constraint, namely a ‘boundedness’ of the characteristic function, and also forms the basis of the computer program GNOM (Semenyuk and Svergun, 1991; Svergun, 1992) that was used for the determination of distance distribution functions from SAXS data throughout this thesis. In addition, GNOM also employs a regularisation parameter,  $\alpha$  that allows the user to compensate for the quality of fit to the data and the smoothness of the  $p(r)$  function (Svergun, 1992). The choice of  $\alpha$  is crucial for the stability and validity of the solution: too small a value leads to unstable solutions, while too large a value results in systematic deviations from the experimental data. GNOM performs an automatic ‘visual’ search, with the solution characterized by a number of perceptual criteria such as discrepancy, systematic deviations, smoothness, compactness and positiveness of  $p(r)$  and solution stability of small changes in  $\alpha$  (Svergun, et al., 1988; 1992). Hence GNOM either finds the optimal solution automatically or in the event of being unable to obtain a good solution, indicates that the user’s assumptions about the system (e.g. the value of  $D_{max}$ ) are incorrect. Generally, the accurate value of  $D_{max}$  is not known, and hence several trial-and-error computations have to be carried out for different values of  $D_{max}$ . GNOM automatically calculates the value of  $I(0)$  by back extrapolating  $I(s)$  as described in equation 2.55 to zero angle. In addition, it uses the  $p(r)$  function to determine  $R_g$  which is less sensitive to interparticle interactions and/or small amounts of aggregate, as the entire scattering curve is used for the calculation rather than just the Guinier region.

As the scattering curves and their corresponding  $p(r)$  functions share the same information content, the particle shape can often be deduced by straightforward visual inspection of  $p(r)$  as shown in Fig. 2.7. For example, the scattering patterns of globular (sphere) and elongated (rod) shapes are significantly different with distinct minima. Very anisometric particles yield featureless scattering curves that decay much more slowly than those of globular particles (Svergun and Koch, 2003).

Historically, three-dimensional information from SAXS data was obtained by building simple low resolution models on a trial-and-error basis. These models consisted mainly of geometrical shapes such as spheres, cylinders or ellipsoids, and the calculated scattering curves and  $p(r)$  functions were compared with the experimental data (Glatter, 1980; Glatter and Kratky, 1982; Kratky and Pilz, 1972). However, more recently calculation of theoretical scattering curves from high-resolution structures obtained from x-ray crystallography or NMR has been exploited for evaluation of SAXS data. This approach entails the inclusion of scattering atoms and the particle excluded volume effects during the scattering curve simulation and is given by the equation,

$$I(s) = \left\langle \left| F_{particle}(s) - \rho_0 F_{buffer}(s) \right|^2 \right\rangle_{\Omega} \quad (2.59)$$

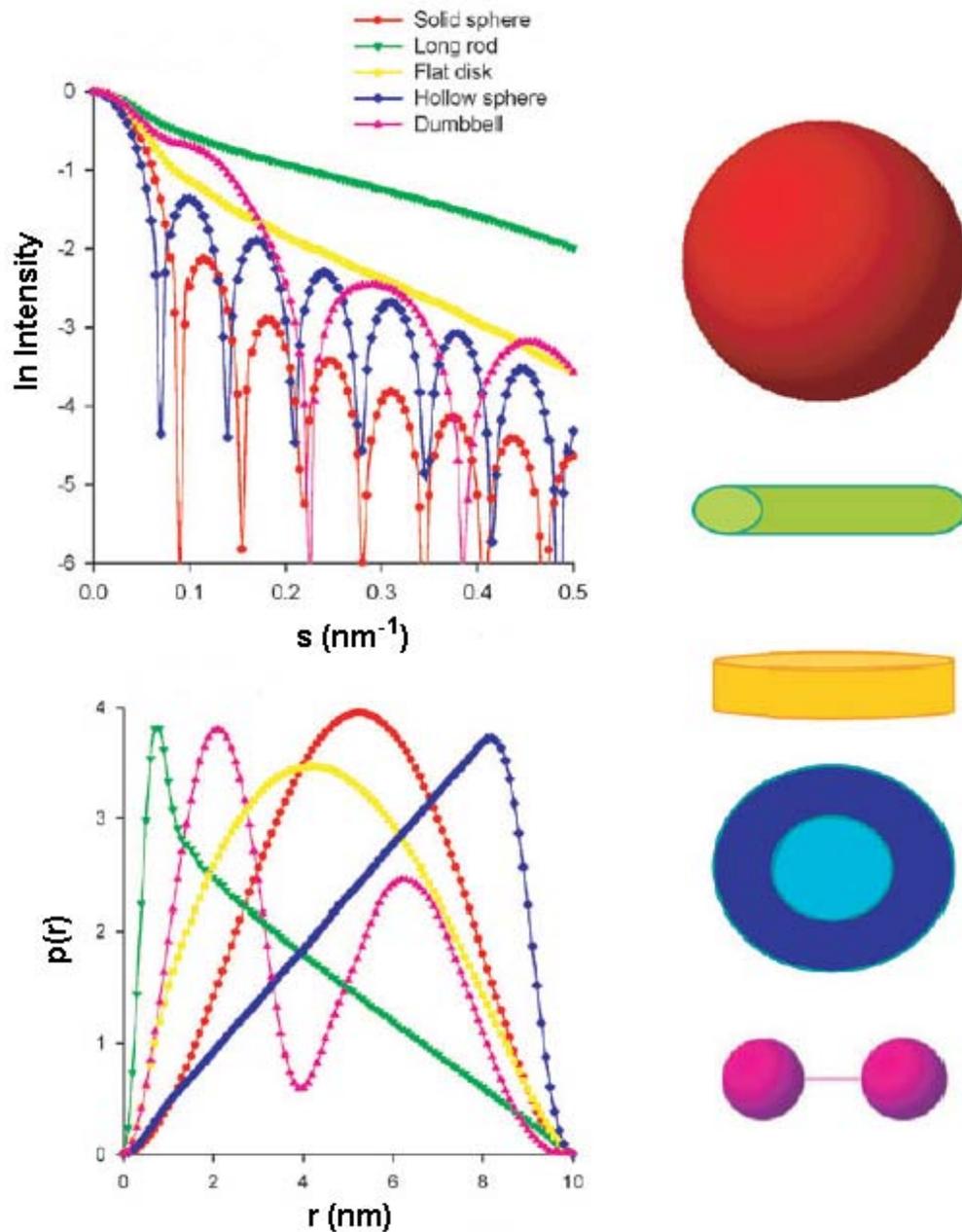
where  $\rho_0$  is the average electron density of the buffer,  $F_{particle}$  and  $F_{buffer}$  represent the scattering amplitudes of the particle and reference buffer, respectively, and the symbol  $\langle \rangle_{\Omega}$  denotes the spherical average.

The primary problem in the scattering curve calculation is the estimation of the contribution to the excluded volume and various methods have been developed to deal with this problem. Most notable, is the spherical harmonics approach of Lattman (1989) that was further improved by Svergun and co-workers (1995) by the inclusion of a hydration shell of about 0.3 nm thickness and having a different electron density ( $\rho_h$ ) to both that of the particle and the buffer:

$$I(s) = \left\langle \left| A_a(s) - \rho_b A_{ex}(s) + \delta\rho_h A_h(s) \right|^2 \right\rangle_{\Omega} \quad (2.60)$$

Here,  $A_a$ ,  $A_{ex}$  and  $A_h$  are scattering amplitudes from the particle *in vacuo*, from the excluded volume and from the hydration shell, respectively. The scattering electron densities of the bulk solvent differs from that of the hydration shell, resulting in a non-zero contrast for the shell  $\delta\rho_h = \rho_h - \rho_b$  (Svergun, et al., 1995). This method forms the basis of CRY SOL (Svergun, et al., 1995) that computes the scattering curve of any high- or low-resolution structure and fits it to the experimentally determined data. To improve the fits, CRY SOL allows for

variation in the hydration shell electron density and the average excluded solvent volume per atomic group.



**Figure 2.7 Scattering intensities and  $p(r)$  functions of various geometrical bodies**

Intensity scattering curves and  $p(r)$  distribution function of various geometrical bodies with the same maximum size are shown. While globular spherical particles (red) display bell-shaped  $p(r)$  functions with a maximum at about  $D_{max}/2$ , elongated particles (green) exhibit skewed distributions with a clear maximum at small distances corresponding to the radius of the cross-section. Similarly, a broad maximum with a distance less than  $D_{max}/2$  is displayed by flattened particles (yellow), while hollow particles (blue) show a maximum at distances larger than  $D_{max}/2$ . Particles with well separated subunits, e.g. dumbbell-shaped display multiple minima (pink), the first corresponding to the intrasubunit distances and the others yielding separation between the subunits. Image taken from Svergun and Koch (2003)

### 2.3.1.3 *Ab initio* modelling

The extraction of the three-dimensional model of an object from its one-dimensional scattering curve is a non-trivial problem. In the past, shape modelling from SAXS data was based on trial-and-error, computing the scattering patterns of different geometrical shapes, and that at times included additional information (e.g. from EM or hydrodynamic data).

However, the development of more sophisticated and robust methods over the last 40 years has enabled restoration of more detailed 3D structural information from the analysis of SAXS data. The first *ab initio* approach developed by Stuhrmann (1970) involved the representation of the particle shape by an angular envelope function that was described by a series of spherical coordinates. This method was developed and implemented in the program SASHA (Svergun and Stuhrmann, 1991; Svergun, et al., 1997). SASHA generates unique solutions from random starting configurations indicating the stability of reconstructions modelled from the experimental data (Svergun, et al., 1996). The stability and resolution can be further improved if information on the particle symmetry is available (Vachette and Svergun, 2000).

However, use of the angular envelope function was limited to relatively simple shapes with no internal cavities. Over the years, more comprehensive modelling methods were developed that used bead models via Monte Carlo-like searches. This type of approach was implemented in the program DAMMIN (Svergun, 1999), where a sphere of diameter  $D_{max}$  is filled with closely packed dummy atoms/beads on a hexagonal lattice. Each dummy atom is either part of the particle (index = 1) or the solvent (index = 0), and the shape is thus described by a long binary string of atoms. Starting from a random distribution of atoms, the lowest energy configuration within the specified search space is attained by using a simulated annealing procedure. A penalty term is also implemented during the modelling process to ensure the compactness and connectivity of the resulting shape. Akin to SASHA, specification of the particle symmetry increases the resolution of the reconstructions. DAMMIN has been rigorously tested and evaluated for several experimental systems (Svergun, 1999; Volkov and Svergun, 2003) and a huge number of papers with DAMMIN reconstructions have been

published during the last few years. Although DAMMIN is mostly reliable, it does experience problems with some complicated structures containing internal cavities (like the PDC cores studied in this project) and is unable to satisfactorily restore the shape of the molecule (Volkov and Svergun, 2003).

A similar approach was implemented in the program DALAI\_GA (Chacón, et al., 2000; Chacón, et al., 1998). Here, a sphere of maximum diameter  $D_{max}$  is filled with a large number of closely packed dummy atoms that either belong to the particle or the solvent. From an initial random configuration, the lowest energy configuration that best fits the experimental data is obtained via a genetic algorithm.

A more versatile approach to construct protein models from SAXS has been recently developed by Svergun and co-workers and implemented in the program GASBOR (Petoukhov and Svergun, 2003; Svergun, et al., 2001). The macromolecule is not represented by hexagonally closely packed beads; instead it is characterized by an assembly of dummy residues, DRs (with one dummy residue per amino acid). GASBOR starts with a random distribution of DRs in the spherical search volume specified by  $D_{max}$ . As with DAMMIN, simulated annealing is then used to obtain the energy minimized conformation with best fits to the experimental data. In addition, GASBOR has encoded within it connectivity constraints to result in 'chain-compatible' spatial arrangements of DRs that form protein-like folds. In particular, as  $C_{\alpha}$  atoms of neighbouring amino acids residues are separated by about 0.38 nm in the primary structure, each DR should have two neighbouring DRs each separated from it by 0.38 nm. Apart from particle and solvent DRs GASBOR also introduces a third kind corresponding to the particle hydration shell, where the water molecules bound form part of the primary hydration layer. While SAXS data include a contribution from a hydration layer due to bound water, it is significantly less than the hydration of a sedimenting particle. Furthermore GASBOR uses data at higher scattering angles than DAMMIN ( $s \leq 0.5 \text{ \AA}^{-1}$ ) and therefore results in more detailed, higher-resolution models (Svergun, et al., 2001).

It is evident that different initial random configurations of these Monte Carlo based methods yield multiple solutions (spatial distribution of beads/DRs) with

essentially the same fit to the data (Svergun and Koch, 2003). These independent models can then be superimposed and averaged to obtain the most probable model using the program DAMAVER (Volkov and Svergun, 2003). However, this averaging is reasonable only for similar models. DAMAVER employs the program SUPCOMB (Kozin and Svergun, 2001) that aligns two (high- or low-resolution) models (represented as an ensemble of points) and yields a measure of dissimilarity (represented by a normalised spatial discrepancy (NSD) factor) between them. In this thesis all pairs of independent *ab initio* models were aligned using SUPCOMB, where for every point (bead) in the first model, the minimum value among the distances between this point and all points in the second model was found, with the same being done for the second model. These distances are then added and normalised against the average distances between the neighbouring points for the two models. An NSD value close to 1 is indicative of model similarity. The model having the lowest NSD or smallest average discrepancy was then taken as the reference. All the other models were aligned with the reference model, and a density of DRs was computed and remapped onto a densely packed grid of beads, with each grid point characterised by its occupancy factor (number of beads present at a specific position in all the models). The average model was then constructed by filtering out low occupancy grid points with a threshold corresponding to the excluded particle volume. Although DAMAVER results in an average consensus model with all structural features of the individual models preserved in most cases, it does experience some problems with structures that are elongated and/or possess internal cavities (Volkov and Svergun, 2003).

#### **2.3.1.4 Instrumentation**

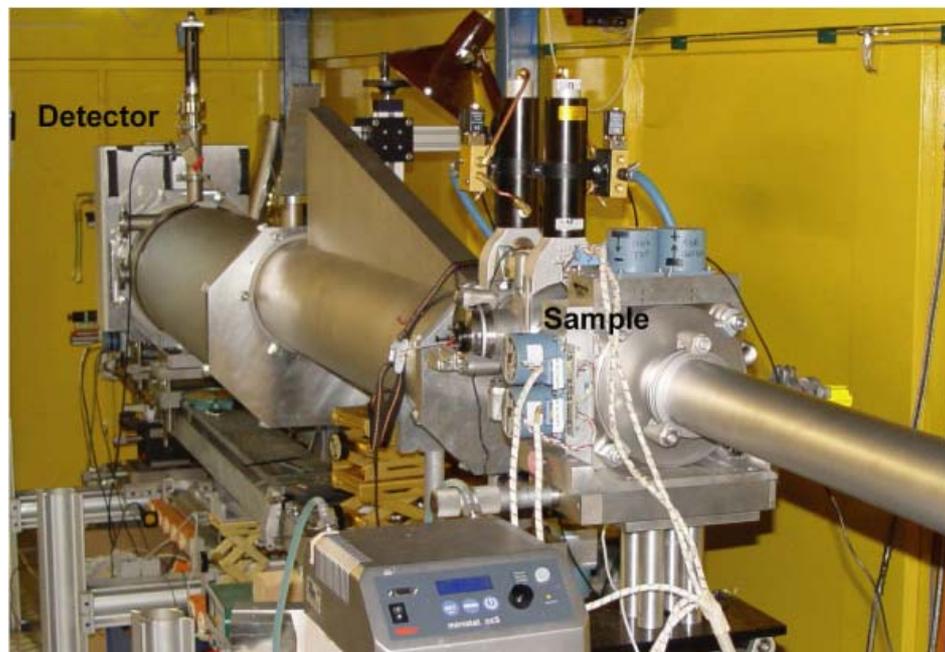
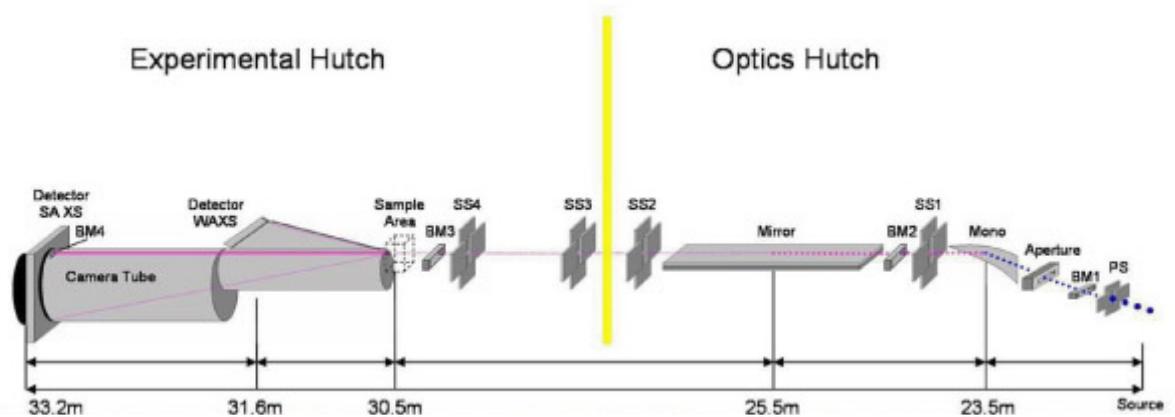
The SAXS work carried out in this project was performed at the x-ray synchrotron sources of EMBL/DESY and SRS. However as meaningful data were obtained only from the former, only the X33 beamline at EMBL/DESY ([http://hasylab.desy.de/facilities/doris\\_iii/beamlines/e6000/index\\_eng.html](http://hasylab.desy.de/facilities/doris_iii/beamlines/e6000/index_eng.html)) is described in the following.

The EMBL/DESY X33 synchrotron is a bending magnet beamline equipped with tunable bent monochromators to obtain a small focus with large photon flux.

The horizontal focus monochromator is a monolithic Si (111) cut crystal that selects a fixed defined wavelength (1.5 Å) from the bending magnet radiation ( $E_c = 16.6$  keV), with the wavelength of the reflected beam defined by Bragg's law (Vachette and Svergun, 2000). Higher harmonics ( $\lambda/2, \lambda/3, \dots$ ) are eliminated by total reflection at the glancing angle using a rhodium coated Zerodur mirror with gravimetrical bending for vertical focus. The typical vertical beam displacement is of the order of 30 mm and hence considerably reduces the background radiation from the storage ring (Vachette and Svergun, 2000). A triangular monochromator (with a fixed energy of 8 keV) is also used for selecting a single Bragg reflection to obtain a more focussed beam (Lemonnier, et al., 1978) (Fig. 2.8). Parasitic scattering around the beam is reduced by using several guard slits made from highly absorbing material like tungsten or tantalum (Vachette and Svergun, 2000).

The SAXS camera comprises three slits (two collimating and one guard) that confine the region around the primary beam, thus defining the beam-stop and the minimum observable scattering angle. The sample position is fixed while the sample-to-detector distances can be adjusted from 0.9-4 m by automatic movement of the MAR 345 detector inside a vacuum tube (Fig. 2.8) ([http://hasylab.desy.de/facilities/doris\\_iii/beamlines/d12\\_embl\\_x33/experimental\\_station/index\\_eng.html](http://hasylab.desy.de/facilities/doris_iii/beamlines/d12_embl_x33/experimental_station/index_eng.html)). The beam size on X33 is typically  $1.2 \times 0.21$  mm<sup>2</sup> (v x h) at the sample position ([http://hasylab.desy.de/facilities/doris\\_iii/beamlines/d12\\_embl\\_x33/beamline/index\\_eng.html](http://hasylab.desy.de/facilities/doris_iii/beamlines/d12_embl_x33/beamline/index_eng.html)). Sample cells used are made of mica windows (thickness 10-20 µm) and are connected to a thermostat (-30 to 220°C) under vacuum to control the temperature of the sample cells during the course of the experiment (Vachette and Svergun, 2000). The X33 beamline consists of two detectors: the MAR 345 image plate detector for SAXS and the linear gas filled delay line detector for wide angle neutron scattering (WAXS). These SAXS and WAXS detectors in turn give resolutions of 700 to 10 Å and 25 to 6 Å, respectively ([http://hasylab.desy.de/facilities/doris\\_iii/beamlines/d12\\_embl\\_x33/experimental\\_station/index\\_eng.html](http://hasylab.desy.de/facilities/doris_iii/beamlines/d12_embl_x33/experimental_station/index_eng.html)). The entire beam layout is made of 2 parts: the optical and the experimental hutch (Fig. 2.8). Entrance to the experimental hutch (where the samples are placed) is possible only when the beam shutters are closed, the opening of which is regulated by an interlock system. The movement of various elements like the slits, sample holder,

detector etc. for instrument alignment and operation are motorised and can be regulated via remote control. Data acquisition is achieved by CANBus control electronics driven by a workstation ([http://hasylab.desy.de/facilities/doris\\_iii/beamlines/d12\\_embl\\_x33/beamline/index\\_eng.html](http://hasylab.desy.de/facilities/doris_iii/beamlines/d12_embl_x33/beamline/index_eng.html)). Further processing of the data obtained was carried out as described in sections 2.3.1.1-2.3.1.3.



**Figure 2.8** Instrumentation layout and experimental hutch of the X33 beamline at the EMBL/DESY synchrotron

Image was taken from [http://hasylab.desy.de/facilities/doris\\_iii/beamlines/d12\\_embl\\_x33/beamline/index\\_eng.html](http://hasylab.desy.de/facilities/doris_iii/beamlines/d12_embl_x33/beamline/index_eng.html)

## 2.3.2 Neutrons

### 2.3.2.1 Small angle neutron scattering (SANS)

Although the physical mechanisms of elastic x-ray (SAXS) and neutron (SANS) scattering by matter are fundamentally different the mathematical formalism and the underlying principles are the same. Hence, all methods described in section 2.3.1 also apply to the treatment and analysis of SANS data. However, additional information on macromolecular complex formation can be obtained from SANS experiments when compared to SAXS.

While x-rays are scattered by the electrons in the particle, the interaction of neutrons is primarily with the nucleus. Neutron scattering depends on the nuclear mass, spin, and on the isotope present (Byron and Gilbert, 2000). The spin here represents the angular momentum of the nucleons (both protons and neutrons have spin 1/2). Nuclear spin results in two types of scattering: coherent and incoherent, the levels of which depend on the net spin of the nucleus concerned. While the coherent scattering provides information on structure and mass, molecular dynamics information is obtained via incoherent scattering (Byron and Gilbert, 2000). Furthermore, neutron scattering can be inelastic or elastic depending on kinetic energy transfer having taken place or not during the reaction, respectively (Svergun and Koch, 2003). SANS studies are based on the coherent elastic scattering of neutrons. As SANS causes no radiation damage, it offers advantages over SAXS. In addition, widely available quartz cuvettes are used as sample containers.

The probability of scattering from a given nucleus is given by its scattering length  $b$  ( $f_n/V$ , in units of  $10^{-15}$  m or fm), that comprises two terms  $f_p$  and  $f_s$  such that  $f_n = f_p + f_s$  (Svergun and Koch, 2003). While the first term corresponds to neutron interaction with the nuclear potential, the second term denotes a scattering length dependent on the alignment of nuclear spins between the incident beam and the macromolecule under study (Stuhrmann, et al., 1986). In SANS  $f_s \cong 0$ , which results in a flat incoherent background (Svergun and Koch, 2003). In contrast to x-rays,  $f_p$  does not increase with the atomic number and is highly sensitive to the isotopic content (Svergun and Koch, 2003) as shown in

Table 2.1. Two major differences between x-ray and neutron scattering lengths are observed from Table 2.1: (1) neutrons are more sensitive to lighter atoms than x-rays and (2) there is a significant difference in the scattering lengths of hydrogen ( $-0.374 \times 10^{-12}$  cm) and deuterium ( $0.667 \times 10^{-12}$  cm) (Svergun and Koch, 2003). The former difference is exploited in neutron crystallography to localise hydrogen atoms in the crystal (Shu, et al., 2000), while the latter provides an effective tool for selective labelling and contrast variation exploited in neutron scattering and diffraction (Chatake, et al., 2003; Snijder, et al., 2003; Wall, et al., 2000; Zaccai and Jacrot, 1983).

Atom	H	D	C	N	O	P	S	Au
Atomic mass	1	2	12	14	16	30	32	197
No. of electrons	1	1	6	7	8	15	16	79
$f_x$ ( $10^{-12}$ cm)	0.282	0.282	1.69	1.97	2.16	3.23	4.51	22.3
$f_n$ ( $10^{-12}$ cm)	-0.374	0.667	0.665	0.940	0.580	0.510	0.28	0.76

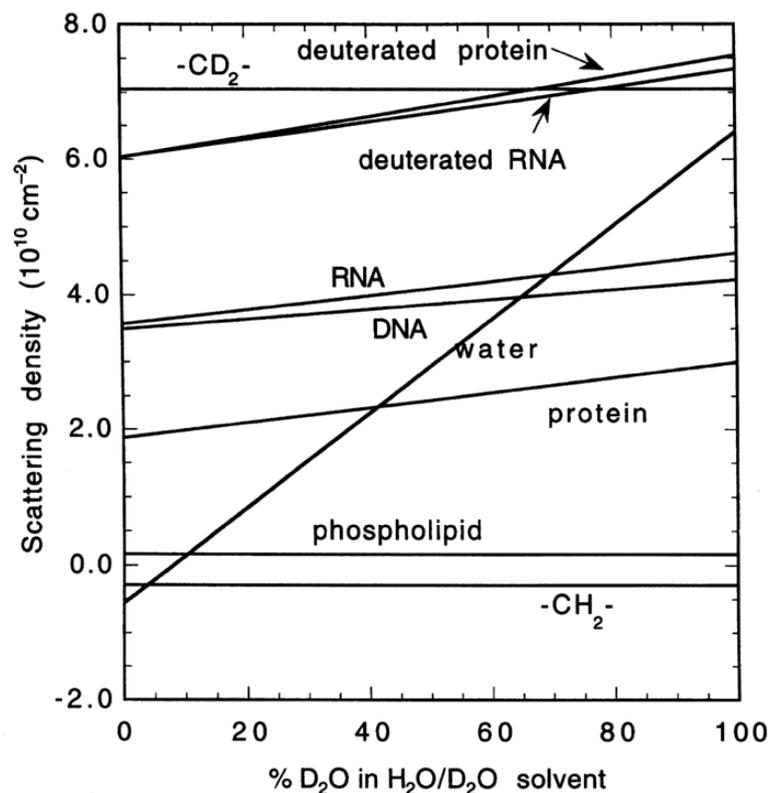
**Table 2.1 X-ray and neutron scattering lengths of some elements**

$f_x$  and  $f_n$  denote the scattering lengths of x-rays and neutrons, respectively.

### 2.3.2.2 Contrast variation

The difference in scattering lengths between hydrogen and deuterium is exploited in contrast match experiments. It is possible to render one or more components of a multicomponent system ‘invisible’ to the incident neutrons. This is achieved by solvent exchange ( $\text{H}_2\text{O}/\text{D}_2\text{O}$ ) (Ibel, 1975) or by selective deuteration of components (Engelman and Moore, 1972). In both cases the scattering density of one component of the macromolecular complex is exactly matched with that of the solvent by varying the  $\text{D}_2\text{O}$  content in the buffer. Solvent exchange is used for analysing interactions between two different molecules in a system, such as the protein-DNA complex, owing to the different match points of protein (40%) and DNA (60%) as shown in Fig. 2.9. In addition, it also enables calculation of molecular weight, the molecular shape at infinite contrast and the internal arrangement of the constituents within the complex. Moreover, the usage of  $\text{D}_2\text{O}$  results in high signal-to-background ratio, therefore experiments can be performed with low concentrations of sample.

On the other hand, selective deuteration of a specific subunit leads to the determination of its relative location within a multisubunit structure allowing, for example, investigation of protein-protein interactions. Typically, this involves the over-expression of one component in deuterated media before being reconstituted with the remaining protonated constituent(s). Exact  $D_2O/H_2O$  ratios can vary considerably for deuterated proteins, depending on the labelling efficiency. Scattering studies using  $D_2O$  are based on the assumption that the structure of the macromolecule in question is not modified by changing the contrast conditions. Mathematically, the use of various  $D_2O/H_2O$  ratios or selective deuteration of components corresponds to effective variation in the so-called contrast that is related to the molecule and solvent scattering densities, which in turn is proportional to the scattering intensity of the molecule.



**Figure 2.9 Scattering length densities of biomolecules in solvents of various  $D_2O/H_2O$  ratios**

The scattering length density for  $H_2O$  and  $D_2O$  are close to  $-5 \times 10^9 \text{ cm}^{-2}$  and  $64 \times 10^9 \text{ cm}^{-2}$ , respectively. The scattering length densities vary in a linear fashion with no scattering observed at about 8%  $D_2O$ . Adapted from Jacrot (1976).

### 2.3.2.3 Instrumentation

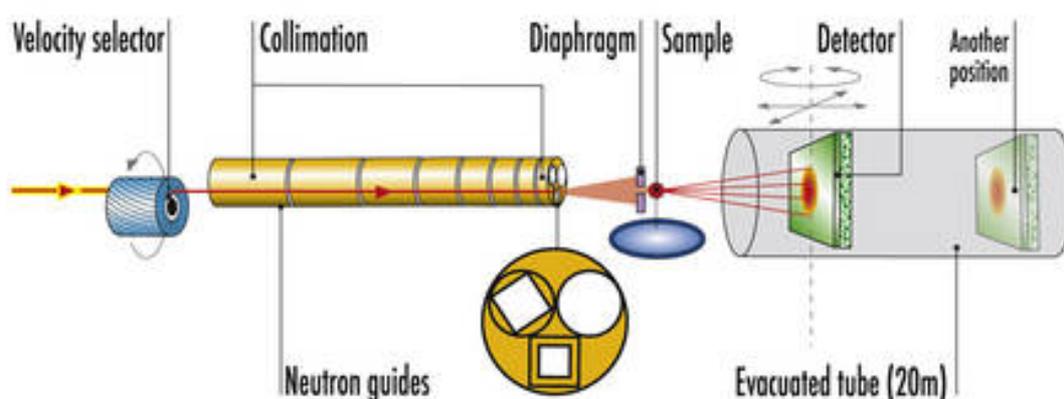
All SANS experiments in this thesis were undertaken at the D22 beamline of the Institut Laue Langevin (ILL) reactor source in Grenoble, France. The neutron wavelength is given by de Broglie's relationship

$$\lambda = \frac{396.6}{v} \quad (2.61)$$

where  $v$  is the group velocity of the neutrons ( $\text{ms}^{-1}$ ). Thermal neutrons with wavelength  $\lambda$  around 0.45-4.0 nm are typically used at the D22 beamline at ILL.

A basic description of the instrumentation on D22 is outlined below (<http://www.ill.eu/d22/characteristics/>). The fission of  $\text{U}^{235}$  in a specially designed reactor core produces high energy neutrons, which then pass from the core into a moderator (such as  $\text{D}_2\text{O}$  or graphite) at 300 K to lose energy via collision (Perkins, 1988). When these neutrons reach thermal equilibrium, further moderation is carried out using a cold source such as liquid deuterium at 25 K. This further enhances the longer wavelengths ( $\lambda = 0.5$  nm) that are useful for solution scattering and aid in the relaxation of the angular restrictions on the neutron camera (Perkins, 1988). The neutrons are then delivered by means of neutron guides from the reactor to the externally located instruments. The first section of the guide is bent in order to eliminate gamma rays and fast neutrons (Perkins, 1988). A high speed velocity selector on a rotating drum selects a narrow band of wavelength (velocity) that is transmitted and further collimated before reaching the sample (Fig. 2.10) (<http://www.ill.eu/instruments-support/instruments-groups/instruments/d22/home/>). The virtual source-to-sample distance is chosen by a collimation system consisting of eight sections. Each section comprises three tubes (<http://www.ill.eu/d22/characteristics/>), any one of which can be positioned on the beam axis. While the first tube contains a neutron guide of 40 x 55 mm; the second is equipped with an antiparasitic aperture and the third tube bears fixtures for installation of future neutron-optical equipment. D22 possesses the largest area multidetector ( $^3\text{He}$ ) of all SANS stations with an active area  $1 \text{ m}^2$  corresponding to a pixel size of 0.8 x 0.8 cm, i.e. 16 K resolution elements. The detector comprises 128 linear Reuter-Stokes detector tubes arranged vertically with a spacing of 8 mm

(<http://www.ill.eu/d22/characteristics/>). The detector moves inside a 2.5 m wide and 20 m long vacuum tube covering a wide range of sample-to-detector distances from 1.1-17.6 m. This implies a total  $q$  range of  $4 \times 10^{-4}$  to  $0.44 \text{ \AA}^{-1}$  (no detector offset) or  $0.85 \text{ \AA}^{-1}$  (with detector offset) under standard conditions (<http://www.ill.eu/d22/characteristics/>). Neutron sources are much weaker than x-ray sources and even on D22, currently the world's best SANS instrument, the neutron flux does not exceed  $10^8 \text{ neutrons cm}^{-2} \text{ s}^{-1}$  (<http://www.ill.eu/d22/characteristics/>).



**Figure 2.10** Instrumentation layout of D22 at the ILL nuclear reactor source

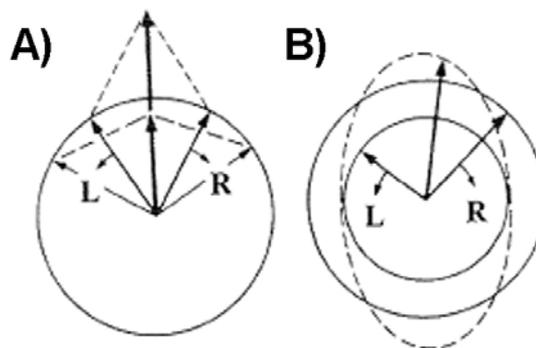
The image was taken from the website <http://www.ill.eu/instruments-support/instruments-groups/instruments/d22/home/>

### 2.3.2.4 Data treatment

The raw data obtained from SANS are reduced by circular averaging of the symmetrical scattering about the beam stop to yield a 2D scattering curve. Data normalisation is then carried out by subtracting the sample scattering curves from their respective buffer scattering curves. Resultant sample scattering curves are in turn corrected and normalised by their corresponding neutron transmissions. In addition, transmission values of sample and buffer are corrected for maximum transmission through an empty sample cell and by a standard scattering material used as reference (such as cadmium in this project). All the above processing of raw data was carried out using GRASP ([http://www.ill.eu/sites/grasp/grasp\\_main.html](http://www.ill.eu/sites/grasp/grasp_main.html)) written by Charles Dewhurst, ILL, to obtain the scattering curves. Further treatment of data to obtain the  $p(r)$  function and *ab initio* models was carried out as described in section 2.3.1.

## 2.4 Circular dichroism (CD)

Since the late 1980s CD has become an increasingly valuable tool to probe the structure of macromolecules in solution. Plane polarised light is made up of two circularly polarised components, one rotating clockwise (right handed, R) and the other counter-clockwise (left handed, L). CD is defined as the differential absorption of these two components (Kelly, et al., 2005). If L and R are equally absorbed, the resulting radiation is plane polarised, while unequal absorption of L and R gives rise to elliptical polarisation (Kelly, et al., 2005) (Fig. 2.11). A CD signal is observed when a chromophore is chiral or optically active. Spectropolarimeters measure the difference in absorbance between the L and R components ( $\Delta A = A_L - A_R$ , usually about  $10^{-3}$  for biological samples), and is represented as ellipticity ( $\theta$ ) in degrees. A simple numerical relationship connects  $\theta$  and  $\Delta A$  given by,  $\theta = 32.98 \Delta A$  (Kelly, et al., 2005). The CD spectrum is obtained with the measurement of dichroism as a function of wavelength.



**Figure 2.11** Origin of the CD effect

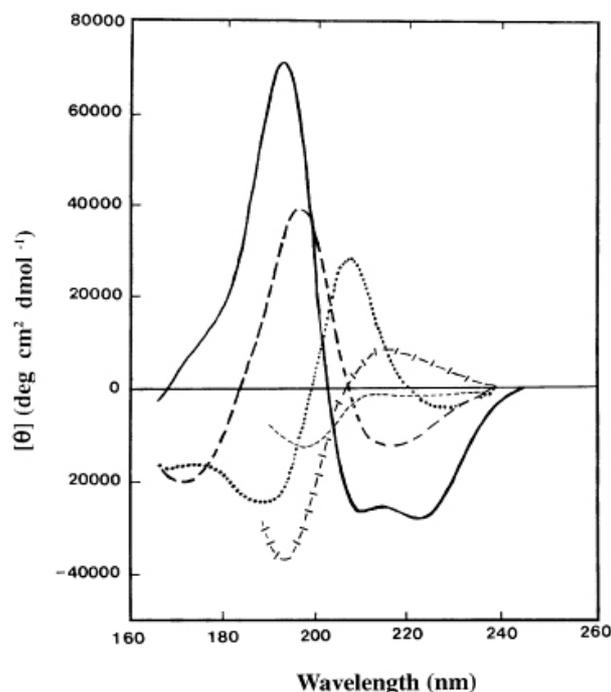
The left (L) and the right (R) circularly polarised components of plane polarised radiation are shown. When **(A)** both components have same amplitude, they combine to generate plane polarised radiation. On **(B)** possessing different amplitudes, L and R combine to give rise to elliptically polarised radiation (represented by the dashed ellipse). Taken from Kelly, et al. (2005).

In proteins, the absorption by various chromophores such as peptide bonds (absorption below 240 nm), aromatic amino acid side chains (absorption in the range 260-320 nm), disulphide bonds (weak absorption around 260 nm) results in specific spectral bands, generating signature patterns that reveal useful structural information. In addition, other non-protein cofactors (e.g. flavins,

haem groups etc.) can also absorb over a wide spectral range (Berova, et al., 2000).

The primary structural information obtained from the CD of proteins is related to its secondary structure and tertiary structural characteristics. While the former is due to the strong absorption of the peptide bonds at 240 nm and below, the latter arises from the aromatic amino acid absorptions between 240-290 nm (Kelly, et al., 2005). In addition, conformational changes in proteins upon ligand binding can also be extensively probed.

Secondary structure estimates are obtained from CD signals in the spectral range of 170-240 nm, also known as the far-UV region. The different types of regular secondary structure found in proteins give rise to characteristic CD spectra in the far-UV as shown in Fig. 2.12.



**Figure 2.12 Far-UV CD of various secondary structural components**

Various types of secondary structures produce characteristic far UV CD spectra as shown.  $\alpha$ -helix (solid line), anti-parallel  $\beta$ -sheet (long dashed line), type I  $\beta$ -turn (dotted line), extended  $3_1$ -helix or poly (Pro) II helix (cross dashed line) and disordered irregular structure (short dashed line). Taken from Kelly, et al. (2005).

The tertiary structure fingerprint on the other hand is obtained from the absorption of the aromatic amino acids in the near-UV region (240-290 nm). Each

of these amino acids tends to have a characteristic profile: tryptophan shows a peak at 290 nm with fine structure between 290-305 nm, tyrosine shows a peak between 275-282 nm with a shoulder at longer wavelengths and phenylalanine shows weak but sharp bands with fine structure between 255-270 nm (Kelly, et al., 2005). The actual shape and magnitude of the near-UV CD spectrum varies from protein to protein depending on the numbers of each aromatic amino acid present, their mobility, the nature of their environment (hydrogen bonding, polarisability etc.) and their spatial disposition within the protein.

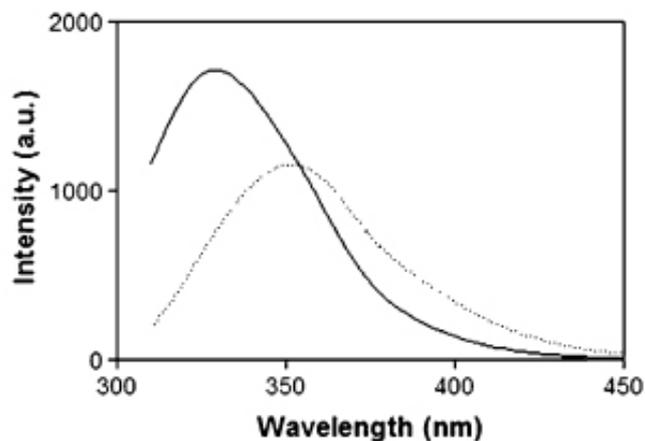
Changes in protein conformation can occur due to the binding of ligands/cofactors or due to folding or unfolding. Structural changes in proteins caused by the binding of ligands have been successfully detected by near- and far-UV CD (Hope, et al., 1996; Pandya, et al., 2004). Refolding of proteins under denatured conditions (in the presence of urea or guanidinium chloride (GdmCl) has also been extensively studied using CD (Dobson, 2004; Jemth, et al., 2004; Radford, 2000). These refolding studies have indicated that while small proteins (less than 100 amino acids) generally fold via a two state mechanism with no intermediates, the folding of larger proteins often involves a multi-stage pathway via intermediates thought to be of the 'molten globule' type, thus offering insights into the overall stability of protein folding (Kelly, et al., 2005). The accurate determination of protein concentration along with high purity and minimal nucleic acid contamination are important pre-requisites for structure determination via CD.

## 2.5 Tryptophan fluorescence

Fluorescence spectroscopy constitutes the most widely used experimental technique in the field of protein folding and conformational dynamics. Fluorescence can be extrinsic or intrinsic, the former due to external probes attached to the molecule, while the latter arises as a result of the number of aromatic residues (tryptophan, tyrosine and phenylalanine) present as well as their environments within the protein of interest. Only intrinsic fluorescence due to tryptophans (Trp) will be described here as it is within the scope of this thesis. Fluorescence has several advantages over other biophysical techniques that probe protein folding: fluorescence signals are sensitive to the local

environment around the Trp and can drastically change under certain circumstances; the signal-to-noise ratio is very high, therefore allowing small quantities of sample to be used; and the time scale of detection is in the low nanosecond range, which is faster than the time scales for folding and conformational transitions (Royer, 2006).

As only a small number of Trps are present in most proteins, specific local structural information can be obtained by exciting them at 295 nm and measuring their fluorescence over a wide spectral range (typically 300-450 nm). Owing to its aromatic character, Trp is often (although not always) found fully or partially buried in the interior hydrophobic parts of the protein, at the interface between subdomains or at the subunit interface in oligomeric systems (Royer, 2006). Upon disruption of the tertiary or quaternary structure, the side chains of the tryptophan become more exposed to solvent as shown in Fig. 2.13. The fully or partially buried tryptophans in the native protein are blue shifted resulting in emissions between 309-335 nm, while solvent exposure brings about a red shift with emission at longer wavelengths (355 nm) (Royer, 2006).



**Figure 2.13 Intrinsic tryptophan fluorescence**

Tryptophan emission of P13<sup>MTCP1</sup> in buffer (full line) indicates blue shift, while in 3 M GdmCl (dotted line) complete solvent exposure of the intrinsic tryptophans brings about a red shift. Image was taken from Royer (2006)

Although the energy of even partially buried tryptophans will invariably shift to the red upon unfolding, it is not possible to predict the quantum yield/total intensity of emission upon solvent exposure. This is primarily due to several amino acid chains, peptide bonds, prosthetic groups (e.g. haem groups, NADH

etc.) or other neighbouring tryptophans present in the protein that act as efficient quenchers of tryptophan fluorescence via excited-state proton or electron transfer (Adams, et al., 2002; Chen and Barkley, 1998). Emission wavelength-dependent fluorescence studies during unfolding or refolding with denaturants like urea/GdmCl is carried out to assess the conformational stability of proteins (Beechem, 1992; Knutson, et al., 1982).

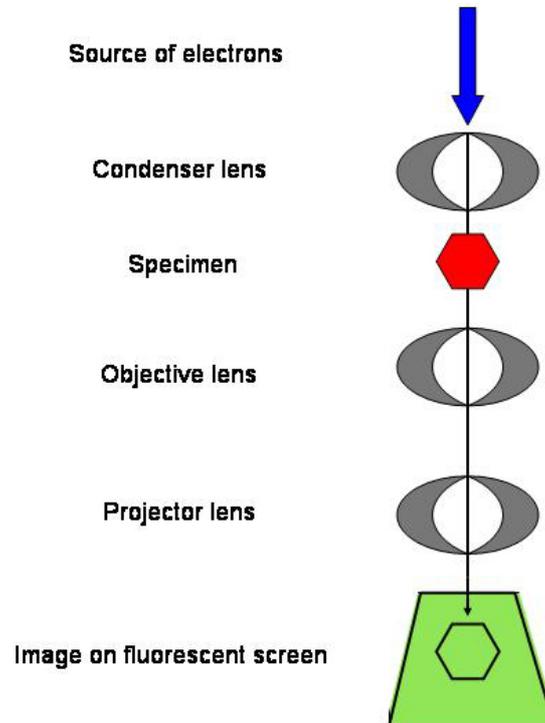
## 2.6 Negative stain electron microscopy (EM)

EM offers structural information complementary to other techniques such as x-ray crystallography, SAXS/SANS and nuclear magnetic resonance (NMR) with resolution for biological specimens in the order of 10 Å (Serdyuk, et al., 2007). The principle of EM is based on the wave like behaviour of electrons. The operation of EM is similar to light microscopy with electrons instead of photons, focused by electromagnetic lenses. The observed structural features in the EM image are due to the interaction of the electrons with the electrostatic potential distribution of the atoms in the sample (Serdyuk, et al., 2007).

EM is used to probe the structural and dynamic aspects of macromolecular assemblies. It can deal with particles of all shapes, ranging from asymmetric ones to helical filaments. In addition, 3D structural reconstructions at nm resolution can be obtained for molecules over a wide molecular weight range (500 kDa-several MDa). However, EM images contain considerable noise that arises due to the statistical fluctuations and inelastic scattering of electrons as they travel through the sample. In EM of biomolecules, the inelastic scattering outnumbers the scattering events by a factor of 3-4 (Serdyuk, et al., 2007). This problem is usually overcome by using energy filters or increasing the accelerating voltage. Current EM microscopes encompass thermoionic or field emission guns that serve as the primary source of electron beams (Serdyuk, et al., 2007).

Negative staining EM is carried out on the transmission electron microscope (TEM), where the interference between electrons that have passed through and interacted with the sample and those that were unaffected by the sample produces a phase contrast image (Fig. 2.14). The interaction of the electrons

with the sample is purely electrostatic, and therefore the EM image reflects the electrostatic potential distribution within the sample. The contrast in the image is approximately proportional to the atomic number of the atoms imaged (Serdyuk, et al., 2007).



**Figure 2.14 Schematic diagram of a transmission electron microscope (TEM)**  
Image was adapted from Serdyuk, et al. (2007)

Negative staining has been widely applied to the study of biological samples. Being relatively simple and rapid, it helps determine the overall morphology of the macromolecule via high-contrast images, whose 3D reconstructions result in structures with about 15 Å resolution (Serdyuk, et al., 2007). As the main constituents of biomolecules, namely oxygen, carbon and nitrogen do not interact strongly with electrons, samples are negatively stained by embedding in an electron-dense material (like ammonium molybdate). As a result the sample in the micrograph appears as a light region surrounded by a dark background originating from the stains. The particle itself is not observed but its surrounding stain is, and hence its structure is inferred from the distribution of the heavy stain (Serdyuk, et al., 2007).

## Chapter 3

# Materials and Methods

### 3.1 Bacterial strains

*E. coli* strain DH5 $\alpha$  (Stratagene, USA) was used for plasmid propagation, while *E. coli* strains BL21 Star (DE3) and BL21 Star pLysS (Invitrogen, UK) were used for protein expression. The BL21 Star strains contain the  $\lambda$  DE3 lysogen that allows high expression levels of T7-regulated genes. Moreover, they also harbour the *rne131* mutation that encodes a truncated RNase E endonuclease, thus lacking the ability to degrade mRNA and resulting in an increase in mRNA stability. Additionally, the pLysS strain contains the pLysS plasmid that encodes for the T7 lysozyme and confers chloramphenicol (Cam<sup>R</sup>) resistance.

#### 3.1.1 Chemicals and standard materials

Some of the other general chemicals used in this study and the companies they were purchased from are listed in Table 3.1.

#### 3.1.2 Preparation of competent cells

Cells were made chemically competent via the calcium chloride protocol as described by Maniatis, et al.(1987b). Typically, *E. coli* strains were streaked on a Lysogeny broth (LB, also commonly known as Luria broth) agar plate and grown overnight at 37°C. Single colonies picked from the plate were cultured overnight in 5 ml LB at 37°C via a shaker incubator. The 5 ml overnight cultures were then employed to inoculate fresh 100 ml LB cultures and grown at 37°C with constant shaking until the optical density (OD<sub>600</sub>) was about 0.5 AU. The cultures were cooled on ice for 5 min and thereafter spun at 655 g, 4°C for 10 min in an Allegra™ 6R centrifuge. The supernatant was discarded and the pellet gently

Chemicals/Kits	Source
20 MC metal chelate resin, 20 HQ high capacity quaternized polyethyleneimine mono anion exchange resin, nitrocellulose membrane (ECL Hybond), thrombin protease.	Amersham, USA
ECL western blotting detection reagents, low molecular weight SDS marker kits, Amplify™.	GE Healthcare, UK
Ampicillin, kanamycin, chloramphenicol, rubidium chloride, lipoic acid, reduced form of nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide (NAD <sup>+</sup> ), leupeptin, deoxyribonuclease (DNase), ethidium bromide, glutaraldehyde, 3-(N-Morpholino) propane sulfonic acid (MOPS), benzamidine, polyethylene glycol 6000 (PEG 6000), Coomassie blue dye, phenylmethanesulphonyl fluoride (PMSF), and N,N,N',N'-Tetramethylethylenediamine (TEMED).	Sigma, UK
Oligonucleotide primers	MWG-Biotech, UK
QIAquick gel extraction kit and QIAexpress® Penta-Anti His HRP kit.	Qiagen, UK
Vent® DNA polymerase, T4 DNA ligase, restriction enzymes.	New England Biolabs (NEB), UK
Wizard® SV DNA Minipreps, restriction enzymes, dNTP mix, 10 kb DNA ladders, Taq polymerase, GoFlexi® Taq DNA polymerase, Pfu DNA polymerase and T4 DNA ligase.	Promega, USA
HiPrep 16/60 Sephacryl S-300 high resolution column.	Applied Biosystems, USA
Zinc chloride (ZnCl <sub>2</sub> ).	Acros Organics, USA
[N-ethyl-1- <sup>14</sup> C] maleimide, SOLVABLE™.	Perkin Elmer, USA
Halt™ protease inhibitor single-use cocktail ethylenediaminetetraacetic acid (EDTA) free (100x).	Thermo Scientific, UK
Precast NOVEX Bis-Tris gels, 20x MES SDS running buffer, BL21 Star™ (DE3) and BL21 Star™ (DE3) plysS one shot competent cells, TOPO/TA® cloning kits.	Invitrogen, UK
Protease inhibitor EDTA free mini tablets	Roche, UK
Imidazole, NaCl, KH <sub>2</sub> PO <sub>4</sub> ,	VWR (BDH Ltd), UK
Centricon plus-20/Amicon ultra concentrators (membrane cutoffs of 30 kDa and 100 kDa), 40% (w/v) acrylamide, Tris base, Triton X-100,	Fisher, UK
Bradford reagent	Bio-Rad, UK
Bacto-tryptone, bacto yeast extract, bacto-agar	Formedium Ltd, UK
1,4-Dithiothreitol (DTT), Isopropyl-beta-D-thiogalactopyranoside (IPTG)	Melford Laboratories Ltd, UK
Sodium azide (NaN <sub>3</sub> )	Fluka, UK
Polyclonal rabbit antisera of E2 and E3BP	Generated in-house at the University of Glasgow.

**Table 3.1 List of common chemicals used and their sources**

suspended in 20 ml of Buffer 1 (100 mM RbCl, 10 mM CaCl<sub>2</sub>, 50 mM MnCl<sub>2</sub>, 15% (v/v) glycerol, 30 mM CH<sub>3</sub>COOK, pH 5.8). The suspended pellet was re-spun as before and the pellet resuspended this time in 2 ml of Buffer 2 (75 mM CaCl<sub>2</sub>, 10 mM RbCl, 15% (v/v) glycerol, 10 mM MOPS, pH 6.5) via gentle pipette mixing. This mixture was then chilled for 15 min on ice and then dispensed into 100/200 µl aliquots for storage at -80°C.

### **3.1.3 Bacterial transformation**

To 50 µl of competent cells, 1 µl of the plasmid DNA was added and the sample chilled for 30 min on ice. Cells were thereafter subjected to heat shock at 42°C for 30 s (BL21 star (DE3)) or 90 s ((DE3) pLysS/DH5α) and then returned to ice for 2 min. To these heat shocked cells, 450 µl of sterile LB was added and the cell suspension was incubated at 37°C for 40-60 min with continuous shaking. The suspensions were then plated (~100-200 µl) on LB plates containing the appropriate antibiotic and incubated at 37°C for 14-16 h.

### **3.1.4 Bacterial media**

All the experiments in this study used LB as the primary medium for bacterial growth unless stated otherwise. LB was prepared by combining 10 g bacto tryptone, 10 g NaCl and 5 g bacto yeast extract per litre of distilled water. For LB agar plates, 15 g bacto-agar was additionally added to the LB. All media were autoclaved before use. The broth and LB agar plates were supplemented with ampicillin (100 µg/ml), kanamycin (25 µg/ml) and chloramphenicol (34 µg/ml) wherever necessary.

### **3.1.5 Oligonucleotide primers**

All PCR primers with the required restriction sites for cloning were ordered from MWG-Biotech, UK (section 4.2.1). The primers were designed to have a higher GC (~65%) than AT content especially in the regions flanking the restriction sites to increase the stability of binding during the amplification process.

### **3.1.6 Ethanol precipitation of DNA**

Ethanol precipitation is an efficient method to concentrate DNA, typically for sequencing. 3 M  $C_2H_3O_2Na$  (50  $\mu$ l) and cold ethanol (300  $\mu$ l) were added sequentially to 50  $\mu$ l purified DNA and kept on ice for 30 min or overnight at 20°C. The precipitated DNA was centrifuged at 14,000 g (15 min) in a benchtop centrifuge and the pellet resuspended in 250  $\mu$ l ice cold 70% (v/v) ethanol. The suspension was subjected to another cycle of centrifugation as above. The supernatant of this spin was discarded and the tubes were left to air dry for 20-30 min at room temperature. Ethanol precipitated DNA samples were sent to MWG-Biotech for sequencing and the results analysed via Gene-Jockey software.

## **3.2 Molecular Biology**

### **3.2.1 Plasmid preparation**

From 10 ml overnight cultures, plasmid DNA was extracted using the Promega Wizard SV Miniprep DNA gel purification kit according to manufacturer's instructions. DNA was eluted in 100  $\mu$ l nuclease-free water and stored at -20°C. The yield and purity of the plasmids was assessed via agarose gel electrophoresis. Plasmids used in this study are listed in Table 3.2.

### **3.2.2 Agarose gel electrophoresis**

Agarose gel electrophoresis was used to assess the purity and yield of the plasmid DNA as described by Maniatis et al. (1987a). Typically, 1% (w/v) agarose gels were run in 1x TAE buffer (1 mM EDTA, 40 mM acetic acid, 40 mM Tris, pH 7.5) with 5  $\mu$ l DNA samples mixed with agarose loading dye (0.25% (w/v) bromophenol blue, 0.25% (v/v) xylene cyanol FF, 15% (w/v) Ficoll). Gels were run at 100 V/250 mA and were subsequently stained with ethidium bromide (0.5  $\mu$ g/ml) for 30 min before visualising on the UV transilluminator.

### 3.2.3 DNA extraction from agarose gels

DNA bands cut from the agarose gels were extracted using the QIAquick gel extraction kit (Qiagen) according to manufacturer's instructions. DNA was eluted in 40 µl elution buffer (10 mM Tris-HCl, pH 8.5) and its quality and yield analysed by agarose gel electrophoresis.

Plasmid Name	Vector	Insert	Source
rE3	pET14b	Human E3, aa 36-509	A. Brown (2002)
rE211b	pET11b	Human E2, aa 87-613	A. Brown (2002)
rE214b	pET14b	Human E2, aa 87-613	A. Brown (2002)
rE3BP-28b	pET28b	Human E3BP, aa 54-501	A. Brown (2002)
tE2Li19-11b	pET11b	Human E2, aa 398-613	S. Vijayakrishnan
tE2Li19-14b	pET14b	Human E2, aa 398-613	S. Vijayakrishnan
tE2Li27-14b	pET14b	Human E2, aa 390-613	S. Vijayakrishnan
tE3BPLi30-28b	pET28b	Human E3BP, aa 245-501	S. Vijayakrishnan

**Table 3.2** Plasmids used in this study

### 3.2.4 Polymerase chain reaction

Polymerase chain reactions were carried out in a PTC 100<sup>TM</sup> thermocycler (Genetic Research Instrumentation, UK), typically in 50/100 µl volumes. The PCR reaction mixture consisted of 10x buffer (specific to the polymerase used), dNTP mix, 5' and 3' primers, DNA polymerase (Vent/GoFlexi *Taq/Pfu*), DNA template and sterile water (details of the reaction amounts and cycle times are described in section.4.2.1). Vent and *Pfu* are high fidelity polymerases possessing dual functionality, serving both as a polymerase and a highly efficient proofreading 3'→5' exonuclease. Thus, they have a 5-15 fold greater fidelity in comparison with the *Taq* polymerase and facilitate accurate amplification of target DNA over a wide size range (0.1-20 kb). The GoFlexi *Taq* polymerase is robust and allows for optimisation of magnesium used in the reaction mix, a critical step for ensuring good yield and sensitivity. PCR products thus obtained were subjected to direct cloning or alternatively cloned via TOPO/ TA kits (Invitrogen).

### **3.2.5 Restriction digestion and plasmid ligation**

Restriction digests were performed before recombining the required DNA inserts and plasmids (50  $\mu$ l) or for diagnostic purposes (10  $\mu$ l). Restriction enzymes were purchased from Promega/New England Biolabs and used as per manufacturer's instructions. Digestions were performed at 37°C for 3.5 h and the digests were subsequently analysed by 1% (w/v) agarose gel electrophoresis following subsequent purification of the digested gel bands of inserts and vectors. Ligation was carried out at various insert:vector ratios with T4 ligase (Promega/New England Biolabs) and the samples incubated 30 min/overnight at room temperature. Ligated mixtures were then transformed into chemically competent *E. coli* DH5 $\alpha$  cells and grown overnight on LB ampicillin/kanamycin plates at 37°C. Details of the restriction enzymes and ligation conditions used in this study are described in section 4.2.1.

## **3.3 Protein methods**

### **3.3.1 Over-expression and solubility**

Overnight cultures (20 ml) were set-up by inoculation of LB with the appropriate antibiotic(s) (Table 3.3) and a single bacterial colony obtained from transformation or previously made glycerol stocks stored at -80°C. Over-expressions were carried out in 500 ml LB cultures inoculated with 20 ml of the overnight cultures and required antibiotics. Bacteria were grown at 37°C to an  $A_{600}$  of 0.6-0.8 and subsequently induced with 1 mM IPTG for 3 h at 30°C or overnight at 18°C/15°C. At induction rE2/E3BP, rE2 and rE3BP-28b (see Chapter 4) were further supplemented with lipoic acid at a final concentration of 50  $\mu$ g/ml in order to maximise insertion of the cofactor. Cells were then harvested by centrifugation at 10,000 g for 12 min at 4°C (JA14 rotor, Beckman J2-21 centrifuge) and the pellets stored at -20°C. Culture samples (1 ml) taken prior to ( $t_0$ ) and after induction ( $t_4$ ,  $t_6$ ,  $t_{17}$ ), were centrifuged at 15,300 g (Sanyo Microcentaur benchtop centrifuge, UK) for 10 min, resuspended in SDS loading buffer and checked for over-expression by SDS-PAGE (see section 3.3.6.1).

Antibiotic	Plasmid	Stock solution	Concentration ( $\mu\text{g/ml}$ )
Ampicillin	pET 14b, pET11b	100 mg/ml in dH <sub>2</sub> O	100
Kanamycin	pET28b	25 mg/ml in dH <sub>2</sub> O	25
Chloramphenicol	pLysS	34 mg/ml in ethanol	34

**Table 3.3 Antibiotics and their final concentrations in over-expression studies**

Protein was solubilised by lysing pelleted cells (50 ml culture) in metal chelate binding buffer (about 3 ml) using a French press (4 times, 750 psi, Thermo Electron Corporation, UK). Lysed cells were then subjected to centrifugation at 15,300 g for 15 min (Sanyo Microcentaur benchtop centrifuge, USA). The cell debris (insoluble fraction) was removed from the protein supernatant (soluble fraction), solubilised in SDS loading buffer and protein solubility assessed by SDS-PAGE (see section 3.3.6.1).

### **3.3.2 Protein purification**

Details of all buffers used during protein purifications can be found in Chapter 4, (section 4.2.3).

#### **3.3.2.1 Cell lysis**

Frozen cell pellets were resuspended in 20 ml of metal chelate binding buffer (20 ml/500 ml of bacterial culture) and supplemented with Complete EDTA-free protease inhibitor tablets, DNase and Halt protease inhibitor cocktail (10  $\mu\text{l/ml}$  of binding buffer). Cells were lysed in a French pressure cell at 950 psi (4 times) and subsequently centrifuged at 10,000 g for 15 min at 4°C (JA17 rotor, Beckman J2-MC centrifuge). The soluble lysate was removed from the insoluble pellet and filtered via a 0.2  $\mu\text{m}$  syringe filter (Sartorius, UK) prior to chromatography.

#### **3.3.2.2 Metal chelate affinity chromatography**

All His-tagged proteins were initially purified via metal chelate affinity chromatography on the BioCAD Sprint or BioCAD 700E workstations (Applied Biosystems, USA). Purification was carried out as follows: a 20MC (metal chelate)

column (manually packed in-house, Applied Biosystems USA) was buffer equilibrated and protein samples purified at a flow rate of 10 ml/min. The column was cleaned thoroughly with alternate washes of 4 column volumes (CV) of strip solution (50 mM EDTA, 1 M NaCl) and distilled water (dH<sub>2</sub>O) in succession for 3 cycles to remove any existing impurities from the column. The column was then primed with 35 CV of metal ions (0.1 M ZnCl<sub>2</sub>) followed by a 5 CV dH<sub>2</sub>O wash. Unbound zinc ions were then removed by a 7 CV wash step with 0.5 M NaCl, followed by equilibration with 5 CV of elution buffer (100 mM NaCl, 500 mM imidazole, 50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 6.0) and 7 CV of the binding buffer (100 mM NaCl, 10 mM imidazole, 50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 8.0). Protein samples were injected onto the equilibrated 20MC column in 5 ml aliquots, with each injection followed by a 2 CV binding buffer wash step. Following the last injection, the column was further washed with 8 CV of binding buffer and thereafter the protein eluted employing a 0-100% (8 CV) gradient step of elution buffer. Elution fractions (1.5 ml) were collected and analysed by SDS-PAGE. Pure protein fractions were then pooled and either subjected to dialysis or concentration for further purification.

### **3.3.2.3 High-capacity HQ anion exchange chromatography**

Pooled pure protein fractions from the metal chelate chromatography were buffer exchanged into dialysis buffer (2 mM EDTA, 450 mM NaCl, 25 mM Tris HCl, pH 7.5), and then subjected to anion exchange chromatography (for removal of DNA) using a high-capacity quarternized polyethyleneimine 20HQ column (Applied Biosystems, USA). The column was initially washed with 10 CV of dH<sub>2</sub>O and 5 CV of 0.2 M NaOH at 5 ml/min to remove any pre-existing impurities. This was followed by a 10 ml/min 10 CV dH<sub>2</sub>O wash and 8 CV of elution buffer (2 mM EDTA, 2 M NaCl, 25 mM Tris, pH 7.5) to remove contaminants that failed to elute in the prior NaOH wash. The column was then washed with 10 CV of dH<sub>2</sub>O followed by equilibration with 15 CV of binding buffer (2 mM EDTA, 25 mM Tris HCl, pH 7.5) at 10 ml/min. Protein was loaded onto the column via the column tubing at 10 ml/min with simultaneous collection of the flow through. The column was then subsequently washed with 10-12 CV of binding buffer and protein eluted in 2 ml fractions via a 0-100% gradient of 100 CV elution buffer. Owing to buffer exchange in high salt prior to HQ anion exchange chromatography, most of the protein (95%) passed into the flow-through with a

very small amount eluting from the column (5%). Thus, the high salt buffer exchange prior to the anion exchange chromatography was instrumental in selective removal of DNA (that remained tightly bound to the 20HQ column and was subsequently eluted), from the protein that came into the flow-through. Protein purification was analysed using SDS-PAGE and peak fractions and flow-through were pooled and concentrated for gel filtration.

#### **3.3.2.4 Gel filtration chromatography**

Gel filtration was carried out on a Sephacryl S-300 column (Amersham, USA) equilibrated with 1 CV (~120 ml) GFC buffer at 1 ml/min. The GFC buffer was either PEB (2 mM EDTA, 0.01% (w/v)  $\text{NaN}_3$ , 50 mM  $\text{KH}_2\text{PO}_4$ , pH 7.5) or PEBS100 (2 mM EDTA, 0.01% (w/v)  $\text{NaN}_3$ , 100 mM NaCl, 50 mM  $\text{KH}_2\text{PO}_4$ , pH 7.5) depending on the protein being purified (for details see section 4.2.3). Elution fractions from earlier purifications were pooled and concentrated to approximately 1 ml and injected onto the Sephacryl S-300 column. Protein elutes were collected as 2 ml fractions in GFC buffer at 1 ml/min over 1.2 CV.

#### **3.3.2.5 PDC purification from bovine heart**

PDC was purified from bovine heart essentially as described by Stanley and Perham (1980) with some modifications: 600 g of bovine heart was blended for 7 min in 800 ml of extraction buffer (2.7 mM EDTA, 3% (v/v) Triton X-100, 100  $\mu\text{M}$  DTT, 1 mM PMSF, 1 mM benzamidine, 0.2% (v/v) silicone anti-foam A, 50 mM MOPS, pH 7.0) and the final volume made up to 2 l. The homogenate was then centrifuged at 10,000 g for 20 min and the pelleted debris discarded. The supernatant fractions were pooled, its pH lowered to 6.45 with 10% (v/v) acetic acid and subjected to a first round of PEG precipitation by the addition of 0.12 vol of 35% (w/v) PEG 6000 and left stirring for 30 min on ice. This supernatant was then centrifuged at 20,000 g for 15 min, the pellets resuspended in 300 ml homogenisation buffer (2.7 mM EDTA, 1% (v/v) Triton X-100, 1.5  $\mu\text{M}$  leupeptin, 1 mM PMSF, 1 mM benzamidine, 50 mM MOPS, pH 6.8) and further clarified by centrifugation at 29,000 g for 40 min. The supernatant was filtered through 8 layers of muslin after which 0.013 vol of 1 M  $\text{MgCl}_2$  and 0.05 volumes of 1 M  $\text{Na}_2\text{HPO}_4$ , pH 6.3 was added dropwise. The pH was prevented from falling below

pH 6.8 by adjustment with 0.5 M NaOH. Adjustment to pH 6.45 with 10% (v/v) acetic acid was then carried out prior to a second round of PEG precipitation (0.12 vol of 35% (w/v) PEG 6000) and left stirring for 30 min on ice. This supernatant was centrifuged at 25,000 g for 10 min and the pellets resuspended in 150-180 ml of homogenisation buffer. Protease inhibitors (1 mM PMSF, 1 mM benzamidine, 1.5  $\mu$ M leupeptin, 100  $\mu$ M DTT) were added to the suspension which was stored overnight at 4°C. The following day, the suspension was rehomogenised before centrifuging at 29,000 g for 60 min. The pH of the supernatant collected was lowered to 6.45 with 10% (v/v) acetic acid and the third and final round of PEG precipitation (0.06 vol of 35% (w/v) PEG 6000) was carried out followed by constant stirring for 30 min on ice. This was followed by centrifugation at 29,000 g for 10 min with the pellet of this spin containing highly-enriched OGDC. PDC was obtained by further ultracentrifugation (Beckman Coulter Optima XL-80K ultracentrifuge) of the supernatant in a Ti 70 rotor at 116,000 g for 180 min. The OGDC and PDC pellets were resuspended in storage buffer (2 mM EDTA, 0.01% NaN<sub>3</sub>, 50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) overnight at 4°C. The solubilised OGDC and PDC were homogenised the next day and protein concentrations determined using the Bradford or Biuret assays. PDC and OGDC were stored in small aliquots at a final concentration of 10 mg/ml in 50% (v/v) glycerol at -20°C

### **3.3.2.6 Sucrose gradient centrifugation**

Discontinuous sucrose gradients were carried out as described by Rahmatullah, et al. (1989b) with the following changes: 4 ml of 20% (w/v) sucrose, 2 ml of 10% (w/v) sucrose and 2 ml of 5% (w/v) sucrose in PEBS2M buffer (2 M NaCl, 2 mM EDTA, 0.01% (w/v) NaN<sub>3</sub>, 50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) were layered one above the other. Protein samples (5-8 ml) were layered on the sucrose gradients and subjected to ultracentrifugation in a Beckman Ti 70 rotor at 182,000 g at 4°C for 180 min. Supernatant fractions (1 ml) were collected from the top using a peristaltic pump, and the pellet suspended in the required buffer.

### **3.3.3 Dialysis**

Buffer exchange of proteins was achieved by dialysing against 5 l of the required buffer at 4°C, with a minimum of 3 buffer changes at 2 h intervals. Dialysis was carried out using a 12,000-14,000 MW cutoff (MWCO) dialysis membrane (size 9, Dia 36/32”, Medicell International Ltd, UK). The dialysis membranes were pre-treated as described by Bollag, et al. (1996a).

### **3.3.4 Protein concentration**

Proteins were concentrated using the Centricon Plus-20/Amicon Ultra concentrators (Fisher) with the required MWCO, namely 100,000 MWCO for rE2/E3BP, rE2 and truncated tLi19/tLi30 cores; 30,000 MWCO for rE3 and rE3BP-28b; 10,000 MWCO for truncated C-terminal E3BP (tLi30). Samples were centrifuged at 3,000 rpm at 4°C in an Allegra 6R centrifuge (Beckman Coulter, UK) for several hours till the desired volume was achieved.

### **3.3.5 Polyacrylamide gel electrophoresis (PAGE)**

#### **3.3.5.1 Sodium dodecyl sulphate PAGE (SDS-PAGE)**

SDS-PAGE as described by Laemmli (1970) was used to assess the purity and concentration of proteins. Laemmli gels comprising a 4-6% stacking gel and a 10-15% resolving gel were prepared (small: 9 cm x 9 cm x 1.5 mm, large: 17 cm x 15 cm x 2 mm). Alternatively, pre-cast 4-12% gradient Bis-Tris NuPAGE gels (Invitrogen) were used according to manufacturer’s instructions. Protein samples were suspended in Laemmli sample buffer to which 150 mM DTT was added prior to boiling at 100°C for 13 min, with subsequent loading of samples (10-20 µl) onto the gel. While the homemade gels were run in SDS running buffer (Laemmli, 1970) at 400 V, 50-65 mA per gel depending on the the gel size, the NuPAGE gels were run in 2-(N-morpholino) ethane sulfonic acid (MES) running buffer (Invitrogen) at 200 V, 125 mA per gel. Gel and buffer recipes are listed in table 3.4. Gels were stained with 0.1% (w/v) Coomassie Brilliant Blue (G-250), 10% (v/v) acetic acid, 50% (v/v) methanol for about 30 min and subsequently destained with 10% (v/v) acetic acid, 10% (v/v) methanol overnight.

Recipe	Ingredients
Homemade SDS resolving gel	8-15% (w/v) acrylamide, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate (APS), 0.1% TEMED, 0.5 M Tris-HCl, pH 8.8
Homemade SDS stacking gel	4-6% (w/v) acrylamide, 0.01% (w/v) SDS, 0.1% APS, 0.1% (w/v) TEMED, 0.06 M Tris-HCl, pH 6.8
1x Laemmli SDS sample buffer	2% (w/v) SDS, 10% (w/v) sucrose, trace of pyronin Y dye, 62.5 mM Tris-HCl, pH 6.8
1x SDS running buffer	192 mM glycine, 1% (w/v) SDS, 1% (w/v) SDS, 25 mM Tris-HCl, pH 8.8
1x MES running buffer (Invitrogen)	1 M MES, 1 M Tris-base, 69.3 mM SDS, 20.5 mM EDTA
4-12% Bis-Tris gradient gel (Invitrogen)	Proprietary

**Table 3.4 SDS-PAGE gel and buffer recipes used in this study.**

### 3.3.5.2 Native PAGE

Native PAGE (Bollag, et al., 1996b) was carried out to determine the oligomeric state of the protein. This employed separation of proteins based on their size and charge. Stacking (5% (w/v) acrylamide, 0.1% APS, 0.1% (w/v) TEMED, 0.06 M Tris-HCl, pH 6.8) and resolving gels (6% (w/v) acrylamide, 0.1% (w/v) ammonium persulphate (APS), 0.1% TEMED, 0.5 M Tris-HCl, pH 8.8) were prepared and run with 1x Tris-glycine running buffer (192 mM glycine, 25 mM Tris-HCl, pH 8.8) at 200 V and a constant current of 50 mA for 3 h. Typically, protein samples were resuspended in 5x sample buffer (50% (v/v) glycerol, 0.05% (w/v) bromophenol blue, 312.5 mM Tris-HCl, pH 6.8) before loading onto the gel. Alternatively the NuPAGE Novex Tris-glycine 4% native gels (Invitrogen) were used with 1x Tris-glycine running buffer (Invitrogen) at 125 V, 6-12 mA. Gels were stained with Coomassie Brilliant Blue for 30-60 min at room temperature and subsequently destained overnight in 10% (v/v) acetic acid, 10% (v/v) methanol to enable clear protein band visualisation.

### **3.3.6 *Glutaraldehyde crosslinking***

Glutaraldehyde is an amine-reactive homobifunctional crosslinker. It reacts with primary amine groups in the protein sidechain (e.g. amines of lysine) resulting in inter- and intra-molecular covalent bonds. These covalent interactions result in the formation of higher state oligomers that are usually assessed by SDS-PAGE. Proteins (20-30  $\mu\text{g}$ ) were treated with 2% (v/v) glutaraldehyde (Sigma) and incubated for 15-20 min at room temperature. Samples were then suspended in SDS loading buffer and their oligomeric state resolved by SDS-PAGE on a 4-12% gradient gel.

### **3.3.7 *Western blotting***

Proteins resolved by SDS-PAGE were electrophoretically transferred onto nitrocellulose membrane (ECL Hybond, Amersham) at 30 V, 170 mA using the X Cell II™ Blot module (Invitrogen, UK). Transfer was carried out in the presence of transfer buffer (192 mM glycine, 20% (v/v) methanol, 25 mM Tris-HCl, pH 7.2) and pre-stained marker (BioRad, USA). The use of pre-stained marker helped confirm efficient transfer onto the nitrocellulose membrane. The methodology varied slightly from this point depending on the antibody probe being used as follows.

#### **3.3.7.1 *His-tag antibody***

The nitrocellulose membrane was washed rigorously with TBS buffer (150 mM NaCl, 10 mM Tris-HCl, pH 7.5) three times at 10 min intervals with constant shaking. The washed blot was then immersed in 10 ml blocking buffer (0.1 g blocking reagent (Qiagen), 1 ml blocking reagent buffer (Qiagen), 1% (v/v) Tween-20) with constant shaking for 1 h to remove excess binding sites. This was followed by 3 cycles of TBST buffer (20 mM Tris-HCl, 500 mM NaCl, 0.05% (v/v) Tween-20, 0.2% (v/v) Triton X-100) and a single wash of TBS buffer, each at 10 min intervals. The membrane was then incubated with a 1:2000 dilution of the QIAexpress Penta-His HRP conjugate antibody (Qiagen, UK) for 1 h with vigorous shaking. The Penta-His HRP conjugate antibody has the secondary horseradish peroxidase (HRP) antibody attached to it; hence no secondary antibody is

required. After the antibody incubation, the blot is washed with 3 cycles of TBST and 1 cycle of TBS (10 min incubations). Detection was carried out using the ECL Western blotting reagents (GE Healthcare) according to manufacturer's instructions and x-ray films (Kodak, UK) were developed at various exposures.

### **3.3.7.2 Other antibodies**

The nitrocellulose membrane was incubated in blocking buffer (15 mM NaCl, 5% (w/v) non-fat milk, 0.2% (v/v) Tween-20, 20 mM Tris-HCl, pH 7.2) for 1 h at room temperature with constant shaking. It was then subjected to two washes each for 10 min, with wash buffer (15 mM NaCl, 1% (w/v) non-fat milk, 20 mM Tris-HCl, pH 7.2). This was followed by incubation with primary antibody solution (15 mM NaCl, 1% (w/v) non-fat milk, 0.1% (v/v) Tween-20, 20 mM Tris-HCl, pH 7.2 at a 1:1000-5000 dilution of the primary antibody) at room temperature for 1 h. Excess antibody was removed by 4 wash cycles of wash buffer every 10 min, followed by incubation with anti-rabbit IgG secondary antibody conjugated to horseradish peroxidase (150 mM NaCl, 1% (w/v) non-fat milk, 20 mM Tris-HCl, pH 7.2 at a 1:2000-20000 dilution) for 1 h at room temperature. The membrane was then incubated (15 min) with wash buffer followed by a 30 min wash in high salt solution (150 mM NaCl, 20 mM Tris-HCl, pH 7.2). Detection was carried out with the ECL Western blot reagents (GE Healthcare, UK) and x-ray films (Kodak, UK) developed for suitable exposure times.

### **3.3.8 *Determination of protein concentration***

All three methods (given below) used for protein concentration determination resulted in concentration estimates that were consistent with each other. Concentration measurements by Bradford and Biuret were in good agreement (< 5%) with the  $A_{280}$  (using the calculated extinction coefficient).

#### **3.3.8.1 Protein absorbance**

Protein absorbance was measured on an Ultrospec 4300 Pro UV/vis spectrophotometer over a range of 200-600 nm. Extinction coefficients were

calculated from the protein sequences via the PROTPARAM program of the EXPASY suite (<http://expasy.org/tools/protparam.html>). The molar protein concentrations were determined by dividing the net absorbance (that measured at 280 nm subtracted from that at 310 nm) by the molar extinction coefficient at 280 nm. While all protein concentrations were based on protein absorbance at 280 nm, E3 protein concentration was measured via FAD absorbance at 450 nm. As the protein sequence of bovine E2/E3BP is unknown, the extinction coefficient of recombinant E2/E3BP was used instead to estimate its concentration. Concentrations in mg/ml were calculated by multiplying the molar concentrations by the respective molecular weights of the proteins. The extinction coefficients and molecular weights of the various proteins are listed in Table 3.5.

Protein	Molar extinction coefficient ' $\epsilon$ ' at 280 nm ( $M^{-1}cm^{-1}$ )	Molecular weight (Da)
Recombinant E2/E3BP (48/12 model)	1,951,320	$3.55 \times 10^6$
Recombinant E2	2,124,600	$3.74 \times 10^6$
Truncated tLi19/tLi30	862,200	$1.67 \times 10^6$
rE3BP-28b	20,970 (monomer), 41,940 (dimer)	51,636 (monomer), 103,272 (dimer)
tLi30	9,970 (monomer), 19,940 (dimer)	31,274 (monomer), 62,548 (dimer)
Bovine E2/E3BP	Not known	Not known
FAD	11,300	507,000
Recombinant E3	22,600	106,000

**Table 3.5** Extinction coefficients and molecular mass of proteins used in this study

### 3.3.8.2 Bradford assay

Bio-rad assays were performed to estimate protein concentrations of PDC and OGDC isolated from bovine heart. This method involves the conversion of

Coomassie Brilliant Blue G-250 from red to blue on protein binding, resulting in an increase in absorbance at 595 nm. Standard curves were plotted using BSA and IgG from which unknown protein concentrations were determined. Measurements were performed on a Shimadzu UV-2100 PC UV-Vis scanning spectrophotometer.

### **3.3.8.3 Biuret method**

The Biuret method was also used to determine protein concentration. The underlying principle is the formation of a purple complex between the protein peptide backbone and the copper ions in the Biuret reagent under alkaline conditions. Biuret reagent was prepared by initially mixing 79.7 mM  $\text{KNaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$ , 30 mM  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  and 75.3 mM KI in 100 ml of 0.2 M NaOH, and making up the final volume to 250 ml with  $\text{dH}_2\text{O}$ . While this method is extremely accurate with no interfering agents to affect protein concentration determination, it does require large amounts of material. Standard IgG (1-5 mg/ml) absorbance curves at 550 nm were measured using a Shimadzu UV-2100 PC UV-Vis scanning spectrophotometer. Unknown protein samples (1 ml) were mixed with 2 ml of Biuret reagent in 4 ml plastic cuvettes (Fisher, UK). Samples were thoroughly mixed by inverting cuvettes a few times followed by a 20 min incubation at room temperature. The absorbance (at 550 nm) of the protein-Biuret mix was determined and converted to a concentration via the IgG standard curve. Concentrations of unknown proteins were then determined from this calibrated IgG standard curve.

## **3.3.9 Determination of Radioactivity**

### **3.3.9.1 [ $^{14}\text{C}$ ] N-ethyl maleimide (NEM) labelling**

Protein samples (rE2/E3BP, rE3, bE2/E3BP and bE3) were radio-labelled with [N-ethyl-1- $^{14}\text{C}$ ] maleimide ([ $^{14}\text{C}$ ]-NEM) (0.1 mCi/ml, 33.2 mCi/mmol, PerkinElmer). 3-4  $\mu\text{l}$  of [ $^{14}\text{C}$ ]-NEM labelled protein (0.64 mM) were mixed with 0.5 mM  $\text{NAD}^+$  or 1 mM NADH and incubated at room temperature for 30 min. Reactions were terminated by the addition of 50 mM DTT and further incubated for 12 min at room temperature, after which 6  $\mu\text{l}$  SDS loading buffer was added. The labelled

samples were either immediately assessed by SDS-PAGE or stored at  $-80^{\circ}\text{C}$  for future use. SDS-PAGE was carried out on 8% (w/v) slab gels (17 cm x 15 cm x 2 mm) at 400 V and 65 mA. The SDS loading dye was left to run off the edge of the slab gels for at least 1 h, after which it was stained with Coomassie brilliant blue and destained overnight (section 3.3.6.1).

### **3.3.9.2 Scintillation counting**

The required protein bands were excised from the destained radio-labelled gel (section 3.3.11.1) and solubilised overnight in 1 ml of SOLVABLE™ (PerkinElmer, UK) at  $37^{\circ}\text{C}$ . Gel pieces of similar dimensions were cut from blank tracks and used to estimate background radioactivity. Solubilised 1 ml aqueous samples were transferred to 20 ml capped plastic scintillation vials, followed by the addition of 6 ml of EcoScint™ A scintillant (National Diagnostics, UK). Samples were thoroughly mixed by vortexing and counted using a Beckman LS 6500 scintillation counter, with a count period of 10 min/sample.

### **3.3.9.3 Fluorography**

[ $^{14}\text{C}$ ]-NEM radiolabelled samples were run on 8% (w/v) SDS-PAGE gels, stained and fixed overnight in destain solution (see section 3.3.11.1). Fluorography was performed as described by Chamberlain (1979) with the following modifications. The overnight fixed gels were immersed in about 150 ml of Amplify™ (GE Healthcare) and incubated for 30 min at room temperature. Gels were then directly dried under vacuum on Whatman No. 3 filter paper at  $80^{\circ}\text{C}$  for 2 h. The dried gel was left to cool for about 30 min before transferring it to a cassette for storage at  $-80^{\circ}\text{C}$ . Fluorographs were developed on x-ray films (Kodak, UK) using a Kodak X-OMAT S processor for exposure times ranging from 1 to 14 days.

## **3.4 Biophysical methods**

### **3.4.1 *Buffer density and viscosity calculations***

The densities and viscosities of buffers used in analytical ultracentrifugation (AUC) were calculated using the program SEDNTERP (Laue, et al., 1992) (<http://>

www.jphilo.mailway.com/) and listed in Table 3.6.

Buffer	Buffer composition	Temperature (°C)	Density 'ρ' (g/ml)	Viscosity 'η' (Poise)
PEB	2 mM EDTA, 0.01% NaN <sub>3</sub> , 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5	4	1.00774	0.01567
		20	1.00596	0.01002
PEBS100	2 mM EDTA, 0.01% NaN <sub>3</sub> , 100 mM NaCl, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5	4	1.01188	0.016141
		20	1.01009	0.010322

**Table 3.6** Calculated densities and viscosities of buffers used in AUC.

### 3.4.2 Sedimentation velocity (SV)

Sedimentation velocity (SV) experiments were carried out in a Beckman Coulter Optima XL-I analytical ultracentrifuge (Palo Alto, USA) using an An-50 Ti 8 hole rotor. Samples (360 µl) of various concentrations were loaded into 12 mm path length, charcoal-filled, epon double sector centrepieces, spun at 20,000 rpm at 4°C and a series of scans collected using either interference optics or absorbance optics or a combination of both. PEB or PEBS100 buffer were used as reference solvents depending on the experiment in question. Data were recorded over a radial range of 6.0-7.25 cm, and a radial step size of 0.002 was used in the case of absorbance optics. In the case of interference optics, 400-471 scans were recorded, 1 min apart (depending on the protein, details are described in the individual Chapters) and the laser delay was adjusted prior to the run to obtain high quality interference fringes. Data were analysed using SEDFIT (Schuck, 2000; Schuck, et al., 2002). Sedimentation boundaries were initially modelled as size based distribution of Lamm equation solutions using the  $c(s)$  analysis. Apparent sedimentation coefficients were further obtained via the non-interacting discrete species model that employs finite element analysis (see section 2.2.1.3 for details). The apparent sedimentation coefficients obtained at 4°C were converted into  $s$  values at 20°C ( $S_{20,w}$ ) via standard theory (see section 2.2.1.2.2). The sedimentation coefficient at infinite dilution  $s_{20,w}^0$  was then determined from the y intercept of a plot of sedimentation coefficient ( $S_{20,w}$ ) against concentration (nM or mg/ml). As data obtained from both SV absorbance

and interference data yielded the same results, only interference data have been presented in this thesis unless stated otherwise.

### **3.4.3 Sedimentation equilibrium (SE)**

Sedimentation equilibrium (SE) experiments were conducted in a Beckman Coulter Optima XL-I analytical ultracentrifuge (Palo Alto, USA) using an An-50 Ti 8 hole rotor at speeds of 3,000, 5,000 and 7,000 rpm. All experiments were carried out at 4°C with protein samples (80 µl) at various concentrations loaded into 12 mm path length charcoal-filled, epon double sector centrepieces. PEB or PEBS100 buffer was used as the reference solvent depending on the experiment in question (details are described in the various Chapters). After an initial delay period of 16-20 h, a series of scans (12-15) separated by 3 h was recorded using interference optics. Data were recorded over a radial range of 6.8-7.25 cm with the laser delay adjusted before the run. The program WinMATCH ([www.biotech.uconn.edu/auf/](http://www.biotech.uconn.edu/auf/)) was used to confirm that the system had reached equilibrium by measuring no appreciable change in the root mean square deviation ( $\text{rmsd} < 0.1 \times 10^{-2}$ ) as a function of time. SE data were analysed using the programs SEDPHAT (Schuck, 2003; Vistica, et al., 2004) and the Beckman Coulter XL-A/XL-I software that uses MicroCal Origin v6.0. Single data analysis was performed for every concentration within a single speed to obtain an apparent whole-cell weight average molecular weight,  $M_{app}$ . SE analysis in MicroCal Origin v6.0 included the second virial coefficient  $B$  that represents a measure of non-ideality in the system. Initial starting estimates of  $B$  were calculated using the program COVOL (Harding, et al., 1999) (<http://www.nottingham.ac.uk/ncmh/unit/method.html>) based on the surface charge and molecular mass of the protein. However, inclusion of  $B$  worsened the fits and as a result Origin v6.0 was not used for further SE analysis. Alternatively, SEDPHAT was used to carry out single fit analysis at every sample concentration via the non-interacting species model that allows fitting a maximum of four species. The average molecular weight independent of concentration was determined by plotting  $1/M_{app}$  vs concentration (nM or mg/ml), with the y intercept denoting the whole-cell average molecular weight  $M_w^0$ , at infinite dilution.

### 3.4.4 Small angle x-ray scattering (SAXS)

Small angle x-ray scattering experiments were carried out on beamline X33 of the EMBL/Deutsches Elektronen Synchrotron (DESY) at Hamburg and station 2.1 of the SRS Daresbury Laboratory, UK.

The X33 beamline at EMBL/DESY is a bending magnet beamline, with beam currents of 80-240 mA, a fixed wavelength of 1.54 Å and an electron energy of 2 GeV. Data were collected in mica sample holders for various sample concentrations (details are given in Chapter 3) and temperatures (10°C, 20°C and 37°C) at a sample-to-detector distance of 4 m, over a momentum transfer range of  $0.08 < s < 4.97 \text{ nm}^{-1}$  ( $s = 4\pi \sin \theta / \lambda$ ), where  $2\theta$  and  $\lambda$  denote the scattering angle and x-ray wavelength, respectively. The 345 mm 2D MAR image plate detector was calibrated using BSA as a standard prior to the experiment. In addition, all data scans were recorded over a period of 4 min each, with scattering data of buffer (PEBS100) collected before each sample run. Collected data were integrated, normalised to the main incident beam and detector response, and processed using the program PRIMUS (<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html>) (Konarev, et al., 2003). Radiation damage was assessed at the various temperatures by checking for a change in the normalised scattering intensity between the first and last scans. Scattering curves unaffected by aggregation were then averaged, buffer subtracted and scaled for concentration using PRIMUS. The final average buffer subtracted curve was fed into GNOM (Semenyuk and Svergun, 1991; Svergun, 1992) and the  $p(r)$  distance distribution plots and maximum dimensions,  $D_{max}$  of the proteins were determined. The radius of gyration  $R_g$  was obtained from the Guinier approximation (in PRIMUS) and GNOM.

SAXS experiments at station 2.1 of the SRS Daresbury, UK (operating at a fixed current and wavelength of 250 mA and 1.54 Å) was conducted in a similar manner as described above. However, owing to smaller detector size, two camera lengths (2.25 m and 6.25 m) were used to cover a momentum transfer range of  $0.00017 < s < 0.0057 \text{ nm}^{-1}$ . The response of the quadrant detector used was calibrated using iron as the radiation source. All experiments were conducted at 10°C with various protein concentrations (see section 3.2.4 for

details) and the data acquired over a period of 30 min. Data obtained were integrated, normalised, checked for sample aggregation and processed using XOTOKO (Boulin, et al., 1988). Protein scattering curves were buffer subtracted, and thereafter the low angle region of the low concentration curve was merged with the high angle region of the high concentration curve using Origin v 6.0. Effective merging of the curves was ascertained by scaling appropriately, resulting in good superimposition in the overlap region. The scaled scans were then averaged (PRIMUS) to produce a final scattering curve for each protein sample. Further data analysis via Guinier approximation and GNOM were performed as detailed above.

### **3.4.5 Small angle neutron scattering (SANS)**

SANS was conducted at the D22 beamline at the Institut Laue-Langevin (ILL) in Grenoble, France with a neutron wavelength of 6 Å. Samples were first measured in protein buffer (PEB/PEBS100) at detector distances of 4 m (low angle scattering) and 14 m (high angle scattering), and subsequently dialysed overnight in 40% and 100% deuterium buffer (D<sub>2</sub>O PEB), before re-measuring scattering at the same detector distances. All experiments were conducted in 1 mm path length quartz cuvettes at 4°C. Transmission and scattering data of buffer and sample at both detector distances (4 m and 14 m) were collected over a time period of 4 min and 15 min, respectively thus covering an overall  $s$  range of  $0.0034 < s < 0.143 \text{ \AA}^{-1}$ . Data were recorded on a large two dimensional area gas detector and detector response was calibrated with background, water and cadmium transmissions. More details on the instrumentation can be found in section 2.3.2. While the Unix program MAD was used to control data acquisition, the GUI program GRAS<sub>ans</sub>P written by Charles Dewhurst, ILL, ([http://www.ill.eu/sites/grasp/grasp\\_main.html](http://www.ill.eu/sites/grasp/grasp_main.html)) was employed for graphical inspection, analysis and reduction of the raw data. The  $p(r)$  distance distribution, maximum dimension ( $D_{max}$ ) and radius of gyration  $R_g$  of protein samples were obtained from programs PRIMUS and GNOM. All SANS experiments were carried out in collaboration with Dr. Phil Callow, ILL, Grenoble, France.

### **3.4.6 Negative staining electron microscopy (EM)**

Negative staining EM was conducted on a 1200 EX scanning microscope (JOEL, Japan), with an operating magnification of 30,000x and acceleration voltage of 120 kV. Samples were prepared as follows. Carbon coated copper grids (Agar Scientific Ltd., UK) were initially ionized in an ion chamber, after which 5  $\mu$ l of protein was added to the non-shiny plastic surface of the grids. The grids were then sequentially washed with three 50  $\mu$ l droplets of distilled water and one 50  $\mu$ l droplet of 2% (w/v) ammonium molybdate, pH 7.2 (negative stain, Agar Scientific Ltd., UK), all placed on a sheet of parafilm. These washes required the protein to stay afloat on the droplets for a couple of seconds without complete immersion. Excess liquid was then drained using a strip of Whatman filter paper at the edge of the grid. Grids were left to dry for a few min after which they were scanned to obtain good resolution micrographs. Stained micrographs were recorded at high magnifications (90,000x-120,000x) on Kodak S0163 film, developed at 5000 dots/inch (Nikon 4000 scanner) and the images processed using the software Digita Micrograph. Data collection and processing were performed in collaboration with Dr. David Bhella, MRC Virology Unit at the University of Glasgow.

### **3.4.7 Circular dichroism (CD)**

CD studies were performed on rE2/E3BP, rE2 and tLi19/tLi30 core samples in either PB or PEBS100 buffer on a JASCO J-810 spectropolarimeter at 25°C. Experiments were carried out at 1 nm bandwidth and scan speeds and response times of 50 nm/min and 0.5 s (far-UV), and 10 nm/min and 2 s (near-UV), respectively. Quartz cells of 0.5 cm and 0.02 cm were used for the near- and far-UV, thus covering a wavelength range of 250-320 nm (near-UV) and 180-260 nm (far UV), respectively. Data were analysed using the online data algorithm server, Dichroweb (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) in order to obtain estimates of the secondary structure content. In addition, CD experiments (far and near UV) were carried out in the presence of denaturant (0-6M GdmCl) at room temperature to assess the stability of the rE2, rE2/E3BP and tLi19/tLi30 cores. CD experiments were conducted and analysed in collaboration with Dr. Sharon Kelly, University of Glasgow.

### **3.4.8 Tryptophan fluorescence**

Tryptophan fluorescence was monitored for rE2, rE2/E3BP and tLi19/tLi30 core samples during chemical denaturation in the presence of increasing concentrations of GdmCl (0 to 6M). Proteins in PEB buffer were excited at 295 nm and the emission monitored over a range of 320-380 nm using a PerkinElmer L550B spectrophotometer at room temperature. Quartz cells (1 cm path length) were used and data analysed by software provided by the manufacturer. All fluorescence experiments were conducted and data analysed by Dr. Sharon Kelly, University of Glasgow.

## **3.5 Computational methods**

### **3.5.1 Sequence alignments**

Protein sequence alignments were carried out with the Unix version of Clustal, known as ClustalW (<http://www.ebi.gla.ac.uk/clustalw>) (Chenna, et al., 2003; Larkin, et al., 2007). Amino acid sequences were supplied in FASTA format and alignment was performed using default parameters.

### **3.5.2 *Ab initio* modelling of SAXS and SANS data**

Initial attempts on *ab initio* reconstructions employed the program DAMMIN (Petoukhov and Svergun, 2003). However, fits to the experimental data were unsatisfactory probably owing to the large size of the dummy atoms that prevent restoration of the detailed structure. DAMMIN is also known to experience problems with detailed structure restoration of complex hollow structures (Volkov and Svergun, 2003). Therefore, subsequently all *ab initio* reconstructions of molecular envelopes from SAXS and SANS data carried out in this study were generated using the program GASBOR (Petoukhov and Svergun, 2003; Svergun, et al., 2001). GASBOR restores the structure as a collection of dummy residues (DR), with one DR per amino acid. With the search volume, particle symmetry and various penalties applied, GASBOR starts with a random configuration of DRs and via a simulated annealing procedure, gradually results in a meaningful

configuration having the lowest energy. However, as the proteins in this study are massive, GASBOR runs are extremely time intensive (10 days on a Desktop PC). To overcome this problem, 100-200 simulations were conducted per protein using the computer grid system SCOTGRID ([www.scotgrid.ac.uk/](http://www.scotgrid.ac.uk/)). Managed by the Physics department at the University of Glasgow, SCOTGRID is a powerful computer hub comprising over 150 computer nodes (Dual core Opteron 280 processor, 2.4 GHz) that are connected as a grid and fed into 8 powerful servers and 10 large data storage nodes. Each simulation on SCOTGRID took about 7 days to complete on a single computer node. GASBOR models were superimposed and averaged by DAMAVER (Volkov and Svergun, 2003). Further details on the *ab initio* modelling can be found in sections 2.3.1.3, 5.2.6 and 6.2.4.

### **3.5.3 Homology modelling**

A homology model of the recombinant human truncated E2 (tE2) was obtained from SWISS-MODELLER server (<http://swissmodel.expasy.org/>) (Arnold, et al., 2006; Guex and Peitsch, 1997; Peitsch, et al., 2000; Schwede, et al., 2003). The E2 crystal structure of *B. stearothermophilus* PDC (PDB ID 1B5S) (Izard, et al., 1999) was used as the template and the model submitted via the project mode in the program Swiss-PDBViewer (spdbv; <http://www.expasy.org/spdbv>) (Guex and Peitsch, 1997). The amino acid sequence (FASTA format) was loaded into spdbv along with the bacterial E2 crystal structure template (1B5S.pdb), and fitted to the template via the 'Fit raw sequence command'. After checking for good alignment the project was submitted via the 'submit modelling request' option in spdbv. The human tE2 model obtained from SWISS-MODEL was a pentamer and the complete oligomeric 60-mer was built using crystallographic symmetry in Pymol (Delano Scientific, USA). All structures were visualised using VMD (<http://www.ks.uiuc.edu/Research/vmd/>) (Humphrey, et al., 1996), Swiss-PDBViewer (Guex and Peitsch, 1997) and Pymol (Delano Scientific, USA).

### **3.5.4 Hydrodynamic modelling**

The programs HYDRO++ (García de la Torre, et al., 2007; García de la Torre, et al., 1994) and SUPCW (Spotorno, et al., 1997) were used to calculate the hydrodynamic properties of atomic resolution structures. HYDRO++ is an

improved version of an earlier program HYDRO with enhanced calculations for rotational properties and intrinsic viscosities. Hydrodynamic parameters including the sedimentation coefficient, were obtained by providing a hydrodynamic bead model (HBM) generated from atomic coordinates as an input to HYDRO++ and SUPCW. HBMs generated using programs AtoB (Byron, 1997) implemented within SOMO (Rai, et al., 2005) and TRANS2VORONOI (developed by M. Nöllmann, Centre de Biochimie Structurale Montpellier, France) were subsequently used as input for HYDRO++ and SUPCW, respectively.

The HBM essentially is a representation of the macromolecule by an assembly of beads (or spheres) for which hydrodynamic parameters are computed. In this thesis, AtoB was used to construct HBMs from *ab initio* models obtained from SAS data as well as atomic resolution structures, while TRANS2VORONOI employed only the *ab initio* SAS models for its HBM generation. The atomic coordinates of the macromolecule supplied to AtoB are enclosed within a cuboid whose length equals the maximum dimension of the molecule. The cuboid is in turn divided into a number of cubes of user-specified dimension (the nominal resolution of the HBM), and spheres with radii proportional to the cube root of the mass contained within that cube are positioned at the centre of gravity (Byron, 2008). The HBMs thus obtained from AtoB comprised overlapping beads of equal radii. TRANS2VORONOI is essentially an extended version of AtoB and employs a similar cubic grid for bead modelling. Anhydrous sedimentation coefficients obtained from HYDRO++ and SUPCW, were adjusted for hydration as described by hydrodynamic theory (see section 2.2.1.2.2, equation 2.33) and compared to observed experimental values. More details on the modelling procedures can be found in the methods sections of Chapters 5 and 6.

### **3.5.5 Calculation of scattering curves from crystal structures**

Scattering curves for high-resolution crystal structures or models and their fits to experimental scattering curves were obtained using the program CRY SOL (Svergun, et al., 1995) or CRYSON (Svergun, et al., 1998) with default parameter settings. The solvent contribution in scattering of a macromolecule consists of two terms, one due to the excluded volume (i.e. volume inaccessible to the

solvent) and the second due to the difference in scattering density of the solvent in the hydration shell and in the bulk (see section 2.3.1.2). CRY SOL and CRY SON take into account this scattering from the excluded volume as well as the hydration shell. CRY SON computes an angular envelope function of the particle with a 0.3 nm thick hydration layer with variable density (Koch, et al., 2003). In addition, the use of multipole expansion analysis considerably speeds up the computation process. Thus, on supplying the atomic coordinates, CRY SOL predicts the scattering profile or fits the experimental data by adjusting the particle excluded volume and density of the hydration layer (Koch, et al., 2003). CRY SON is an analogue of CRY SOL that computes neutron scattering curves taking into account contributions due to H/D exchange.

### **3.5.6 Superimposition of *ab initio* and crystal structure models**

Superimposition of high-resolution atomic and low-resolution solution structures was done using the program SUPCOMB (Kozin and Svergun, 2001) employing default parameters. SUPCOMB allows superimposition of multiple low-resolution structures with each other as well as with high-resolution structures. It provides the best fit by maximising the search volume and is ideal for superimposition of structures having high similarity with variable numbers of atoms or between high- and low-resolution models. The models are represented as an ensemble of points and the superimposition is achieved by minimising a dissimilarity measure between two models, termed the normalized spatial discrepancy (NSD). For every point in the first model, the minimum value among the distances between this point and all points in the second model is found, and the same is done for the points in the second model (Koch, et al., 2003). All these distances are then added and normalized against the average distances between neighbouring points for the two models. The NSD value is close to 1 if two models are very similar (Koch, et al., 2003).

### **3.5.7 Mathematical modelling of core subunit organisation**

The subunit organisation of the E2/E3BP core was mathematically modelled in

collaboration with Prof. David Gilbert (Bioinformatics, University of Glasgow) and Prof. Peter Kropholler (Department of Mathematics, University of Glasgow). The algorithm involves modelling of the subunit organisation based on the trimeric units that make up the icosahedral E2/E3BP core. As the E2/E3BP core is a 60-meric icosahedron, it comprises 20 trimers that form the vertices of the pentagonal dodecahedron and 30 edges that bridge these trimers together. In the 60-meric E2 core, all the 20 trimers are homotrimers comprising only E2 subunits. However, the introduction of E3BP into the E2 only core (on the basis of the substitution model) results in heterotrimers (2E2 and 1E3BP) being formed in addition to the E2 homotrimers. Two important constraints are placed during the modelling process. The first one being that only one E3BP is present per heterotrimer and the second one being that the interaction of E3BP is mediated only by its dimerisation on the core surface, i.e. an E3BP that is part of one heterotrimer can only interact with a neighbouring E3BP that is part of another heterotrimer. No direct interactions between E3BP and E2 are permitted between neighbouring trimers. On this basis, one can theoretically envisage core subunit organisation ranging from 60E2 + 0E3BP to a maximum of 40E2 + 20E3BP. The faces of the dodecahedron are defined in terms of the number of possible E2 edges with the above constraints. Thus, the 3-edge, 4-edge and 5-edge faces of E2 are represented as unknown variables X, Y and Z. Linear equations based on these variables with the above constraints are then simultaneously solved to give a list of solutions for X, Y and Z for the various E2+E3BP subunit organisations. The linear equations used for modelling are given below,

$$X + Y + Z = 12$$

$$40 \leq 3X + 4Y + 5Z \leq 60$$

Solving these linear equations simultaneously results in the equation,

$$4 \leq Y + 2Z \leq 24$$

where Y is even

The algorithm was implemented using a logic programming language called Sicstus Prolog (<http://www.sics.se/isl/sicstuswww/site/index.html>) that uses constraints over finite domains. The solutions were generated by the constraint solver.

### 3.6 Nomenclature for stoichiometry experiments used in this study

The nomenclature adopted for the various stoichiometry experiments conducted in this project were slightly different and is explained below.

Although stoichiometry experiments were performed on the rE2/E3BP:rE3 complex, the binding stoichiometry (or ratio) refers to only stoichiometry between monomeric E3BP and dimeric E3 (E3BP:E3) that bind tightly to each other. In addition, concentration estimation of rE2/E3BP was based on the 48E2+12E3BP model.

- Stoichiometry experiments conducted using AUC and GFC are represented as ratios ranging from 3:1 to 1:3. In this case the ratio represents the number of rE3BP:rE3. i.e. a ratio of 1:1 represents equal numbers of rE3 and rE3BP, i.e. 12rE3BP:12rE3. Similarly, while 3:1 represents 12rE3BP:4rE3, the ratio 1:3 corresponds to 12rE3BP:36rE3.
- Experiments conducted by SANS employed deuterated E3 (dE3). A 1:1 ratio of E3BP:E3 (as described above) in this case was represented as 12+12 (see Fig. 7.4), implying a subunit composition of 12dE3 dimers per 12E3BPs (or per E2/E3BP core). SANS was carried out covering a range of E3BP+dE3 subunit compositions namely 12E3BP+2dE3 (12+2) to 12E3BP+18dE3 (12+18). The ratio of observed dE3 per rE2/E3BP core is represented as dE3:rE2/E3BP. While 12:1 represents 12dE3 dimers bound per rE2/E3BP core (i.e. 12dE3 dimers per 12E3BP (12+12), indicating a 1:1 binding ratio of dE3:E3BP), 10:1 corresponds to 10dE3 dimers per core (i.e. 10dE3 dimers per 12E3BP (12+10) or 10dE3:12E3BP)
- Stoichiometry experiments conducted via [14C]-NEM radiolabelling follows a similar nomenclature as that carried out for the AUC and GFC data. Ratios in this case correspond to the ratio of rE3BP (or core):rE3, i.e. 2:1 represents 12E3BP:6E3, while 10:1 corresponds to 12E3BP:1.2E3

## Chapter 4

# Cloning, over-expression and purification

### 4.1 Introduction

Early research in biochemistry involved extraction and purification of proteins and enzymes from natural sources by procedures that were often extremely time consuming and laborious yielding relatively small amounts of protein. However, the arrival of recombinant DNA technology in the 1970s has revolutionised the face of molecular biology and served as the foundation for the rapid development of modern biomedical research on the analysis of clinically relevant proteins. Apart from serving as a pivotal scientific research tool, recombinant DNA technology has also had profound impacts on the diagnosis and treatment of genetic disorders. Several clinically important proteins like somatostatin (Itakura, et al., 1977), human growth hormone (Goeddel, et al., 1979a), insulin (Goeddel, et al., 1979b), and follicle stimulating hormone (Olijve, et al., 1996) have been generated as recombinant products in large amounts and subsequently obtained at high levels of purity, thus advancing and enabling effective treatment of diseases.

Genes encoding recombinant proteins are usually cloned into *E. coli*, yeast or mammalian vectors and then introduced into bacterial, yeast or mammalian cells. Many different bacterial expression systems are commercially available today for large-scale expression and purification of recombinant proteins in a cost-effective manner. Moreover, the greatest advantage of this technology is in the absolute control over the DNA sequence to be engineered, thus allowing for all types of genetic manipulations. Current major applications of recombinant DNA technology include *in vitro* mutagenesis studies, gene therapy, protein engineering and large scale production, structural and functional genomics, studies on single domain or multi-domain proteins in solution and medical diagnostics.

While protein engineering has been made possible by the advent of recombinant technology, large scale purification of recombinantly overexpressed polypeptides has been facilitated by the development of several protein tags and sophisticated chromatographic techniques. Protein tags (e.g. His tag) or fusion proteins (such as maltose binding protein (MBP), glutathione S-transferase (GST)) are usually added onto the N- or C-termini of recombinant proteins aiding in shorter isolation times and greater efficiency in purification and detection. While MBP and GST fusion proteins are employed routinely for purification of small proteins or individual domains, the majority of protein purifications include the presence of a short His-tag. These tags comprise 6-10 histidine residues that are fused to the N- or C-termini of the protein via short linker regions. His-tag proteins are purified by metal chelate affinity chromatography exploiting the strong binding affinity of histidine residues for metal ions like  $\text{Ni}^{2+}$ ,  $\text{Zn}^{2+}$  or  $\text{Co}^{2+}$ . The affinity column is packed with a negatively charged resin such as imidoacetate that binds to the loaded metal ions used for priming the column. These metal ions in turn interact with the His-tagged protein of interest and elution of the protein is achieved by using a strong metal binding competitor like imidazole. All the recombinant proteins purified in this study were N-terminally His-tagged and purified by an initial metal chelate affinity chromatography step using  $\text{Zn}^{2+}$  as the metal ion. The sole protein complex to be purified from a natural source in this project was the bE2/E3BP core assembly that was obtained in large amounts from PDC purified from bovine heart.

PDC represents the largest multi-enzyme system to be recombinantly produced to date. The individual constituent enzymes of PDC (E1, E3, [E2 or E2/E3BP]), from various organisms have been successfully over-expressed by recombinant DNA technology. These include PDC from human (Brautigam, et al., 2006; Ciszak, et al., 2001; 2003; 2006; Hiromasa, et al., 2004; Quinn, et al., 1993; Smolle, et al., 2006), pig (Toyoda, et al., 1998a), *S. cerevisiae* (Stoops, et al., 1992; 1997; Toyoda, et al., 1998b), *B. stearothermophilus* (Allen, et al., 2005; Domingo, et al., 1999; Izard, et al., 1999; Kalia, et al., 1993; Lessard, et al., 1998; Lessard and Perham, 1994; Mande, et al., 1996), *E. coli* (Allen, et al., 1989; Arjunan, et al., 2002; Green, et al., 1995), *Streptococcus faecalis* (Allen and Perham, 1991), *Streptococcus pneumoniae* (Hakansson and Smith, 2007), *A. vinelandii* (Hengeveld, et al., 1997; Schulze, et al., 1991a; 1991b) and parasitic

nematodes (Chen, et al., 1998; 1999; Harmych, et al., 2002; Huang and Komuniecki, 1997; Huang, et al., 1998; Klingbeil, et al., 1996). As full-length E2 and E3BP constructs of human PDC used in this study encompass the flexible N-terminal SBDs and LDs that are not conducive to high resolution structural studies, truncated constructs comprising only the CTDs of E2 and E3BP were designed for facilitating structure determination by solution scattering and x-ray crystallography. This chapter describes the cloning, expression and purification of these truncated CTD constructs of E2 and E3BP as well as the large-scale overexpression and purification of the full-length versions of the various proteins used in this study. Over-expression of all proteins in this study was highly successful, with high yields of soluble proteins obtained.

## 4.2 Materials and methods

### 4.2.1 Cloning of C-terminal constructs: *tE2* and *tE3BP*

Primers were designed for PCR amplification of the C-terminal regions of mature E2 and E3BP (*tE2* and *tE3BP*) lacking the N-terminal LDs and relevant SBDs. However, successful cloning of these constructs and subsequent attempts at protein over-expression resulted in the production of insoluble products. Hence, primers for truncated *tE2* and *tE3BP* were re-designed and typically encompassed the C-terminal region and several additional residues of the preceding linker region that connected the SBD and CTD of these proteins. While *tE2* was cloned via the TOPO/TA cloning kit (Invitrogen), *tE3BP* was successfully cloned directly into the vector of choice via the classical cloning approach. Primers to enable isolation of two clones of *tE2*: Li19 (19 residues of the linker region, aa 398-613) and Li27 (27 residues of the linker region, aa 390-613), and two clones of *tE3BP*: Li30 (30 residues of the linker region, aa 245-501) and Li58 (58 residues of the linker region, aa 217-501) were obtained from MWG Biotech (UK). While *tE2* clones included restriction sites *NdeI* and *BamHI* for insertion into vectors pET11b and pET14b, the *tE3BP* clones comprised restriction sites *BamHI* and *XhoI* for site-directed cloning into pET28b. The various forward and

reverse primers and their restriction sites (coloured and underlined) are shown in Fig. 4.1.

Truncated tE2 Li19 with *NdeI* (cyan) and *BamHI* (blue) restriction sites

**Forward:** 5' - CCT AGT CAT ATG GCT CCT GCT CCG GC – 3'

**Reverse:** 5' - AGT TAG GGA TCC TTA CAA CAA CAT AGT GAT AG – 3'

Truncated tE2 Li27 with *NdeI* (cyan) and *BamHI* (blue) restriction sites

**Forward:** 5' - AAG AAG CAT ATG GAC TCT TTT GTG CCT AGT AAA GTT GC – 3'

**Reverse:** 5' - AGT TAG GGA TCC TTA CAA CAA CAT AGT GAT AG – 3'

Truncated tE3BP Li30 with *BamHI* (blue) and *XhoI* (violet) restriction sites

**Forward:** 5' - CCC ACA GGA TCC GTC GCC CCT ACA GGC CAC AGC – 3'

**Reverse:** 5' - TGA AGC CTC GAG CTA GGC AAG TCG GAT AGG – 3'

Truncated tE3BP Li58 with *BamHI* (blue) and *XhoI* (violet) restriction sites

**Forward:** 5' - AAA GAG GGA TCC GCT CAA ACT TGT CCA GTT G – 3'

**Reverse:** 5' - TGA AGC CTC GAG CTA GGC AAG TCG GAT AGG – 3'

**Figure 4.1** Primer sequences for the truncated CTD constructs of rE2 and rE3BP

PCR reactions were setup typically in 50-100 µl volumes. Initial PCRs employed the *Pfu* polymerase. However, owing to greater yields and higher stability, subsequent PCR reactions used either Vent or GoFlexi *Taq* polymerase unless stated otherwise. A typical 50 µl reaction included 50-100 ng of template DNA, 5 µl MgCl<sub>2</sub> (25 mM), 10x reaction buffer (5 µl), 1 µl dNTP mix (0.25 mM each of dATP, dCTP, dGTP and dTTP), 1 µl each of specific forward and reverse primers (300 nM) pre-heated to 95°C for 5 min, 1 µl Vent/GoFlexi *Taq* Polymerase (3.5 U/µl) and sterile water (dH<sub>2</sub>O). PCR cycles are shown in Table 4.1.

Step	Temperature (°C)	Time
1. Initial denaturation	95	2 min
2. Denaturation	95	1 min
3. Annealing	50	30 s
4. Extension	72	4 min
Repeat steps 2-4 for 24 cycles		
5. Extension	72	5 min
6. Cooling	4	24 h

**Table 4.1** PCR cycle reactions

PCR products obtained were analysed on a 1% (w/v) agarose gel before extracting the DNA from the gel. From this point onwards, cloning proceeded via TOPO/TA (as done for tE2) or via the direct approach (as done for E3BP). In this project, cloning was carried out for both clones of tE2 (Li19 and Li27) and only one clone of tE3BP (Li30).

- PCR amplified products of tE2 (Li19 and Li27) obtained using GoFlexi *Taq* polymerase were subjected to further TOPO/TA cloning (Invitrogen). The TOPO cloning system serves as a quick one-step method for direct insertion of PCR products containing 3' deoxyadenosine (A) overhangs (created by *Taq* polymerase) into the linearized PCR 2.1-TOPO vector having overhanging 3' deoxythymidine (T) by means of direct ligation. Typically, this involved a reaction mixture (total vol 6 µl) comprising 2 µl freshly purified PCR product (with 3' A overhangs), salt solution (1 µl), 1 µl TOPO vector (5 ng) and 2 µl sterile dH<sub>2</sub>O. Following a 25 min incubation at room temperature, 2 µl of the mix was transformed into chemically competent TOP10 cells (supplied as part of the kit). Positive colonies were selected and purified (Wizard *Plus* mini prep kit) from a 10 ml overnight LB culture (section 3.2.1).
- The PCR amplified inserts of tE3BP (Li30) obtained using the Vent polymerase were purified and subjected to further digestion.

Purified inserts (Li19, Li27, Li30) and the recipient plasmids (pET11b, pET14b, pET28b) were digested with restriction enzymes *NdeI* and *BamHI* (for Li19 and Li27, NEB) and *BamHI* and *XhoI* (for Li30, Promega). Typically, 40 µl of plasmids and purified inserts were digested in a 50 µl reaction volume with 3 µl of the appropriate restriction enzymes (10 U/µl), 5 µl of 10x buffer (specific to the restriction enzyme pair and supplied by the manufacturer) and sterile dH<sub>2</sub>O at 37°C for 3.5 h. Digested DNA (inserts and vectors) was extracted and purified (section 3.2.1 and 3.2.3) before carrying out 20 µl vol ligations at various insert:vector ratios.

Digested products corresponding to truncated E2 (Li19 and Li27) were ligated into vectors pET11b and pET14b respectively at insert:vector ratios of 1:1, 3:1 and 5:1 using 1.5  $\mu$ l of NEB T4 ligase (3 U/ $\mu$ l) and 2  $\mu$ l 10x ligase buffer, followed by incubation for 20 min at room temperature. Similarly, digested truncated E3BP (Li30) was ligated into vector pET28b at insert:vector ratios of 1.6:1, 3.6:1 and 5.6:1 with 1  $\mu$ l Promega T4 ligase and 2  $\mu$ l 10x ligase buffer at room temperature with overnight incubation. All ligation mixtures were subsequently transformed into *E. coli* DH5 $\alpha$  cells. Negative controls in the ligations contained vector DNA only.

Diagnostic digests were performed on the transformed purified ligated products using restriction enzymes *NdeI-BamHI* (for the tE2 clones) or *BamHI-XhoI* (for the tE3BP clone). Reactions were set up with 1  $\mu$ g DNA, sterile dH<sub>2</sub>O and 5 U of enzymes in 10x Promega buffer D (for tE3BP-Li30) or 10x NEB buffer 4 (for tE2-Li19/Li27) at 37°C for 2.5 h and subsequently analysed by agarose gel electrophoresis. The recombinant plasmids were then sent to MWG for sequencing to confirm the exact coding sequence.

## 4.2.2 Protein over-expression

All over-expression plasmids were transformed into BL21 star (DE3) cells except recombinant E3 (rE3) that was transformed into BL21 (DE3) pLysS cells as outlined in Table 4.2. Protein over-expression of recombinant proteins rE3, rE3BP-28b, rE2, tLi19-14b and co-expression of recombinant cores rE2/E3BP and truncated core tLi19/tLi30 were carried out using the standard procedure described in section 3.3.1.

Plasmid	<i>E. coli</i> strain	Antibiotic <sup>R</sup>	Temp (°C)	Time (h)	IPTG (mM)	Lipoic acid ( $\mu$ g/ml)
rE2	BL21 (DE3) star	Amp	18	17	1	50
rE2/E3BP	BL21 (DE3) star	Amp, Kan	15	17	0.5	50
rE3BP-28b	BL21 (DE3) star	Kan	15	17	0.5	50
rE3	BL21 (DE3) pLysS	Amp, Cam	30	4	1	None
tLi19/tLi30	BL21 (DE3) star	Amp, Kan	18	17	1	None
tLi19-14b	BL21 (DE3) star	Amp	18	17	1	None

**Table 4.2** *E. coli* bacterial strains and conditions for protein over-expression

The various antibiotics used namely, ampicillin (Amp), kanamycin (Kan) and chloramphenicol (Cam) are denoted.

While synthesis of all proteins was induced with IPTG, cultures for rE2/E3BP, rE2 and rE3BP-28b production were additionally supplemented with lipoic acid prior to induction. Over-expressions were analysed on SDS-PAGE and solubility tested as outlined in section 3.3.1.

### **4.2.3 Protein purification**

All recombinant over-expressed proteins were subjected to 2-3 rounds of chromatographic purification to obtain good yields of highly pure samples. Proteins rE3, dE3, rE3BP-28b and tLi19/tLi30 were purified by a combination of metal chelate and gel filtration chromatography. The purification of full-length cores namely, rE2/E3BP and rE2 involved a sequence of metal chelate, anion exchange and gel filtration chromatography steps. Metal chelate chromatography was carried out using 0.1 M ZnCl<sub>2</sub> as the protein binding agent while a Sephacryl S-300 column was employed for gel filtration chromatography. Technical details of the various chromatographic techniques are described in sections 3.3.2.2 to 3.3.2.4. Details of the buffers used for the various chromatographic steps are outlined in Table 4.3. Bovine E2/E3BP (bE2/E3BP) was obtained from three rounds of sucrose gradient centrifugation of purified bovine heart PDC (refer sections 3.3.2.5 and 3.3.2.6) based on the method developed by McCartney (1998). Pellets of bE2/E3BP obtained from the first two rounds of sucrose gradient centrifugation were homogenised in PEBS2M buffer (2 M NaCl, 2 mM EDTA, 0.01% (w/v) NaN<sub>3</sub>, 50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4), while the pellet obtained from the final round of centrifugation was suspended in PEB buffer (2 mM EDTA, 0.01% (w/v) NaN<sub>3</sub>, 50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) and subsequently used for further biochemical and biophysical studies.

## **4.3 Results**

### **4.3.1 Cloning of truncated clones tE2 and tE3BP**

The full length clones ORFs for native E2 and E3BP have been previously cloned into vectors pET11b/pET14b and pET28b, respectively (Brown, 2002a). However, in addition to the CTD these full-length E2 and E3BP constructs include the

Protein	Purification	Binding buffer	Elution buffer	Elution gradient
rE2/E3BP	MC	100 mM NaCl, 10 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 8.0	100 mM NaCl, 500 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 6.0	0-100% 8 CV
	HQ	2 mM EDTA, 25 mM Tris HCl, pH 7.5	2 mM EDTA, 2 M NaCl, 25 mM Tris, pH 7.5	0-100% 100 CV
	GFC	2 mM EDTA, 0.01% (w/v) NaN <sub>3</sub> , 100 mM NaCl, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5		
rE2	MC	100 mM NaCl, 10 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 8.0	100 mM NaCl, 500 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 6.0	0-100% 8 CV
	HQ	2 mM EDTA, 25 mM Tris, pH 7.5	2 mM EDTA, 2 M NaCl, 25 mM Tris, pH 7.5	0-100% 100 CV
	GFC	2 mM EDTA, 0.01% (w/v) NaN <sub>3</sub> , 100 mM NaCl, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5		
tE2/tE3BP (tLi19/tLi30)	MC	100 mM NaCl, 10 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 8.0	100 mM NaCl, 500 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 6.0	0-100% 8 CV
	GFC	2 mM EDTA, 0.01% (w/v) NaN <sub>3</sub> , 100 mM NaCl, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5		
rE3BP-28b	MC	100 mM NaCl, 10 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 8.0	100 mM NaCl, 500 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 6.0	0-100% 8 CV
	GFC	2 mM EDTA, 0.01% (w/v) NaN <sub>3</sub> , 100 mM NaCl, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5		
rE3	MC	100 mM NaCl, 10 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 8.0	100 mM NaCl, 500 mM imidazole, 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 6.0	0-100% 8 CV
	GFC	2 mM EDTA, 0.01% (w/v) NaN <sub>3</sub> , 50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5		

**Table 4.3 Purification buffers used in this project**

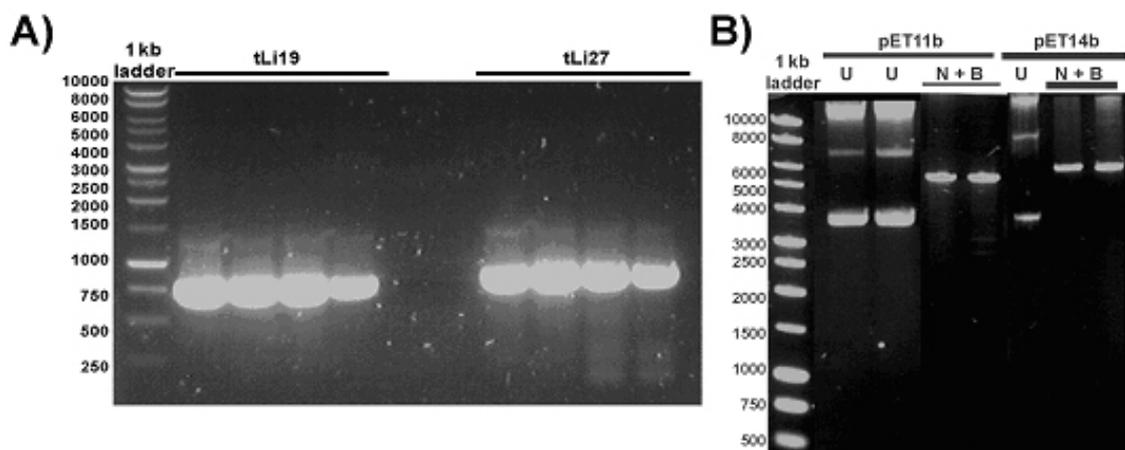
CV is column volume while MC, HQ and GFC denote metal chelate, anion exchange and gel filtration chromatographies.

highly flexible SBD and LDs that are not favourable for high resolution structural studies. Therefore, truncated constructs encompassing only the CTD with a few residues of the preceding linker were successfully cloned into the following vectors: tE2 constructs Li19 and Li27 into pET11b and pET14b, and the single tE3BP construct tLi30 into pET28b.

#### 4.3.1.1 Cloning of tE2: tLi19 and tLi27

Amplification of the DNA sequences corresponding to residues 290-613 and 298-613 was successful and yielded the expected 753bp tLi19 and 777 bp tLi27 products (Fig. 4.2A).

The amplified truncated products were then successfully cloned into the TOPO vector. Purified DNA inserts tLi19 and tLi27 from the TOPO ligations were then digested along with pET11b and pET14b vectors (Fig. 4.2B) using restriction enzymes *NdeI* and *BamHI*.



**Figure 4.2 Cloning of truncated E2 constructs – tLi19 and tLi27**

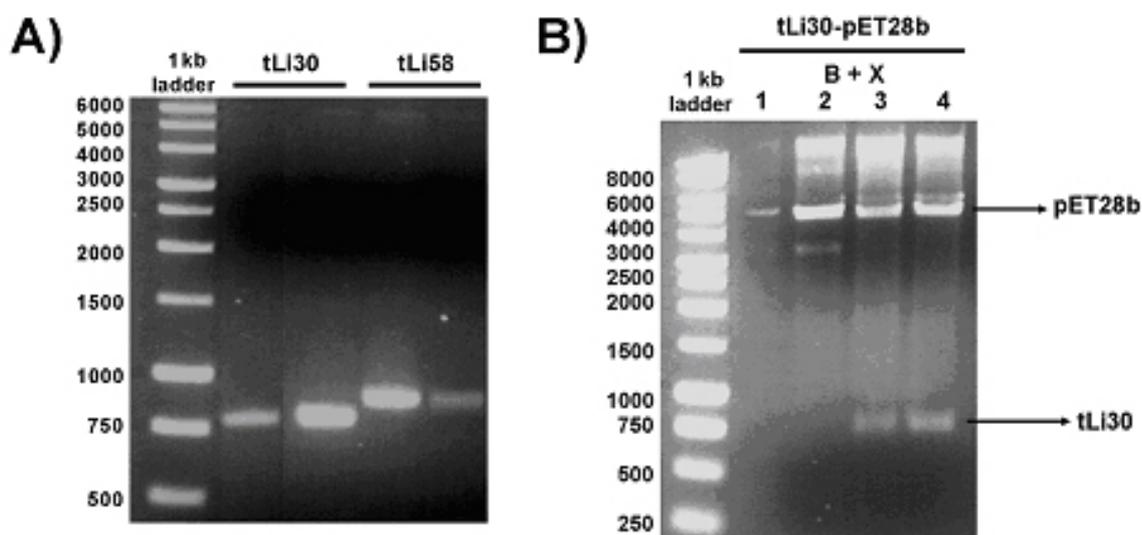
(A) PCR amplified products of tLi19 (aa 290-613) and tLi27 (aa 298-613). (B) Undigested (U) and digested pET11b and pET14b with enzymes *NdeI* and *BamHI* (N + B) is shown. Band sizes are indicated in bp.

Subsequent ligation of digested inserts (tLi19 and tLi27) and vectors (pET11b and pET14b) resulted in products tLi19-pET11b, tLi19-pET14b, tLi27-pET11b and tLi27-pET14b. The presence of the inserts was confirmed by diagnostic restriction digests with *NdeI* and *BamHI*. Furthermore analysis of the digests on a 1% (w/v) agarose gel indicated successful cloning, resulting in single bands of

753 bp and 777 bp: tLi19 and tLi27, respectively. This was further confirmed by DNA sequencing (see Appendix).

#### 4.3.1.2 Cloning of tLi30

The tLi30 DNA sequence corresponding to residues 245-501 was successfully amplified by PCR using Vent polymerase (Fig. 4.3A).



**Figure 4.3 Cloning of the truncated E3BP tLi30 construct**

**(A)** PCR amplified products tLi30 (aa 245-501) and tLi58 (aa 217-501). **(B)** Diagnostic digests of cloned product tLi30-pET28b with *Bam*HI and *Xho*I (B + X), resulting in 775 bp tLi30 are indicated. Sizes of bands are indicated in bp.

A slightly larger construct tLi58 (aa 217-501) was also amplified, but due to cloning difficulties and lack of time it was not pursued in this study. The PCR amplified inserts were digested with enzymes *Bam*HI and *Xho*I and subsequently ligated into digested pET28b via a direct cloning strategy (Fig. 4.3B), resulting in the ligated product tLi30-pET28b. Successful cloning of tLi30 into pET28b was confirmed by setting up diagnostic digests with restriction enzymes *Bam*HI and *Xho*I that yielded the expected 775 bp product (Fig. 4.3C).

DNA sequencing of the tE3BP clone (tLi30) confirmed the absence of any mutations introduced into the nucleotide sequences (see Appendix).

## **4.3.2 Protein over-expression and purification**

### **4.3.2.1 Full-length recombinant E2/E3BP core (rE2/E3BP)**

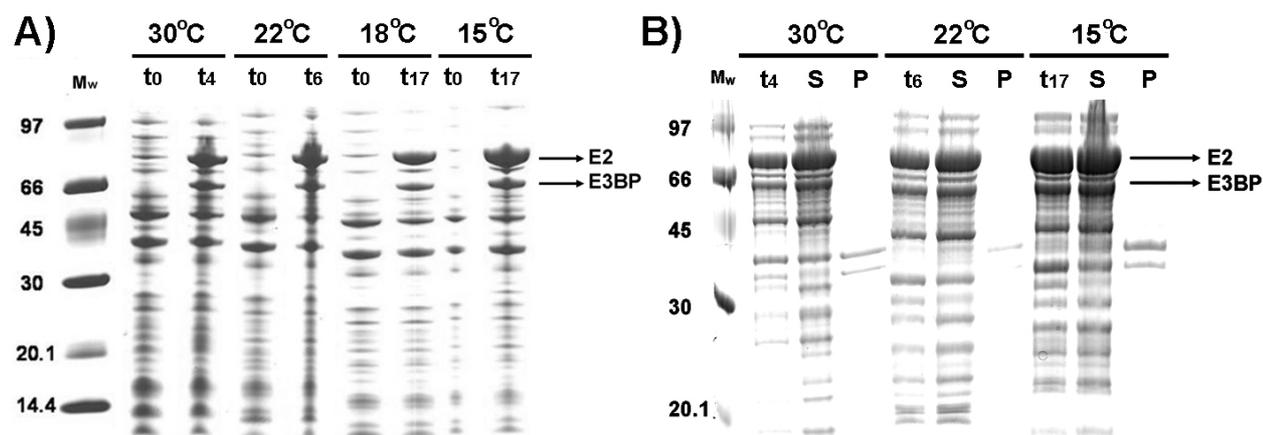
Previous experiments have revealed the essential requirement for co-expression (dual expression) of rE2 and rE3BP in order to generate a functional rE2/E3BP core. This was achieved by co-transformation of two different plasmids having the two genes of interest (rE2 in pET11b and rE3BP in pET28b) into *E. coli*. The two components, rE2 and rE3BP integrate to form a functional rE2/E3BP assembly, but do not co-associate if mixed post-translationally. Over-expression of rE2/E3BP was performed at various temperatures (30°C, 22°C, 18°C and 15°C) and IPTG concentrations (0.5 mM, 0.8 mM, 1 mM) for 4 h (30°C) or 6 h (22°C) or 17 h (18°C, 15°C) to find the best conditions for obtaining high yields of protein. While varying the IPTG concentration had no significant effect on expression, an interesting temperature dependent trend was observed, resulting in an increase in over-expression with decrease in temperature (Fig. 4.4A). In particular a relative increase in expression yields of E2 in comparison with E3BP was clearly observed with decrease in temperature. Although the various IPTG concentrations did not significantly alter the expression profile at the various temperatures, the highest levels of expression were consistently observed at an IPTG concentration of 0.5 mM (data not shown).

Solubility studies carried out across the temperature range of 30°C-15°C showed a similar trend with increasing yields of soluble protein as the temperature was lowered, thus complementing the over-expression data (Fig. 4.4B).

The E2 and E3BP migrate more slowly on an SDS-PAGE gel than would be expected for 62 kDa and 50 kDa proteins, respectively. This has been attributed to the elongated or swollen nature of the LDs (Bleile, et al., 1979) and/or anomalies in electrophoretic mobility induced by the inter-domain linkers rich in alanine and proline (Guest, et al., 1985; Miles, et al., 1988; Spencer, et al., 1984).

Purification of high quality rE2/E3BP was time consuming involving three chromatography steps. Over-expressed rE2/E3BP at low temperatures (15°C)

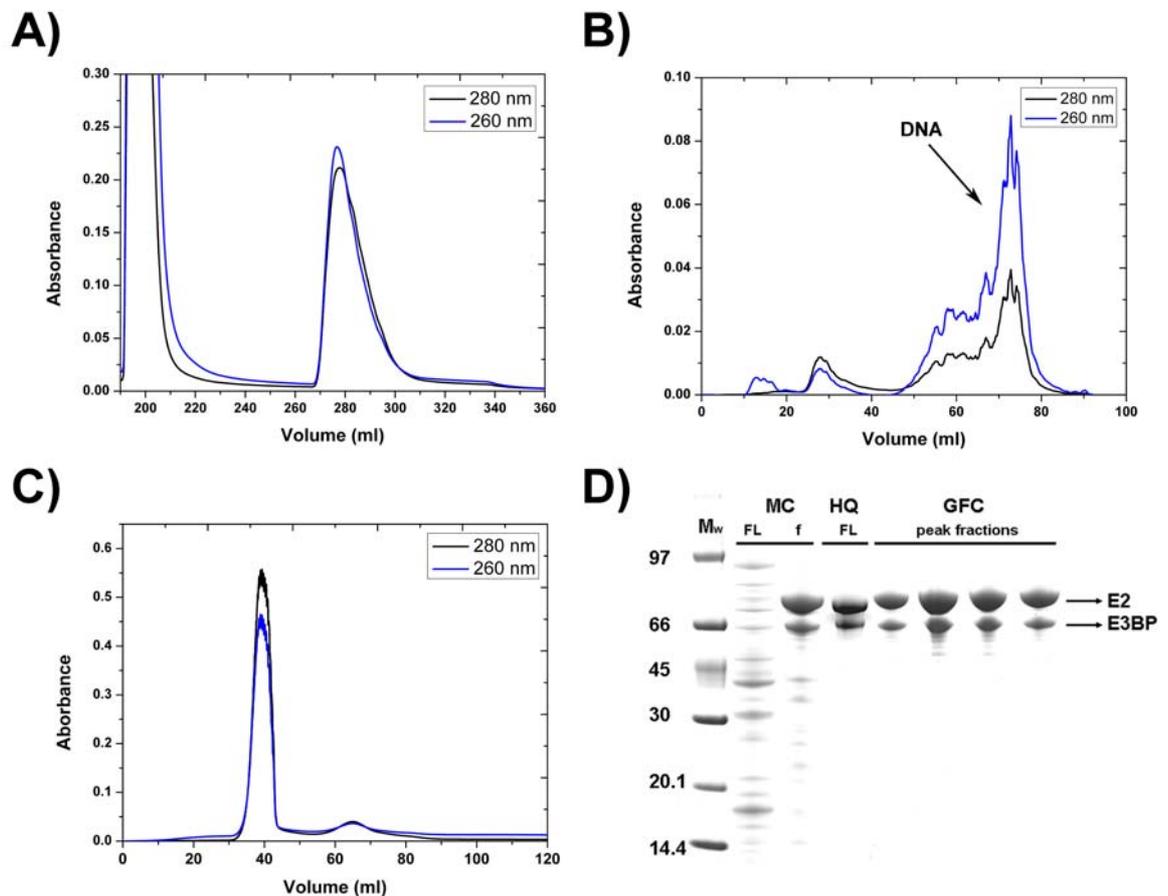
resulted in no precipitation or aggregation of the core during purification. Initial purification was carried out by metal chelate chromatography (Fig. 4.5A).



**Figure 4.4 Over-expression and solubility of rE2/E3BP at various temperatures**  
**(A)** rE2/E3BP expression was induced with 0.5 mM IPTG at various temperatures ranging from 30°C to 15°C, and culture samples taken prior to induction ( $t_0$ ) and 4, 6 or 17 h after induction ( $t_4$ ,  $t_6$ ,  $t_{17}$ ) depending on the temperature. **(B)** Solubility of rE2/E3BP was tested with cell pellets grown at temperatures 30°C, 22°C and 15°C with 0.5 mM IPTG. All soluble protein is present in the supernatant (S), similar to  $t_4$ ,  $t_6$  or  $t_{17}$  and none is observed in the pellet (P) at any of these temperatures. Molecular weight of marker proteins ( $M_w$ ) is shown in kDa.

The metal chelate step of rE2/E3BP purification exploits the strong interaction between rE2 and rE3BP, as rE3BP alone possesses an N-terminal His-tag. Hence, rE2/E3BP core is selectively purified from any unbound rE3BP or rE2. However, protein purified at this stage contained significant amounts of other contaminant proteins and DNA. The DNA was removed by anion exchange chromatography via a modified approach. Pooled protein from the metal chelate chromatography was extensively dialysed in anion exchange binding buffer and high salt (450 mM NaCl, see section 3.3.2.3 for details). Owing to the differential binding affinity of DNA and rE2/E3BP for the column resin, while rE2/E3BP passed into the flow-through, DNA remained bound to the column, thus ensuring selective removal of DNA from the purified protein (Fig. 4.5B). Any excess E3BP and contaminants present at this stage were removed by a final gel filtration purification step on a Sephacryl S-300 column (Fig. 4.5C). SDS-PAGE shows pure rE2/E3BP in the final preparation (Fig. 4.5D). It is to be noted that even after ion exchange purification, GFC-purified rE2/E3BP still contained residual traces of DNA (ratio of  $A_{280}/A_{260} = 1.8$  for a pure protein, while in the case of rE2/E3BP  $A_{280}/A_{260} \sim$

1.2). Hence, the presence of DNA may have some slight effect on the results presented in the later chapters.



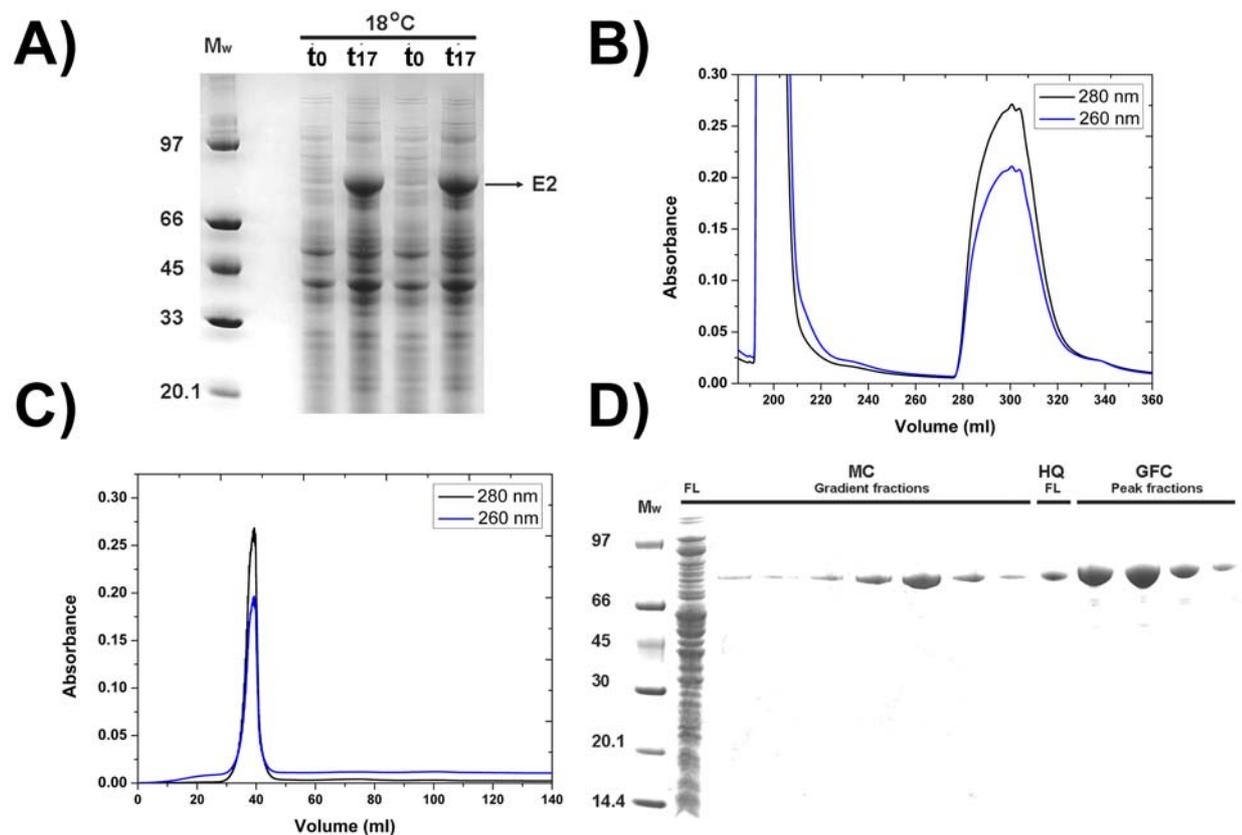
**Figure 4.5 Purification of rE2/E3BP**

(A) Metal chelate chromatography of rE2/E3BP eluted from 250-360 ml with an 8CV gradient of 0.5 M imidazole. Absorbances at 280 nm (black) and 260 nm (blue) are indicated. (B) Anion exchange chromatography of rE2/E3BP ensured selective removal of DNA from the protein that passed into the flow-through (data not shown). Elution was carried out over a vol of 0-100 ml with a 100CV gradient of 2M NaCl. (C) Gel filtration profile of rE2/E3BP being eluted at the void volume (about 40 ml). (D) SDS-PAGE analysis of flow-through (FL) and protein elution fractions (f) of metal chelate (MC), anion exchange (HQ) and gel filtration (GFC) are shown. Molecular weight of marker proteins ( $M_w$ ) is shown in kDa.

#### 4.3.2.2 Full-length recombinant E2 core (rE2)

Over-expression of rE2 was successfully carried out at 18°C after induction with 1 mM IPTG for 17 h producing large yields of over-expressed protein as observed in Fig. 4.6A. The solubility of rE2 was determined by lysing cells grown at 18°C. SDS-PAGE analysis indicated all rE2 to be present in the soluble fraction (data not shown).

Purification of rE2 was very similar to the rE2/E3BP purification. Protein purified initially by metal chelate chromatography had minor DNA contamination (Fig. 4.6B). Non-specifically bound DNA was selectively removed by a high salt dialysis step prior to anion exchange chromatography. rE2 fractions passing into the flow-through of the anion exchange column was then pooled and finally purified by gel filtration thus removing any minor protein contaminants present at this stage (Fig 4.6C). SDS-PAGE indicates anomalous migration of high yields of pure rE2 obtained from the final purification step, with bands running slightly higher than that predicted for a 62 kDa protein (Fig. 4.6D).



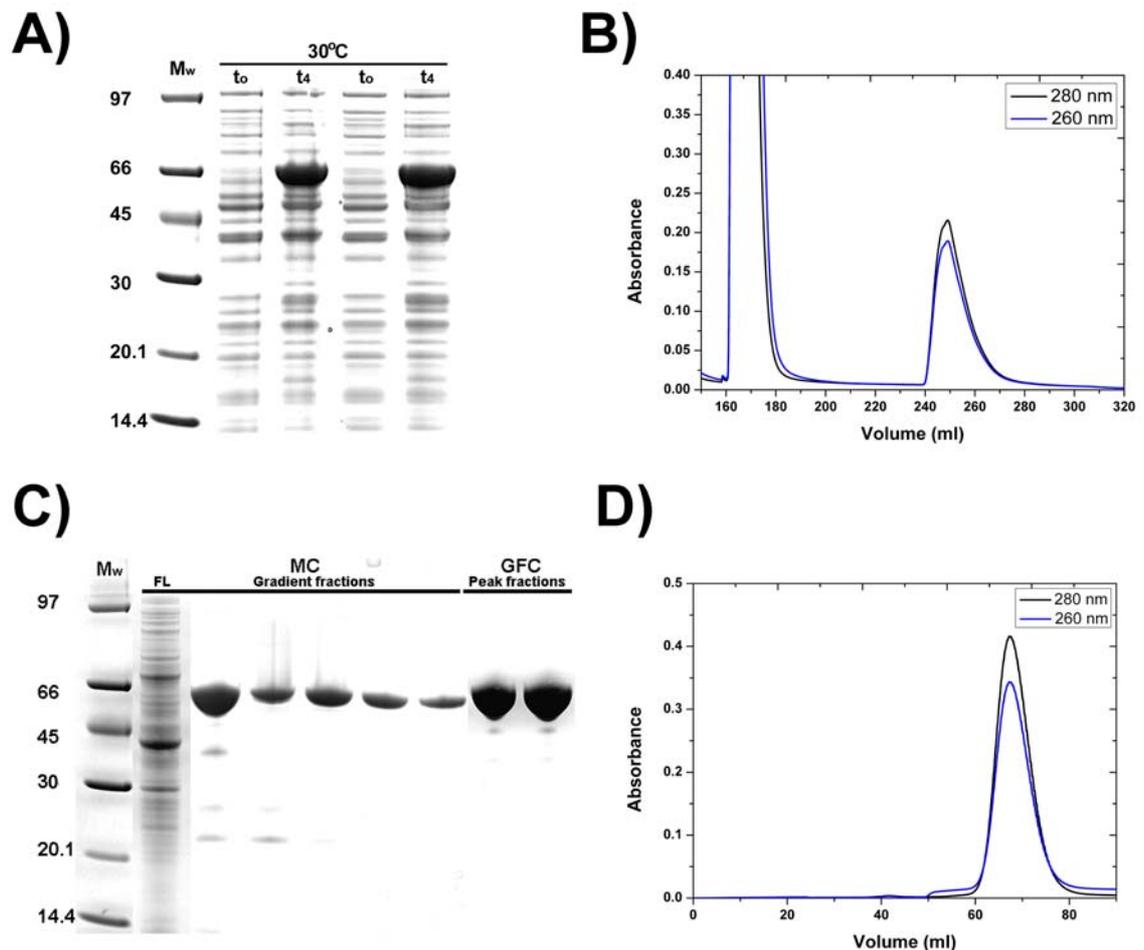
**Figure 4.6 Purification of rE2**

(A) Over-expression profile of rE2 at 18°C and 1 mM IPTG. Protein samples collected before (to) and after (t<sub>17</sub>) induction are shown. (B) Metal chelate chromatography of rE2. Bound material was eluted with a 0.5 M imidazole gradient (8CV) over a vol of 260-360 ml. (C) Size exclusion chromatography of rE2 protein elution at the void volume. (D) SDS-PAGE analysis of rE2 protein fractions obtained from metal chelate (MC) and gel filtration (GFC) along with rE2 flow-through (FL) collected from the anion exchange (HQ) column are shown. Molecular weights of marker proteins (Mw) are shown in kDa.

### 4.3.2.3 Recombinant E3 (rE3 and dE3)

Over-expression of rE3 was induced with 1 mM IPTG at 30°C for 4 h. Subsequent analysis by SDS-PAGE revealed good yields of over-expressed protein as seen in Fig. 4.7A.

Purification of rE3 was straightforward. It was first purified by metal chelate chromatography (Fig. 4.7B), and pooled protein fractions were then subjected to gel filtration to eliminate minor contaminants from the previous purification step. SDS-PAGE revealed greater than 98% pure protein in the final preparation (Figs 4.7C and 4.7D).



**Figure 4.7 Purification of rE3**

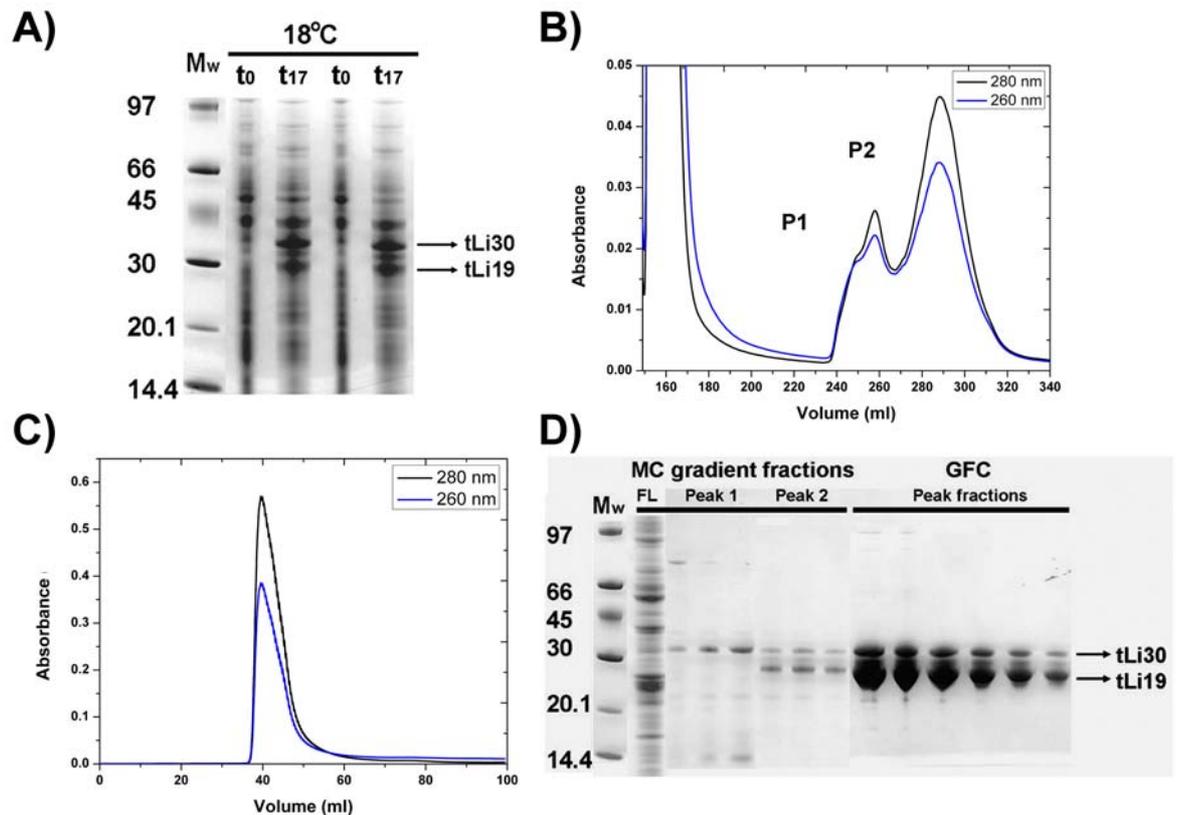
(A) Stable over-expression of rE3 at 30°C. Protein samples collected before (to) and after (t4) induction are shown. (B) Metal chelate chromatography of rE3. Protein elution was carried out with an 8CV gradient of 0.5 M imidazole over the vol 220-320 ml. (C) SDS-PAGE of flow-through (FL) and pooled metal chelate (MC) fractions of rE3. Pure rE3 protein fractions from gel filtration chromatography (GFC) are also shown (molecular weights of protein marker (Mw) are indicated in kDa). (D) Gel filtration profile of rE3.

Deuterated E3 (dE3) was overexpressed by Dr. Phil Callow at the Institut Laue Langevin (ILL), Grenoble France as follows. The plasmid encoding rE3 was transformed into BL21 (DE3) cells and subsequently used for over-expression. Bacterial cultures were grown in Infors minimal medium using an Infors fermentation system at 30°C to an absorbance at 600 nm ( $A_{600}$ ) of 15. Glycerol was used as the carbon source to deuterate E3 to 75% (contrast matched with 100%  $D_2O$ ). Water ( $H_2O$ ) in the buffer was replaced with 85%  $D_2O$ . The dE3 cell pellets were then purified at the University of Glasgow in a manner similar to rE3 (as described above) via two chromatographic steps namely, metal chelate and gel filtration. The purity and yield of purified dE3 was thereafter assessed by SDS-PAGE.

#### **4.3.2.4 Truncated recombinant E2/E3BP core (tE2/tE3BP)**

Although truncated constructs of E2 (tLi19, tLi27) and E3BP (tLi30) were produced successfully, co-expression of truncated core was carried out using only constructs tLi19 and tLi30 unless stated otherwise, resulting in the formation of a truncated core tLi19/tLi30. This was achieved by the co-transformation of two different plasmids encompassing the two genes of interest (tLi19-pET11b and tLi30-pET28b) into *E. coli*. Over-expression carried out at 18°C for 17 h with 1 mM IPTG produced stable high yields of tLi19/tLi30 (Fig. 4.8A).

Only the E3BP (tLi30) is His-tagged and by virtue of its natural co-association with E2 (tLi19), purification of the tLi19/tLi30 core was achieved by a standard two-step process; an initial metal chelate chromatography step, followed by gel filtration chromatography. In contrast to full-length rE2/E3BP, the truncated tLi19/tLi30 contained negligible DNA contamination and gave a two peak profile during metal chelate chromatography, with the first peak being free tLi30 (E3BP) (Fig. 4.8B). Therefore, in the absence of significant DNA contamination, protein fractions from the second peak of the metal chelate column were pooled, concentrated and directly subjected to a final gel filtration step to remove residual protein contaminants and any excess tLi30 (Fig. 4.8C). Analysis by SDS-PAGE shows good yields of pure protein as seen in Fig. 4.8D.



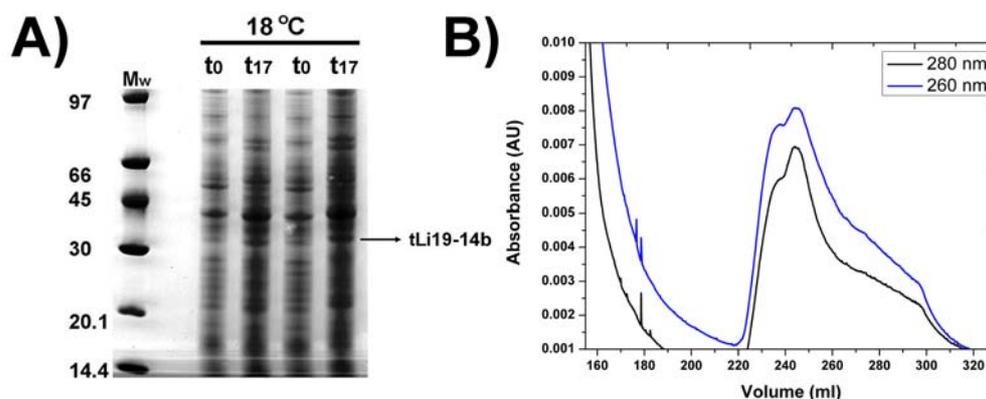
**Figure 4.8 Purification of tLi19/tLi30 core**

(A) Over-expression of tLi19/tLi30 at 18°C, 1 mM IPTG. Protein samples collected before (to) and after (t17) induction are shown. (B) Metal chelate chromatography of tLi19/tLi30 shows a dual peak elution profile, peak 1 (P1) being free E3BP (tLi30) and peak 2 (P2) being tLi19/tLi30. Protein elution was achieved using an 8CV gradient of 0.5 M imidazole over a vol of 220-340 ml. (C) Void volume protein elution of tLi19/tLi30 by gel filtration chromatography. (D) SDS-PAGE analysis of metal chelate (MC) flow-through (FL), protein (Peaks 1 and 2) fractions and gel filtration (GFC) fractions of purified tLi19/tLi30. Molecular weight of marker proteins (Mw) is shown in kDa.

#### 4.3.2.5 Truncated recombinant E2 core (tE2)

Truncated construct tLi19 was cloned into pET14b to produce recombinant truncated E2 core 'tLi19-14b'. Over-expression studies at 18°C with 1 mM IPTG for 17 h and subsequent SDS-PAGE analysis revealed poor expression (Fig. 4.9A). Despite several expression trials during which growth media, temperature and IPTG concentration were varied, it was not possible to successfully pinpoint conditions that resulted in high yields of over-expressed tLi19-14b. Initial metal chelate purification resulted in very low yields of protein with large amounts of DNA contamination (Fig. 4.9B). Poor protein yields may be a result of the linker region (preceeding the CTD) not being long enough to produce highly stable and soluble protein. The importance of the length of the CTD-preceeding linker

region towards the stability of the bacterial core has been previously shown (Titman, 2005). Owing to unsatisfactory over-expression and purification these constructs were not employed for subsequent experiments in this study.

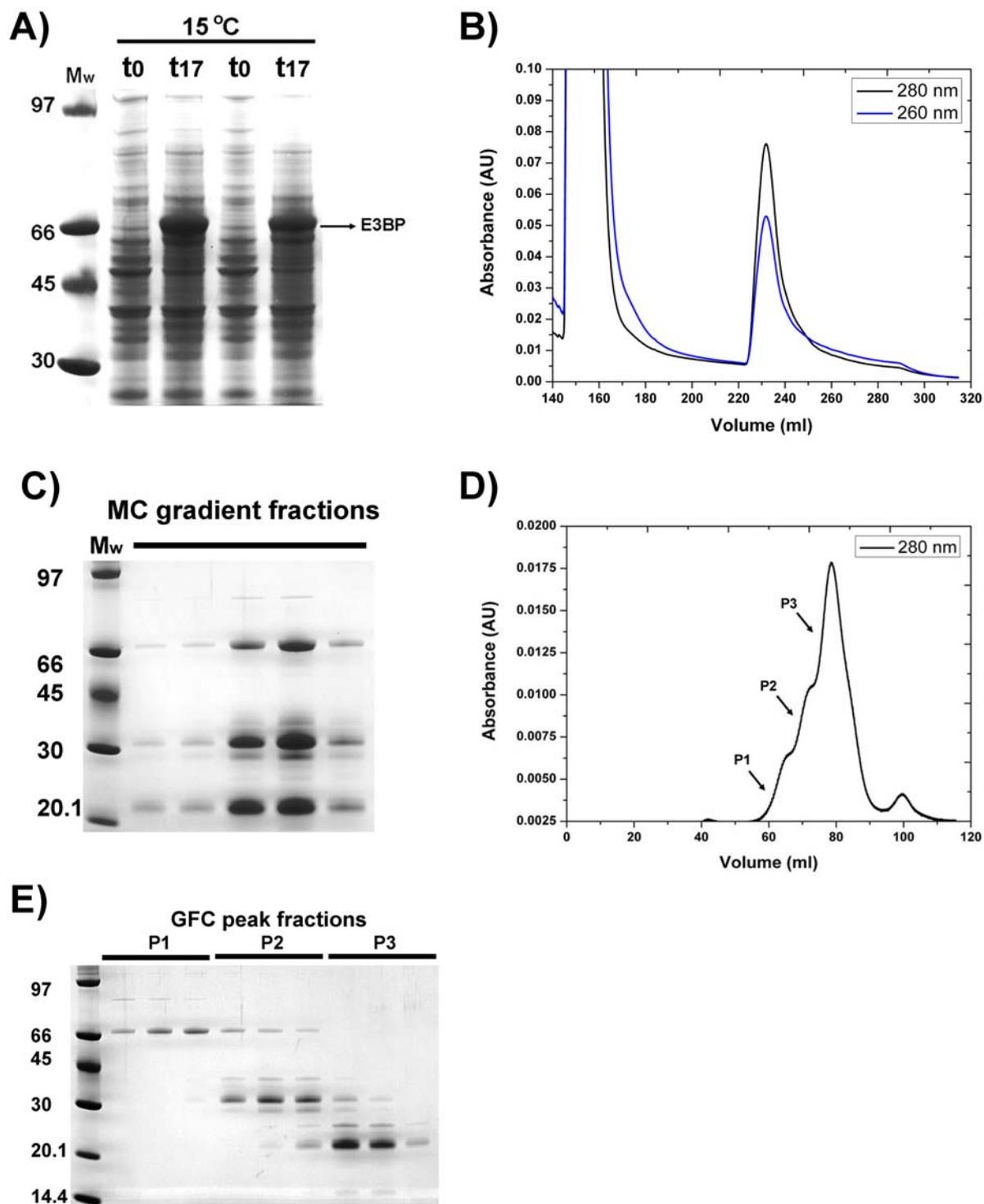


**Figure 4.9 Purification of tLi19-14b**

(A) Expression studies showed very poor expression of tLi19-14b at 18°C. Protein samples collected before (to) and after (t<sub>17</sub>) induction are denoted. (B) Metal chelate chromatography of tLi19-14b produced very low amounts of protein contaminated with significant amounts of DNA. Bound tLi19-14b was eluted with a 0.5 M gradient (8CV) over a vol of 200-330 ml. Molecular weights of marker proteins (Mw) are shown in kDa.

#### 4.3.2.6 Full-length recombinant E3BP-28b

Full-length His-tagged rE3BP cloned into pET28b was expressed at a low temperature (15°C), 0.5 mM IPTG for 17 h. Levels of over-expression were analysed by SDS-PAGE, indicating good yields of protein (Fig. 4.10A). On its own rE3BP has a tendency to precipitate on purification. However, over-expression of rE3BP at low temperatures ensured no precipitation, thus rendering greater solubility and possibly by improving folding and stability during over-expression. Purification of rE3BP was similar to purification of rE3. Following initial metal chelate affinity purification (Figs. 4.10B, 4.10C), a gel filtration (GFC) step was performed to remove residual protein contaminants (Figs. 4.10D, 4.10E). SDS-PAGE of the multi-peak GFC profile showed the presence of intact rE3BP in peak 1 (P1) and several low molecular weight bands in the latter peaks 2 and 3 (P2 and P3). Western blotting confirmed these bands to correspond to various truncated species of rE3BP harbouring the N-terminal domains formed as the result of proteolytic cleavage (data not shown). Interestingly, despite being a monomer of 50 kDa, rE3BP elutes from the gel filtration column (P1) at a fraction volume close to that of a dimer (about 70 ml,  $M_r$  of rE3BP homodimer ~



**Figure 4.10 Purification of rE3BP-28b**

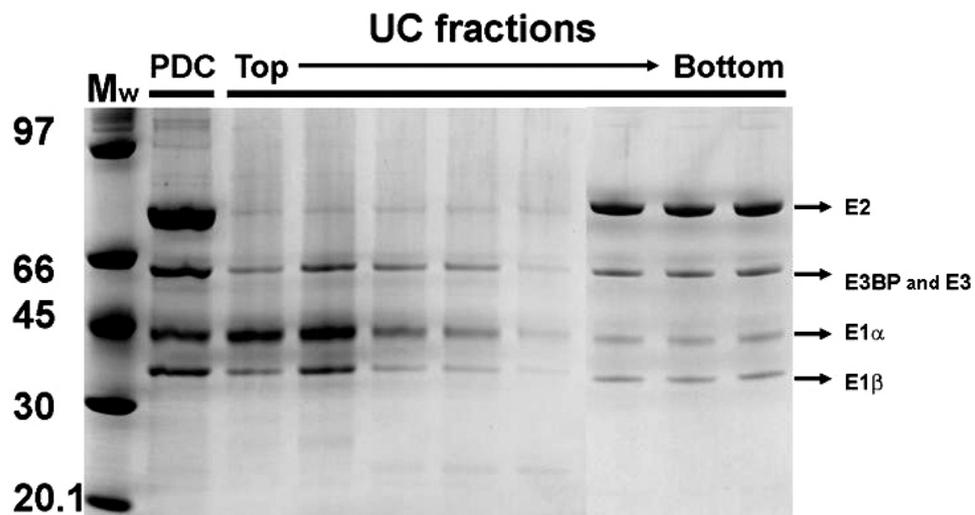
(A) Over-expression of rE3BP at 15°C with samples collected before (t<sub>0</sub>) and after (t<sub>17</sub>) induction. (B + C) Metal chelate chromatography of rE3BP. Protein fractions were eluted using a 0.5 M imidazole gradient (8CV) over a vol 210-320 ml and subsequently run on SDS-PAGE. (D + E) Gel filtration chromatography of rE3BP. GFC peak fractions analysed by SDS-PAGE showed the presence of pure rE3BP only in peak 1 (P1), and contaminating truncated forms of rE3BP in peaks 2 and 3 (P2 and P3). Molecular weights of marker proteins (M<sub>w</sub>) are indicated in kDa.

105 kDa) similar to rE3; this may be attributed to the elongated structure of rE3BP.

#### 4.3.2.7 Purification of bovine E2/E3BP core (bE2/E3BP)

Bovine E2/E3BP core (bE2/E3BP) was obtained from previously purified bovine PDC (see section 3.3.2.5) by sucrose gradient centrifugation (refer to section 3.3.2.6 for details). The first round of centrifugation resulted in the top gradient fractions enriched with E3 and E1. However as the bE2/E3BP pellet retained a significant amount of E1, it was further subjected to two additional rounds of centrifugation. Furthermore, loss of E3 was checked by immunoblotting of bE2/E3BP (data not shown).

Homogenised bE2/E3BP core obtained after three steps of centrifugation had most of the E1 removed. However, a slight contamination of 10-15% E1 was always observed in the final purified preparation of bE2/E3BP, and its total elimination was very difficult to achieve. The purified bE2/E3BP core was either used directly in experiments or stored in 50% (v/v) glycerol at -20°C for future use.



**Figure 4.11 Purification of bE2/E3BP**

SDS-PAGE of intact PDC sucrose gradient centrifugation (UC) fractions. The top (containing E1 and E3) and the bottom (comprising pellets of bE2/E3BP core from three rounds of centrifugation) fractions are shown. The positions of the various proteins are indicated and molecular weights of marker proteins (Mw) are shown in kDa.

## 4.4 Discussion

All proteins used in this study were successfully over-expressed and purified. Studies have shown that low temperature expression improves plasmid stability by decreasing the activity of T7 RNA polymerase and thereby providing more time for the protein to fold correctly (Zhang, et al., 2003). In addition, low temperatures may prevent ribosome pausing and dissociation along with the concomitant decrease in rates of elongation and protein synthesis. These lowered rates of translation at low temperatures may be comparable to that occurring *in vivo* in mammalian cells, therefore resulting in the production of correctly folded proteins exhibiting greater stability.

Over-expression of all proteins in this project was carried out at low temperatures and resulted in good yields of highly soluble protein. All recombinant full-length and truncated rE2/E3BP cores employed the co-expression of rE2 and rE3BP, a mandatory requirement for the generation of a functional core. The truncated sequences of tE2 and tE3BP were successfully cloned and purified for use in low- and high-resolution structural studies, namely solution scattering and x-ray crystallography. This is the first time that a truncated human E2/E3BP core has been successfully produced using recombinant technology for structure determination studies. Interestingly, inclusion of a few residues from the preceding linker regions into the truncated constructs seems to be critical for the solubility and probably architectural stability of the tE2/tE3BP core. This complements previous studies in bacterial PDC showing the importance of the inner linker regions towards folding and architectural stability of the E2 core assembly (Titman, 2005).

The His-tag is present only on the E3BP subunit, and therefore purification of the rE2/E3BP (full-length and truncated) core was achieved by exploiting the natural co-association between E2 and E3BP. Furthermore, the oligomeric state of isolated rE3BP was investigated: E3BP forms low molecular weight species as observed from its GFC profile, in direct contrast to rE2 that is capable of forming a 60-meric core on its own. Despite some proteins being susceptible to proteolytic cleavage, typical purification yields obtained for the majority of proteins was 3-5 mg/l with exceptions being rE3 (30-40 mg/l) and bE2/E3BP core

producing 100-150 mg of native PDC per kg of bovine heart. The purified proteins were then used for the various biophysical and biochemical experiments as described in the future chapters.

## Chapter 5

# Characterisation of the full-length cores: E2/E3BP and E2

### 5.1 Introduction

PDC contains an oligomeric E2 assembly of distinctive morphology that forms the central core, thereby providing the framework for interaction with peripheral enzymes E1 and E3. In Gram-negative bacteria like *E. coli*, the core is octahedral comprising 24 E2 molecules with cubic (432) symmetry. In contrast, in eukaryotes and Gram-positive bacteria e.g. *B. stearothermophilus*, the core assembly is larger with 60 E2 molecules arranged in a pentagonal dodecahedron displaying 532 icosahedral symmetry (Reed and Oliver, 1968; 1982; Wagenknecht, et al., 1991). The high resolution crystal structure of the truncated core from *B. stearothermophilus* (PDB ID 1B5S) provides clear insights into its subunit organisation, with the basic building blocks, namely E2 trimers located at the 20 vertices of the icosahedron (Izard, et al., 1999).

Eukaryotic PDCs contain an additional accessory subunit, E3 binding protein (E3BP, previously termed protein X) that is also involved in the formation of the icosahedral core (De Marcucci and Lindsay, 1985; Jilka, et al., 1986; Rahmatullah, et al., 1989b). Densitometry and radiolabelling studies on SDS-PAGE gels of purified bovine E2/E3BP indicated the presence of 12 molecules of E3BP per core, suggestive of an icosahedral E2/E3BP assembly comprising 60 copies of E2 and 12 copies of E3BP (Sanderson, et al., 1996b). Similarly, Maeng and colleagues (1994) showed the binding of 12-15 E3BP molecules to the PDC core of *S. cerevisiae*. As an icosahedron has 12 pentagonal faces, it was proposed that one copy of E3BP was positioned on each face. Comparative cryo-EM studies on truncated yeast E2 and E2/E3BP cores confirmed the suggested localisation of 12 E3BPs, with the binding of each E3BP on the inside of the pentagonal faces, close to the tips of the E2 trimers (Stoops, et al., 1997; Zhou, et al., 2001b).

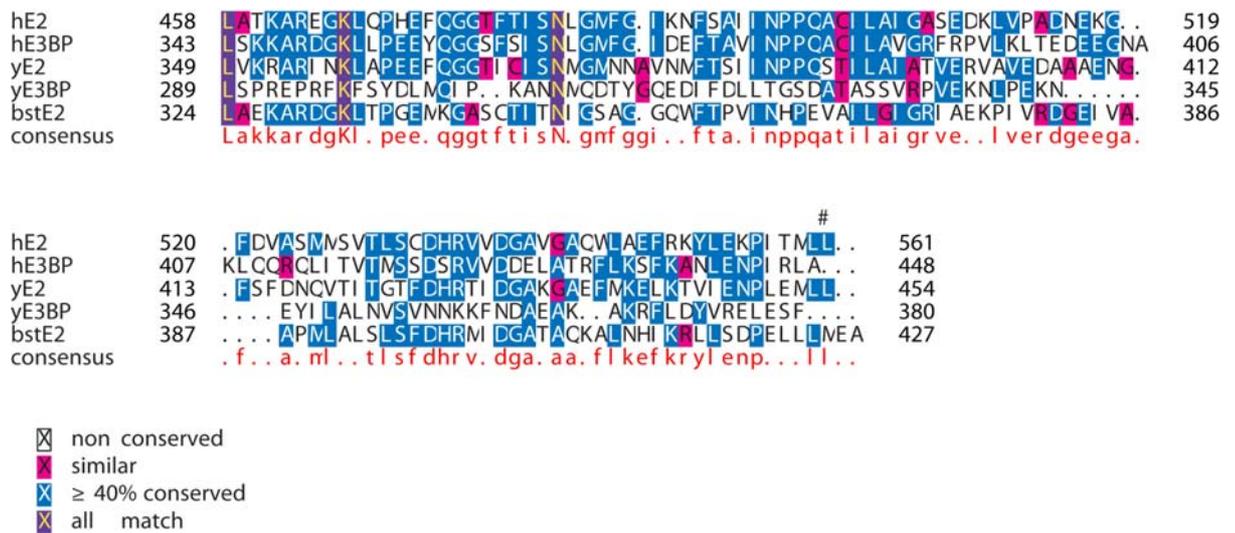
However, recent work based on AUC and SAXS by Hiromasa and co-workers (2004) has led to the concept of an alternative 'substitution model' for mammalian rE2/E3BP core organisation. Instead of 12 E3BPs binding to the 12 faces ('addition model', 60E2+12E3BP), the new model proposes the replacement of 12 E2 polypeptides by 12 E3BP subunits, resulting in a substituted 48E2+12E3BP core assembly.

Human E2 and E3BP have a similar modular domain organisation. E2 and E3BP each possess peripherally extended N-terminal LD(s), two on E2 and one on E3BP. Each LD carries a key lysine residue located within a highly conserved DKA motif at the tip of a type I  $\beta$ -turn (Wallis and Perham, 1994). The LD is followed by an SBD and finally a CTD that is involved in core assembly. While the active site (DHRXXDG) for the acetyltransferase reaction is located on the CTD of E2 (Radford, et al., 1987), it is absent from the CTD of E3BP, rendering it incapable of catalysis. The modular domains are connected by flexible linker regions (rich in alanine, proline and a few charged amino acids) that allow the LDs to access all active sites in turn during the catalytic cycle. The amino acid sequences of the inner LD, SBD and CTD of E2 are 46%, 38% and 50% identical to those of E3BP. Variation in the subunit organisation of the eukaryotic PDC cores has been attributed to the dissimilarity of their E3BP sequences (Fig. 5.1).

While in bacteria, E1 and E3 compete for binding to E2, specific binding is observed in eukaryotes: E1 binds to E2-SBD and, the equivalent SBD of E3BP binds to E3. E2 cores of patients who do not make E3BP retain a residual affinity for E3 binding (Marsac, et al., 1993). The overall complex activity in these patients is 10-20% of the controls, probably because the E2-SBD appears to have retained a limited capacity to bind to E3. Interestingly, while mammalian E2 can form a 60-meric core on its own, isolated E3BP is unable to do the same.

Although a considerable amount of structural and biochemical information is available for the different components of mammalian PDC, relatively little is known about the mammalian E2/E3BP core, including its precise subunit composition and stoichiometry of binding to E1 and E3.





**Figure 5.1 Sequence alignment of the domains of eukaryotic E2 and E3BP**

ClustalW (<http://www.ebi.gla.ac.uk/clustalw>) alignment of the amino acid sequences of E2 and E3BP [human (h), yeast (y) and *B. steraothermophilus* (bst)]. Lipoyl domains (LD), linker regions (H1 and H2), subunit binding domains (SBD) and inner C-terminal domains (CTD) of human E2 and E3BP along with the consensus sequence (red) are indicated. Sequence numbers on the right indicate the end of domains, and the lipoylation site (\*) and key residues mediating the ‘ball and socket’ interaction (# and †) between trimeric units are also indicated. The image was created using TEXshade (<http://www.ctan.org/tex-archive/help/Catalogue/entries/texshade.html>).

Moreover, as the mammalian E2/E3BP core provides the basic structural framework for ensuring efficient functioning of PDC, as well as being significantly different from its bacterial counterpart (E2), elucidating its structure and operation at the molecular level is essential for a complete understanding of this mammoth molecular machine.

The research described in this chapter sets out to achieve this goal by conducting detailed *in vitro* biophysical and biochemical characterisation of the full-length recombinant E2/E3BP and E2 cores. The architecture and subunit organisation of the rE2/E3BP and rE2 cores was investigated using a range of biophysical methods including, AUC, SAXS, SANS and negative stain EM. Furthermore, bE2/E3BP core was characterised biophysically and the data compared with those determined for the rE2/E3BP core. Low resolution structures of rE2/E3BP and bE2/E3BP were obtained by SAXS.

The icosahedral 60-meric cores are composed of 20 trimeric units. In the case of the bacterial E2 core: these 20 trimers are composed solely of E2 (homotrimers),

while, in the substitution model, the mammalian E2/E3BP core is made up of a mixture of homotrimers (all E2) and heterotrimers (2E2+1E3BP). Interactions between these trimeric building blocks are the primary factor responsible for the overall quarternary structure and stability of the core. This chapter also investigates the stability of the E2 core on introduction of E3BP via comparative CD and tryptophan fluorescence studies in the presence of the chemical denaturant GdmCl.

## **5.2 Materials and methods**

### **5.2.1 Protein purification**

rE2/E3BP and rE2 core assemblies were over-expressed and purified as outlined in sections 4.2.2 and 4.2.3. Extraction of bE2/E3BP was performed as detailed in sections 3.3.2.5, 3.3.2.6 and 4.2.3. All cores (rE2/E3BP, bE2/E3BP and rE2) were solubilised in either PEB or PEBS100 buffer (see section 4.2.3) after the final gel filtration step and immediately used in biophysical studies. All concentration measurements for rE2/E3BP used in this study were based on the 48E2+12E3BP substitution model, unless stated otherwise.

### **5.2.2 Sedimentation velocity**

SV experiments for all cores (rE2/E3BP, rE2, bE2/E3BP) were conducted with PEBS or PEBS100 as reference solvents (section 3.4.2). Sedimentation data from purified rE2/E3BP in PEBS100 buffer were recorded at 4°C and a rotor speed of 20,000 rpm using interference optics. A total of 450 scans, 1 min apart were recorded employing sample volumes of 360 µl loaded into 12 mm double sector centrepieces. The concentrations of samples covered a range from 76 nM to 306 nM (rE2/E3BP), 60 nM to 260 nM (rE2) and 1 mg/ml to 3.5 mg/ml (bE2/E3BP), respectively. All concentration estimates of bE2/E3BP carried out in this thesis are expressed in standard nomenclature (mg/ml), as the exact sequence of bE2/E3BP is not known; hence determination of its accurate molar concentration is not possible. The laser delay was adjusted prior to the run to obtain high quality interference fringes.

All sedimentation profiles were analysed using SEDFIT (Schuck, 2000) (see sections 2.2.1.3 and 3.4.2 for details). The partial specific volume  $\bar{v}$  of rE2/E3BP and rE2 calculated from their amino acid compositions using SEDNTERP (Laue, et al., 1992) (<http://www.jphilo.mailway.com/>) was 0.744 ml/g at 20°C. Although the recombinant and bovine E2/E3BP cores are expected to have similar amino acid sequences, primary sequences of bovine E2 and E3BP of purified bE2/E3BP are not available. Hence the value of  $\bar{v}$  for rE2/E3BP (0.744 ml/g) was also adopted for bE2/E3BP. Initial analysis by fitting the Lamm equation solutions to the SV data via the differential sedimentation coefficient distribution,  $c(s)$ , was performed to obtain a species distribution profile. The concentrations and sedimentation coefficients retrieved from the  $c(s)$  analysis were then used as the basis for the non-interacting discrete species model in SEDFIT that uses a finite element approach to determine the apparent sedimentation coefficients. These coefficients were then corrected to standard conditions of temperature and solvent (see section 2.2.1.2.2 and section 3.4.2 for details) and subsequently extrapolated to infinite dilution to obtain a sedimentation coefficient independent of concentration,  $s_{20,w}^0$ .

### **5.2.3 Sedimentation equilibrium**

SE data were collected for rE2/E3BP, rE2 and bE2/E3BP using interference optics with PEB or PEBS100 as the reference solvents. All samples were used within 24 h of purification. SE was carried out at 4°C with samples (80  $\mu$ l) loaded into 12 mm double sector centrepieces. Initial SE runs were conducted at rotor speeds of 3,000, 5,000 and 7,000 rpm. However, analysis of the data at 5,000 and 7,000 rpm was unsatisfactory yielding very steep exponential equilibrium solute distributions indicating the formation of sample pellets at the bottom of the cell. Hence, all subsequent SE studies were performed at 3,000 rpm, unless stated otherwise. After an initial delay period of 20-24 h to allow the system to reach equilibrium 12-15 scans were recorded 3 h apart for samples with concentrations between 154 nM to 461 nM (rE2/E3BP), 60 nM to 260 nM (rE2) and 1-5 mg/ml (bE2/E3BP), respectively. Attainment of equilibrium for all runs was ascertained using WinMATCH ([www.biotech.uconn.edu/auf](http://www.biotech.uconn.edu/auf)), when a minimal change in concentration gradients was observed via the measurement of the root mean square deviation ( $\text{rmsd} < 0.1 \times 10^{-2}$ ) of scans recorded 3 h apart.

SE data from all the cores were analysed using the non-interacting species model in SEDPHAT (Vistica, et al., 2004) as described in sections 2.2.2.2 and 3.4.3. Single fits at all sample concentrations were carried out to determine the whole-cell weight average molecular weight  $M_{app}$  for each sample concentration. The value of the whole-cell weight average molecular weight independent of concentration,  $M_w^0$ , was determined by plotting  $1/M_{app}$  against sample concentration (in nM or mg/ml), where  $M_w^0$  is the inverse of the y intercept. Alternatively, similar values for  $M_{app}$  and  $M_w^0$  were obtained via single fit analysis using the Beckman XL-A/XL-I software implemented in MicroCal ORIGIN. Additionally, this software allows for consideration of non-ideality by means of the second virial coefficient  $B$ . However, including fitting of  $B$  for rE2/E3BP and rE2 seemed to worsen the fits (data not shown) with uncertain estimates of  $M_{app}$ . Hence, all SE analysis reported in this study was undertaken with SEDPHAT, unless stated otherwise.

#### **5.2.4 Small angle x-ray scattering**

SAXS data for rE2/E3BP and bE2/E3BP were collected on beamline X33 of the Deutsches Elektronen Synchrotron (DESY) at the EMBL outstation in Hamburg, Germany with a single camera length of 4 m. Scattering curves were obtained for rE2/E3BP concentrations of 0.5, 2.95 and 3.8 mg/ml, while bE2/E3BP data were recorded at 0.65 and 1.14 mg/ml. Scattering experiments were conducted at three temperatures (10°C, 20°C and 37°C) to look for temperature-mediated structural changes. Scattering data for rE2 were obtained on beamline 2.1 at the SRS Daresbury, UK using dual camera lengths of 2.25 and 6.25 m and sample concentrations of 0.6 and 1.2 mg/ml. More details are given in section 3.4.4.

#### **5.2.5 Small angle neutron scattering**

SANS of rE2/E3BP and rE2 was carried out at the D22 beamline at the Institut Laue-Langevin (ILL) in Grenoble, France with camera lengths of 4 and 14 m. Scattering data were recorded at 4°C at protein concentrations of 3.69 and 6.84 mg/ml for rE2/E3BP and rE2, respectively. More details can be found in section 3.4.5.

## 5.2.6 *Ab initio* modelling of SAXS and SANS data

Initial *ab initio* reconstructions of rE2/E3BP and bE2/E3BP were carried out using the program DAMMIN (Petoukhov and Svergun, 2003). Although several models were obtained from a series of DAMMIN trials, none of these provided a satisfactory fit to the experimental scattering curve. This may be attributed to the large size of the dummy atoms used to fill the search space, resulting in failure to restore a detailed internal structure. Moreover, DAMMIN is known to have difficulty in restoring hollow structures (Petoukhov and Svergun, 2003). In this case the hollow nature of the cores coupled with icosahedral symmetry constraints posed too great a problem for DAMMIN. Therefore, all subsequent model reconstructions were performed using GASBOR (Svergun, et al., 2001) to restore biologically relevant models with very good fits. Since both cores possess dodecahedral morphology, the PICO symmetry option was enabled during the GASBOR modelling along with no application of penalties for the disconnectivity and peripheral constraints (for details refer to section 3.5.2). All runs were carried out in batch mode using the computer grid system SCOTGRID.

Averaging 10 *ab initio* models of rE2/E3BP and bE2/E3BP to obtain a consensus structure for each using DAMAVER (Volkov and Svergun, 2003) posed huge problems on account of the existence of hollow cavities and elongated density (from the SBDs and LDs) within the same structure. As the number of amino acids (9879) in rE2/E3BP and bE2/E3BP cores greatly exceeds the maximum limit of dummy residues available (8000) to model in GASBOR, it was assumed that the *ab initio* restorations would be limited and result in E2/E3BP models occupying volumes much lower than the actual core volume. Interestingly, despite the greatly reduced number of dummy residues, there appears to be just enough residues per *ab initio* model (7879) to fit the data well, as volumes obtained from CRY SOL (Svergun, et al., 1995) ( $4.3\text{-}4.37 \times 10^6 \text{ \AA}^3$ ) for each reconstruction compare favourably with the calculated volumes of the cores ( $4.38 \times 10^6 \text{ \AA}^3$ ). Therefore, representing a consensus model of rE2/E3BP and bE2/E3BP cores by the superimposition of 10 *ab initio* GASBOR models each seems a rational approach to overcome the problem of structure restoration encountered with DAMAVER.

## 5.2.7 Hydrodynamic modelling

The GASBOR *ab initio* model of rE2/E3BP core comprising 7879 dummy residues was converted to a hydrodynamic bead model (HBM) comprising 1990 beads employing a 15 Å cubic grid using the AtoB algorithm (Byron, 1997) implemented within SOMO (Rai, et al., 2005). Hydrodynamic parameters, namely the anhydrous sedimentation coefficient  $s_0$  and translational diffusion coefficient  $D_t$  of the HBM were then calculated using HYDRO++ (García de la Torre, et al., 2007; García de la Torre, et al., 1994). Alternatively, a HBM with no bead overlaps utilising a 7 Å cubic grid was generated using the program TRANS2VORONOI, an extended version of AtoB, developed by M. Nöllmann (Centre de Biochimie Structurale Montpellier, France), resulting in a total of 4343 beads. Identification of the percentage of surface exposed and buried beads in this HBM was performed using the ASAB1 utility of SOMO. Hydrodynamic computations from the resulting HBM were then independently performed using the SUPCW subroutine of BEAMS (Spotorno, et al., 1997). The HBMs used for calculations with HYDRO++ were corrected for the exact volume of rE2/E3BP, an estimate obtained from theoretical calculations taking into account the molecular mass  $M$ , partial specific volume  $\bar{v}$  and Avogadro's number  $N_A$  ( $6.023 \times 10^{23} \text{ mol}^{-1}$ ). The anhydrous sedimentation coefficients obtained from HYDRO++ and SUPCW were converted to their equivalent hydrated values employing a conversion factor of 0.8664 derived from hydrodynamic theory (refer to section 2.2.1.2.2 for details). All models obtained from AtoB and TRANS2VORONOI were visualised with one of several programs: FreeWRL (<http://freewrl.sourceforge.net>), Pymol (Delano Scientific, USA), Rasmol (Sayle and Milner-White, 1995) (<http://openrasmol.org>) or Visual Beads (<http://leonardo.fcu.um.es/macromol/programs/visualbeads0b/visualbeads.htm>). Further details on the modelling can be found in section 3.5.4.

## 5.2.8 Circular dichroism

Far-UV (190-260 nm) and near-UV (250-320 nm) CD analyses were performed on rE2/E3BP and rE2 cores at 25°C in the presence of increasing amounts of GdmCl (see section 3.4.8 for technical details). The cores were suspended in PEB buffer and data recorded at sample concentrations of 0.2 mg/ml and 0.9 mg/ml for the

far- and near-UV, respectively. The GdmCl concentration was accurately measured via refractrometry prior to each run. The percentage total change during unfolding/dissociation monitored at 285 nm (near-UV) and 222 nm (far-UV) in the presence of increasing chemical denaturant was calculated via the following equation

$$\% \text{total change on unfolding} = \frac{E_n - E_{\min}}{E_{\max} - E_{\min}} \times 100$$

where  $E_{\max}$  is the ellipticity of the completely unfolded protein at 6 M GdmCl at 285/222 nm,  $E_{\min}$  is the ellipticity of the native folded protein at 0 M GdmCl at 285/222 nm and  $E_n$  is the ellipticity obtained at every GdmCl concentration at 285/222 nm, with n representing the various GdmCl concentrations. The extent of unfolding/dissociation independent of sample concentration for both rE2/E3BP and rE2 was obtained by plotting the %total change on unfolding against GdmCl concentration (in M). Additionally, the Gibbs free energy of the reaction,  $\Delta G$ , at standard temperature (293 K) and pressure (1 atm) was computed via the following equation

$$\Delta G = -RT \ln K = -2434.83 \ln K$$

where free energy  $\Delta G$  is in J/mol,  $R$  is the universal gas constant (8.31451 J/mol K<sup>-1</sup>),  $T$  is the temperature in K (293 K) and  $K$  is the equilibrium constant defined by the expression,

$$K = \left[ \frac{\% \text{total unfolded protein}}{\% \text{total folded protein}} \right]$$

The free energy  $\Delta G^0$  at infinite dilution was determined from the y intercept of a plot of  $\Delta G$  against GdmCl concentration (in M).

### 5.2.9 Tryptophan fluorescence

Changes in tryptophan fluorescence for rE2/E3BP and rE2 were monitored at 25°C in the presence of increasing GdmCl concentrations. Fluorescence emissions

for both cores were recorded over a spectral range of 310-450 nm at sample concentrations of 0.2 mg/ml (refer to section 3.4.8 for technical details).

### **5.2.10 Negative stain electron microscopy**

Negative stain EM images were obtained for rE2/E3BP, bE2/E3BP and rE2 cores at very high magnification. Protein samples for imaging were prepared as described in detail in section 3.4.6. Images of rE2/E3BP and rE2 cores were recorded (magnification 90,000x and 120,000x) at sample concentrations of 70  $\mu\text{g/ml}$  and 61.5  $\mu\text{g/ml}$ , respectively. A bE2/E3BP protein concentration of 1 mg/ml was used to record images at a magnification of 90,000x and 120,000x.

## **5.3 Results and modelling**

### **5.3.1 Recombinant E2/E3BP core (rE2/E3BP)**

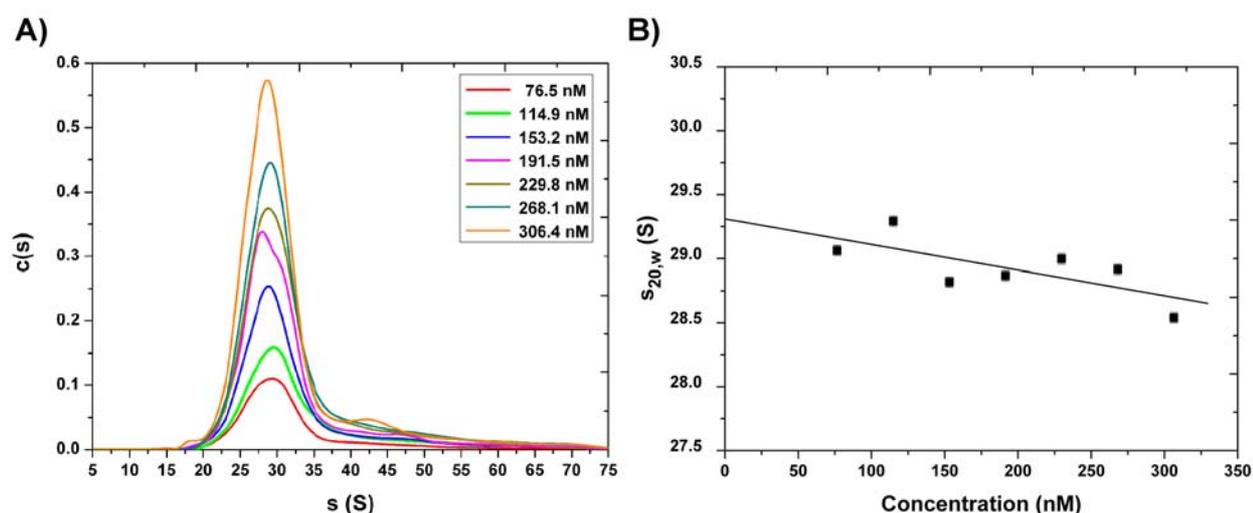
#### **5.3.1.1 Shape determination of rE2/E3BP core**

##### *5.3.1.1.1 SV AUC*

Sedimentation velocity of rE2/E3BP (Fig. 5.2) confirmed the presence of a main species consistent with a 60-meric core (Fig. 5.2A). Additionally, a small amount ( $\leq 5\%$  of the total protein concentration) of high molecular weight species was also observed denoted by the tail in Fig. 5.2A. Weight-average sedimentation coefficients  $s_w$  were determined to be 29 and 43 S by integration of each peak in the  $c(s)$  analysis. The peak around 29 S is thought to correspond to intact rE2/E3BP core and its  $s_w$  compares favourably with the sedimentation coefficient of 31.8 S obtained by Hiromasa and co-workers (2004). The observed tail around 43 S may correspond to the presence of aggregates or a small fraction of possible dimer that was persistent in all SV runs. This trailing edge comprising the faster sedimenting species has been observed before (Hiromasa, et al., 2004). Moreover, the 29 S peak is quite broad and probably reflects inherent heterogeneity due to the presence of E2/E3BP cores of various sizes and/or subunit composition, more about which is discussed in Chapter 7. Therefore, it is

easy to envisage potential difficulties in producing a well-defined core representing a uniform single species.

Finite element analysis on the SV interference data with the non-discrete species model in SEDFIT yielded sedimentation coefficients for rE2/E3BP at all experimental concentrations. These sedimentation coefficients were then extrapolated to infinite dilution, giving a concentration independent sedimentation coefficient of  $s_{20,w}^0 = 29.3 \pm 0.04$  S (Fig. 5.2B).



**Figure 5.2 Sedimentation velocity analysis of rE2/E3BP core**

**(A)**  $c(s)$  distribution from SV interference data for rE2/E3BP. The axes  $s$  and  $c(s)$  denote the sedimentation coefficient and the  $c(s)$  distribution model (from SEDFIT analysis), respectively. **(B)** Determination of concentration independent  $s_{20,w}^0$  ( $29.3 \pm 0.04$  S) for intact rE2/E3BP core obtained from  $c(s)$  analysis. Error bars are shown, but are not clearly visible owing to their small size.

The calculated value of  $f/f_0$  (obtained from the ratio  $s_0/s_{20,w}^0$ , i.e. ratio of the sedimentation coefficient of an anhydrous sphere of equivalent mass and volume as that of rE2/E3BP,  $s_0$  and the experimentally measured sedimentation coefficient,  $s_{20,w}^0$ ) that measures the molecular elongation (and/or hydration) for rE2/E3BP is 2.69, consistent with values obtained from the SEDFIT analysis (2.56) of the SV data. The large value of  $f/f_0$  indicates a structure with large hydrodynamic radius and deviation from spherical symmetry. Calculated values of the various hydrodynamic parameters from the SV analysis of rE2/E3BP are listed in Table 5.1.

$M$ (Da)	$\bar{v}$ (ml/g)	$s_{20,w}^0$ (S)	$f/f_0$	$R_0$ (nm)	$R_s$ (nm)	$D_s$ (Å)	$D_t$ (cm <sup>2</sup> /s)
3,551,100	0.744	29.3 <sup>a</sup> 79.0 <sup>b</sup>	2.69	10.2	27.3	546	$8.10 \times 10^{-8}$

**Table 5.1 Hydrodynamic parameters for rE2/E3BP derived from SV data**

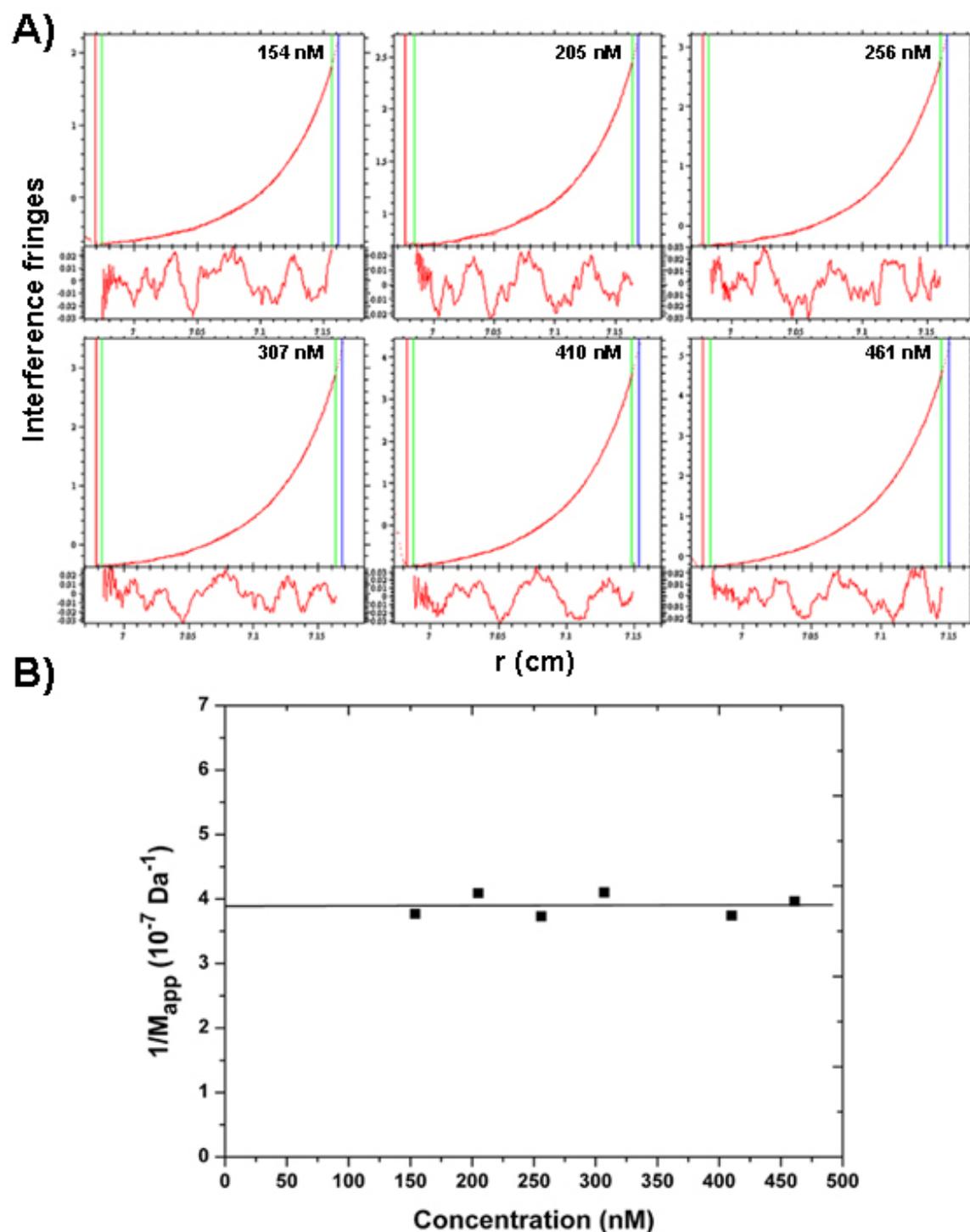
<sup>a</sup>experimental and <sup>b</sup>calculated sedimentation coefficients.  $M$  is the molecular mass based on the amino acid composition,  $\bar{v}$  is the calculated partial specific volume at 20°C,  $s_{20,w}^0$  is the sedimentation coefficient at infinite dilution,  $R_0$  is the radius of the anhydrous sphere of equivalent mass and specific volume as rE2/E3BP,  $R_s$  is the hydrodynamic or Stokes radius,  $D_s$  is the diameter of the particle obtained from the Stokes radius and  $D_t$  is the translational diffusion coefficient.

### 5.3.1.1.2 SE AUC

SE data obtained for rE2/E3BP show a very small increase in the inverse of the apparent whole-cell weight-average molecular weight,  $1/M_{app}$  with increasing sample concentration, indicating minimal non-ideality in these water-filled, negatively charged cores (Fig. 5.3B). Global analysis of SE data in SEDPHAT (Schuck, 2003; Vistica, et al., 2004) with a single species model yielded very poor fits that were greatly improved upon using a two species model as seen in Fig. 5.3A, yielding molecular weights of 2.48 MDa and 6.01 MDa, respectively. Molecular weight of the predominant species (i.e. the first species, consistent with 60-meric rE2/E3BP core) independent of concentration,  $M_w^0$ , was determined to be  $2.57 \pm 0.24$  MDa, considerably lower than the value predicted from the amino acid sequence of 3.55 MDa (Fig. 5.3B).

This discrepancy between the experimental and calculated molecular weights could reflect the formation of incomplete cores or proteolytic cleavage of the N-terminal arms of rE2/E3BP over the experimental time period (~3 days) required for the SE runs. However, electron microscopy images (see section 5.3.1.4.1) point towards the presence of complete cores, indicating proteolytic cleavage to be the more likely explanation. The rE2/E3BP core is known to be highly susceptible to proteolysis over time, particularly the flexible linker regions connecting the various domains. In addition, GFC of an old preparation of rE2/E3BP (4-5 days) resulted in elution of several low molecular weight products

(apart from the void volume elution; data not shown), indicating the high susceptibility of rE2/E3BP towards proteolytic cleavage.

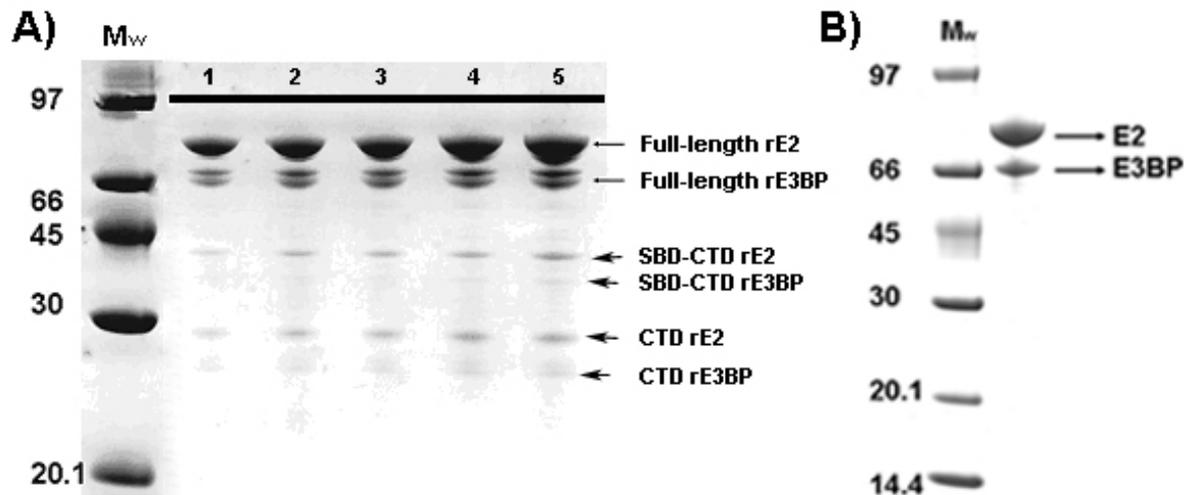


**Figure 5.3 Sedimentation equilibrium analysis of rE2/E3BP**

(A) SE interference data (dotted lines) at 3,000 rpm were fit (smooth lines) with a two species model in SEDPHAT (2.48 MDa and 6.01 MDa).  $M_{app}$  was determined at each sample concentration. Residual plots of various samples are shown on different scales.

(B) Extrapolation to zero of the  $1/M_{app}$  vs concentration plot yielded the molecular mass,  $M_w^0$  independent of concentration for the predominant species.

This significant proteolytic cleavage is apparent from SDS-PAGE of post-SE rE2/E3BP samples (Fig. 5.4). Apart from the expected monomeric rE2 (61 kDa) and rE3BP (50 kDa), SDS-PAGE also reveals possible truncated products of E2 and E3BP with molecular weights of 29 kDa (CTD cores) and 43 kDa (SBD-CTD cores).



**Figure 5.4 SDS-PAGE of post-SE rE2/E3BP core samples**

(A) Samples of rE2/E3BP after the SE run were analysed by SDS-PAGE (lanes 1-5 represent various concentrations) and show several bands arising from proteolytic cleavage of rE2 and rE3BP (indicated by arrows). (B) Freshly purified rE2/E3BP is shown for comparison. Molecular weights of marker proteins ( $M_w$ ) are indicated in kDa.

However, there may be other minor proteolytic products not clearly visible on SDS-PAGE due to the relative insensitivity of the Coomassie blue stain. Additionally, estimation of the extent of proteolysis is difficult and prone to significant error, as the intact core has a larger molecular weight and therefore a higher capacity to bind to the dye as compared to the smaller proteolytic products. Owing to proteolysis, the samples may comprise a concoction of cores. Table 5.2 indicates several ways to obtain a 2.57 MDa rE2/E3BP core (from E2 and E3BP subunits possessing different domains), reinforcing the issue of complexity and inherent variation in core organisation. It is important to note that these are only a few of the several combinations that could exist in the actual sample.

The weight-average sedimentation coefficient of 43 S obtained from SV data for the faster sedimenting species was initially thought to correspond to a possible rE2/E3BP dimer or non-specific aggregate. However, fitting of SE data yielded a

% of rE2/E3BP with all domains (full core)	% of rE2/E3BP with only CTD domain (truncated core)	% of truncated rE2/E3BP having only SBD-CTD domains	Resultant molecular weight $M$ (MDa)
100	-	-	3.55
-	100	-	1.67
-	-	100	2.64
48	52	-	<b>2.57</b>
40	17	33	<b>2.57</b>
43	47	10	<b>2.57</b>

**Table 5.2 Variable core organisation of rE2/E3BP and its effect on molecular weight**

Some possible combinations towards obtaining the SE measured molecular weight 'M' (blue) of rE2/E3BP from a mixture of cores comprising various domains (full and truncated) are shown.

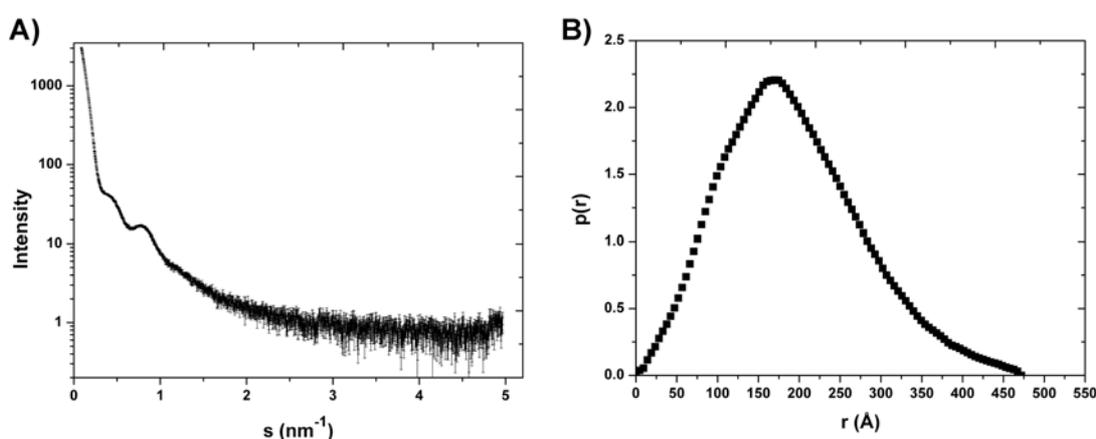
molecular weight independent of concentration,  $M_w^0$ , of  $5.71 \pm 0.11$  MDa for the second species (data not shown), which is lower than expected for intact rE2/E3BP dimer (7.1 MDa), but in rough agreement with the predicted mass of a rE2/E3BP dimer (5.14 MDa) that comprises monomers of 2.57 MDa, as obtained from SE data. However, fitting with the self-association model in SEDPHAT was unsatisfactory, pointing instead to the presence of aggregates rather than higher order oligomers. Despite several attempts to minimise the extent of proteolytic degradation during purification by inclusion of protease inhibitors at all stages, the molecular mass of the core gave consistent values in the range of 2.50-2.70 MDa.

### 5.3.1.2 Solution structure of rE2/E3BP core

#### 5.3.1.2.1 SAXS, SANS and *ab initio* modelling

SAXS curves for purified rE2/E3BP were acquired at three different concentrations and temperatures. Inspection of the low angle region of the scattering curves revealed a tendency for rE2/E3BP to aggregate at concentrations greater than 3 mg/ml. In addition, the SAXS curve at the lowest concentration measured (0.5 mg/ml) was too noisy to yield satisfactory estimates of molecular weight at all temperatures (10°C, 20°C, 37°C). However, the scattering curve obtained at 2.95 mg/ml was devoid of aggregates or any inter-particle interference effects (Fig. 5.5A) and was hence used for *ab initio* modelling. The molecular weight estimate for rE2/E3BP obtained by

extrapolation of scattering intensity to zero angle,  $I(0)$ , was 3.17 MDa, slightly lower than the calculated mass of 3.55 MDa from the protein sequence and higher than the mass determined by SE (2.57 MDa). This discrepancy in the molecular weight may be due to experimental errors, as molecular weight estimates in SAXS are not particularly accurate for very small or large molecules (Vachette and Svergun, 2000). Guinier analysis and GNOM (Semenyuk and Svergun, 1991; Svergun, 1992) were used to determine the radius of gyration,  $R_g$ , to be  $147 \pm 1 \text{ \AA}$  and  $148 \pm 1 \text{ \AA}$ , respectively. Additionally, no temperature induced conformational changes or radiation induced effects were observed for the core samples.



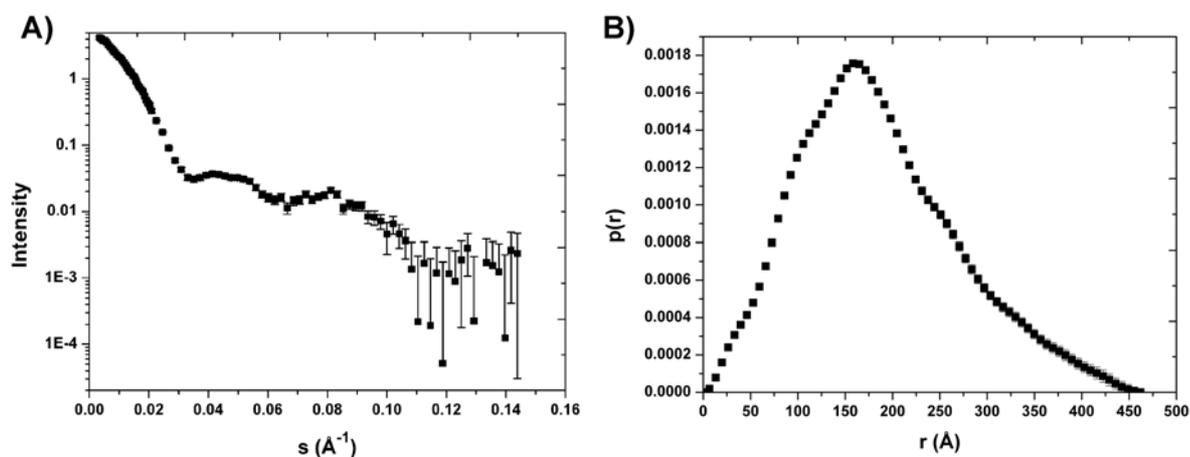
**Figure 5.5 Small angle x-ray scattering of rE2/E3BP**

The scattering curve for rE2/E3BP is shown in (A). The x-axis  $s$  denotes the scattering angle. The distance distribution function,  $p(r)$  vs  $r$  (B) was calculated using GNOM resulting in a  $D_{max}$  of 472 Å. Error bars are shown, but not clearly visible owing to their small size.

The distance distribution function for rE2/E3BP core,  $p(r)$  (Fig. 5.5B) shows a ‘bell like’ shape. A slight deviation from a perfect Gaussian distribution symbolic of a spherical molecule, and the minor elongation in shape at large distances (high values of  $r$ ) is readily observed. The maximum dimension of the particle,  $D_{max}$ , was determined to be 472 Å. Although the value of  $R_g$  ( $147 \pm 1 \text{ \AA}$ ) is consistent with the published value of  $151 \pm 2 \text{ \AA}$  (Hiromasa, et al., 2004), the  $D_{max}$  determined in this study is somewhat higher than the value ( $420 \pm 10 \text{ \AA}$ ) previously reported by Hiromasa and colleagues (2004). Variation in these values probably reflects the different approaches employed, but more importantly may be due to the size variation observed in eukaryotic PDC cores as a result of

variable distances between the trimeric building blocks, giving rise to the “breathing” of the core (Zhou, et al., 2001a).

SANS curves for rE2/E3BP were obtained at a single concentration of 3.69 mg/ml owing to sample constraints (Fig. 5.6A). The molecular weight from the extrapolation of scattering intensity to zero angle,  $I(0)$ , was estimated to be 3.66 MDa, slightly elevated with respect to the calculated mass of 3.55 MDa, the elevation being well within experimental error. Moreover, this estimated mass (3.66 MDa) is much higher than the masses obtained by SE (2.57 MDa) and SAXS (3.17 MDa). The  $R_g$  determined using the Guinier approximation is  $143 \pm 2 \text{ \AA}$ . The distribution function  $p(r)$  is shown in Fig. 5.6B. The  $D_{max}$  and  $R_g$  calculated from the  $p(r)$  function were found to be  $462 \text{ \AA}$  and  $144 \pm 2 \text{ \AA}$ , respectively, agreeing favourably with the value obtained from the Guinier approximation. Thus, the values of  $R_g$  and  $D_{max}$  obtained from SANS are in close agreement with the SAXS values.



**Figure 5.6 Small angle neutron scattering of rE2/E3BP**

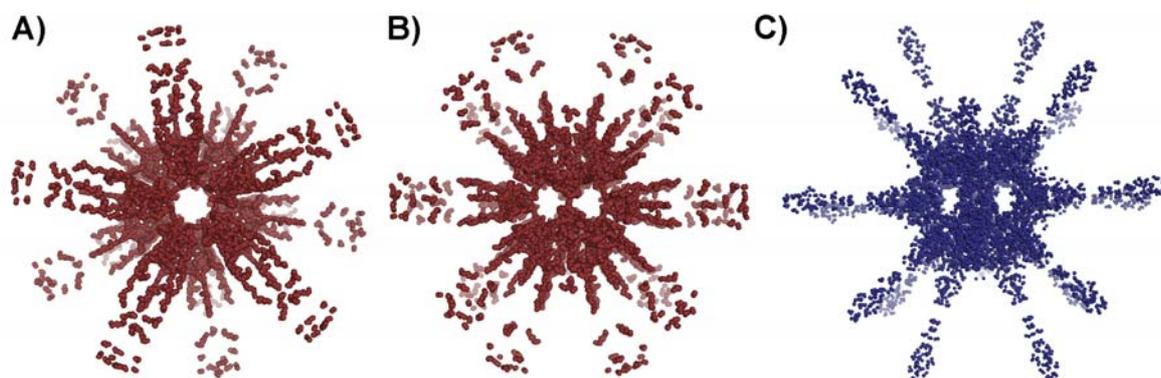
(A) Intensity scattering curve with error bars for rE2/E3BP is denoted. The x-axis  $s$  denotes the scattering angle. (B) A  $D_{max}$  of  $462 \text{ \AA}$  was calculated from the distance distribution function,  $p(r)$ , of rE2/E3BP using GNOM.

*Ab initio* models of rE2/E3BP (from SAXS and SANS data) were produced with the program GASBOR (Petoukhov and Svergun, 2003; Svergun, et al., 2001), employing icosahedral symmetry and various penalty constraints during the modelling process (see section 3.5.4). The *ab initio* reconstruction of rE2/E3BP was not entirely straightforward owing to its massive size and large number of

amino acid residues exceeding the maximum limit of dummy atoms/residues that scattering shape restoration programs like DAMMIN (Svergun, 1999) and

GASBOR can currently accommodate. Therefore, the modelling was performed employing fewer atoms than are present in the actual system. Initial *ab initio* reconstructions conducted with DAMMIN did not fit the experimental scattering curve satisfactorily and could not restore a biologically meaningful structure (see section 5.2.1.6). The algorithm in DAMMIN is known to pose considerable problems in restoring hollow structures (Volkov and Svergun, 2003). Thankfully, GASBOR modelling proved more fruitful, largely by virtue of the dummy residue having a smaller size, thereby allowing restoration of a more detailed internal structure for the core. Several GASBOR runs were conducted to obtain a consistent model: reconstructions were extremely time intensive (7-9 days/run). As GASBOR tries to best fit the experimental data with the imposed constraints, there was no unique solution and several structural models that satisfied the search volume were generated, but not all were biologically realistic. For instance, models were considered unrealistic if their appearance deviated substantially from that suggested by EM of mammalian PDC (Wagenknecht, et al., 1991), i.e. dodecahedral structures possessing large hollow cavities and empty pentagonal faces. Discarding the non-biologically relevant models, only those that appeared similar to published and our own EM images (section 5.3.1.4.1) were chosen for further analysis to obtain a consensus model for rE2/E3BP.

*Ab initio* models of rE2/E3BP obtained from SAXS and SANS are in good agreement with each other, with structural features such as hollow cavities and flanking arms observed in both models (Fig. 5.7). The positions of rE3BP within the core cannot be ascertained from the SAXS *ab initio* models. However, all reconstructions of the filtered subset result in structures having a maximum dimension of 472 Å and hollow empty faces, thus clearly lending support to the 48+12 (E2+E3BP) substitution model of subunit organisation. Greater electron density is observed within the inner part of the core as compared with the peripheral flanking arms. This may possibly be a consequence of the GASBOR algorithm favouring compactness (i.e. stacking greater number of dummy residues within the centre), modulated via peripheral



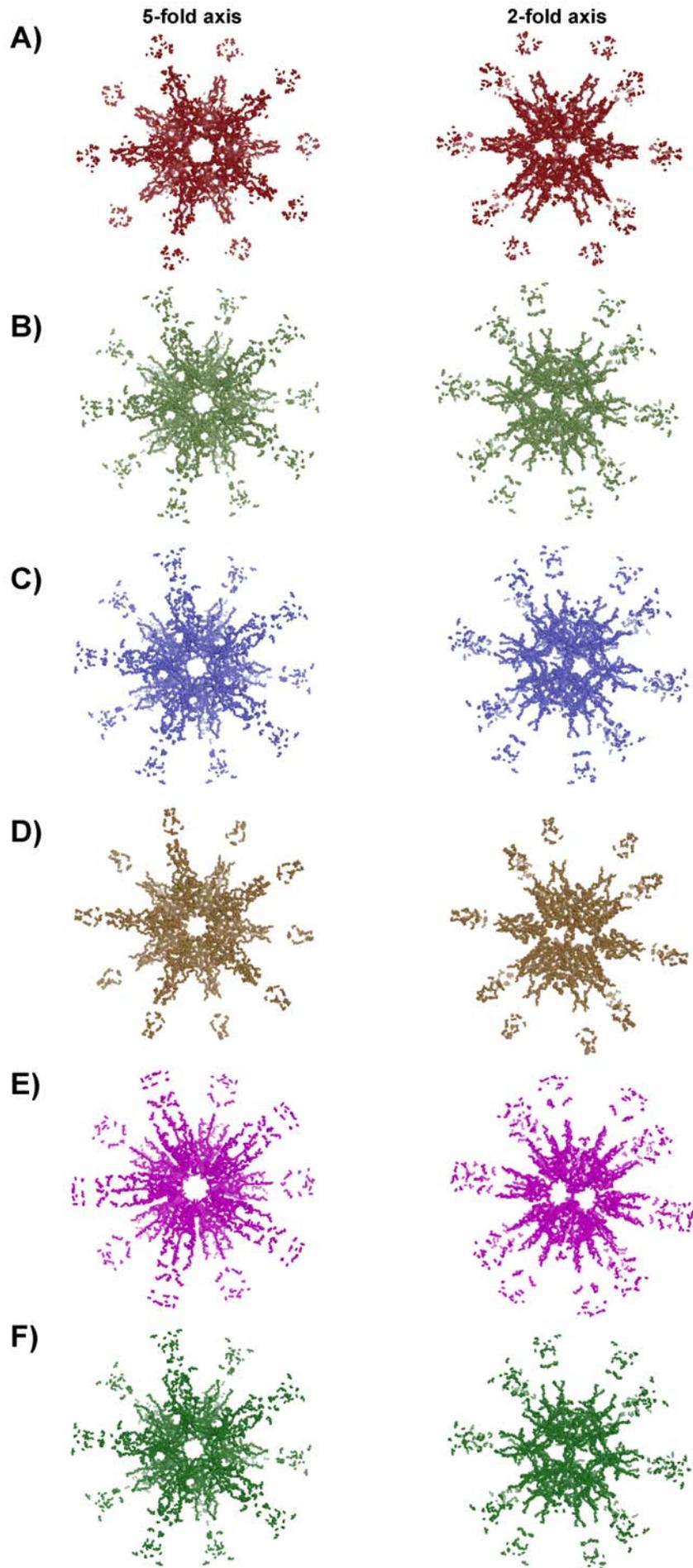
**Figure 5.7** *Ab initio* models of the rE2/E3BP core

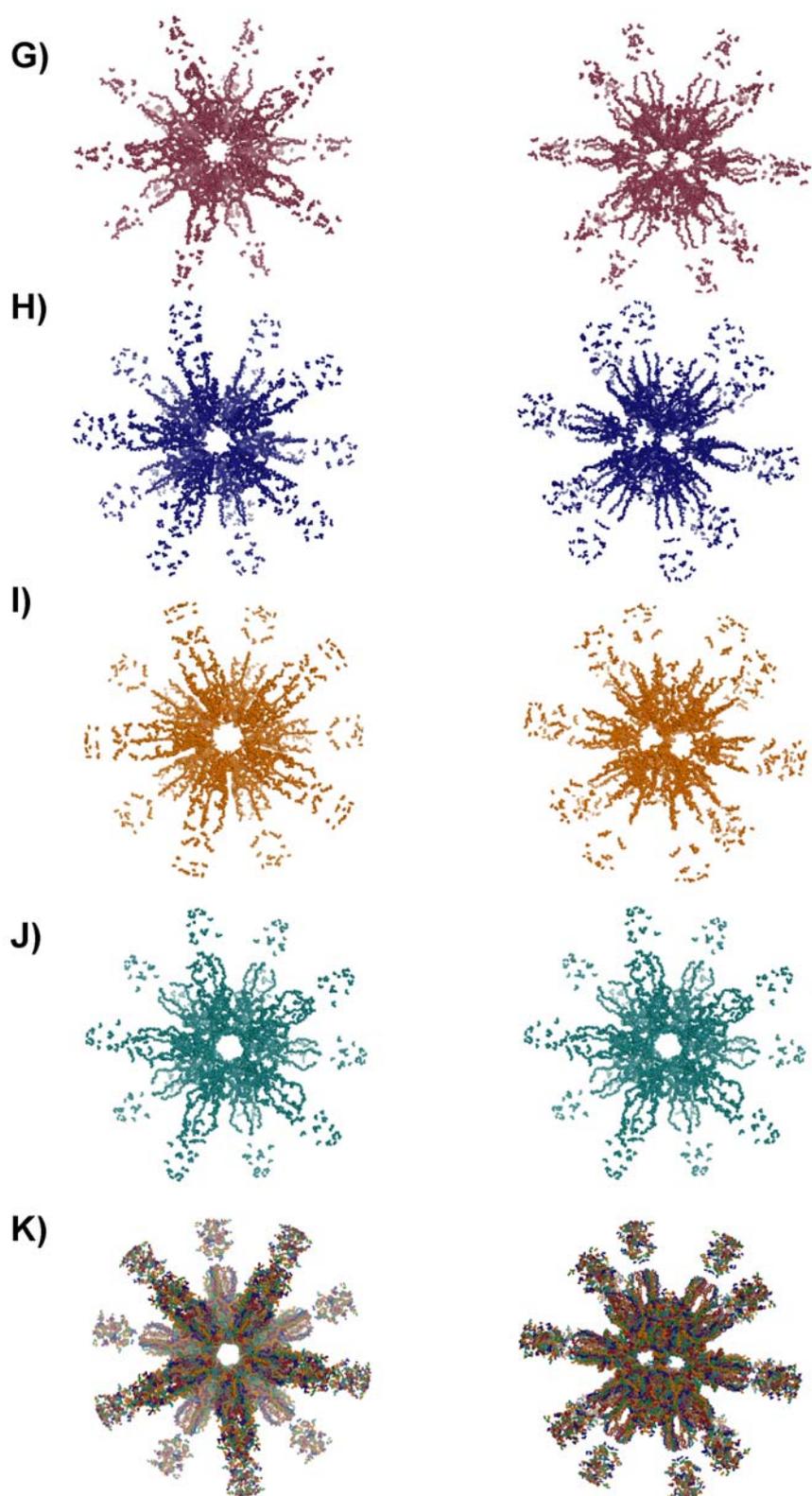
A single GASBOR reconstruction of rE2/E3BP obtained from SAXS data shown along the (A) 5-fold and (B) 2-fold axes of symmetry. (C) Complementary single GASBOR model of rE2/E3BP obtained from SANS data.

and bond length constraints that were used in this study.

On averaging 10 *ab initio* reconstructions via DAMAVER (Volkov and Svergun, 2003), it was impossible to retrieve an average structure that preserved the intrinsic features of each of the individual reconstructions. The averaged rE2/E3BP core thus obtained was a solid icosahedron, and did not possess any hollow internal cavities or flanking peripheral arms as seen in the individual models. In the case of both full-length cores (rE2/E3BP and bE2/E3BP), the algorithm in DAMAVER seems unable to deal with the complexity of several large cavities and externally elongated domains within the same structure. Although averaged models generated by DAMAVER are consistent with the individual *ab initio* models in most cases reported in the literature, there are some instances when averaging has little effect and shows systematic deviations from the initial shape (Volkov and Svergun, 2003).

However, as volumes obtained from CRY SOL (Svergun, et al., 1995) for each *ab initio* reconstruction compare favourably with the calculated volume for rE2/E3BP, a consensus model of rE2/E3BP was generated by the superimposition of 10 *ab initio* GASBOR models (Figs 5.8A to 5.8J) (refer to section 5.2.1.6 for details). It is interesting to note that the superimposition increases the overall electron density, particularly at the peripheral arms, whilst preserving the structural features such as the hollow cavities throughout all the models (Fig. 5.8K).

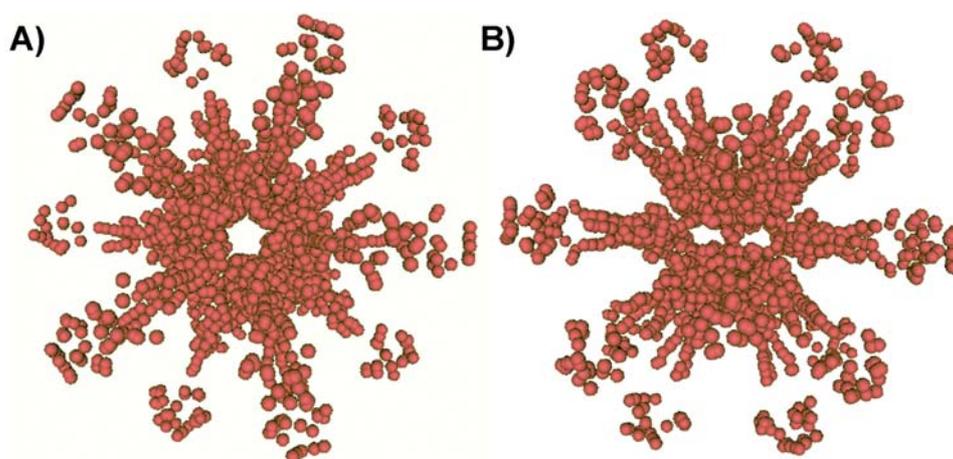




**Figure 5.8 Superimposition of 10 *ab initio* GASBOR models of rE2/E3BP core**  
10 *ab initio* GASBOR reconstructions (A to J) of rE2/E3BP were superimposed to give rise to a consensus model of rE2/E3BP (K) preserving all the key structural features of the core. All models are shown along the 2-fold and the 5-fold axes of symmetry, with the diameter of each model being 472 Å.

### 5.3.1.2.2 Hydrodynamic modelling

Hydrodynamic bead models (HBMs) of rE2/E3BP were independently generated from the *ab initio* GASBOR model using programs AtoB and TRANS2VORONOI as described in section 5.2.1.7 (Fig. 5.9). Assuming hydration of 0.4 g/g, a hydrated sedimentation coefficient of 28.3 S for rE2/E3BP was calculated using HYDRO++, in excellent agreement with the experimental value of 29.3 S obtained from SV data.



**Figure 5.9** AtoB generated model of rE2/E3BP core

The rE2/E3BP model generated by AtoB retains the key features of the parent structure (i.e. Fig. 5.8K). Shown along the (A) 5-fold and (B) 2-fold axes of symmetry for convenience.

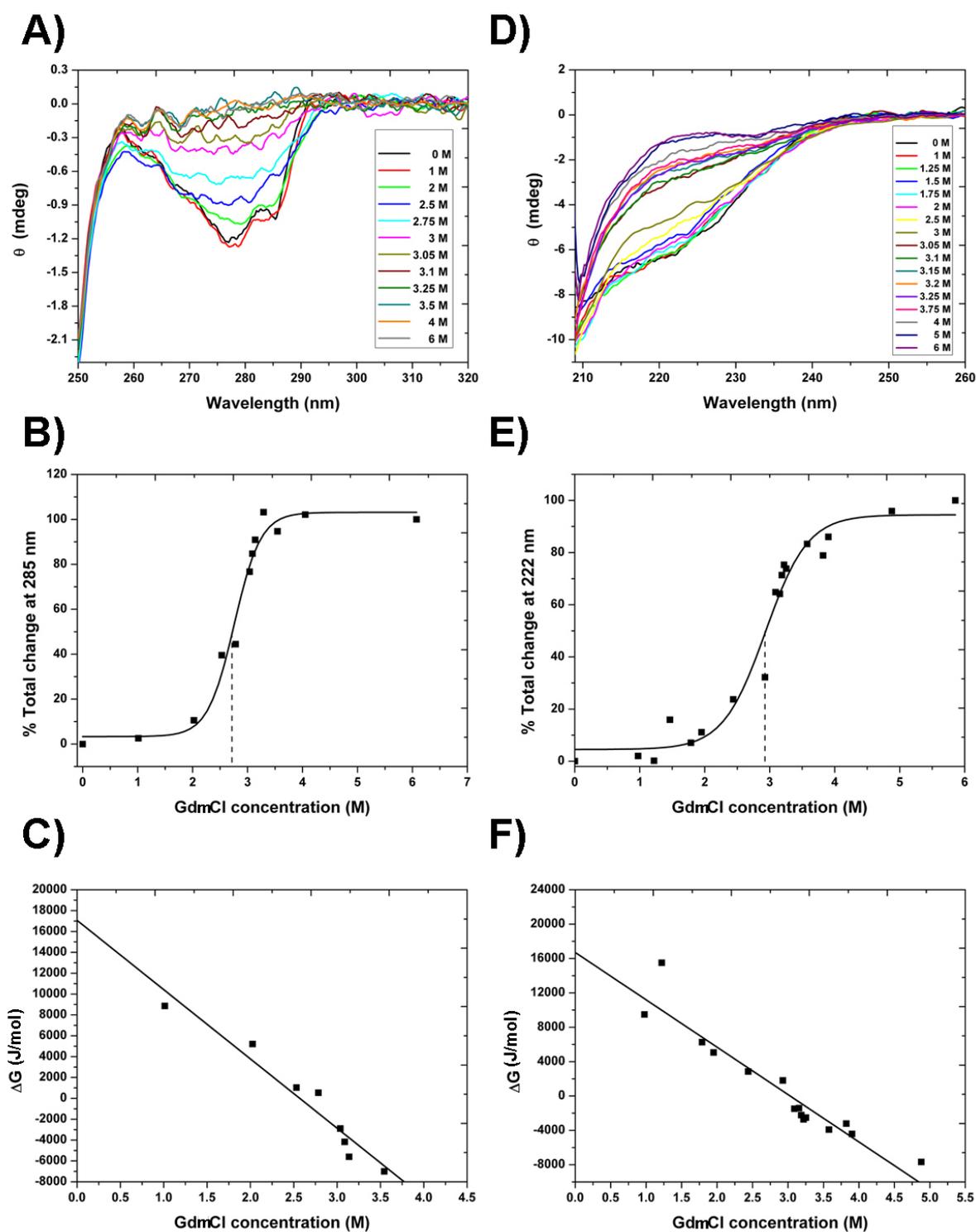
### 5.3.1.3 Stability of rE2/E3BP core

The trimeric building blocks that make up the cores of PDC serve as the key element for core stability (Izard, et al., 1999). Cryo-EM studies on the yeast E2 core revealed variation in the inter-trimer distances resulting in ‘breathing’ of the core (Zhou, et al., 2001a). Interestingly, integration of E3BP as an additional core component in eukaryotic PDCs may result in structural changes that mediate inter-trimer distances and overall core stability. As we have been able to produce recombinant PDC cores successfully, a comparative study on core stability of full-length rE2/E3BP, rE2 (section 5.3.3.3) and truncated E2/E3BP (section 6.3.5) cores was conducted using CD and fluorescence in the presence of the chemical denaturant GdmCl.

The CD spectrum of a protein in the far-UV region (180-260 nm) provides a quantitative estimate of the secondary structure content, while the near-UV spectrum (250-320 nm) is sensitive to aspects of tertiary structure related to aromatic amino acid residue contributions.

CD spectra of rE2/E3BP in the presence of increasing concentrations of GdmCl resulted in loss of minima at 285 nm and 222 nm (Figs 5.10A and 5.10D). At high concentrations, GdmCl absorbs strongly at wavelengths below 210 nm, therefore data below this wavelength were excluded from the analysis. Unfolding/dissociation was monitored by measuring changes in the CD signals at 222 nm and 285 nm. The 285 nm signal was more sensitive to GdmCl-induced unfolding/dissociation than its 222 nm counterpart, indicating a higher degree of structural perturbation of rE2/E3BP in the near-UV region. The mid-points of unfolding were 2.70 M GdmCl at 285 nm and 2.92 M GdmCl at 222 nm (Figs 5.10 B, 5.10E). This is consistent with the hypothesis that low concentrations of GdmCl cause overall destabilisation of local regions of tertiary structure, while the secondary structure is more resistant, requiring higher concentrations of GdmCl to unfold the structure completely. The peak at 285 nm corresponds to the spectral contribution from aromatic amino acid residues (such as tryptophan and tyrosine). This contribution is rapidly lost with increasing GdmCl. This may be attributed to conformational changes in the environment of aromatic residues resulting in alterations to the overall tertiary structure during unfolding/dissociation. In addition, the 'sigmoidal' curves suggest two-state unfolding (Fig. 5.10E). Earlier refolding studies on bE2/E3BP core demonstrated changes in the quaternary structure resulting in the formation of low molecular weight intermediates corresponding to trimers and monomers during the unfolding event (McCartney, et al., 1997).

The Gibbs free energy of unfolding of a protein,  $\Delta G$ , is assumed to be linearly dependent on the denaturant concentration. For a two-state reversible unfolding model (native state  $\rightleftharpoons$  unfolded state), the Gibbs free energy independent of concentration,  $\Delta G^0$ , can be obtained from the y-intercept of the plot of  $\Delta G$  versus denaturant concentration.  $\Delta G^0$  of rE2/E3BP unfolding was determined to be 17 kJ/mol and 16.2 kJ/mol from the near- and far-UV data, respectively (Figs 5.10C and 5.10F), thus indicating almost comparable extents of destabilisation

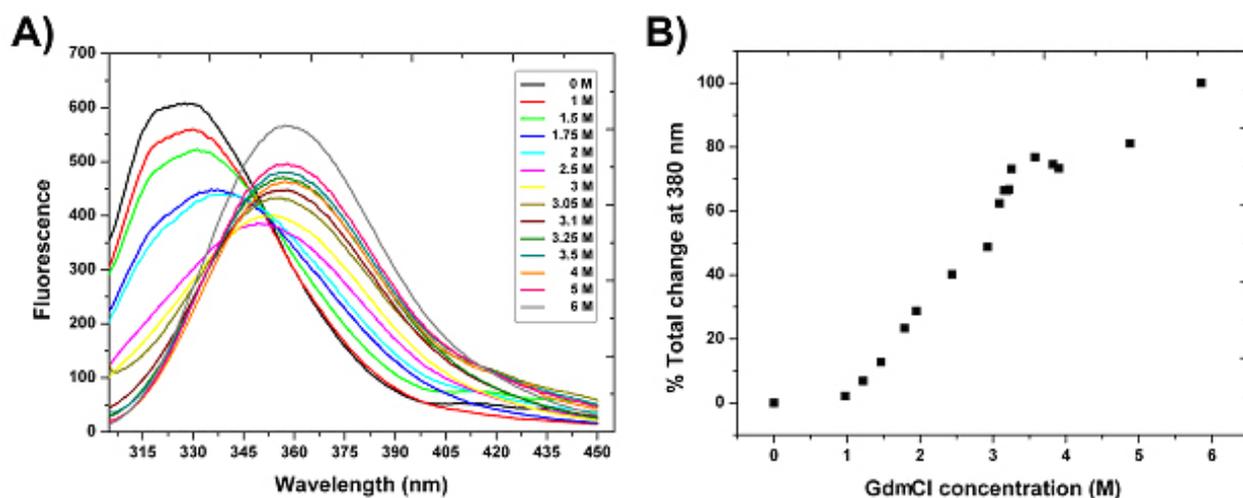


**Figure 5.10** CD spectra of GdmCl-induced denaturation of rE2/E3BP

(A) Near-UV and (D) far-UV CD spectra of rE2/E3BP recorded in the presence of increasing quantities of GdmCl are shown. Best fits to the stability curves denoting the percentage of unfolding of rE2/E3BP via (B) near-UV and (E) far-UV CD monitored at fixed wavelengths of 285 and 222 nm are given. Free energy plots of unfolding of rE2/E3BP in the (C) near- and (F) far-UV regions result in  $\Delta G^0$  values of 17 kJ/mol and 16.2 kJ/mol, respectively.

of the secondary and tertiary structure, and a similar mode of unfolding of the core.

Fluorescence spectra were recorded by monitoring the intrinsic fluorescence of tryptophans present in rE2/E3BP. A total of 7 tryptophans are found in the rE2/E3BP core, 4 in the rE2 subunit (1 outer LD, 1 inner LD, 2 CTD) and 3 in the rE3BP subunit (1 LD, 1 inner linker, 1 CTD) of the core. GdmCl-induced fluorescence emission spectra of rE2/E3BP recorded in the range 310-450 nm are shown in Fig. 5.11A. The maximum emission of rE2/E3BP in the fully folded native state in the absence of GdmCl is observed at 328 nm, indicative of tryptophans buried in the core. However, with the addition of increasing amounts of GdmCl, a gradual red-shift is observed.



**Figure 5.11 GdmCl-induced change in the fluorescence emission spectrum of rE2/E3BP**

(A) Fluorescence emission of tryptophans (excited at 295 nm) monitored at wavelengths 310-450 nm in the presence of increasing amounts of the chemical denaturant GdmCl is shown. (B) The percentage change of unfolding monitored at 380 nm as a function of GdmCl concentration indicates 50% of unfolded rE2/E3BP core at 2.6 M GdmCl. This wavelength (380 nm) was chosen as it showed a progressive trend from 0 M to 6 M GdmCl.

Interestingly, this red shift begins very early on with the addition of just 1.5 M GdmCl resulting in the shift of the maximum emission peak to 334 nm, suggesting local dissociation and partial exposure of tryptophans to the solvent. On increasing the concentration of GdmCl further, a dramatic shift of the maximum emission peak to 352 nm is observed at 2.5 M GdmCl, and this alters progressively to a maximum of 360 nm by 6 M GdmCl. The significant shift

towards 360 nm from 2.5 M GdmCl onwards suggests complete exposure of tryptophans to the solvent brought about by the overall unfolding of rE2/E3BP. As the fluorescence emission signal arises from all the tryptophans present in the rE2/E3BP core, contribution from specific tryptophans towards this signal as a consequence of unfolding or dissociation is difficult to assess at this stage.

Early changes in the observed tryptophan fluorescence between 0-2 M GdmCl may be indicative of local perturbation in structure and/or formation of trimeric intermediates owing to the partial solvent exposure of tryptophans prior to any major disruption of secondary structure. This is consistent with previous studies on bE2/E3BP demonstrating the formation of trimeric intermediates between 1.8-2 M GdmCl (McCartney, et al., 1997). This is immediately followed by a phase of rapid denaturation with about 50% unfolded/dissociated rE2/E3BP observed by 2.6 M GdmCl (Fig. 5.11B), corresponding to an overall unfolding event with major loss of quaternary and secondary structure of the core.

### **5.3.1.4 Towards a high-resolution structure**

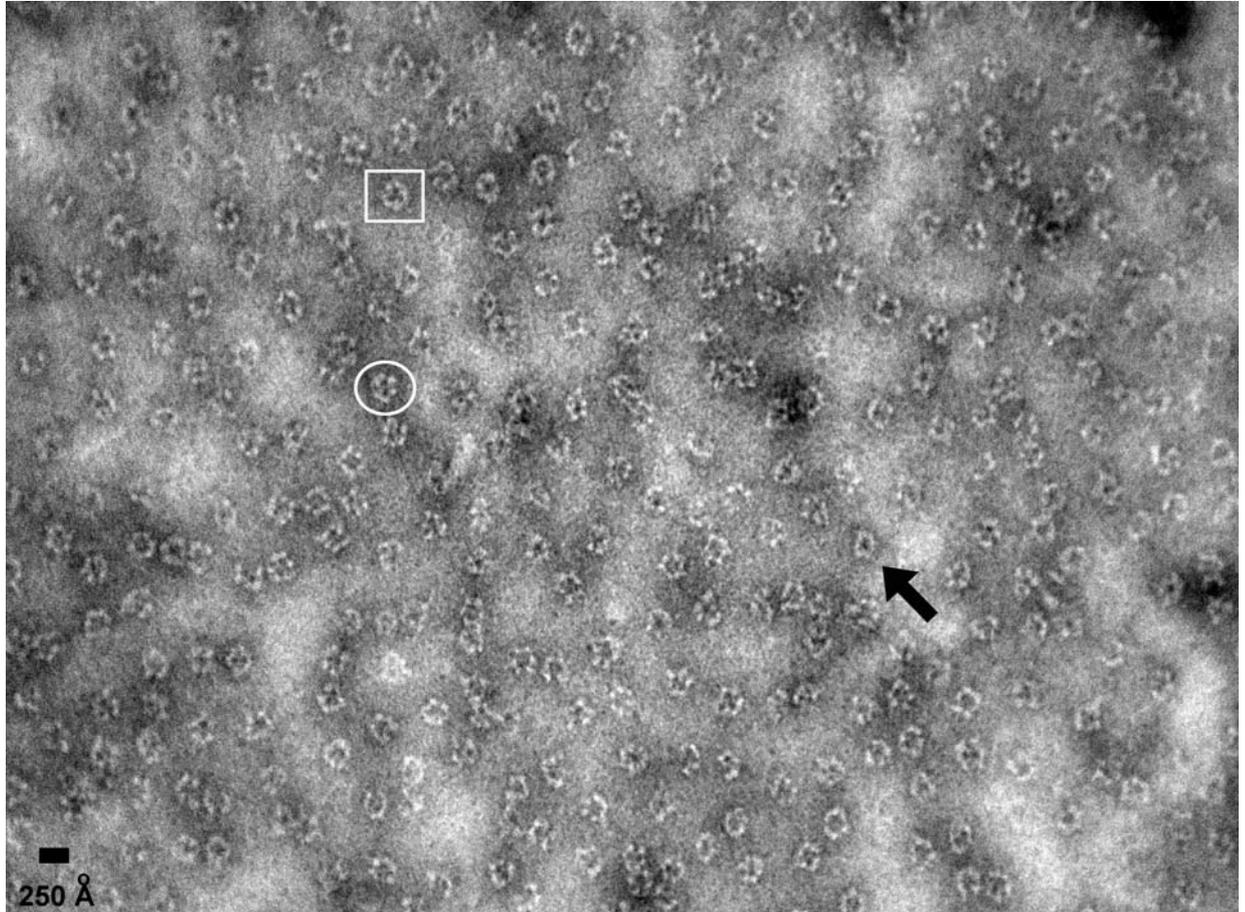
#### *5.3.1.4.1 Negative stain EM*

Negative stain EM of rE2/E3BP reveals a uniform distribution of well formed icosahedral core structures with empty pentagonal faces (Fig. 5.12), consistent with previous EM data for bovine heart PDC (Wagenknecht, et al., 1991). As the flexible N-terminal arms of the core are likely to be fixed by the stain, they are poorly resolved on the EM micrograph. The observation of empty pentagonal faces with no density above or below them provides strong support for the 'substitution model' of the rE2/E3BP assembly. Core structures are clearly visible, exhibiting the underlying 5-, 3-, and 2-fold structural symmetry (Fig. 5.12); although the 5-fold cores are more common than the 3- and 2-fold structures, consistent with cryo-EM data for bE2/E3BP (Wagenknecht, et al., 1991).

#### *5.3.1.4.2 Preliminary crystallisation trials*

Crystallisation of purified rE2/E3BP was also attempted, as a first step towards a high resolution structure. However, initial crystallisation screens yielded only

salt crystals. It is well known that highly mobile regions of proteins significantly hinder the process of crystallisation (Dale, et al., 2003; Sousa, 1995), and it is likely that the flexible N-terminal arms of the rE2/E3BP core hinder the production of protein crystals.



**Figure 5.12 Negative stain EM image of rE2/E3BP**

This negative stain image of rE2/E3BP shows uniformly distributed complete icosahedral cores with empty (pentagonal) faces. Orientation of cores along the 5- (arrow), 3- (square) and 2-fold (circle) axes along with the scale is also indicated.

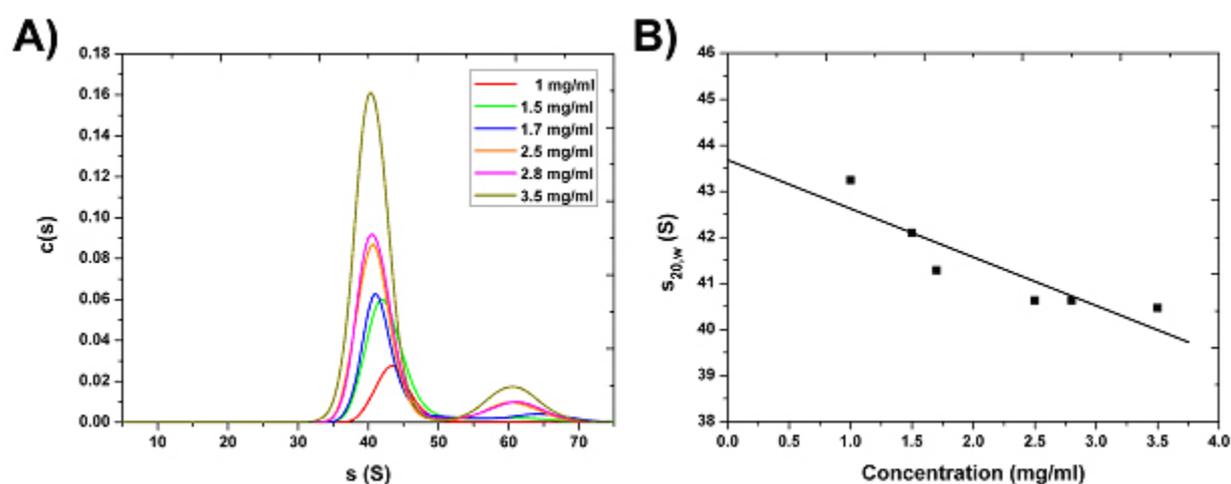
### **5.3.2 Bovine E2/E3BP core (bE2/E3BP)**

#### **5.3.2.1 Shape determination of bE2/E3BP**

##### *5.3.2.1.1 SV AUC*

It proved impossible to obtain homogenous bE2/E3BP core, and protein samples used for AUC analysis were contaminated by a small fraction of E1 (10-20%), as

estimated from visualisation after SDS-PAGE (see section 4.3.2.7) and densitometric analysis. SV experiments were conducted for various concentrations of bE2/E3BP (with E1) using interference optics and the sedimentation profiles modelled using  $c(s)$  analysis in SEDFIT (Schuck, 2000). The  $c(s)$  analysis shows a dual peak profile; a dominant peak with a weight average sedimentation coefficient of 41 S and a slightly smaller peak at 60 S (Fig. 5.13A). Finite element analysis with a two-species model yielded sedimentation coefficients at all experimental concentrations. These were then extrapolated to give a sedimentation coefficient for the major peak of bE2/E3BP independent of concentration,  $s_{20,w}^0 = 43.7 \pm 0.03$  S (Fig. 5.13B). The major peak has a sedimentation coefficient significantly greater than the already published value of 35 S for the bovine E2/E3BP core, but this is consistent with an expected increase due to bound E1. The smaller peak (60 S) may correspond to possible dimers of bE2/E3BP or aggregates, consistent with observations made by Roche and co-workers (1993). Hence, the major peak with sedimentation coefficient of 43.7 S represents 60-meric bE2/E3BP populated with small amounts of bound E1.



**Figure 5.13 Sedimentation velocity analysis of bE2/E3BP core**

(A)  $c(s)$  analysis for a range of concentrations of purified bE2/E3BP core. The axes  $s$  and  $c(s)$  denote the sedimentation coefficient and the  $c(s)$  distribution model (from SEDFIT analysis), respectively. (B) Determination of  $s_{20,w}^0$  for the major peak of bE2/E3BP. Error bars are shown, but are not clearly visible owing to their small size.

### 5.3.2.1.2 SE AUC

As the exact amino acid composition of bE2/E3BP is not known, it was assumed that bE2/E3BP is similar to the recombinant rE2/E3BP core in terms of subunit composition, extinction coefficient and other solvent properties for the purpose of this study.

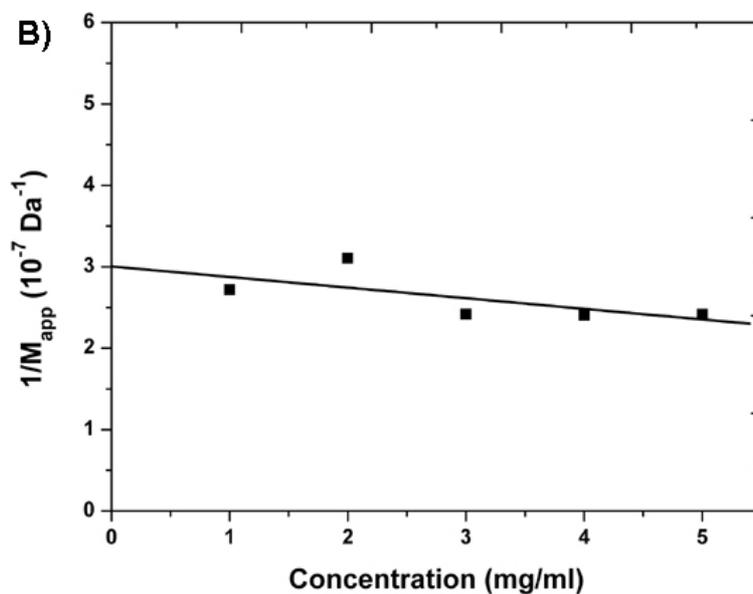
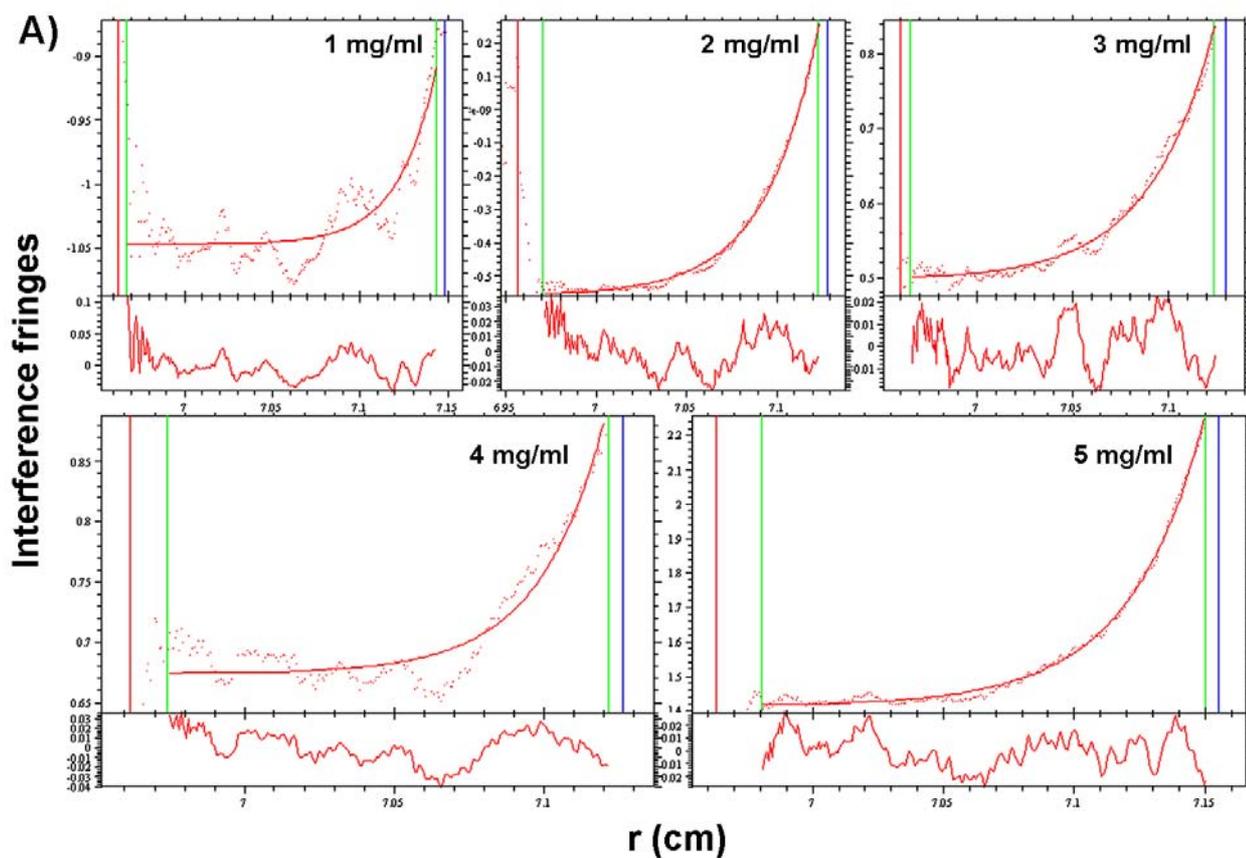
SE data obtained for bE2/E3BP were fitted with a two species model using the discrete species module in SEDPHAT (Schuck, 2003; Vistica, et al., 2004), yielding molecular masses of 3.59 MDa and 5.69 MDa for the first (predominant) and second (minor) species, respectively (Fig. 5.14A). A decrease in  $1/M_{app}$  of the predominant species with increasing sample concentration was clearly observed, indicative of self-association and/or aggregate formation (Fig. 5.14B). The molecular weight of the major species of bE2/E3BP independent of concentration,  $M_w^0$ , was determined to be  $3.33 \pm 0.6$  MDa (Fig. 5.14B).

Although this value compares favourably with the molecular weight of recombinant E2/E3BP core as predicted from the amino acid sequence (3.55 MDa), it is considerably less than that expected for bE2/E3BP bound with 20% E1 (~4.47 MDa). This may be due to proteolytic degradation of the N-terminal arms of bE2/E3BP over the long time course of the SE experiment resulting in a low molecular weight core. Moreover, global fitting of the SE data yielded a molecular mass of 5.69 MDa for the second minor species which is lower than expected for intact bE2/E3BP (bound E1) dimer. Furthermore fitting to the self-association model in SEDPHAT was unsatisfactory, indicating the presence of large aggregates, rather than higher order oligomers.

It is to be noted that the SE data was extremely noisy for some bE2/E3BP samples. This may be due to protein sticking on the walls of the sample cells or time invariant noise that arises owing to dust or scratches on the sapphire windows, variations in the interference optics etc. Therefore, SE experiments of bE2/E3BP will have to be conducted again in order to ensure reliability of data.

### 5.3.2.2 *Ab initio* solution structure of bE2/E3BP

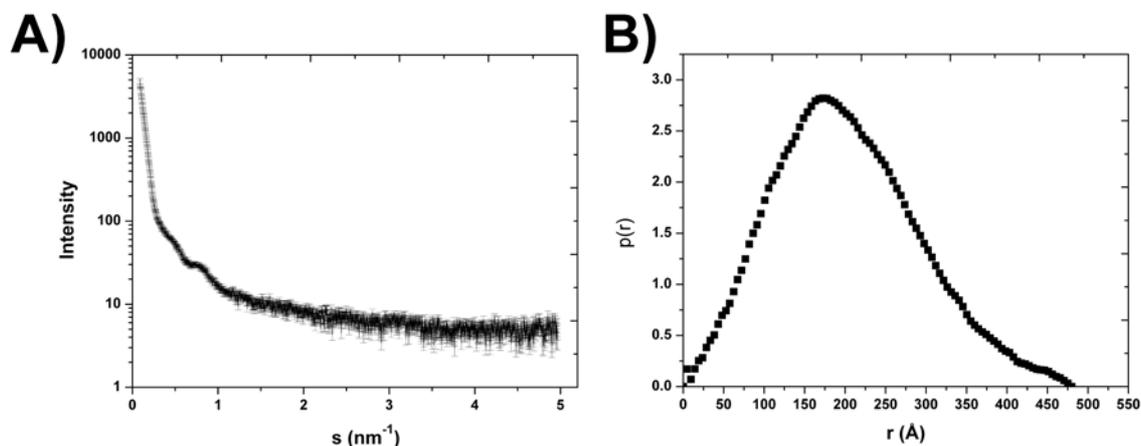
SAXS data for purified bE2/E3BP were acquired at sample concentrations of 0.65 and 1.14 mg/ml at temperatures 10°C and 37°C. However, measurements at 37°C



**Figure 5.14 Sedimentation equilibrium analysis of bE2/E3BP**

(A) SE interference data (dotted) were analysed with SEDPHAT to obtain best fits (smooth) and  $M_{app}$  at the various sample concentrations. Residual plots of the various samples (1-5 mg/ml) are represented on different scales. (B)  $1/M_{app}$  was plotted for the predominant species as a function of sample concentration and extrapolated to zero concentration to yield  $M_w^0$ .

indicated sample aggregation. Scattering data obtained at the lower concentration (0.65 mg/ml) at 10°C revealed no apparent aggregates (Fig. 5.15A) and were employed for further *ab initio* modelling. An attempt to estimate the molecular weight of bE2/E3BP obtained by extrapolation of scattering intensity to zero angle,  $I(0)$ , did not provide satisfactory results.



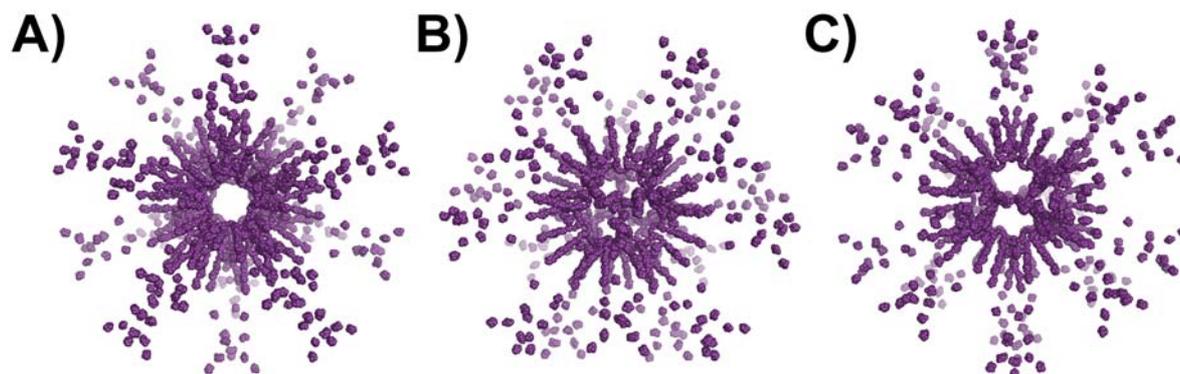
**Figure 5.15 Small angle x-ray scattering of bE2/E3BP core**

(A) The scattering curve for bE2/E3BP is shown. The x-axis  $s$  denotes the scattering angle. (B) The distance distribution function,  $p(r)$ , calculated using GNOM results in a maximum diameter,  $D_{max}$ , of 480 Å. Error bars are shown in both plots, but not clearly visible in (B) owing to their small size.

The radius of gyration  $R_g$  determined using the Guinier approximation is  $158 \pm 1$  Å. The particle distribution function  $p(r)$  is shown in Fig. 5.15B. The  $D_{max}$  is 480 Å and  $R_g$  calculated from the  $p(r)$  function is  $156 \pm 1$  Å which agrees well with the value obtained using the Guinier approximation. As the binding site of E1 is situated close to the inner part of the core, the maximum particle diameter of bE2/E3BP is unaffected in the presence of bound E1. The  $p(r)$  curves and values of  $R_g$  and  $D_{max}$  determined in this study from scattering data of bE2/E3BP and rE2/E3BP compare favourably with each other, indicating a high degree of similarity in the molecular shape of these cores.

*Ab initio* shape restoration of bE2/E3BP using GASBOR (Petoukhov and Svergun, 2003; Svergun, et al., 2001) produced an icosahedral symmetrical core with empty pentagonal faces, further confirming the structural similarity with the rE2/E3BP core (Fig. 5.16). *Ab initio* modelling of bE2/E3BP was carried out in a manner similar to that for rE2/E3BP by imposing similar penalties and

icosahedral symmetry during the modelling process. More details on the modelling process can be found in section 5.3.1.2.



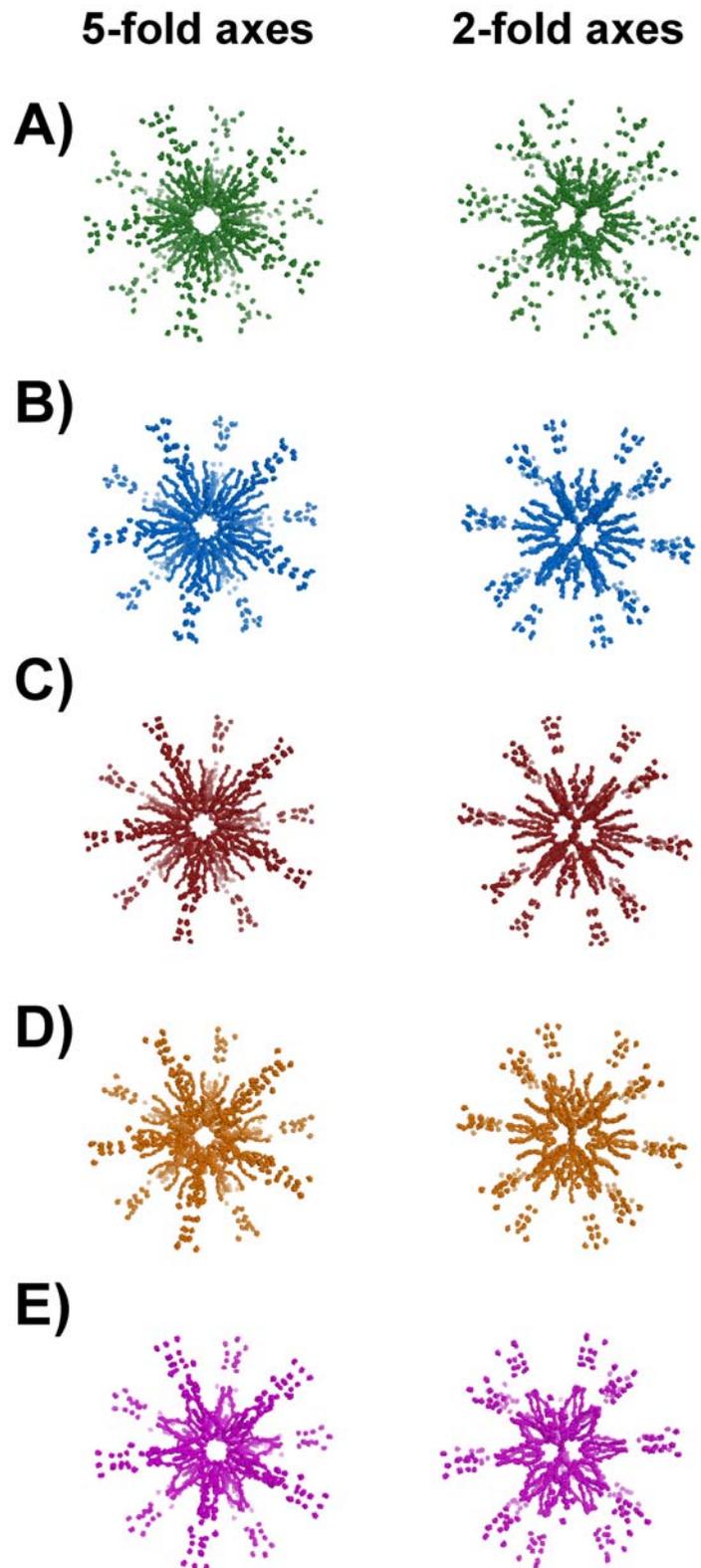
**Figure 5.16** *Ab initio* models of bE2/E3BP core

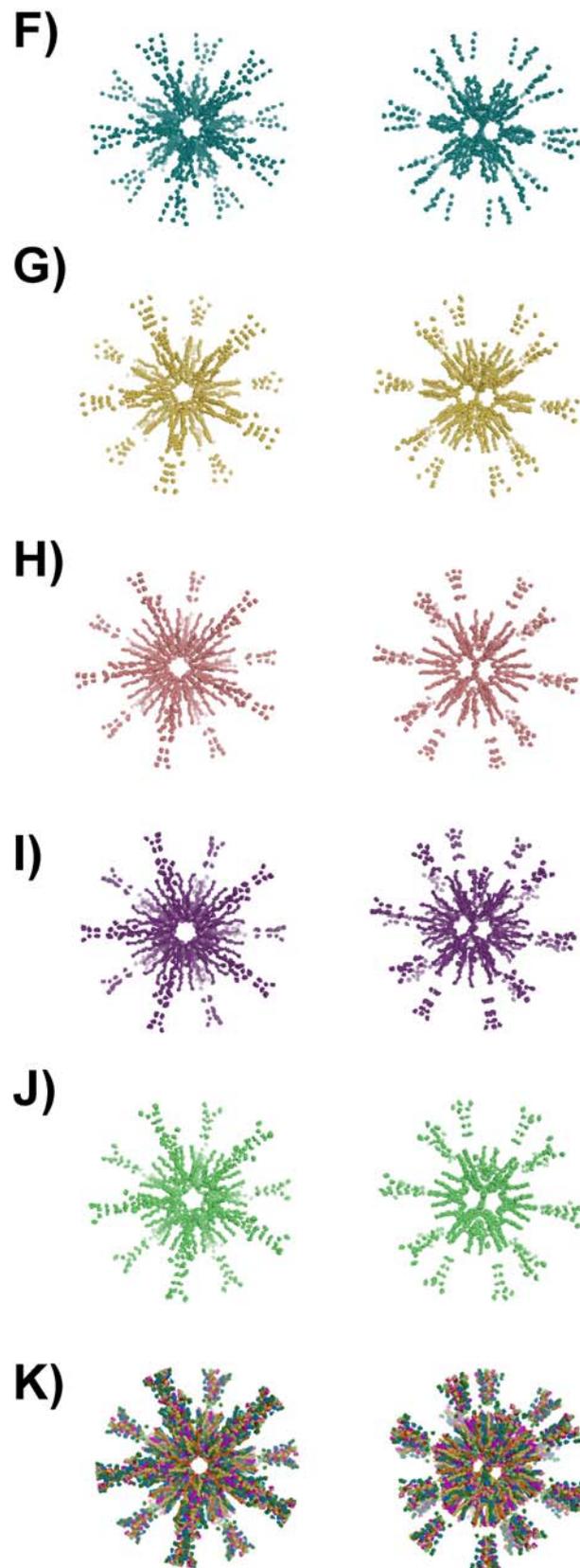
A single *ab initio* GASBOR reconstruction of bE2/E3BP obtained from SAXS data shown along the (A) 5-fold and (B) 3-fold and (C) 2-fold axes of symmetry.

All *ab initio* models of the filtered subset of bE2/E3BP generated structures having empty pentagonal faces and a maximum dimension of 480 Å, similar to the rE2/E3BP core. It was not possible to obtain an average structure using DAMAVER similar to the rE2/E3BP core (see section 5.3.1.2.1) (Volkov and Svergun, 2003), as the algorithm was unable to handle the structural complexity (hollow faces and elongated arms) posed by the modelled data. Therefore, akin to the modelling process for rE2/E3BP described in section 5.3.1.2, 10 *ab initio* GASBOR models of bE2/E3BP (Figs 5.17A-5.17J) were superimposed to obtain a consensus model, thus increasing the overall electron density and preserving all structural features of the core (Fig. 5.17K).

### 5.3.2.3 Electron microscopy of bE2/E3BP core

Negative stain EM of the bE2/E3BP core complements previous observations of the dodecahedral morphology exhibited by eukaryotic PDC cores (Fig. 5.18) (Wagenknecht, et al., 1991; Yu, et al., 2008). Despite the lack of uniform staining, icosahedral cores along the 2-fold, 3-fold and 5-fold axes are observed. However, the pentagonal faces of bE2/E3BP are not as clearly visible as those observed in the EM micrograph for rE2/E3BP (Fig. 5.12). This may be due to E1 contamination in the purified preparation of bE2/E3BP. The heterotetrameric E1 has dimensions of 70-100 Å (Hayakawa, et al., 1969; Junger and Reinauer, 1971)

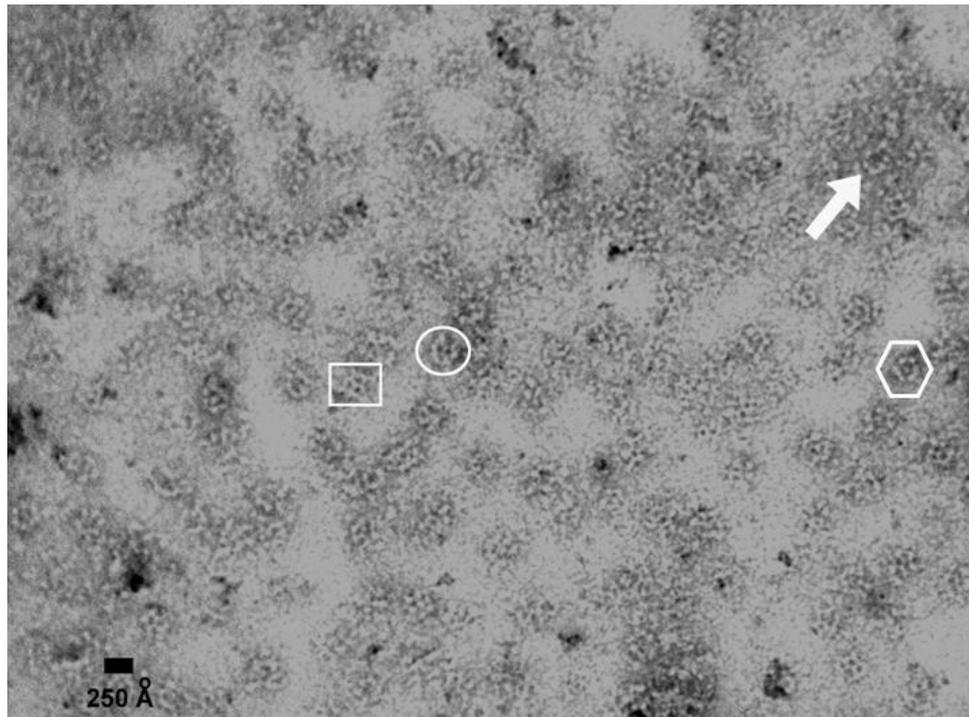




**Figure 5.17 Superimposition of 10 *ab initio* GASBOR models of bE2/E3BP**

The superimposition of 10 *ab initio* GASBOR models of bE2/E3BP (A to J) resulted in a consensus model of bE2/E3BP (K). Models are shown along the 5-fold and 2-fold axes, with maximum dimension of 480 Å.

and binds to the SBD of E2 at a distance of about 60 Å from the inner bE2/E3BP core across the pentagonal faces (Wagenknecht, et al., 1991). However, the hollow pentagonal face has a diameter of only about 50 Å (Wagenknecht, et al., 1991), smaller than the maximum dimension of E1, hence contaminating E1 seems to partially obscure these pentagonal faces. This is consistent with previous observations by Zhou and co-workers (2001b) based on cryo-EM of bovine kidney PDC complexed with E1.



**Figure 5.18 Electron micrograph of bE2/E3BP**

Negative stain image of bE2/E3BP indicates some icosahedral structures with empty pentagonal faces along their 5- (arrow), 3- (square) and 2-fold (circle) axes. bE2/E3BP core with bound E1 (hexagonal) is also shown. Scale bar of 250 Å is denoted.

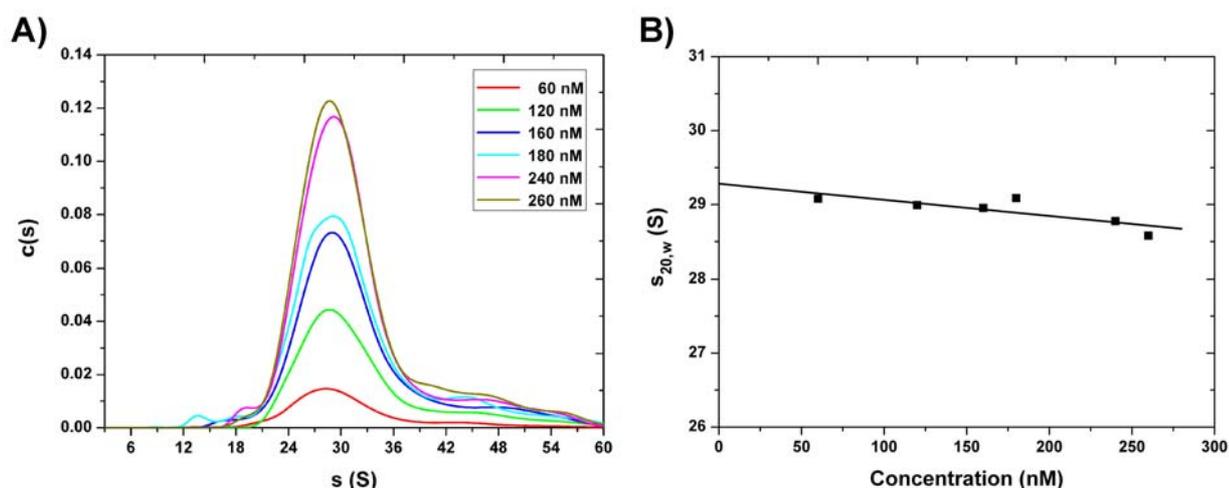
### **5.3.3 Recombinant E2 core (rE2)**

#### **5.3.3.1 Shape determination of rE2**

##### **5.3.3.1.1 SV AUC**

SV data show that rE2 is largely monodisperse at all concentrations, with a prominent major peak corresponding to intact 60-meric core, and a trailing edge of high molecular weight species that could possibly be rE2 dimer and/or

aggregates (Fig. 5.19A). Weight average sedimentation coefficients of 28.5 S and 47 S were obtained by integration of each peak in the  $c(s)$  analysis. Finite element analysis of the SV data with a two-species model yielded sedimentation coefficients at all experimental concentrations. These were then extrapolated to infinite dilution to give  $s_{20,w}^0 = 29.3 \pm 0.02$  S for the intact rE2 core (Fig. 5.19B), slightly lower than the value of 36.0 S reported by Hiromasa and co-workers (2004). This may reflect the different approaches employed in sample preparation and data analysis.



**Figure 5.19 Sedimentation velocity analysis of rE2**

(A)  $c(s)$  distribution derived from SV interference data collected over a range of rE2 concentrations. The axes  $s$  and  $c(s)$  denote the sedimentation coefficient and the  $c(s)$  distribution model (from SEDFIT analysis), respectively. (B) Determination of  $s_{20,w}^0$  of rE2 using finite element analysis in SEDFIT. Error bars are shown, but are not clearly visible owing to their small size.

The value of  $f/f_0$  for rE2 calculated in a manner similar to rE2/E3BP (see section 5.3.1.1.1) was found to be 2.79 and indicates significant deviation from spherical symmetry, consequently resulting in a core with large hydrodynamic radius,  $R_s$ . Some of the hydrodynamic values calculated for rE2 are tabulated below (Table 5.3).

### 5.3.3.1.2 SE AUC

SE data obtained for rE2 show a small increase in  $1/M_{app}$  with increasing sample concentration, possibly indicative of some non-ideality owing to the presence of

$M$ (Da)	$\bar{v}$ (ml/g)	$s_{20,w}^0$ (S)	$f/f_0$	$R_0$ (nm)	$R_s$ (nm)	$D_s$ (Å)	$D_t$ (cm <sup>2</sup> /s)
3,741,780	0.744	29.3 <sup>a</sup> 81.9 <sup>b</sup>	2.79	10.33	28.8	576	7.68 x 10 <sup>-8</sup>

**Table 5.3 Hydrodynamic parameters of rE2 derived from SV analysis**

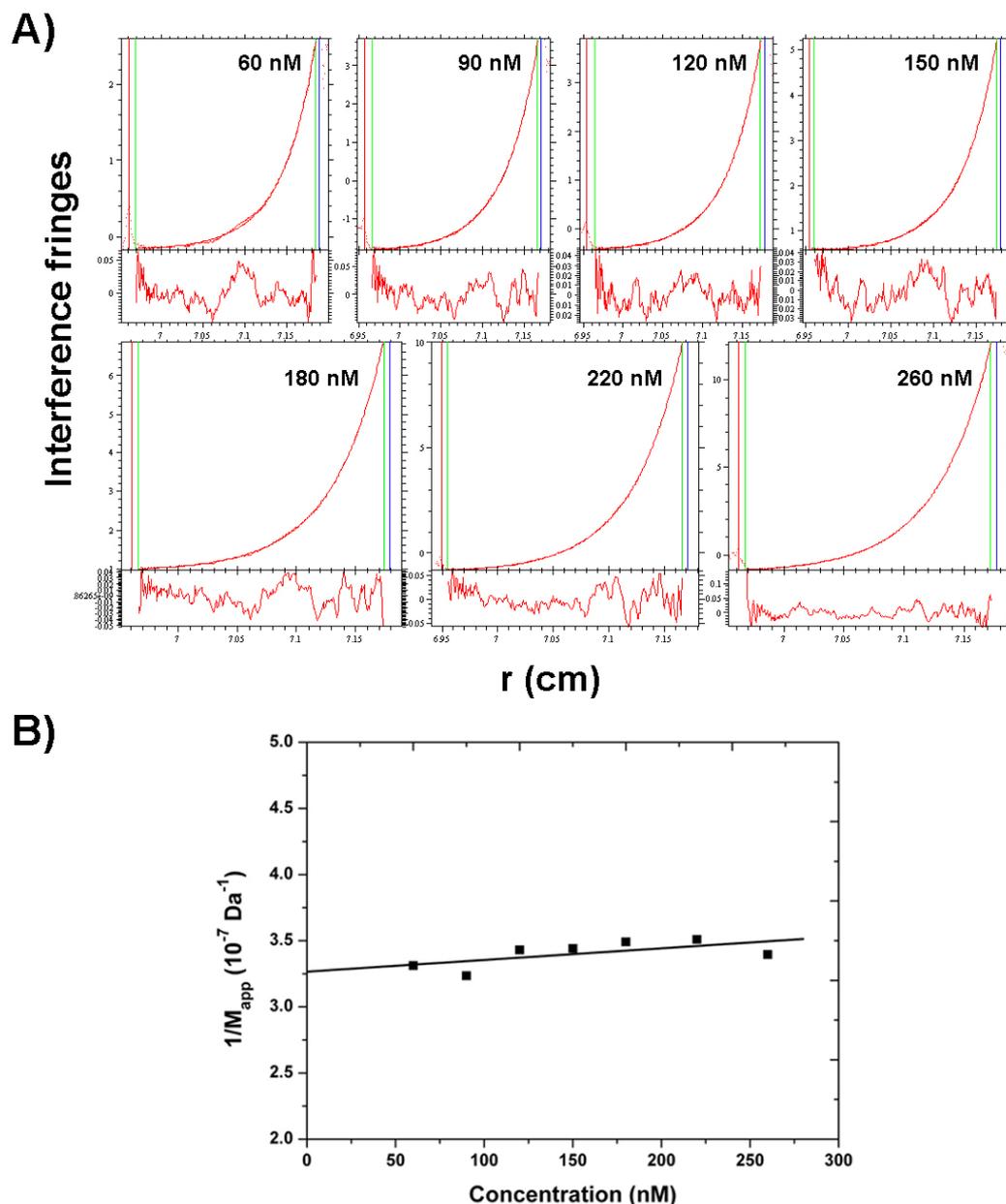
<sup>a</sup>experimental and <sup>b</sup>calculated sedimentation coefficients.  $M$  is the molecular mass based on the amino acid composition,  $\bar{v}$  is the calculated partial specific volume at 20°C,  $s_{20,w}^0$  is the sedimentation coefficient at infinite dilution,  $R_0$  is the unhydrated radius of a sphere of equivalent mass and partial specific volume as that of rE2,  $R_s$  and  $D_s$  are the hydrodynamic radius and its equivalent diameter.  $D_t$  is the translational diffusion coefficient.

hollow solvent-filled cavities found in these cores. SE data were globally analysed using a two species model in SEDPHAT (Schuck, 2003; Vistica, et al., 2004) as shown in Fig. 5.20A and yielded molecular masses of 2.78 MDa and 4.59 MDa for the first (predominant) and second (minor) species, respectively.

The molecular weight of rE2 (the predominant species) independent of concentration,  $M_w^0$ , was determined to be  $3.06 \pm 0.25$  MDa, lower than its predicted molecular weight (3.74 MDa) (Fig. 5.20B). This is probably due to a small fraction of incomplete cores or proteolytic degradation of rE2 over the long time period of the SE experiment (3-4 days); in particular the linker regions connecting the various outer N-terminal domains are highly susceptible to proteolysis. Proteolytic cleavage is also apparent in samples of rE2 analysed by SDS-PAGE after the SE run (Fig 5.21). SDS-PAGE indicates several low molecular weight bands that may correspond to various forms of cleaved rE2 (Fig. 5.21). However, the intact core being a larger molecule has a higher capacity to bind to the Coomassie blue stain than the smaller proteolytic products, thereby posing great difficulties in estimating the amount of truncated cores present in the samples.

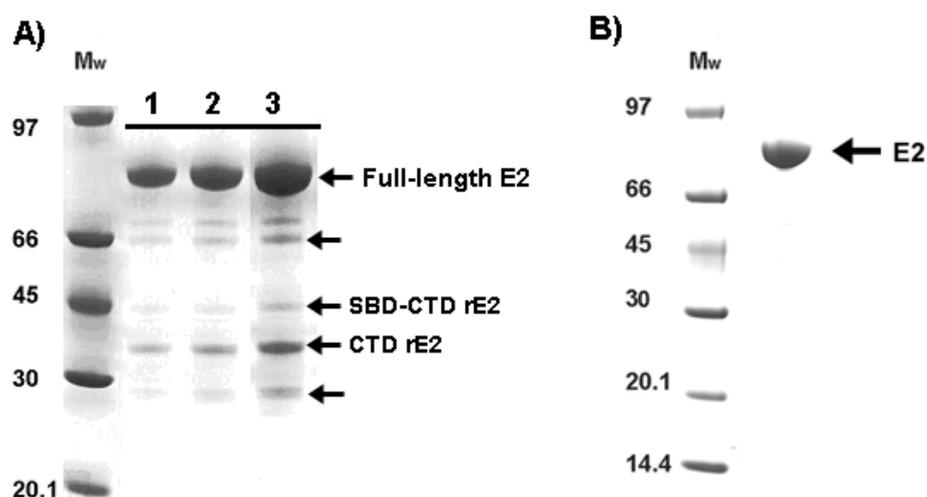
As proteolysis of 60-meric rE2 can result in cleaved cores comprising various domains, the actual sample may be a mixture of rE2 cores. As a result several possibilities exist towards formation of a 3.04 MDa rE2 core, some of which are listed in Table 5.4. Therefore, determination of the exact molecular weight of

rE2 is difficult owing to the underlying heterogeneity and inherent variation in core organisation.



**Figure 5.20 Sedimentation equilibrium analysis of rE2**

**(A)** Best fits (smooth lines) to the SE interference data (dotted lines) for rE2 at 3,000 rpm using a two species model in SEDPHAT. Residual plots of all the samples are represented on different scales.  $M_{app}$  was determined at each sample concentration. **(B)** Extrapolation to zero of the  $1/M_{app}$  vs concentration plot of the predominant species yields the molecular mass,  $M_w^0$  independent of concentration for rE2.



**Figure 5.21 SDS-PAGE of post SE samples of rE2**

(A) Post SE samples of rE2 were analysed on SDS-PAGE (Lanes 1-4) and show several bands relating to proteolytic products (indicated by arrows) of rE2. The different lanes represent various concentrations of samples used. (B) Freshly purified rE2 is shown for comparison. Molecular weights of marker proteins (Mw) are indicated in kDa.

% of rE2 with all domains (full core)	% of CTD rE2 cores	% of truncated SBD-CTD rE2 cores	Resultant molecular weight $M$ (MDa)
100	-	-	3.74
-	100	-	1.74
-	-	100	2.79
65	35	-	3.04
26	-	74	3.04
50	21	29	3.04

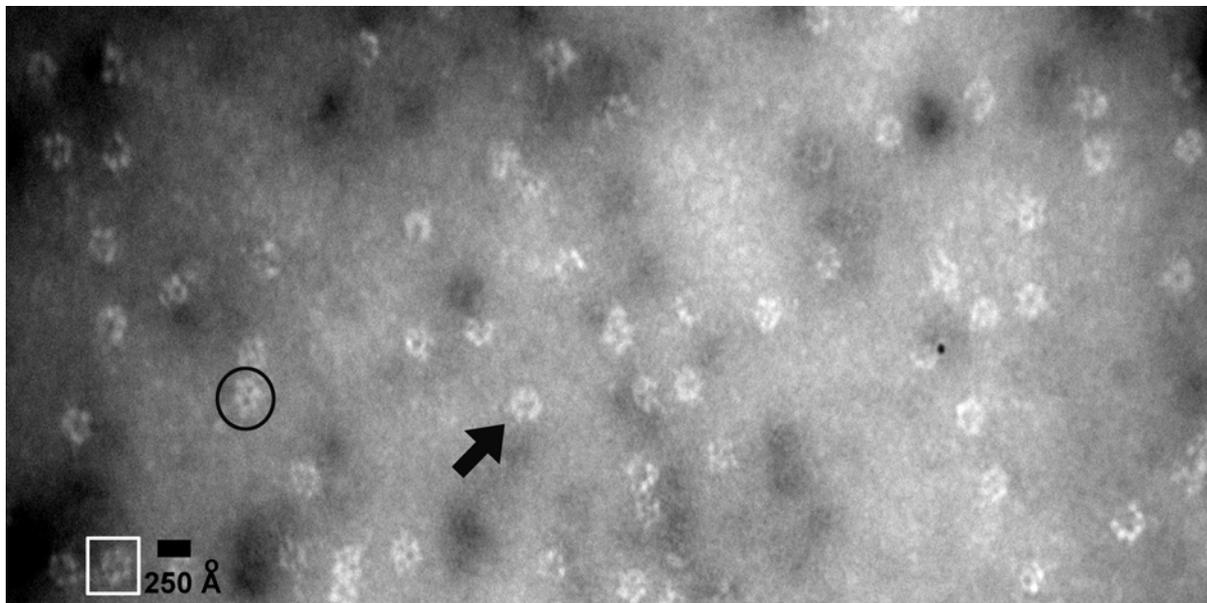
**Table 5.4 Variable core organisation of rE2 and its effect on molecular weight**

Possible combinations of the different rE2 cores possessing various domains (full and truncated) resulting in a molecular weight consistent with the value obtained from SE data (blue) for rE2.

Global fitting of SE data yielded a molecular weight of 4.59 MDa for the second minor species (i.e. 47 S peak from rE2 SV analysis) lower than that expected for 60-meric rE2 dimer (data not shown). Moreover, data modelled with the self association model in SEDHAT did not yield satisfactory results, indicating the presence of aggregates rather than rE2 dimers or other higher order oligomers.

### 5.3.3.1.3 Negative stain EM

Negative staining EM of rE2 showed the presence of icosahedral cores with empty pentagonal faces (Fig. 5.22), in agreement with previously observed EM data for rE2/E3BP and bE2/E3BP (sections 5.3.1.4.1 and 5.3.2.3), and published cryo-EM data on the bE2/E3BP core (Wagenknecht, et al., 1991).

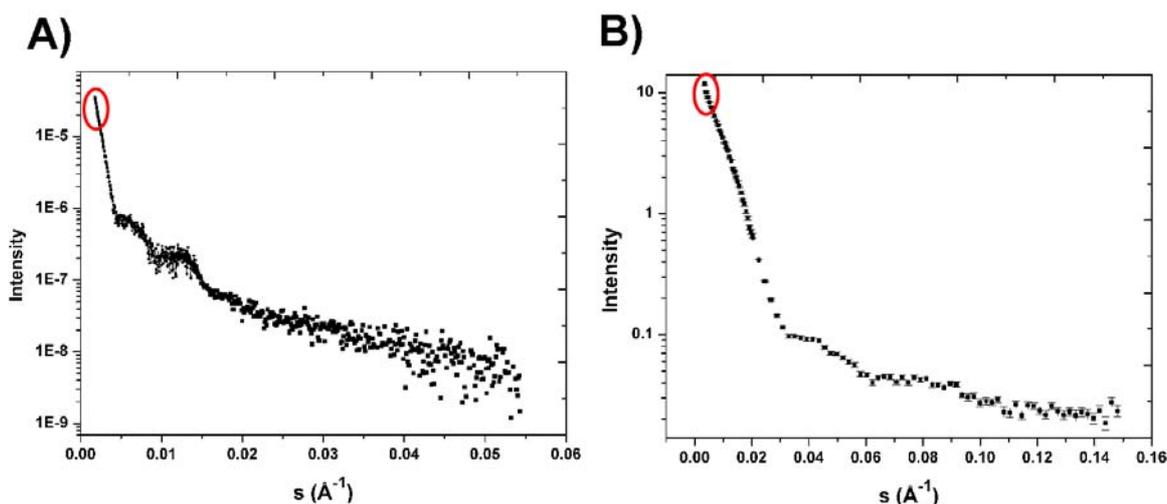


**Figure 5.22 Negative stain image of rE2**

This negative stain image of rE2 reflects icosahedral core structures with empty pentagonal faces along the 5-fold (arrow), 3-fold (square) and 2-fold (circle) axes of symmetry. Scale bar of 250 Å is denoted.

### 5.3.3.2 Solution structure of rE2 core

SAXS and SANS analysis of rE2 was problematic, owing to the formation of aggregate species as can be seen from the steep slope of the scattering curves at very low angles (Fig. 5.23). During the SAXS experiments, individual scattering frames were checked for x-ray induced aggregation, but no such process was observed. Therefore, it seems likely that aggregates had formed prior to the scattering experiments, perhaps as a consequence of high protein concentration. The data could not be fitted satisfactorily using GNOM (Semenyuk and Svergun, 1991; Svergun, 1992) to yield sensible  $p(r)$  distribution curves, especially at low angles, thus further indicating the polydisperse nature of the samples. Using Guinier analysis, the radius of gyration,  $R_g$ , was determined to be  $218 \pm 5$  Å and



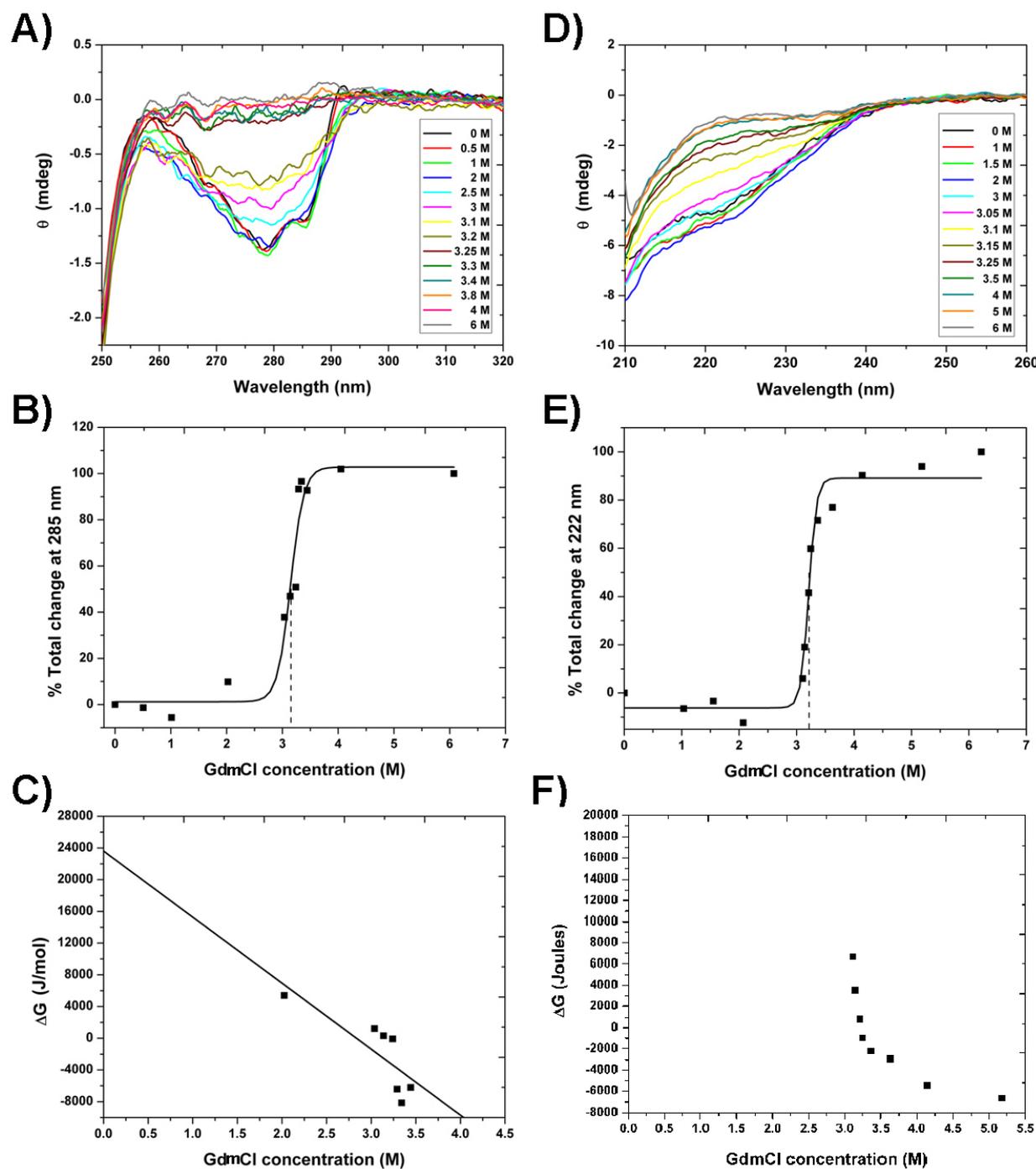
**Figure 5.23 Small angle scattering curves of rE2**

Scattering curves for rE2 obtained by (A) SAXS and (B) SANS show steep slopes at low angles (red circle), a usual sign of aggregates. The x-axis  $s$  denotes the scattering angle.

$222 \pm 6$  Å for the SAXS sample concentrations of 0.6 mg/ml and 1.2 mg/ml respectively, while the SANS data from the sample at 6.84 mg/ml yielded an  $R_g$  of  $235 \pm 8$  Å. In comparison, rE2/E3BP that has a similar structure to rE2 has an  $R_g$  of only 143 Å. Owing to these problems *ab initio* modelling of rE2 was discontinued at this stage.

### 5.3.3.3 Stability of the rE2 core

The stability of rE2 core was assessed by chemical denaturation with GdmCl. Unfolding/dissociation of rE2 observed by near- and far-UV CD occurred with increasing concentration of GdmCl at 25°C with complete loss of structure observed by 6 M GdmCl (Figs 5.24A and 5.24D). The extent of unfolding or dissociation of rE2 was monitored at wavelengths of 285 nm and 222 nm (Figs 5.24B and 5.24E). Negative changes in ellipticity are observed between 0-1 M GdmCl in the near-UV and 0-2 M GdmCl in the far-UV CD, respectively. This is an indicator of the loosening of protein structure as a result of solvent interaction. Between concentrations of 2-3 M GdmCl, some loss of structure was detected followed by gradual unfolding with increasing concentrations of GdmCl. This unfolding trend is consistent with previous refolding studies carried out on the bovine bE2/E3BP core (McCartney, et al., 1997). The initial structural change (2-3 M GdmCl) may be attributed to the overall dissociation of the 60 meric rE2



**Figure 5.24** CD analysis of GdmCl-induced denaturation of rE2

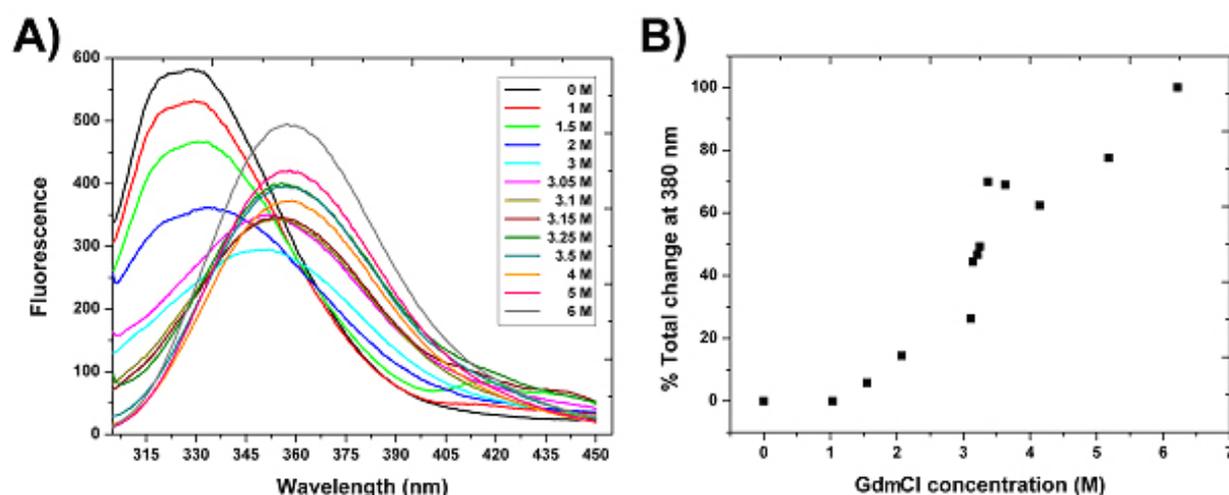
The stability of rE2 was assessed by near- (A) and far-UV (D) CD in the presence of increasing amounts of GdmCl. The rate of unfolding is represented by means of change in ellipticity monitored at wavelengths 285 nm (near-UV) (B) and 222 nm (far-UV) (E). The Gibbs free energy ( $\Delta G$ ) plots are shown for the unfolding of rE2 studied using near- (C) and far-UV (F) CD.

core into trimers, which are then further disassembled to monomers before complete unfolding ( $> 3$  M GdmCl). Hence, both the near- and far-UV CD indicate

3 M GdmCl) may be attributed to the overall dissociation of the 60 meric rE2 core into trimers, which are then further disassembled to monomers before complete unfolding ( $> 3$  M GdmCl). Hence, both the near- and far-UV CD indicate that the unfolding of rE2 does not follow a simple two state (native state  $\rightarrow$  unfolded state) mechanism.

The midpoint of unfolding for rE2 is 3.19 M and 3.21 M GdmCl for the near- and far-UV CD, respectively (Figs 5.24B and 5.24D). The free energy,  $\Delta G$  plots are shown in Figs 5.24C and 5.24F and the data indicate that the two-state model is not strictly applicable and that there may be at least one intermediate state in the unfolding process. Any estimates of  $\Delta G$ , such as those presented on the basis of a two-state analysis, can only be regarded as indicative. A more detailed analysis requires additional data points on the unfolding curve and the use of more elaborate models involving intermediate states.  $\Delta G$  values of 23.6 kJ/mol and 17.1 kJ/mol were obtained from the near- and far-UV CD of rE2.

Fluorescence spectra of rE2 were also recorded in the presence of GdmCl. A gradual shift of the maximum emission peak towards higher wavelengths is readily observed with increasing concentrations of GdmCl (Fig. 5.25A).



**Figure 5.25** GdmCl-induced change in the fluorescence emission spectrum of rE2 **(A)** Fluorescence emission of tryptophans (excited at 295 nm) monitored at wavelengths 310–450 nm in the presence of increasing amounts of GdmCl. **(B)** The percentage change of unfolding monitored at 380 nm indicates 50% of rE2 core unfolded by 3.1 M GdmCl.

A total of 4 tryptophans are present in monomeric rE2; 1 in the outer LD, 1 in the inner LD and 2 in the CTD. At 0 M GdmCl, the maximum peak of emission is seen to be at 331 nm indicating the semi-buried nature of the tryptophans. At 3 M GdmCl, a significant peak shift to 351 nm is observed (Fig. 5.25A). Interestingly, this dramatic shift occurs at a lower concentration of GdmCl (2.5 M) in rE2/E3BP as previously observed (section 5.3.1.3). With further increase in GdmCl concentration, the emission peak is gradually red shifted to a maximum of 360 nm at 6 M GdmCl, indicating solvent exposure of all tryptophans as a result of complete unfolding/dissociation of the rE2 core. The extent of unfolding monitored at 380 nm, may reflect a gradual trend of dissociation into potential trimeric intermediates with local regions of structural perturbations up to 3 M GdmCl, followed by rapid denaturation towards the completely unfolded state (Fig. 5.25B). This trend of GdmCl-induced rE2 unfolding by fluorescence correlates well with that observed in the near-UV CD data (Fig. 5.24B).

## 5.4 Discussion

The results presented in this chapter strongly support the ‘substitution model’ of subunit organisation of human PDC core proposed by Hiromasa and co-workers (2004) in which the subunit stoichiometry is 48E2+12E3BP. However, from our current knowledge of E2/E3BP core assembly, it is difficult to envisage why cores of uniform subunit composition (48E2+12E3BP) should be formed in which E3BP is distributed symmetrically around the core surface, and the authors fail to clearly explain this aspect of their model. The 60-meric cores are assembled from basic trimeric units that serve as building blocks to form the 20 vertices of the dodecahedron (Izard, et al., 1999). On this basis the 48E2+12E3BP model would imply core formation from a mixture of 12 heterotrimers (trimers with 2E2+1E3BP) and 8 homotrimers (trimers with 3E2). However, previous densitometry studies on eukaryotic E2/E3BP core, suggested the binding of 10-15 E3s to the core at maximal occupancy (Maeng, et al., 1996; Sanderson, et al., 1996b). Moreover, E3BP is proposed to exist as weakly interacting dimers on the surface of the core (Hiromasa, et al., 2004). On this basis the number of E3BPs bound in the E2 core can theoretically range from 0 (an E2-only core) to a maximum of 20 (maximum number of E3BPs included in the core), consequently resulting in a range of 0-20 heterotrimers that assemble to form a 60 meric core.

Data obtained from our AUC studies point towards a possible inherent variation in the subunit composition of rE2/E3BP, reflected in the broad peaks obtained from SV and radiolabelling data (Chapter 7, section 7.3.2.1.2). This aspect of broad variation in subunit distribution of the core is further supported by quasielastic light scattering (QELS) studies on bE2/E3BP core (Roche, et al., 1993).

Evidence confirming the substitution model of core organisation is obtained from SE studies by virtue of the molecular mass of rE2/E3BP being lower than rE2; substitution of E3BP into 60-meric E2 core entails reduction of the overall molecular weight because E3BP has a greater mass (50 KDa) compared with E2 (61 KDa). Both rE2 and rE2/E3BP cores have high frictional ratios ( $f/f_0$ ) resulting in structures having large hydrodynamic radii,  $R_s$ . The contribution to the large hydrodynamic size of these complexes is largely due to the solvent exposed elongated outer linker and LDs of E2, and the presence of alanine and proline residues in the linker regions (Roche, et al., 1993) that produces significant drag. Interestingly, it is observed that the  $f/f_0$  and  $R_s$  values obtained for rE2 in our study are higher than for rE2/E3BP; in direct contrast with previously observed data from AUC studies on these cores (Hiromasa, et al., 2004). Furthermore, the effects due to non-ideality, characterised by the large excluded volume, net charge and high degree of solvation are higher for rE2 than for rE2/E3BP from SE studies, in contrast with data obtained by Hiromasa and colleagues (2004). As these core structures are massive and highly solvated, it is likely that they would have high excluded volumes, indicating molecular crowding in the samples. The non-ideality exhibited by the 60-meric rE2 is indicative of a large excluded volume that may be a consequence of the heavy electrostatic repulsion between the highly anionic LDs of rE2 on the surface of the core. On the other hand, introduction of E3BP into a 60-meric E2 core (via substitution of E2 subunits) will result in these heavily charged E2 LDs being replaced by E3BP LDs, possibly giving rise to a significant change in the overall electrostatic nature on the core surface, and accordingly the non-ideality of the rE2/E3BP core. This also has major implications for the higher frictional coefficient of rE2 than rE2/E3BP. The highly anionic nature of the E2 LDs, the large number of amino acids (>120) in the flexible linker regions of E2 that connect the various domains and the high content of alanine and proline

residues in these linkers are thought to be the major factors contributing towards the large hydrodynamic radius, and as a result indirectly to the high frictional ratio of the rE2 core (Roche, et al., 1993) .

Data reported by Hiromasa and colleagues (2004) involves analysis of SE data using the Beckman XL-A/XL-I software implemented in MicroCal ORIGIN. Their SV data show the presence of slight aggregates and this may have profound effects if not fitted satisfactorily. Although their data suggest high non-ideality for the human cores (E2/E3BP being specifically greater than E2) and use of the first virial coefficient,  $B_2$  in their data analysis, they fail to mention how exactly  $B_2$  was implemented during data fitting and how consistent it was with its experimentally determined counterpart. In addition, the authors have performed their SE experiments at rotor speeds below 3,000 rpm, at which the AUC rotor is known to precess slightly.

QELS and cryo-EM studies on the bovine heart and kidney PDC established the presence of the icosahedral structural framework possessing pentagonal faces and a maximum dimension of 401 Å for the E2/E3BP cores (Roche, et al., 1993; Wagenknecht, et al., 1991). However, owing to the high flexibility of the outer N-terminal domains of the E2/E3BP cores, it was extremely difficult to resolve them in the EM micrographs. Additionally, only the value of the maximum dimensions of bovine heart PDC and its associated E2/E3BP core were obtained from the QELS studies. The structural models presented in this chapter constitute the first solution structures for full-length rE2/E3BP and bE2/E3BP cores obtained by SAXS. Both rE2/E3BP and bE2/E3BP show excellent structural similarity between each other, and with previously published data (Wagenknecht, et al., 1991). Both SAXS structures (rE2/E3BP and bE2/E3BP) reveal an inner core with empty pentagonal faces and elongated arms emanating from the icosahedral core, lending further support to the 'substitution model' of subunit organisation. This is in direct contrast to the structure of the yeast E2/E3BP core, wherein 12 E3BP are located within the 12 pentagonal openings of the icosahedral E2-only core, giving rise to the 'addition model' of 60E2+12E3BP (Stoops, et al., 1997). Moreover, the dodecahedral morphology with empty pentagonal faces was further confirmed by negative stain EM of bovine and recombinant E2/E3BP cores.

Only 13-16% of the exterior volume (volume outside the E2/E3BP core volume, occupied by E3, E1, kinases, phosphatases and outer domains of E2 and E3BP) is occupied by the peripherally located constituents, while greater than 80% of the total volume (exterior + core volume) is actually occupied by the solvent (Wagenknecht, et al., 1991). Hence, being characteristic of large solvent filled near-spherical like particles, these cores have high frictional coefficients and large hydrodynamic radii. A direct consequence of this is the comparatively low value of radius of gyration,  $R_g$  as observed for these cores, indicative of the outer parts of these structures being lightly populated. The  $R_g$  values of SAXS structures of rE2/E3BP (144-148 Å) and bE2/E3BP (155 Å) compare favourably with previously published values of 150 Å and 153 Å for recombinant E2/E3BP and E2 cores (Hiromasa, et al., 2004). However, the maximum dimensions of rE2/E3BP (462-472 Å) and bE2/E3BP (480 Å) in solution are somewhat larger than previously reported values of 401 Å and 420 Å for the bovine and recombinant E2/E3BP cores, respectively (Hiromasa, et al., 2004; Roche, et al., 1993). However, these values are well within the 20% size variation observed for mammalian PDC cores (Borgnia, et al., 2004; Kong, et al., 2003; Zhou, et al., 2001a). These differences in size have been linked to conformational changes in the trimeric units that are connected to each other via hydrophobic 'ball and socket' points, giving rise to thermally driven expansions and contractions, i.e. 'breathing' of the core (Borgnia, et al., 2004; Kong, et al., 2003; Zhou, et al., 2001a). Variation in core size, flexibility and subunit composition of rE2/E3BP has significant implications for the catalytic mechanism and regulation of the human PDC complex. As SAXS data for E2/E3BP yield only a time-averaged structure and are unable to distinguish between the E2 and E3BP subunits, more experiments are necessary to elucidate the extent of variation in the number and positions of E3BP bound within the core. Cryo-EM and cryo electron tomography of E2/E3BP core labelled with E3 or specific antibody that binds to E3BP should be able to provide more definitive insight towards understanding this problem of variation and inherent protein dynamics. Additionally, comparative studies to evaluate differences in the distribution of 'breathing sizes' and conformational flexibility observed between the rE2 and rE2/E3BP cores will shed more light on into their regulation, in particular of the rE2 core that is observed in patients lacking E3BP, while still retaining a residual capacity to bind E3 (Marsac, et al., 1993).

The dodecahedral core is made up of 20 trimers that act as building blocks. Based on the 'substitution model' of subunit organisation, human E2/E3BP core would comprise a mixture of homotrimers (3E2s) and heterotrimers (2E2+1E3BP) depending on the number of E3BP subunits that are integrated into the core. Alternatively, the 60-meric E2 core will consist of only 20 homotrimers. In contrast to the intensive interdigitated associations among the monomers within each trimer, the contact between adjacent trimers is very limited (Izard, et al., 1999). Therefore, the inter-trimer contacts are comparatively weaker than the intra-trimer contacts. The addition of E3BP within a trimer is thought to further weaken the interaction between neighbouring trimers, as E3BP does not possess the key residues involved in the hydrophobic 'ball and socket' connection that holds neighbouring trimers together (Hiromasa, et al., 2004). Data presented in this chapter on the assessment of core stability via unfolding/dissociation studies monitored by CD and fluorescence are in agreement with the above hypothesis. Treatment of rE2/E3BP with denaturant revealed lower amounts of GdmCl was needed to induce destabilisation by way of unfolding/dissociation, as compared to the rE2 core. A similar trend was observed with the GdmCl-induced fluorescence of these cores, and can be attributed to the relatively weaker inter-trimer contacts present in rE2/E3BP as compared to rE2. Initial changes in rE2/E3BP and rE2 core fluorescence (at low concentrations of GdmCl) prior to large-scale disruption of secondary structure reflect the formation of potential trimers and dissociation of local regions of quaternary structure. Thus, these data strongly suggest that the introduction of E3BP into the E2 core has a natural tendency to destabilise the core. However, further experiments are necessary to substantiate this hypothesis. In addition, it will be of great interest to determine the extent and involvement of specific domains of E2 and E3BP in this core destabilisation.

## Chapter 6

# Characterisation of the truncated C-terminal

## PDC core: tE2/tE3BP

### 6.1 Introduction

PDCs are amongst the largest ( $M_r \sim 4-9 \times 10^6$ ) and most complicated multimeric assemblies described to date and serve as a paradigm for analysis of protein-protein interactions, substrate channelling and active site coupling. A central feature of these complexes is the 24-meric (Gram negative organisms) or 60-meric (eukaryotes and some Gram positive organisms) E2 core with morphologies of a cube or pentagonal dodecahedron, respectively (Guest, et al., 1989; Patel and Roche, 1990; Reed and Oliver, 1968; 1982; Wagenknecht, et al., 1991). The E2 core has a structural and functional role in the overall organisation of the complex. In yeast and mammalian PDC an additional component, E3BP is present, and is also involved in core formation (De Marcucci and Lindsay, 1985; Jilka, et al., 1986; Rahmatullah, et al., 1989b). Both E2 and E3BP are multidomain polypeptides with high sequence similarity comprising LDs, SBD and an inner CTD that associates to form the icosahedral core.

Apart from their exceptional size, these complexes have some of the most unusual features found in biology. E2 serves as the catalytic unit of the core involved in the acetyltransferase activity, and is one of two oligomeric enzymes known to be organised as a pentagonal dodecahedron. Additionally, the inner CTD dodecahedral core has an unusual feature: the trimers at each of its 20 vertices are interconnected by 30 flexible bridges that enable the core to 'breathe' as evidenced by the 20% size variability (Zhou, et al., 2001a). This breathing property is thought to be a common feature of PDC cores, suggesting the importance of protein dynamics for optimal functioning of these massive complexes. Moreover, the dodecahedral morphology favours a synchronous change in the length of the bridges that are responsible for the observed size

variation (Zhou, et al., 2001a). In addition to its catalytic role, E2/E3BP also provides the structural scaffold to which its cognate partner enzymes (E1 and E3) are tethered at specific locations.

The amino acid sequence of of human E2 CTD is 50% identical to that of human E3BP CTD and 35% identical to that of *B. stearrowthermophilus* E2 CTD (Harris, et al., 1997). Each of the E2 trimers forming the dodecahedron has three active site channels located at the inter-subunit interfaces and serves as a conduit for the passage of acetyl groups (Mattevi, et al., 1992a). The CTD domain of E3BP associates with the CTD of E2 that forms the dodecahedron and this integration is reported to vary between yeast and mammalian PDC (Hiromasa, et al., 2004; Stoops, et al., 1992; Stoops, et al., 1997). This variation has been attributed to the differences between their E3BP sequences in comparison with their cognate E2s. Several structures of the individual domains of E2-PDC have been solved to high resolution by x-ray crystallography or NMR spectroscopy. Most notably of interest to this chapter, are the structures of the truncated CTD E2 (tE2) of *A. vinelandii* (Mattevi, et al., 1992a) and *B. stearrowthermophilus* (Izard, et al., 1999) that have been determined by x-ray crystallography. A recent cryo-EM reconstruction of truncated human E2 has also been determined by Yu and co-workers (2008). These structural data indicate significant evolutionary diversity imparting unique regulatory properties despite a common overall morphology and conserved enzymatic activities.

Eukaryotic PDCs vary from their bacterial counterparts in many respects, two key aspects being the integration of E3BP into the inner core framework and the complex regulatory mechanisms associated with this massive assembly. However, there is no structure available for the human E2/E3BP complex to date. This is probably related to its massive size, significant flexibility of the N-terminal flanking arms of E2 and E3BP and the subunit heterogeneity. Furthermore, with only limited structural data currently available on the truncated bacterial and human E2 cores, a detailed understanding of the architecture and functional properties of the human E2/E3BP core remains unclear.

The work described in this chapter aims to address this problem by gaining greater insights into the organisation of the human truncated CTD E2/E3BP core.

Truncated E2/E3BP (tE2/tE3BP) core was characterised with AUC, SANS, CD and EM in order to extend our knowledge of its overall architecture and key topographical features. Attempts were also made to crystallise the truncated E2/E3BP core to enable atomic resolution structure determination. Where possible, tE2/E3BP was compared with full-length rE2/E3BP and rE2 cores.

## 6.2 Materials and methods

### 6.2.1 Protein purification

Truncated E2/E3BP 'tLi19/tLi30' (i.e. core composed of tLi19 and tLi30 constructs, that comprise 19 aa and 30 aa, respectively of the linker region preceding the CTD) was purified as outlined in section 4.3.2.4. Gel filtered tLi19/tLi30 was then buffer exchanged into PEBS100 buffer (see section 3.3.2.4) and immediately used for biophysical studies. All protein concentration measurements for tLi19/tLi30 were based on a 48E2+12E3BP subunit stoichiometry, unless stated otherwise.

### 6.2.2 Sedimentation velocity

SV experiments were performed as previously described in section 3.4.2. Sedimentation data for tLi19/tLi30 in PEBS100 buffer were recorded at 4°C and a rotor speed of 20,000 rpm using interference optics. Samples (360 µl) were loaded into 12 mm double sector centrepieces, and a total of 471 scans, 1 min apart were recorded. Sample concentrations covered a range from 116 nM to 730 nM. Sedimentation profiles were analysed using SEDFIT (Schuck, 2000) (refer to sections 2.2.1.3 and 3.4.2 for details). The partial specific volume  $\bar{v}$  of tLi19/tLi30 calculated by SEDNTERP (<http://www.jphilo.mailway.com/>) (Laue, et al., 1992) based on the amino acid composition was 0.746 ml/g at 20°C. After initial  $c(s)$  analysis, finite element analysis in SEDFIT was used to determine the sedimentation coefficients of all samples. These were extrapolated to infinite dilution to give a sedimentation coefficient independent of concentration,  $s_{20,w}^0$ .

### 6.2.3 Sedimentation equilibrium

SE data were collected for tLi19/tLi30 using interference optics. All samples used for the SE experiments were analysed within 24 h of purification. Samples (80  $\mu$ l) were loaded into 12 mm double sector centrepieces and PEBS100 was used as the reference buffer. Further details can be found in section 3.4.3. SE runs for tLi19/tLi30 were carried out at 4°C and a rotor speed of 3,000 rpm. After an initial delay of 18-20 h to allow the system to approach equilibrium, 15 scans were recorded 3 h apart for samples ranging in concentration from 251 nM to 726 nM. Attainment of equilibrium was ascertained with WinMATCH ([www.biotech.uconn.edu/auf](http://www.biotech.uconn.edu/auf)). SE data were analysed using single fit analysis with SEDPHAT (Vistica, et al., 2004) as described in sections 2.2.2.2 and 3.4.3. In order to determine the molecular weight of tLi19/tLi30 independent of concentration,  $M_w^0$ , the whole cell weight average molecular weight  $M_{app}$ , determined at each sample concentration, was extrapolated to infinite dilution via the  $1/M_{app}$  vs concentration plot.

### 6.2.4 SANS and *ab initio* modelling

SANS data for tLi19/tLi30 were collected on beamline D22 of the Institut Laue-Langevin (ILL) in Grenoble, France. Scattering curves were obtained at 4°C for protein concentrations of 0.27, 0.59 and 1.19 mg/ml. Two camera lengths of 4 and 10 m were used. More details can be found in section 3.4.5.

*Ab initio* reconstructions of tLi19/tLi30 were generated using GASBOR (Svergun, et al., 2001) in batch mode via the computer grid system, SCOTGRID. Icosahedral (PICO) symmetry was applied with no constraint on the peripheral penalty during the modelling process. More details on modelling can be found in section 3.5.2. The *ab initio* models of tLi19/tLi30 were averaged using DAMAVER (Volkov and Svergun, 2003) to obtain a consensus average structure.

### 6.2.5 Homology and hydrodynamic modelling

A homology model of human truncated E2 (tE2) was obtained as described in section 3.5.3. The human tE2 homology model was generated with the *B.*

*stearothermophilus* tE2 crystal structure (PDB ID 1B5S) (Izard, et al., 1999) as the template, and superimposed onto the *ab initio* GASBOR model of tLi19/tLi30 using SUPCOMB (Kozin and Svergun, 2001).

The *ab initio* model of tLi19/tLi30 was converted into a hydrodynamic bead model (HBM) comprising 1986 beads using a 5.7 Å cubic grid in the AtoB module (Byron, 1997) within SOMO (Rai, et al., 2005). Hydrodynamic parameters for the HBM were then calculated using HYDRO++ (García de la Torre, et al., 2007; García de la Torre, et al., 1994). The radii of beads in the HBM used for calculations with HYDRO++ were adjusted to give the volume of tLi19/tLi30 obtained from theoretical calculations. Additionally, an HBM was generated using TRANS2VORONOI (developed by M. Nöllmann, Centre de Biochimie Structurale Montpellier, France) with a 7 Å cubic grid, resulting in a total of 1426 beads. After identification of surface exposed and buried beads using the ASAB1 module in SOMO, hydrodynamic computations for the HBM were independently performed using SUPCW (Rai, et al., 2005; Spotorno, et al., 1997). Anhydrous sedimentation coefficients obtained from HYDRO++ and SUPCW were converted to hydrated values via the conversion factor of 0.8666 derived from hydrodynamic theory (refer to 2.2.1.2.2 for details). The models obtained from AtoB and TRANS2VORONOI were visualised via Pymol (Delano Scientific, USA) and Rasmol (Sayle and Milner-White, 1995) (<http://openrasmol.org>). Further details on the modelling can be found in section 3.5.4.

### **6.2.6 Negative stain electron microscopy**

Samples for imaging were prepared as described in detail in section 3.4.6. The negative stain micrograph of tLi19/tLi30 was recorded at high magnifications of 90,000 and 120,000 at a sample concentration of 53.5 µg/ml.

### **6.2.7 CD and tryptophan fluorescence**

Near-UV (250-320 nm) CD was conducted on purified tLi19/tLi30 (0.9 mg/ml) in PEB buffer at 25°C in the presence of increasing concentrations of GdmCl as described in section 3.4.7. GdmCl concentrations were measured by refractrometry. The extent of unfolding or dissociation in the presence of

increasing denaturant was monitored at 285 nm by calculating the percentage total change in ellipticity. More details on the calculations can be found in section 3.4.7. Additionally, the Gibbs free energy at infinite dilution,  $\Delta G^0$ , of the unfolding reaction of tLi19/tLi30 at standard conditions of temperature and pressure was also computed by plotting the values of  $\Delta G$  as a function of GdmCl concentration (see section 3.4.7 for details).

Changes in tryptophan fluorescence for tLi19/tLi30 were monitored at 25°C in the presence of increasing levels of GdmCl. Fluorescence emissions were recorded over the spectral range 310-450 nm at a sample concentration of 0.2 mg/ml (refer to section 3.4.8 for technical details).

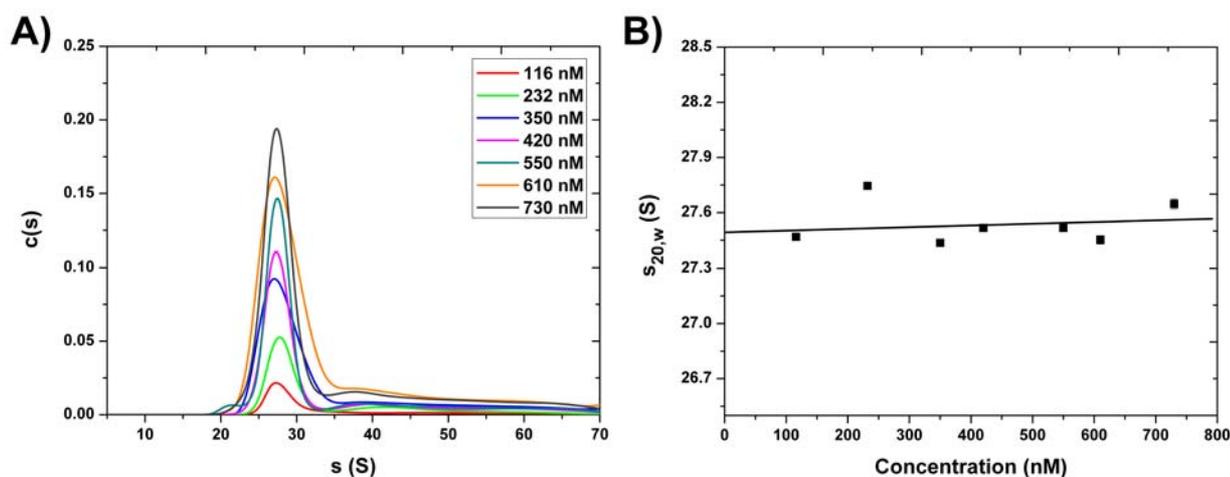
## 6.3 Results and modelling

### 6.3.1 Solution shape determination of tE2/tE3BP

#### 6.3.1.1 Sedimentation velocity analysis

SV experiments conducted using interference optics reveal that the purified tLi19/tLi30 sample is not entirely monodisperse (Fig. 6.1). Sedimentation profiles modelled by  $c(s)$  analysis in SEDFIT (Schuck, 2000) indicate a distinctive main peak with a weight average sedimentation coefficient of 27 S (Fig. 6.1A). Additionally, a long trailing edge is evident with a minor peak at about 38 S. While the major peak may correspond to intact 60-meric tLi19/tLi30, the long tail could be attributed to the presence of low amounts of tLi19/tLi30 dimers, a small percentage of incomplete cores or aggregates. However, all these were negligible when compared to the total amount of intact tLi19/tLi30 (>96% of total protein concentration). Finite element analysis was conducted to determine the sedimentation coefficients of tLi19/tLi30 at all sample concentrations. Each dataset was modeled via the non-interacting discrete species model based on the  $c(s)$  distribution. The resultant values of  $s_{20,w}$  were then extrapolated to infinite dilution to give  $s_{20,w}^0 = 27.5 \pm 0.31$  S (Fig. 6.1B). As the crystal structure of truncated E2 of *B. stearothermophilus* has been solved (Izard, et al., 1999), its sedimentation coefficient was calculated using the

program HYDRO++ (García de la Torre, et al., 2007; García de la Torre, et al., 1994). The atomic coordinates of the crystal structure of bacterial tE2 were initially converted to an HBM using AtoB (Byron, 1997), that in turn was employed in the calculation of the sedimentation coefficient using HYDRO++ (García de la Torre, et al., 2007; García de la Torre, et al., 1994). On applying the hydration factor (0.8666), a sedimentation coefficient of 29.9 S was obtained for bacterial tE2, higher than the experimentally determined value of 27.5 S for tLi19/tLi30.



**Figure 6.1 Sedimentation velocity analysis of tLi19/tLi30**

(A)  $c(s)$  distribution derived from SV data for various concentrations of tLi19/tLi30. The axes  $s$  and  $c(s)$  denote the sedimentation coefficient and the  $c(s)$  distribution model (from SEDFIT analysis), respectively. (B) Determination of  $s_{20,w}^0$  ( $= 27.5$  S) for the intact 60-meric tLi19/tLi30 core. Error bars are shown, but are not clearly visible owing to their small size.

This difference may indicate a more flexible, open conformation of tLi19/tLi30 in solution in comparison to the bacterial tE2 crystal structure, stemming from so-called crystal packing effects. More importantly, it may also reflect slight differences in the overall structures arising as a consequence of the integration of E3BP into the human E2 core, as opposed to the bacterial core composed solely of E2 subunits. Some of the hydrodynamic parameters calculated from the SV data for tLi19/tLi30 are listed in Table 6.1. The  $f/f_0$  of tLi19/tLi30 is 1.73, smaller than the values obtained for full-length rE2/E3BP and rE2 cores (sections 5.3.1.1.1, 5.3.3.1.1). Additionally, this  $f/f_0$  of 1.73 predicts a hydrodynamic radius,  $R_s$  of 13.7 nm. The  $D_s$  of 27.4 nm is slightly larger than previously reported values of the average diameter ranging from 21-24 nm (Junger and

Reinauer, 1971; Reed and Oliver, 1968; Wagenknecht, et al., 1991). However, this elevated value for the core diameter is well within the 20% variability attributed to the ‘breathing’ of the inner core (Zhou, et al., 2001a). At its N terminus, tLi19/tLi30 includes the 19 and 30 amino acid segments of the inner linker regions preceding the CTDs of E2 and E3BP, respectively. These extra linker regions are presumed to be located on the outside of the dodecahedron and may contribute slightly to the hydrodynamic diameter of tLi19/tLi30.

$M$ (Da)	$\bar{v}$ (ml/g)	$s_{20,w}^0$ (S)	$f/f_0$	$R_0$ (nm)	$R_s$ (nm)	$D_s$ (Å)	$D_t$ (cm <sup>2</sup> /s)
1,671,348	0.7458	27.5 <sup>a</sup> 47.5 <sup>b</sup>	1.73	7.91	13.7	274	$1.569 \times 10^{-7}$

**Table 6.1 Hydrodynamic parameters of tLi19/tLi30 derived from SV analysis**

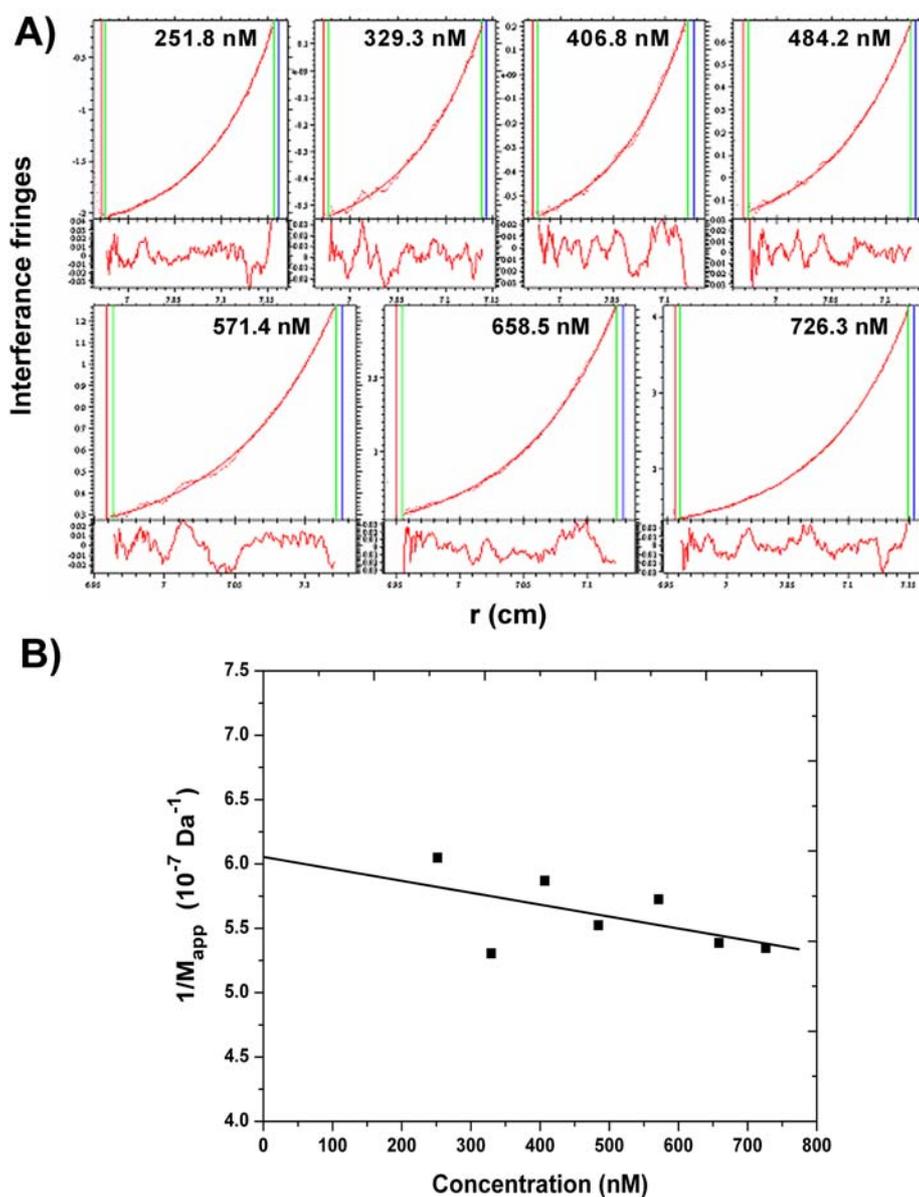
<sup>a</sup>experimental and <sup>b</sup>calculated sedimentation coefficients.  $M$  is the molecular mass of tLi19/tLi30 core obtained from the amino acid composition,  $\bar{v}$  is the calculated partial specific volume at 20°C,  $s_{20,w}^0$  is the sedimentation coefficient at infinite dilution and  $R_0$  is the unhydrated radius of a sphere of equivalent mass and partial specific volume as that of tLi19/tLi30.  $R_s$  and  $D_s$  are the hydrodynamic radius and its equivalent diameter while  $D_t$  is the translational diffusion coefficient.

Interestingly, the  $c(s)$  profile and sedimentation coefficient of tLi19/tLi30 (27.5 S, Fig. 6.1A) is comparable with SV data for full-length rE2/E3BP (29.3 S, section 5.3.1.1.1) and rE2 (29.3 S, section 5.3.3.1.1). Therefore, it can be inferred that the decrease in particle radius or elongation (and as a consequence the decrease in frictional drag) of tLi19/tLi30 is offset by the decrease in its molecular weight, thus causing it to sediment in a fashion similar (at the same speed) to the full-length rE2/E3BP and rE2 cores.

## 6.3.2 Molecular weight determination of tE2/tE3BP

### 6.3.2.1 Sedimentation equilibrium analysis

Analysis of the individual SE data sets using a two species model in SEDPHAT (Schuck, 2003; Vistica, et al., 2004) (Fig. 6.2A) yielded  $M_{app}$ . Extrapolation of  $1/M_{app}$  to infinite dilution for the predominant first species gave a  $M_w^0 = 1.65 \pm$



**Figure 6.2 Sedimentation equilibrium analysis of tLi19/tLi30**

(A) Fits (smooth lines) of the SE interference data (dotted lines) of tLi19/tLi30 at 3,000 rpm using the two species model in SEDPHAT. Residual plots of all samples are represented on different scales.  $M_{app}$  was determined at each sample concentration. (B) Extrapolation to zero of the  $1/M_{app}$  vs concentration plot for the main component yielded the molecular mass,  $M_w^0$  (1.65 MDa) independent of concentration.

0.03 MDa (Fig. 6.2B), in good agreement with the value predicted from the tLi19/tLi30 amino acid sequence (1.67 MDa). The observed decrease in  $1/M_{app}$  with increasing concentration is indicative of self-association or aggregate formation. Global analysis and fitting yielded a molecular weight of 2.29 MDa for the second minor species (data not shown), notably lower than that of possible tLi19/tLi30 dimer. In addition, fitting with the self-association model in

SEDPHAT did not yield satisfactory results, suggesting the formation of aggregates rather than the presence of higher order oligomers.

Attempts were also made to determine the exact molecular weight of tLi19/tLi30 using mass spectrometry (MS) in collaboration with Prof. Carol Robinson at the University of Cambridge. However heterogeneity of the core and limited proteolysis made this unfeasible.

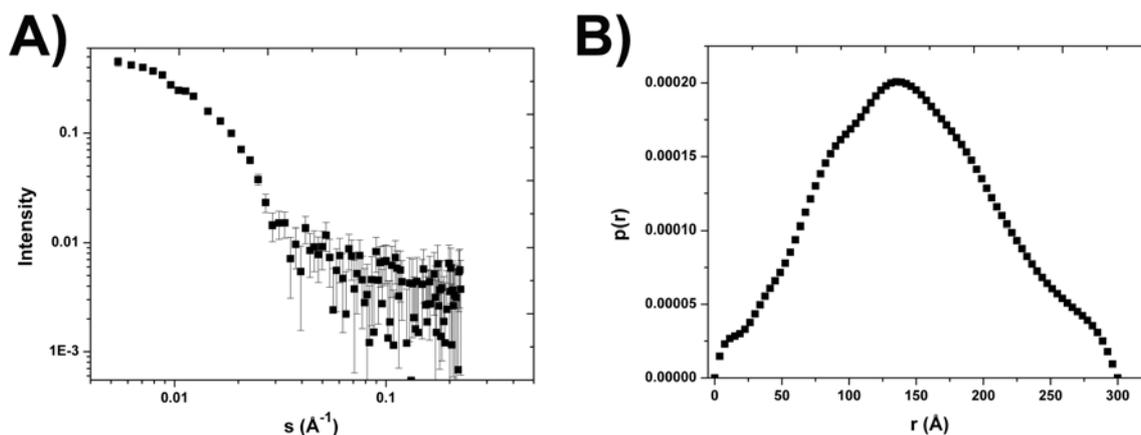
### **6.3.3 Solution structure and modelling of tE2/tE3BP**

#### **6.3.3.1 SANS and *ab initio* modelling**

SANS curves were acquired for purified tLi19/tLi30 at three protein concentrations to account for interparticle interference effects. However, protein aggregation was observed in the low angle region of the SANS curve for the highest sample concentration (1.19 mg/ml), while measurements at the lowest concentration measured (0.27 mg/ml) were too noisy to yield satisfactory estimates of molecular weight. However, the scattering intensity curve obtained at 0.59 mg/ml showed no evidence of aggregates (Fig. 6.3A), and hence was used as the data set for *ab initio* modelling.

The molecular weight estimate obtained for tLi19/tLi30 by extrapolation of scattering intensity to zero angle,  $I(0)$ , was  $1.81 \pm 0.15$  MDa, slightly higher than the mass calculated from protein sequence (1.67 MDa). This difference may be the result of experimental error. However, it may also reflect the minor presence of larger species or aggregates, not apparent from the low angle regions of the intensity scattering curve. The radius of gyration,  $R_g$ , determined using the Guinier approximation is  $107 \pm 2$  Å. The  $p(r)$  distribution function is shown in Fig. 6.3B. The  $D_{max}$  is 300 Å and  $R_g$  calculated from the  $p(r)$  function is  $111 \pm 2$  Å, which agrees well with the value obtained from the Guinier approximation.

The  $D_{max}$  (300 Å) obtained is much higher than the previously published value of 230 Å for bovine tE2/tE3BP (Wagenknecht, et al., 1991). As SANS data yield an average solution structure, it is likely that the experimental sample may comprise tLi19/tLi30 cores of variable diameters along with some aggregates.



**Figure 6.3 Small angle neutron scattering of the tLi19/tLi30 core**

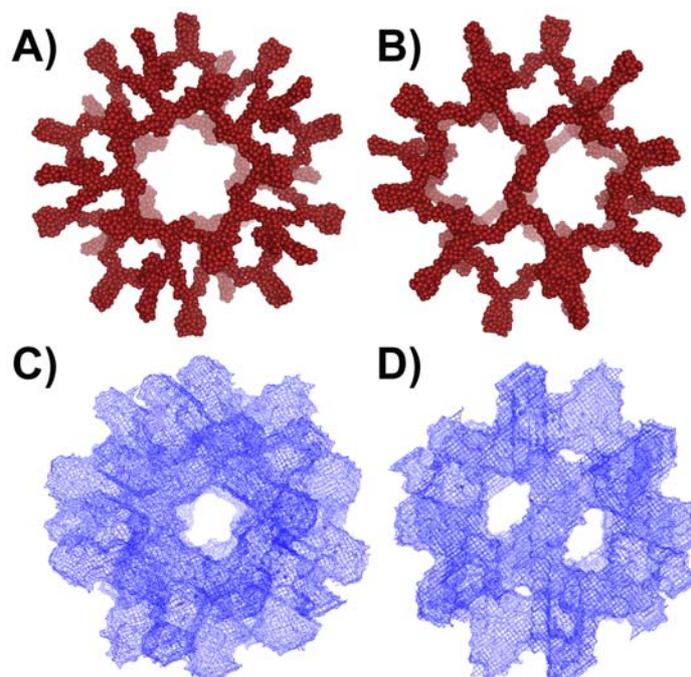
(A) Scattering curve recorded for tLi19/tLi30 at  $4^\circ\text{C}$ . The x-axis  $s$  denotes the scattering angle. (B) The distance distribution function  $p(r)$  was calculated using GNOM.

Hence the elevated  $D_{max}$  of 300  $\text{\AA}$  probably reflects the presence of these minor aggregate species.

*Ab initio* restoration using GASBOR (Petoukhov and Svergun, 2003; Svergun, et al., 2001) produced an icosahedral core with hollow internal cavities and empty pentagonal faces (Figs 6.4A and 6.4B). No density was observed within or outside the pentagonal faces, as would occur if the CTD of E3BP was added to the inner or outer surface of the core, as observed in yeast (Stoops, et al., 1997), thus strongly supporting the ‘substitution’ model of core organisation. Moreover, the solution structure of tLi19/tLi30 is consistent with previously obtained ‘inner core’ SAXS structures of full-length rE2/E3BP and bE2/E3BP as described in Chapter 5 of this study, as well as with the structure of bE2/E3BP obtained from cryo-EM studies (Wagenknecht, et al., 1991). Ten *ab initio* GASBOR models were averaged using DAMAVER (Volkov and Svergun, 2003) to obtain a consensus average structure of tLi19/tLi30 (Figs 6.4C and 6.4D).

### 6.3.3.2 Hydrodynamic modelling

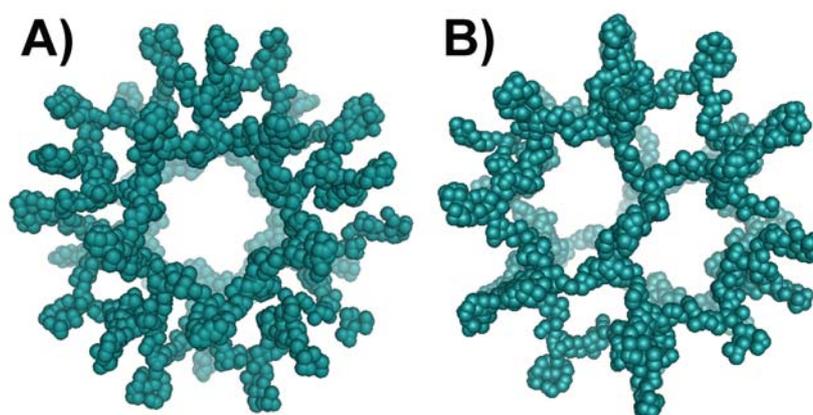
Hydrodynamic bead models (HBMs) of tLi19/tLi30 were generated (from the *ab initio* GASBOR model) using programs AtoB (Byron, 1997) and TRANS2VORONOI (M. Nöllmann, Centre de Biochimie Structurale Montpellier, France) as described in section 6.2.6. Assuming hydration of 0.4 g/g, hydrated sedimentation coefficients of 28.3 S and 28.2 S were calculated using HYDRO++ and SUPCW



**Figure 6.4 Solution structure of the tLi19/tLi30 core**

A single GASBOR *ab initio* model of tLi19/tLi30 is shown along the (A) 5-fold and (B) 2-fold axes of symmetry. Superimposition of 10 *ab initio* models yielded an average consensus structure (blue, wire mesh) shown along the (C) 2-fold and (D) 5-fold axes of symmetry for convenience.

(García de la Torre, et al., 2007; García de la Torre, et al., 1994; Rai, et al., 2005; Spotorno, et al., 1997), respectively, in excellent agreement with the experimental value of 27.5 S obtained from SV data. The HBM obtained from AtoB (Fig. 6.5) retains the hollow icosahedral core structure of the *ab initio* model reconstructions generated from the scattering data.

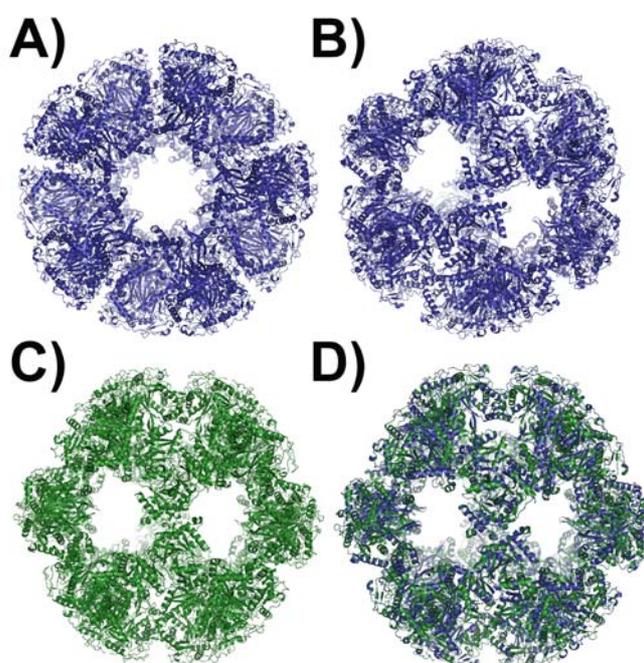


**Figure 6.5 AtoB generated model of tLi19/tLi30**

The model of tLi19/tLi30 generated from AtoB retains the hollow icosahedral structure of its parent *ab initio* GASBOR model. Views are shown along the (A) 5-fold and (B) 2-fold axes of symmetry for convenience.

### 6.3.3.3 Homology modelling

As no structural data are available for the human E2/E3BP core, homology modelling was carried out in addition to AUC and SAXS to gain more insight into the human tE2 (tLi19) structure. The homology model of tLi19 (Figs 6.6A and 6.6B) was generated using the crystal structure of bacterial truncated E2 (Izard, et al., 1999) as the template (Fig. 6.6C) as described in section 3.5.3. Unsurprisingly, superimposition of the homology model of tLi19 and crystal structure of bacterial tE2 indicated a high degree of similarity in the overall secondary, tertiary and quaternary structures as seen in Fig. 6.6D, although tLi19 has a slightly greater diameter.

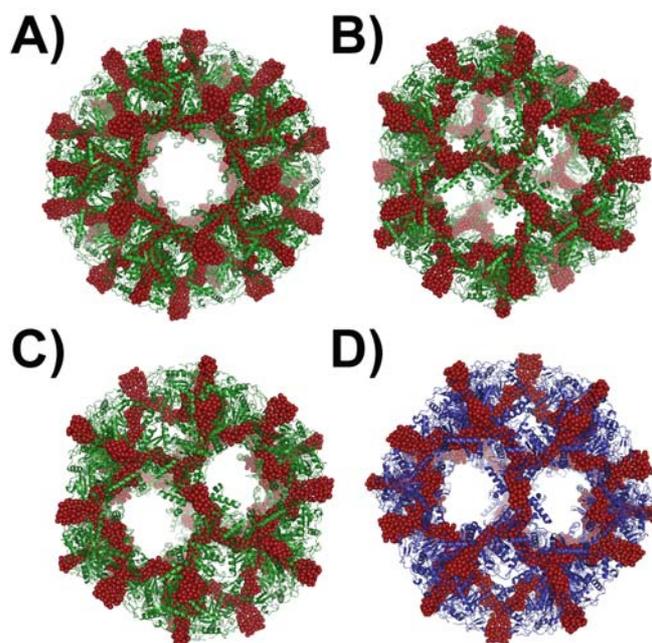


**Figure 6.6 Homology model of the tLi19/tLi30 core**

Homology model of tLi19/tLi30 (blue) shown along the **(A)** 5-fold and **(B)** 2-fold axes was generated via SWISS-MODEL (<http://swissmodel.expasy.org/>) (Arnold, et al., 2006) using the **(C)** crystal structure of *B. stearothermophilus* (green) truncated E2 as template. **(D)** Superimposition of the crystal structure of bacterial tE2 (green) and the homology model of tLi19/tLi30 (blue) is shown.

### 6.3.4 Conserved C-terminal domains: superimposition of x-ray crystal and SAXS solution cores

The *ab initio* model of tLi19/tLi30 obtained from the scattering data was superimposed with the crystal structure of bacterial tE2 (Izard, et al., 1999) using SUPCOMB (Kozin and Svergun, 2001) as shown in Fig. 6.7. The superimposition of the cores along the 5-fold, 3-fold and 2-fold axes of symmetry (Figs 6.7A, 6.7B, 6.7C) shows reasonable overlap of the solution and crystal structures of the truncated cores. In addition, the overlay of the homology model of tLi19 and the *ab initio* SANS model of tLi19/tLi30 indicate good conservation of the gross structural features of the icosahedral core (Fig. 6.7D).



**Figure 6.7 Superimposition of the crystal structure of tE2 and the *ab initio* model of tLi19/tLi30**

The superimposition of the crystal structure of bacterial tE2 (green) with the solution structure of tE2/tE3BP (red) obtained from SANS is shown along the (A) 5-fold, (B) 3-fold and (C) 2-fold axes of symmetry. (D) Reasonable overlay of the homology model of tLi19 (blue) and the *ab initio* model of tLi19/tLi30 (red) is also observed

Thus, the superimpositions indicate that despite variable regulatory functions, the overall icosahedral framework of the CTD cores of human E2/E3BP and bacterial E2 are highly conserved. As atomic structural details cannot be obtained from SANS solution structures, obtaining information pertaining to specific differences in the fine structures of these cores is not possible; in particular the structural variations arising from the integration of E3BP into the human E2 core. However, some structural differences have been reported between the human and bacterial tE2 cores, mainly confined to the hairpin

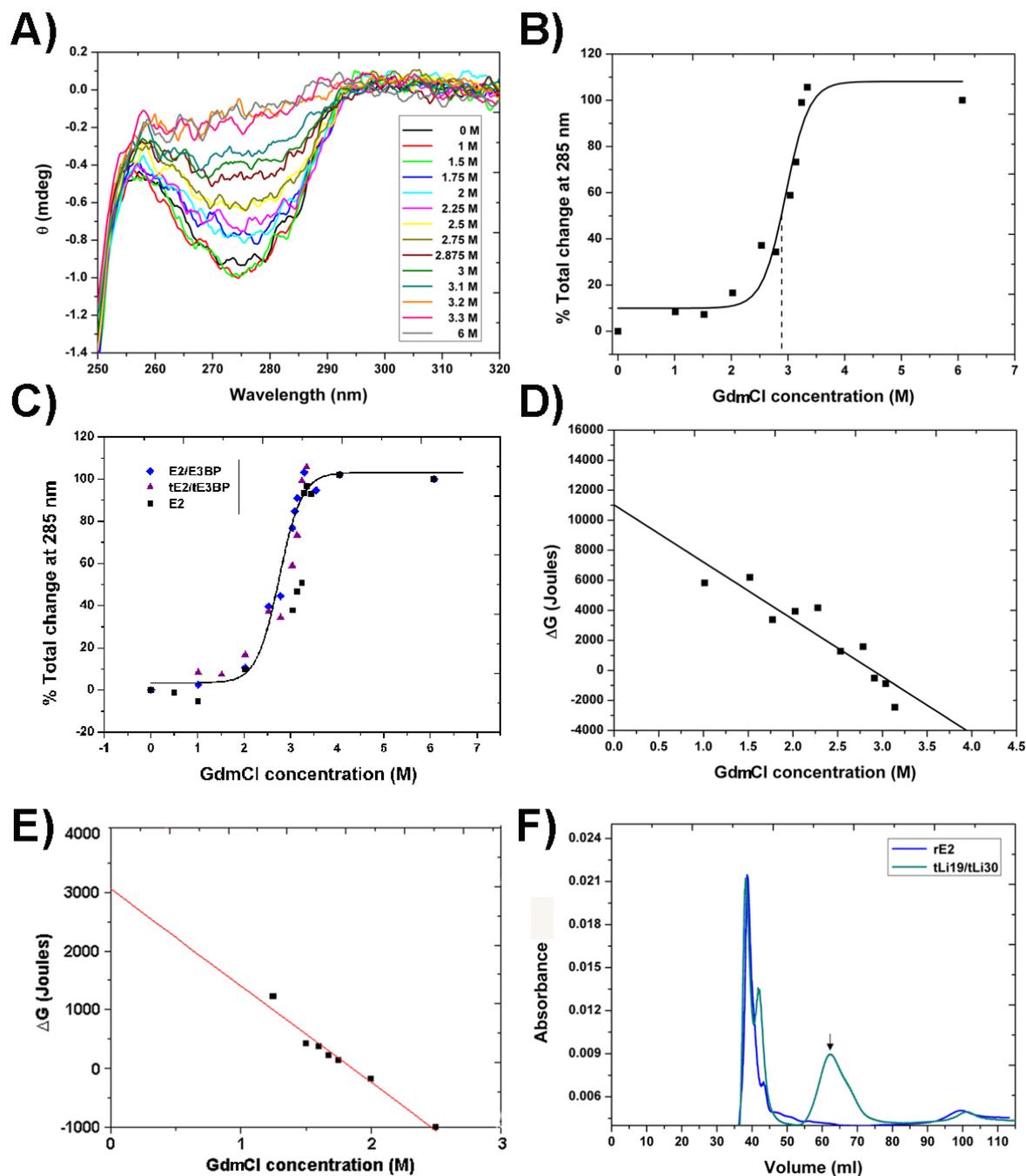
domain and the N-terminal helix connected to the linker region preceding the CTD (Yu, et al., 2008).

### **6.3.5 Stability of the tE2/tE3BP core**

The stability of tLi19/tLi30 was assessed by chemical denaturation with GdmCl observed with near-UV CD and tryptophan fluorescence. The tLi19/tLi30 core assembly was unfolded or dissociated with increasing concentrations of GdmCl, with complete loss of structure taking place by 6 M GdmCl as shown in Fig. 6.8A. Monitoring the total change in ellipticity at 285 nm indicates the midpoint of unfolding at 2.88 M GdmCl (Fig. 6.8B). At concentrations less than 3 M GdmCl a gradual unfolding/dissociation event is evident reflected in the progressive decrease in ellipticity, followed by an abrupt change beyond 3 M GdmCl (Fig. 6.8B). Therefore, the stability curve of tLi19/tLi30 denotes two-state unfolding possibly via some intermediates, consistent with data obtained for rE2/E3BP (section 5.3.1.3), as well as with previously published refolding studies on the bE2/E3BP core (McCartney, et al., 1997). The tLi19/tLi30 unfolding profile is comparable with that of rE2/E3BP and rE2 cores as illustrated in Fig. 6.8C.

Interestingly, the full-length rE2/E3BP core (section 5.3.1.3) is less stable than the tLi19/tLi30 core, indicating the possible contribution of the N-terminal flexible arms towards its destabilisation. The stability curves obtained from near-UV CD for the truncated and full-length cores indicate that rE2 is more stable than tLi19/tLi30, which in turn is more stable than rE2/E3BP (as reflected in their midpoint values of unfolding: 2.7 M (rE2/E3BP), 2.88 M (tLi19/tLi30) and 3.19 M GdmCl (E2), (Fig. 6.8C)).

The free energy,  $\Delta G$  versus GdmCl concentration plot for tLi19/tLi30 clearly indicates the two-state unfolding via some intermediates (Fig. 6.8D). This agrees well with the  $\Delta G$  versus GdmCl concentration plot obtained for full-length rE2/E3BP that shows a similar trend (section 5.3.1.3). However, it is in marked contrast to data for a truncated N-terminal rE2 construct, E2DD that comprises the inner LD and SBD (Fig. 6.8E). The free energy profile for E2DD is completely



**Figure 6.8 GdmCl-induced denaturation of tLi19/tLi30**

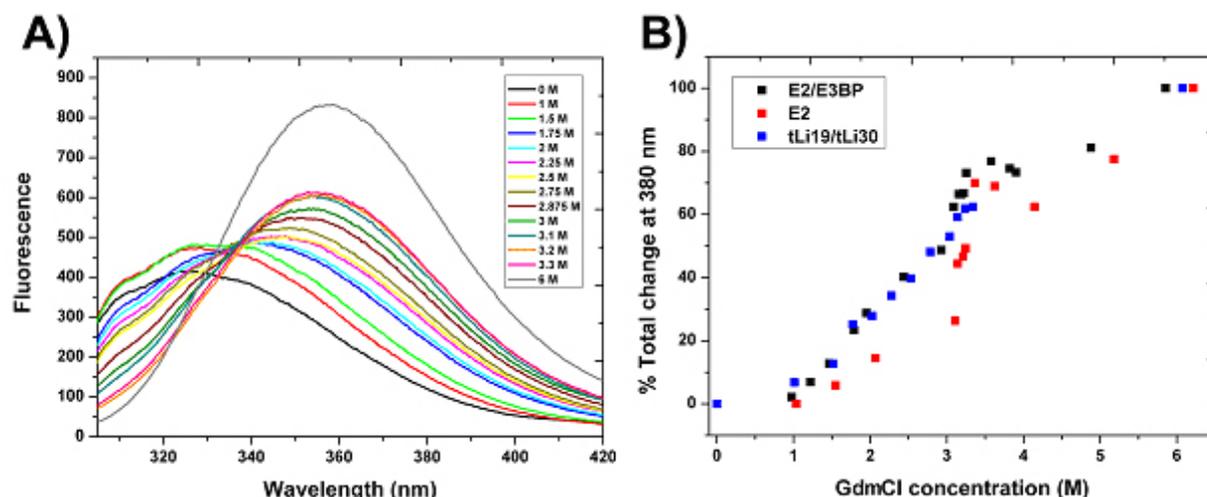
The stability of tLi19/tLi30 was assessed by **(A)** near-UV CD and **(B)** the extent of unfolding was monitored at 285 nm in the presence of increasing amounts of GdmCl. **(C)** Comparative unfolding curves of the full-length cores (rE2/E3BP and rE2) and tLi19/tLi30 assessed by near-UV CD are also shown. Gibbs free energy ( $\Delta G$ ) plots reflecting the two-state unfolding profile for **(D)** tLi19/tLi30 (via intermediates) and the **(E)** truncated construct, E2DD (cooperative), respectively are shown for comparison. **(F)** GFC of purified protein (tLi19/tLi30 and rE2) with 2.85 M GdmCl indicates differences in the elution profile. Elution peaks corresponding to void volume (\*) and putative trimers of tLi19/tLi30 (arrow) are shown.

linear, clearly indicating the cooperativity of unfolding (folded state ----> unfolded state) via no intermediates.  $\Delta G^0$  values for tLi19/tLi30 and E2DD were determined to be 11 kJ/mol and 3.1 kJ/mol, respectively.

In addition, the difference in stability between the E2 and E2/E3BP cores was confirmed by gel filtration of the purified cores in the presence of 2.85 M GdmCl, close to the midpoint of unfolding for tLi19/tLi30 (2.88 M), but relatively far from that for rE2 (3.19 M). Gel filtration profiles indicate void volume elution (40 ml) of the intact 60-meric cores of both tLi19/tLi30 and rE2 as seen in Fig. 6.8F. However, an additional peak at an elution volume of 62 ml is observed only with tLi19/tLi30. Previous refolding studies by McCartney and co-workers (1997) showed the possible formation of trimers and monomers during the unfolding of bE2/E3BP. From its elution volume,  $V_e$ , this additional peak of tLi19/tLi30 has an apparent molecular weight of 185 kDa, consistent with the presence of E2 homotrimers (183 kDa) and/or E2/E3BP heterotrimers (2E2+1E3BP, 174 kDa). It is interesting to note that at 2.85 M GdmCl, while the gel filtered tLi19/tLi30 reveals subunit dissociation into trimers and/or monomers, rE2 remains almost intact (Fig. 6.8F), consistent with the greater stability of the E2-oligomeric assembly.

GdmCl-induced changes in the fluorescence of tLi19/tLi30 display a similar trend to the fluorescence profile of full-length rE2/E3BP. The shift of the maximum emission peak towards higher wavelengths is readily observed with increasing concentrations of GdmCl (Fig. 6.9A). The tLi19/tLi30 core houses 3 tryptophans, 2 in E2 (tLi19) and 1 in E3BP (tLi30).

At 0 M GdmCl, the maximum peak of emission is observed at 328 nm, indicative of tryptophans buried in the core. At 1.75 M GdmCl, a significant peak shift to 344 nm is observed indicating changes in the tryptophan environment brought about by partial dissociation, resulting in limited exposure of the tryptophans to the solvent (Fig. 6.9A). These changes between 0-2 M GdmCl may correspond to the dissociation of the overall quaternary structure of the core into trimeric intermediates. Moreover, this dissociation is solely due to the effect of the inner CTD Trps as the Trps associated with the LDs and SBDs are absent. Further increases in GdmCl concentration results in the shift of the emission peak to 360



**Figure 6.9 GdmCl-induced fluorescence emission spectra of tLi19/tLi30**

(A) The tryptophan fluorescence emission of tLi19/tLi30 monitored at wavelengths 310-450 nm in the presence of GdmCl shows a prominent red shift. (B) The percentage change of unfolding monitored at 380 nm as a function of GdmCl concentration for the full-length (rE2/E3BP and rE2) and truncated (tLi19/tLi30) cores is shown for comparison.

nm by 6M GdmCl, suggesting complete unfolding of the tLi119/tLi30 core. The percentage change in fluorescence monitored at 380 nm indicates marked alterations in the three-dimensional structure from 1.75 M GdmCl onwards for tLi19/tLi30, slightly higher in comparison to the equivalent profiles for rE2/E3BP and rE2 (Fig. 6.9B).

It is also interesting to note that the intensity changes for the full-length (rE2/E3BP and rE2) cores differ from those of tLi19/tLi30 (see Figs 5.11 and 5.25). Full-length rE2/E3BP cores exhibit an initial decrease in the fluorescence intensity at low GdmCl concentrations followed by a gradual increase, and this may primarily be associated with the exposure of the N-terminal domain tryptophans of these cores to the solvent, and thus likely to contribute greatly to the observed red shift. In contrast, the steady increase in the fluorescence intensity of tLi19/tLi30 with increasing GdmCl concentrations is probably due to the gradual exposure of the CTD tryptophans that may be largely quenched in the folded state (Fig. 6.9A).

## **6.3.6 High-resolution structure of tE2/tE3BP: the goal**

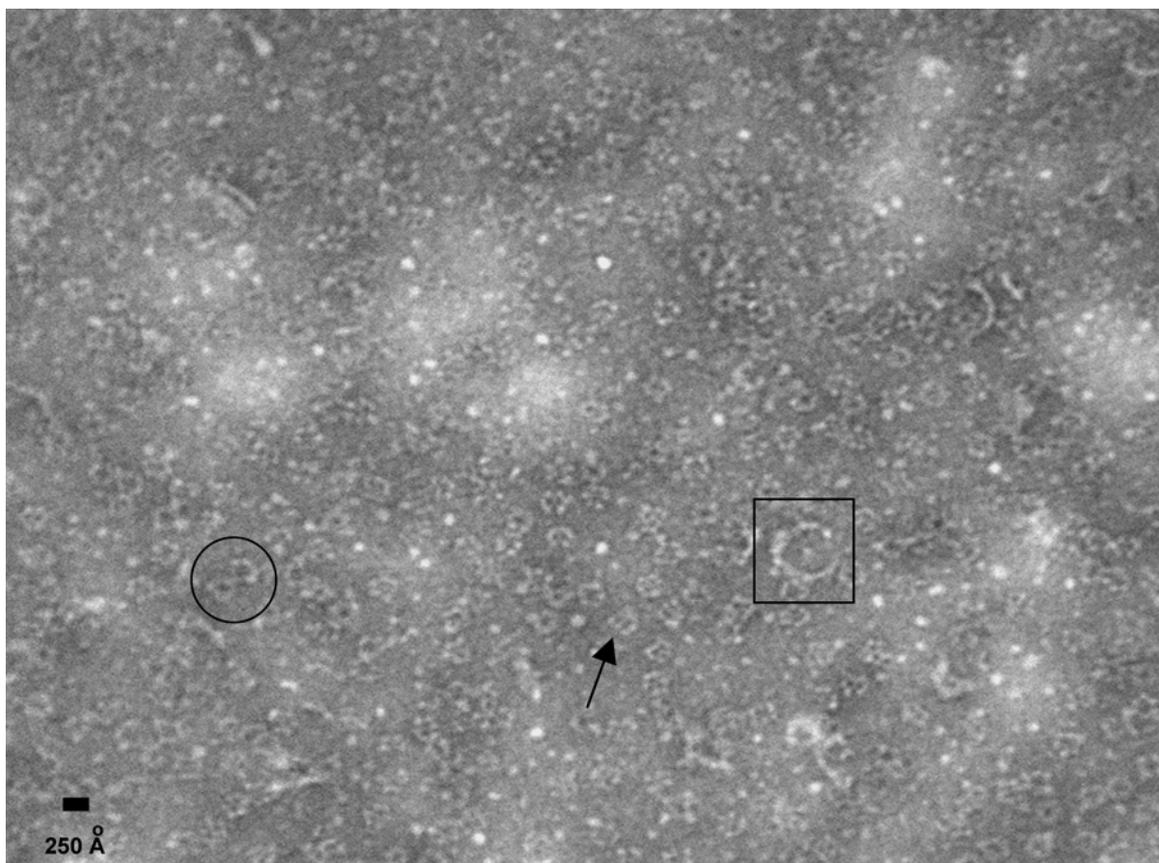
### **6.3.6.1 Negative stain EM of tE2/tE3BP**

The negative stain EM micrograph of tLi19/tLi30 shows a moderate distribution of icosahedral cores with no density in the pentagonal faces (Fig. 6.10), consistent with EM data on the recombinant and bovine E2/E3BP cores (sections 5.3.1.4.1 and 5.3.2.3), as well as with previously reported EM studies on bovine heart PDC (Wagenknecht, et al., 1991). The presence of empty pentagonal faces supports the ‘substitution model’ of E2/E3BP core organisation. However, despite the lack of uniform staining, a reasonable percentage of open-faced icosahedral cores are observed. However, a greater proportion of core clusters can be clearly seen, e.g. dimers (Fig. 6.10). This is in direct contrast to the clear EM image comprising a uniform distribution of full-length rE2/E3BP (section 5.3.1.4.1), and may reflect alterations in the surface charge of the tLi19/tLi30 cores leading to a greater predisposition towards cluster formation. In addition, few huge ‘arc-like’ structures are also visible and may correspond to incomplete cores linking together to form long chain-like structures.

### **6.3.6.2 Crystallisation of truncated E2/E3BP core**

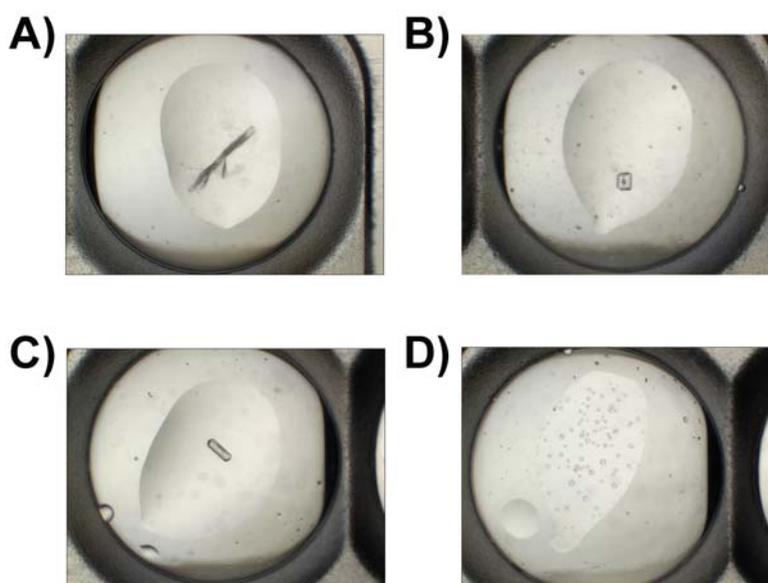
As tLi19/tLi30 is devoid of the flexible N-terminal peripheral arms that pose significant problems for crystallisation, initial attempts at crystallisation of purified tLi19/tLi30 were made as a first step towards determination of its high-resolution structure.

Several conditions yielding crystals were obtained from the initial screens (Fig. 6.11). However, testing those on an in-house Rigaku MicroMax 007 x-ray generator equipped with a Mar345 detector revealed salt crystal formation in the majority of cases (Figs 6.11A, 6.11B, 6.11C). Nevertheless, some protein crystals obtained in the initial screen (crystal condition: 2.2 M  $(\text{NH}_4)_2\text{SO}_4$ , 0.1 M  $\text{Na}_3\text{C}_6\text{H}_5\text{O}_7$ ) diffracted poorly owing to their minute size (Fig. 6.11D). Further work is required to optimize crystal growth conditions to obtain high quality diffracting crystals of a suitable size for x-ray analysis.



**Figure 6.10 Negative stain EM of tLi19/tLi30**

Negative stain image of tLi19/tLi30 confirms the dodecahedral framework of these cores. Apart from isolated core structures (arrow), a small fraction of dimer clusters (circle) and 'arc-like' structures (square) are also observed, indicating the presence of a small population of aberrant structures in the sample.



**Figure 6.11 Crystallisation of tLi19/tLi30**

Initial attempts on tLi19/tLi30 crystallisation resulted in several (A, B, C) salt crystals and possible (D) protein crystals.

## 6.4 Discussion

Evidence presented in this chapter demonstrates the structural characterisation of the first ever recombinant human truncated E2/E3BP core. Although, structures of recombinant human and yeast truncated E2 cores have been determined by Cryo-EM (Stoops, et al., 1992; 1997; Yu, et al., 2008), there is no structure of the functional human E2/E3BP core (or any full-length E2/E3BP core). Human E2/E3BP core has resisted crystallisation to date owing primarily to the intrinsic flexibility of the N-terminal flanking regions of E2 and E3BP and the difficulties in obtaining high yields of truncated E2/E3BP; consequently our understanding of the structure-function relationships in human PDC core has remained limited. Attempts to obtain truncated E2/E3BP have always employed limited proteolysis of the core with trypsin (Rahmatullah, et al., 1989b). Yu and co-workers (2008) have been successful in producing only minimum yields of recombinant human truncated E2 for structural studies. High yields (5-8 mg/ml per L of bacterial culture) of pure recombinant truncated E2/E3BP were successfully obtained in this study and has been structurally characterised as presented in this chapter.

The first solution structure of recombinant human truncated E2/E3BP (tLi19/tLi30) by neutron scattering was also successfully determined. The solution structure indicates a well formed pentagonal dodecahedron with icosahedral symmetry, consistent with previously obtained SAXS structures of rE2/E3BP and bE2/E3BP (sections 5.3.1.2 and 5.3.3.2), and published data on bE2/E3BP (Wagenknecht, et al., 1991). The icosahedral framework of truncated E2/E3BP is further confirmed by negative stain EM. In addition, no density within the pentagonal faces is observed, supporting the 'substitution' model of E2/E3BP core organisation. Interestingly, AUC studies indicate similar rates of sedimentation of the truncated E2/E3BP despite the absence of the N-terminal domains, as compared with the full-length rE2/E3BP and rE2 cores (sections 5.3.1.1.1 and 5.3.3.1.1). This is primarily due to the decrease in molecular mass being offset by the decrease in frictional drag. Moreover, the SANS model reconstruction of tLi19/tLi30 reveals large internal solvent-filled cavities, a characteristic feature of these eukaryotic cores, essential for ensuring their flexibility and catalytic function.

The  $D_{max}$  (300 Å) of truncated E2/E3BP core (obtained by SANS) being higher than the previously observed size variation (225-272 Å) in E2-PDC cores (Kong, et al., 2003; Wagenknecht, et al., 1991; Zhou, et al., 2001a) could possibly reflect a greater extent of 'breathing' in E2/E3BP. As the concept of breathing has been studied only with the yeast E2 core (Zhou, et al., 2001a), it is unclear how breathing modulates the size, flexibility and function of human E2/E3BP and more specifically how the introduction of E3BP affects these processes. It has been proposed that breathing of the core enhances the movement of the lipoyl 'swinging arms' towards the catalytic centres while additionally augmenting substrate channelling and overall rates of catalysis (Zhou, et al., 2001a).

Incorporation of E3BP has been suggested to reduce the symmetry of the core resulting in quasi-icosahedral symmetry (Wagenknecht, et al., 1991). Some structural differences between the truncated human and bacterial E2 cores have also been reported (Yu, et al., 2008). The authors justify these structural changes as being a requirement for the human E2 to accommodate E3BP during core assembly. As SANS cannot distinguish between the E2 and E3BP subunits, it will be of interest to obtain evidence for the localisation of E3BP and its effect on core dynamics. This important question can be addressed by single molecule experiments and/or cryo-EM studies that may provide more definitive information on the localisation and distribution of E3BP within the core assembly.

Superimpositions of the solution structure of tLi19/tLi30 with both the homology model of human tE2 and the crystal structure of bacterial tE2 show good overall similarity in the dodecahedral morphology of these cores, indicative of a conserved inner core structural template. Although recent work on the tE2 core demonstrates that the catalytic centre and trimeric building block unit are conserved across all organisms, subtle structural differences do exist between the human and bacterial truncated E2s, mainly confined to the inner linker region preceding the CTD of E2 (Yu, et al., 2008). These differences have been ascribed to the requirement for E3BP integration during core assembly.

The findings on truncated E2/E3BP core stability presented here strongly support our hypothesis that incorporation of E3BP leads to an overall decrease in core stability. Near-UV CD in the presence of GdmCl clearly shows unfolding via

intermediates, in good agreement with data for the full-length rE2/E3BP core (section 5.3.1.3). Interestingly the truncated E2/E3BP also exhibits greater stability than full-length rE2/E3BP, suggesting a possible contribution of the flexible N-terminal arms in destabilisation. The higher stability of rE2 (compared with tE2/tE3BP) is clearly reflected in their midpoints of unfolding, further supporting the proposition that integration of E3BP into the E2 core is responsible for the decrease in core stability. This is also supported by fluorescence studies on truncated E2/E3BP. Interestingly, previous GdmCl-induced denaturation studies on the truncated E2 core of *B. stearothermophilus* revealed a lower midpoint of unfolding (1.6 M) (Hiromasa, et al., 1998). The difference between the lower (1.6 M) and higher (3.19 M) midpoints of unfolding in the bacterial and human tE2, respectively, may be due to differences in experimental methodology (namely incubation times and temperature). Another possible and more interesting explanation may pertain to the subtle structural differences observed between these two cores, reflecting significant differences in their subunit interactions and overall flexibility. We propose that the extent of size variability or breathing may be considerably higher in the mammalian cores than their prokaryotic counterparts, posing significant implications for overall core flexibility and stability. It would be exciting to validate this hypothesis by cryo-EM or single molecule studies in the near future.

## Chapter 7

# Binding stoichiometry of E2/E3BP:E3 and core variation – an investigation

## 7.1 Introduction

In eukaryotic PDC, the E2/E3BP icosahedral core forms the structural and mechanistic framework for the intact, native assembly, with its complement of E1 and E3 enzymes binding tightly, but non-covalently. The 60-meric E2/E3BP core is composed of 20 trimeric units that are located at the 20 vertices of an icosahedron. Interaction of E3BP with E2 is mediated by its inner CTD in a process that requires the nascent polypeptide chains to come together during core assembly. This was established by previous studies in our laboratory, showing that E2 and E3BP failed to interact in a post-translational fashion, implying a mandatory requirement for integration of E3BP into the core framework during its initial assembly (Singh, 2008). The structure and subunit composition of E2/E3BP is critical to the efficient functioning of PDC. While E3BP is proposed to bind to the 12 faces of the pentagonal dodecahedron in yeast in the ‘addition’ model (60E2+12E3BP) (Stoops, et al., 1997; Zhou, et al., 2001b), it is thought to replace 12E2 subunits in the ‘substitution’ model (48E2+12E3BP) of mammalian PDC (Hiromasa, et al., 2004). These distinctive models of subunit organisation result in variable core composition in terms of its trimeric units. While 20E2 trimers constitute the 60-meric E2 core in yeast, the proposed mammalian E2/E3BP core would comprise a mixture of 8 homotrimers (only E2) and 12 heterotrimers (2E2 and 1E3BP). Moreover, the mammalian core model proposed by Hiromasa and colleagues (2004) suggests that 12E3BPs are symmetrically distributed within the core, although the authors do not clearly explain the grounds for such a distribution. The localisation of E3BP within the core and its overall subunit composition has significant implications for the structural stability, optimal operation and regulation of this macromolecular machine.

The binding stoichiometry of the peripheral enzymes E1 and E3 to the core has proved to be controversial. While in bacteria, both E1 and E3 bind to E2, only E1 binds E2 while E3 binds E3BP in eukaryotic PDCs. In *B. stearothermophilus*, both E1 and E3 bind to E2 with 1:1 stoichiometries in a mutually exclusive fashion (Hipps, et al., 1994; Jung, et al., 2002a; Lessard, et al., 1998). X-ray crystal structures of bacterial E1/E2-SBD (Frank, et al., 2005) and E3/E2-SBD (Mande, et al., 1996) subcomplexes as well as solution studies of these subcomplexes, show clearly that association of a second E2-SBD to either E1 or E3 is not possible, confirming the 1:1 stoichiometric relationship. A recent crystal structure for human E3 bound to E3BP-SBD shows the same 1:1 stoichiometry (Ciszak, et al., 2006). However, a 2:1 stoichiometry between E1:E2 and E3:E3BP has been reported in mammalian PDC, with 30E1 and 6-12E3 bound to the E2/E3BP core at maximal occupancy (Sanderson, et al., 1996b). This was further supported by recent solution studies on human E3 complexed with a truncated E3BP construct comprising the LD and SBD (E3BP-DD) that showed a clear 2:1 (E3BP-DD:E3) binding stoichiometry (Smolle, et al., 2006).

As E1 and E3 compete for overlapping binding sites on E2 in *B. stearothermophilus*, transient interactions governed by weak reversible interactions are probably more advantageous for optimal catalytic efficiency. However, owing to the specialised evolutionary role of E3BP in mammalian PDC and the lack of competition between E1 and E3 for binding sites, a stronger and tighter interaction between the respective subcomplexes might be preferable for efficient operation and regulation of this macromolecular assembly.

This chapter is concerned with the investigation of the stoichiometry of interaction of the E2/E3BP core with E3 using AUC, SANS and SDS-PAGE. Substrate specific radiolabelling of the reduced lipoyl groups has been employed previously to obtain the ratio of E2 and E3BP subunits within the bovine core assembly (Sanderson, et al., 1996b). Comparative radiolabelling of bovine and recombinant E2/E3BP cores using [ $^{14}\text{C}$ ]-NEM and fluorography was carried out in this study in order to determine their precise subunit compositions. In addition, the structural and geometric constraints that govern core assembly and organisation have been explored via mathematical modelling.

## 7.2 Binding stoichiometry of rE2/E3BP:E3

### 7.2.1 *Materials and methods*

#### 7.2.1.1 Protein purification

The full-length rE2/E3BP core was purified as outlined in section 4.3.2.2. Purified rE2/E3BP after gel filtration was exchanged into PEB buffer (see section 3.3.2.4) and immediately used for stoichiometry experiments. All protein concentration measurements of rE2/E3BP were based on the 48E2+12E3BP stoichiometry unless stated otherwise. Recombinant protonated E3 (rE3) was purified by standard chromatography techniques (section 4.3.2.3). Recombinant deuterated E3 (dE3) was provided by Dr. Phil Callow, Institut Laue Langevin (ILL), Grenoble, France and purified as described in section 4.3.2.3. rE3 and dE3 were suspended in PEB buffer and subsequently used for the stoichiometry experiments. The rE2/E3BP:rE3 or rE2/E3BP:dE3 complexes were reconstituted by mixing purified rE2/E3BP and rE3 (or dE3) at various molar ratios in PEB buffer depending on the experiment in question.

#### 7.2.1.2 Sedimentation velocity

For the SV analysis of rE2/E3BP:rE3 complexes, samples were prepared by maintaining the rE2/E3BP concentration at 191.5 nM and varying the concentration of rE3 accordingly to achieve E2/E3BP:E3 ratios of 4:1 (48E2/12E3BP:3E3) to 1:4 (48E2/12E3BP:48E3). All samples in PEB buffer (360  $\mu$ l) were loaded into 12 mm double sector centrepieces. The experiment was carried out at a rotor speed of 20,000 rpm at 4°C. A total of 400 interference scans were recorded, 1 min apart until sedimentation was complete. Sedimentation profiles were analysed using  $c(s)$  analysis in SEDFIT (Schuck, 2000). More details can be found in sections 2.2.1.3 and 3.4.2.

### 7.2.1.3 Gel filtration

Samples of rE2/E3BP:rE3 were prepared by maintaining the rE2/E3BP concentration at 421 nM and varying the rE3 concentrations to satisfy ratios of rE2/E3BP:rE3 ranging from 3:1 (48E2/12E3BP:4E3) to 1:3 (48E2/12E3BP:36E3). All samples were analysed by gel filtration on a Sephacryl S-300 column (Amersham, USA). In addition uncomplexed rE2/E3BP (421 nM) and rE3 (corresponding to 36E3) were also analysed and served as controls.

### 7.2.1.4 SANS contrast variation

rE2/E3BP and dE3 were both thoroughly dialysed in 40% D<sub>2</sub>O and 100% D<sub>2</sub>O to achieve maximum scattering contrast. However, owing to aggregation of the dialysed rE2/E3BP core, SANS data could not be collected for the contrast matched samples. Nevertheless, the scattering signal obtained from protonated rE2/E3BP and rE2/E3BP:dE3 was significantly different and subsequently exploited in stoichiometry studies with undialysed rE2/E3BP:dE3 complex.

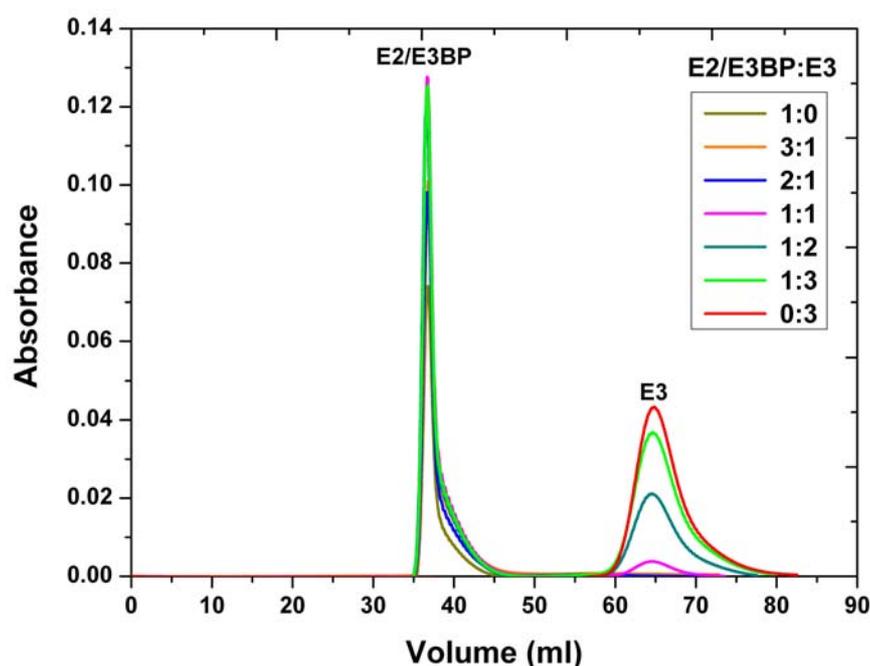
The rE2/E3BP:dE3 complex was prepared by maintaining the concentration of rE2/E3BP at 1.04  $\mu$ M and initially adding dE3 at a concentration appropriate to give an E2/E3BP+E3 subunit composition of 48/12+2. Thereafter dE3 was added stepwise, with every step amounting to the addition of 2 more E3s thus covering the range of 48/12+2 to 48/12+18. Free rE2/E3BP (3.69 mg/ml) and dE3 (11 mg/ml) were employed as controls. Scattering curves were recorded at dual camera lengths (4 and 14 m) at 4°C. Treatment of raw data was carried out as described in section 3.4.5.

## 7.2.2 Results and analysis

### 7.2.2.1 E2/E3BP:E3 binding indicates a 2:1 stoichiometry

#### 7.2.2.1.1 Gel filtration profile of E2/E3BP:E3

The overlaid gel filtration profiles of a series of rE2/E3BP:rE3 mixtures covering a broad range of ratios from 48E2/12E3BP:4E3 (3:1) to 48E2/12E3BP:36E3 (1:3)



**Figure 7.1 Gel filtration analysis of rE2/E3BP:rE3 binding stoichiometry**

Elution profiles from gel filtration chromatography of rE2/E3BP:rE3 at various ratios in the range 3:1 to 1:3, (i.e. 48E2/12E3BP:4E3 to 48E2/12E3BP:36E3) are shown. Uncomplexed rE2/E3BP and rE3 were used as controls. Protein absorbance was measured at 280 nm. Free E3 is observed from a ratio of 1:1 onwards.

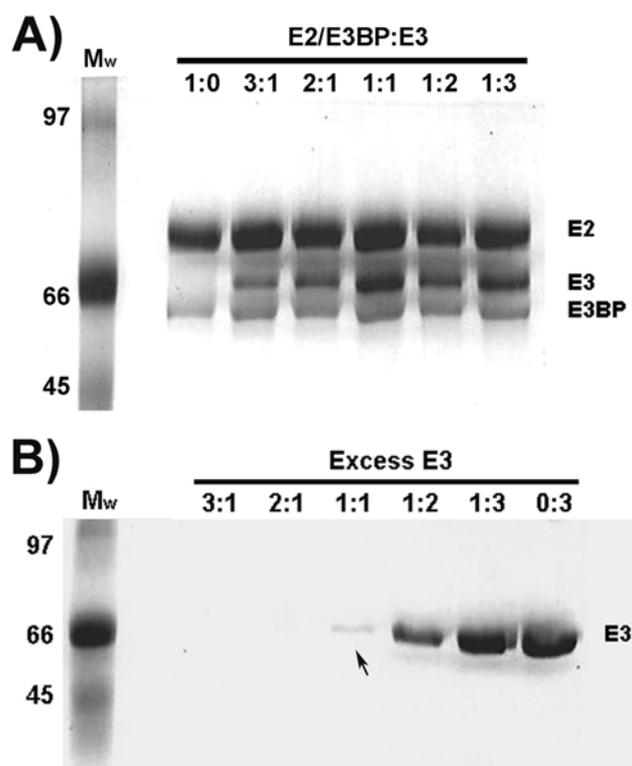
are shown in Fig. 7.1. Uncomplexed rE2/E3BP and rE3 (corresponding to 36E3) were used as controls. The dissociation constant,  $K_d$ , measured for human E3BP:E3 and E3BP-DD:E3 subcomplexes by ITC have been previously reported to be 6.3 nM and 35.7 nM, respectively (Brown, 2002a; Smolle, 2005), indicating high affinity binding between E3BP and E3. Concentrations of E3 and E3BP used in all stoichiometry experiments in this study are significantly greater than the  $K_d$ , ensuring tight and complete binding between E3 and E3BP as well as reliability of all experimental data.

While the void volume (38 ml) elution peaks for uncomplexed and complexed E2/E3BP are clearly seen at all ratios, the appearance of free E3 ( $V_e$  65 ml) is evident only at a ratio of 1:1 and above. Therefore, it can be inferred that at a ratio of 2:1 (48E2/12E3BP:6E3) all the E3BP binding sites are fully saturated, with uncomplexed E3 being observed only at ratios higher than 2:1. These findings support previous data from our laboratory indicating that the

stoichiometry of binding between E3BP and E3 on the core surface is in the region of 2:1 (12E3BP:6E3) (Smolle, et al., 2006). In addition, the height and area of the pure rE3 peak were greater than that observed for rE3 remaining uncomplexed in the 1:3 sample (for which the total concentration of rE3 was equivalent). This further confirms significant but limited binding of rE3 to the rE2/E3BP core.

### 7.2.2.1.2 SDS-PAGE of E2/E3BP:E3

SDS-PAGE analysis (Fig. 7.2) was performed on the elution peak fractions of rE2/E3BP:rE3 complexes and free rE3 obtained from the gel filtration (GFC) stoichiometry studies (section 7.2.2.1.1). SDS-PAGE of void volume (38 ml) GFC samples (rE2/E3BP and rE2/E3BP:E3) clearly shows the presence of all constituent proteins (E2, E3BP and E3) in the complex (Fig. 7.2A). In addition, SDS-PAGE analysis of peak fractions of free rE3 (65 ml) further confirm the presence of excess E3 only at ratios greater than 2:1 (Fig.7.2B).

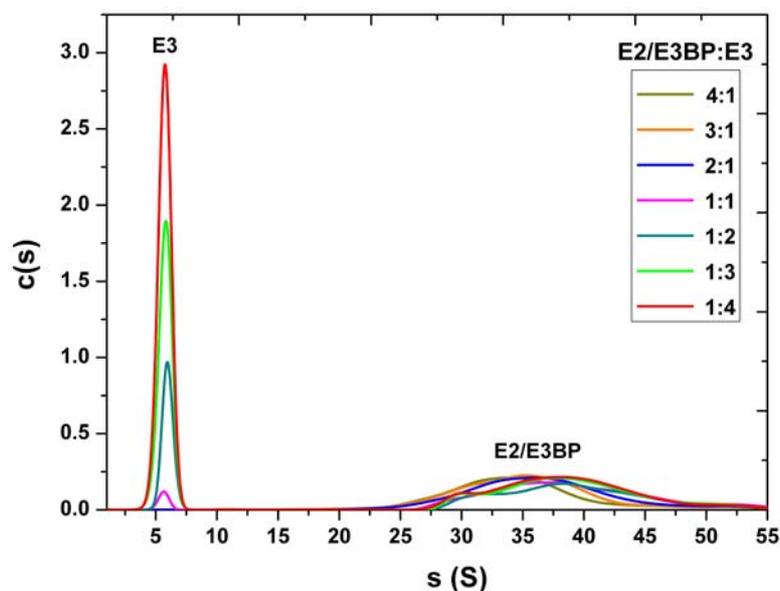


**Figure 7.2 SDS-PAGE analysis of E2/E3BP:E3 stoichiometry**

The GFC peaks of E2/E3BP:E3 stoichiometry samples were analysed by SDS-PAGE. **(A)** Void volume GFC peaks indicate the presence of E3 in the complex. **(B)** GFC peaks of E3 taken at the elution volume of 65 ml show the appearance of excess E3 at a ratio of 1:1 (arrow). Molecular weights (M<sub>w</sub>) are shown in kDa.

### 7.2.2.1.3 SV analysis of E2/E3BP:E3

SV experiments were conducted for the various rE2/E3BP:rE3 ratios and the sedimentation profiles analysed using the  $c(s)$  model in SEDFIT (Schuck, 2000) as shown in Fig. 7.3. The apparent sedimentation coefficient of free E3 is seen to be 5.8 S, in excellent agreement with the previously published value of  $s_{20,w}^0 = 5.9$  S (Smolle, et al., 2006). In section 5.3.1.1.1 the sedimentation coefficient of uncomplexed rE2/E3BP was reported to be 29.3 S. Now, an increase in the sedimentation coefficient of rE2/E3BP upon rE3 binding (35.3 S) is observed. Moreover, Fig. 7.3 also shows the appearance of free E3 at a 1:1 ratio, consistent with the results obtained from GFC stoichiometry studies (section 7.2.2.1.1).



**Figure 7.3 Sedimentation velocity analysis of E2/E3BP:E3 stoichiometry**

$c(s)$  analysis of SV interference data of various E2/E3BP:E3 stoichiometric mixtures indicate the binding stoichiometry to be 2:1. The axes  $s$  and  $c(s)$  denote the sedimentation coefficient and the  $c(s)$  distribution model (from SEDFIT analysis), respectively.

### 7.2.2.2 10E3s bind to the rE2/E3BP core

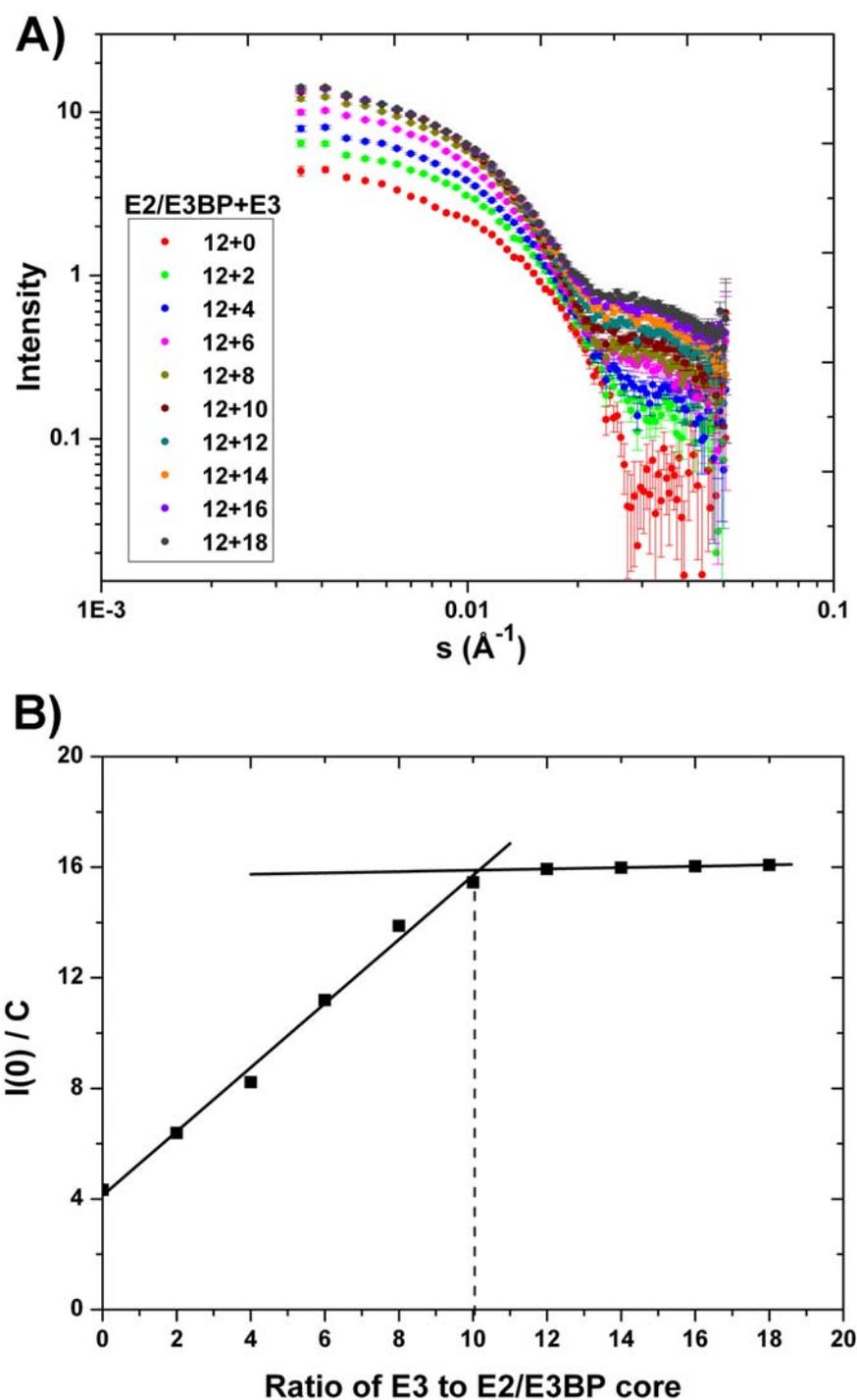
Purified protonated rE2/E3BP was reconstituted with deuterated E3 (dE3) to confirm the binding stoichiometry and accordingly determine the subunit composition of rE2/E3BP using SANS. The significant difference in the interaction

of neutrons with hydrogen and deuterium is exploited in the technique of SANS 'contrast matching'. rE2/E3BP and dE3 were estimated to have matchpoints of 40% and 100%, respectively. Therefore, dialysing both proteins against D<sub>2</sub>O buffers at their matchpoint percentages enables selective scattering of these components. In a solvent containing 40% D<sub>2</sub>O, scattering contribution of rE2/E3BP in reconstituted rE2/E3BP:dE3 is completely matched out. Similarly, when scattering data for rE2/E3BP:dE3 in 100% D<sub>2</sub>O are recorded, the scattering of dE3 in the complex will be the same as that of the solvent, and thus will be matched out. However, dialysis of rE2/E3BP at 40% and 100% D<sub>2</sub>O led to protein aggregation and consequently it was not possible to conduct the stoichiometry analysis with contrast-matched samples.

However, as the scattering of undialysed reconstituted rE2/E3BP:dE3 was significantly higher than uncomplexed rE2/E3BP, SANS stoichiometry studies were successfully performed on undialysed rE2/E3BP:dE3 at various molar ratios by exploiting the difference between the scattering of the deuterated and protonated components (see section 7.2.1.4). Scattering curves obtained for rE2/E3BP and rE2/E3BP:dE3 at various ratios are shown in Fig. 7.4A.

The increase in intensity, particularly in the low angle region can be clearly seen with the addition of E3, resulting in a subunit composition (E3BP+E3) range from 12+0 (uncomplexed rE2/E3BP) to 12+18 (18 dE3 dimers per rE2/E3BP core). The scattering intensity at zero angle,  $I(0)$  and radius of gyration,  $R_g$  were determined using the Guinier approximation and are listed in Table 7.1. An increase in the number of dE3s results in the concomitant increase in  $R_g$  and  $I(0)$  values as expected upon binding of dE3 to rE2/E3BP. Moreover, the addition of dE3 results in a maximum  $R_g$  of 162 Å, consistent with the value of 161.5 Å obtained by Hiromasa and co-workers (2004).

From Table 7.1, it is also apparent that  $I(0)$  tends to decrease at higher concentrations of E3, consistent with a decrease in the average molecular weight of the system upon generation of excess E3. The value of  $I(0)$  normalised for concentration and plotted against the ratio of dE3 to rE2/E3BP core is shown in Fig. 7.4B. Unexpectedly, at 10dE3 dimers per rE2/E3BP core (i.e. 10:1 of dE3:rE2/E3BP) the gradient of the curve changes significantly indicating saturation of binding.



**Figure 7.4 Stoichiometry studies of dE3:rE2/E3BP using SANS**

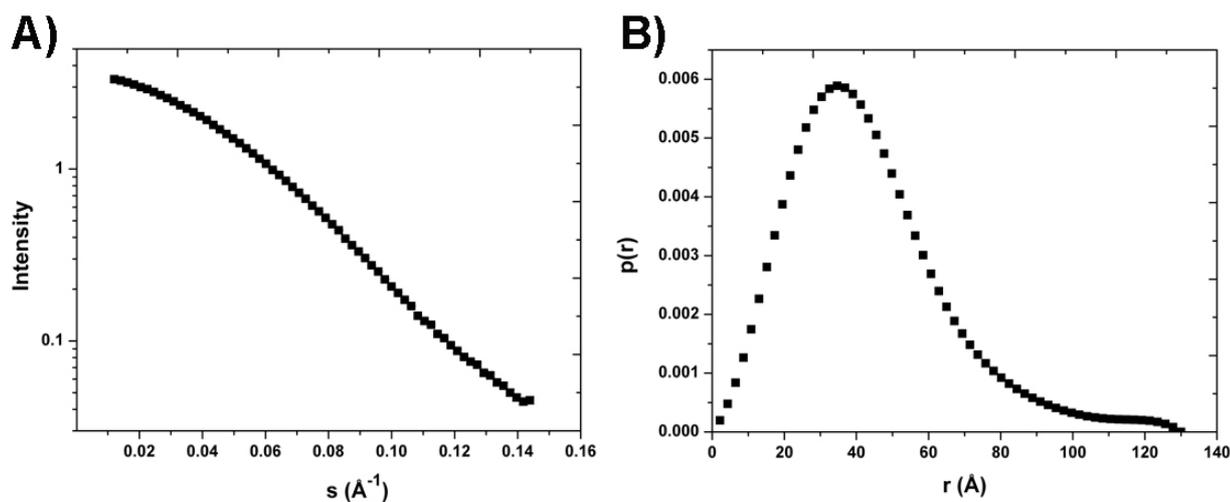
**(A)** The SANS scattering curves of protonated rE2/E3BP complexed with dE3 at various ratios ranging from 0-18 E3 dimers per core were recorded. **(B)**  $I(0)$  normalised for concentration plotted against the ratio of dE3 per core indicates saturation at a ratio of 10:1 of dE3:rE2/E3BP. Error bars are shown but are not visible due to their small size.

Subunit stoichiometry E2/E3BP+E3	Radius of gyration $R_g$ (Å)	Intensity $I(0)$
12+0	143 ± 2	4.3347 ± 1
12+2	145 ± 1	6.3374 ± 1
12+4	151 ± 1	8.0913 ± 1
12+6	157 ± 1	10.922 ± 1
12+8	160 ± 1	13.431 ± 1
12+10	162 ± 1	14.838 ± 1
12+12	162 ± 1	15.183 ± 1
12+14	162 ± 1	15.108 ± 1
12+16	162 ± 1	15.038 ± 1
12+18	162 ± 1	14.965 ± 1

**Table 7.1**  $R_g$  of SANS rE2/E3BP:dE3 mixtures

The observed saturation of binding at 10dE3 dimers per rE2/E3BP core is similar to the previously published value of 12:1 (12E3 dimers per core) (Hiromasa, et al., 2004). The small difference might be accounted for by experimental errors in the precise estimation of protein concentrations or possible aggregation of dE3. It was initially assumed that the observed value of 10:1 (i.e. 10E3+12E3BP) may in fact be 12:1 (i.e. 12E3+12E3BP), indicative of a 1:1 stoichiometry. However, the SANS stoichiometry experiment was conducted twice and gave the same result of 10 dE3 dimers per rE2/E3BP core. Nonetheless, this is an unexpected finding apparently conflicting with previously published data from our lab (Smolle, et al., 2006) and results presented earlier in this chapter. An alternative hypothesis is that the recombinant E2/E3BP core deviates significantly from the current 48E2+12E3BP model and this is tested later in this chapter by standard biochemical techniques.

To rule out problems associated with dE3 aggregation, the scattering curve for free dE3 (11 mg/ml) was obtained (Fig. 7.5A). No signs of aggregation were observed, and from the extrapolation of intensity to zero angle,  $I(0)$ , the molecular weight was estimated to be 110 kDa, in good agreement with the value predicted from the E3 amino acid sequence of 106 kDa. Guinier and GNOM (Semenyuk and Svergun, 1991; Svergun, 1992) analysis were employed to determine the radius of gyration,  $R_g$ , to be  $33 \pm 1$  Å and  $34 \pm 1$  Å, respectively. The distance distribution function,  $p(r)$ , (Fig. 7.5B) reflects the slightly



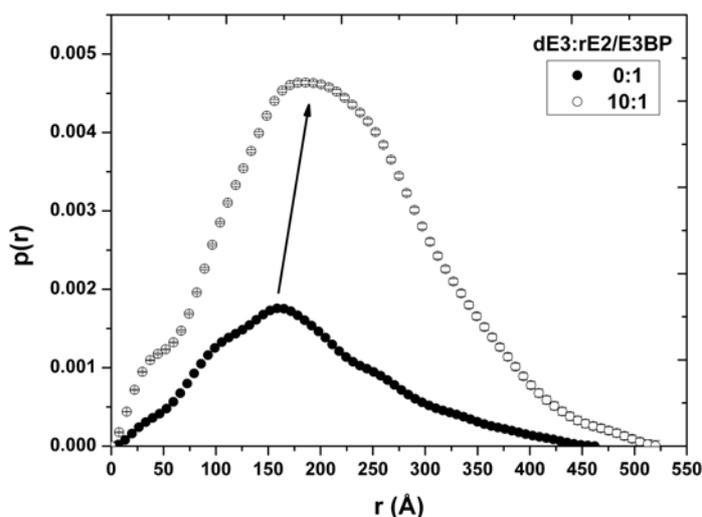
**Figure 7.5 Small angle neutron scattering of deuterated E3**

(A) The scattering curve of uncomplexed dE3 is shown. The x-axis  $s$  denotes the scattering angle. (B) The distance distribution  $p(r)$  calculated using GNOM yields a  $D_{max}$  of 130 Å.

elongated shape of dE3 and was used to estimate the maximum particle dimension,  $D_{max}$  of 130 Å.

The values of  $R_g$  and  $D_{max}$  of dE3 compare well with previously published values obtained for protonated E3 (Smolle, et al., 2006) (38 Å and 130 Å), indicating no major structural changes induced by the deuterium labelling of E3. Moreover, the scattering curve for dE3 reveals a less compact structure in solution relative to the E3 crystal structure, consistent with previously published findings (Smolle, et al., 2006).

Fig. 7.6 shows the change in the  $p(r)$  distribution function of rE2/E3BP with addition of a saturating amount of 10dE3. Binding of dE3 to rE2/E3BP increases  $D_{max}$  from 462 Å to 520 Å. In addition, the peak position of the bell-shaped function is shifted to a higher  $r$ . This indicates a radial outwards shift of mass distribution, with dE3 bound to the exterior of the rE2/E3BP dodecahedron as opposed to the reported distribution of E3 in yeast (Stoops, et al., 1997).



**Figure 7.6** Size distribution  $p(r)$  analysis of dE3:rE2/E3BP at 10:1 saturation

The  $p(r)$  distributions were calculated using GNOM for uncomplexed rE2/E3BP and dE3:rE2/E3BP complex at a binding saturating ratio of 10:1. The radial shift to higher  $r$  upon dE3 binding to the E2/E3BP core is denoted by an arrow. Error bars are shown but are not visible due to their small size.

## 7.3 Subunit composition of the rE2/E3BP core

### 7.3.1 Materials and methods

#### 7.3.1.1 Protein purification

rE2/E3BP and rE3 were purified as described in sections 4.3.2.2 and 4.3.2.3. PDC was isolated from native bovine heart as outlined in section 3.3.2.5. Bovine E3 (bE3) was obtained from sucrose gradient centrifugation of bovine PDC (section 3.3.2.6). All purified proteins were exchanged in PEB buffer (see section 3.3.2.4) and subjected to immediate radiolabelling. The rE2/E3BP:rE3 complex was reconstituted by mixing purified rE2/E3BP and rE3 at various molar ratios.

#### 7.3.1.2 [ $^{14}\text{C}$ ]-NEM radiolabelling and counting

The labelling of purified PDC and rE2/E3BP:rE3 suspended in PEB buffer with [ $^{14}\text{C}$ ]-NEM was carried out as outlined in section 3.3.11.1. While PDC was used at

concentrations of 40 µg, 60 µg and 120 µg, rE2/E3BP (20 µg and 40 µg) was mixed with rE3 at various molar ratios to achieve final stoichiometries of 2:1, 10:1, 20:1 and 40:1 of rE2/E3BP:rE3. All samples were subsequently mixed with NAD<sup>+</sup> (0.5 mM) or NADH (1 mM) and incubated for 10 min at room temperature. Radioactive [<sup>14</sup>C]-NEM (0.65 mM) was then added to all these protein mixtures and subjected to 30 min incubation at room temperature. Reactions were terminated by the addition of 50 mM DTT followed by incubation for 12 min. Thereafter, SDS loading dye was added resulting in final sample volumes of 20-25 µl before being analysed on 8% SDS-PAGE slab gels. More details can be found in section 3.3.11.1. Gels were subsequently stained, destained and the radioactive bands excised and dissolved overnight in SOLVABLE (see section 3.3.11.2). Solubilised samples were then counted using a scintillation counter for 10 min per sample. Radiolabelling experiments were performed on four preparations of PDC and recombinant E2/E3BP:E3 complex. Incorporation of [<sup>14</sup>C] radioisotope into each sample was determined in triplicate. Data were analysed statistically by means of the following equations (Lynch and Walsh, 1997).

$$\text{Ratio of counts} \left( \frac{E2}{E3BP} \right) = \frac{C_{E2}}{C_{E3BP}} \left( 1 + \frac{\sigma_{E3BP}^2}{C_{E3BP}^2} \right) \quad (7.1)$$

$$\text{Variance} \left( \frac{E2}{E3BP} \right) = \frac{C_{E2}^2}{C_{E3BP}^2} \left( \frac{\sigma_{E2}^2}{C_{E2}^2} + \frac{\sigma_{E3BP}^2}{C_{E3BP}^2} \right) \quad (7.2)$$

where  $C_{E2}$  and  $C_{E3BP}$  represent the [<sup>14</sup>C] counts of rE2 and rE3BP,  $\sigma_{E2}$  and  $\sigma_{E3BP}$  are the errors on the [<sup>14</sup>C] counts of rE2 and rE3BP.

### 7.3.1.3 Fluorography

Radiolabelled rE2/E3BP (6 µg) complexed with rE3 at stoichiometric ratios of 10:1 and 20:1 was used for fluorography experiments (see section 3.3.11.1 and 7.3.1.2). SDS-PAGE gels with [<sup>14</sup>C]-NEM radiolabelled rE2/E3BP:rE3 (10:1, 20:1) in the presence of 0.5 mM NAD<sup>+</sup> or 1 mM NADH were stained and fixed prior to drying under vacuum (see section 3.3.11.3). Fluorographs were stored at -80°C and developed on x-ray film after 7 days.

### 7.3.1.4 Mathematical modelling

The subunit organisation of rE2/E3BP core was mathematically modelled as described in detail in section 3.5.7. Possible solutions were obtained for variables X, Y and Z that define the number of 3-edge, 4-edge and 5-edge E2 faces, respectively for core subunit stoichiometries ranging from 40E2+20E3BP to 60E2+0E3BP (60-meric E2 only core). Several solutions were obtained for each subunit stoichiometry based on constraints used during the modelling process, reflecting the various final states possible for each model of core organisation. Each solution is represented as a combination of the variables X, Y and Z, represented as [X, Y, Z].

## 7.3.2 Results and analysis

### 7.3.2.1 Distinctive subunit organisations: rE2/E3BP and bE2/E3BP cores

#### 7.3.2.1.1 Variable composition of rE2+rE3BP based on SANS

As E3 specifically binds to E3BP in eukaryotic PDCs, it serves as a marker for the positions of E3BP within the E2/E3BP core, thereby providing information on the subunit composition of the core. The SANS stoichiometry data for rE2/E3BP:dE3 clearly indicated the presence of 10dE3 dimers per rE2/E3BP core at maximal occupancy (section 7.2.2.2). However, as SANS is a time average technique, the saturation limit of 10dE3s per rE2/E3BP core would represent an average binding ratio. In addition, all other stoichiometry data show a 2:1 (6E3s per rE2/E3BP) binding of rE2/E3BP:rE3 (sections 7.2.2.1.1 and 7.2.2.1.3). Therefore, these data might imply different core compositions deviating significantly from the current 48E2+12E3BP model resulting in distinct apparent stoichiometries for rE2/E3BP:rE3 (1:1 or 2:1) as tabulated in Table 7.2.

2:1	1:1	E3BP:E3 stoichiometry
Number of E3 dimers per rE2/E3BP core		Core composition
<b>10</b>	20	40E2+20E3BP
6	12	48E2+12E3BP

**Table 7.2 Variable substitution core models of the rE2/E3BP core**

The value highlighted in blue denotes the ratio of dE3:rE2/E3BP obtained from SANS stoichiometry data.

### 7.3.2.1.2 [<sup>14</sup>C]-NEM radiolabelling

NADH-induced radiolabelling of E2 and E3BP was performed to measure differences in the subunit composition of the bovine and recombinant E2/E3BP cores. This was achieved by specifically labelling the reduced lipoyl groups of E2 and E3BP with [<sup>14</sup>C]-NEM. Under reducing conditions (NADH) and in the presence of E3, the thiol groups of lipoic acid are covalently modified with NEM. As the E2 and E3BP subunits have 2 and 1 LD, respectively, specific [<sup>14</sup>C] incorporation into the reduced lipoyl groups can yield direct information on their relative abundance within the purified core i.e. provide a direct estimate of E2:E3BP core subunit stoichiometry. Radiolabelling was carried out on triplicate samples of four fresh preparations of bovine PDC and human rE2/E3BP:E3, to prevent any uncertainties owing to batch variation or sample degradation. Additionally, data analysis was based on the assumption of equivalent and complete labelling of both E2 and E3BP subunits.

The expected ratio of radiolabelled [<sup>14</sup>C] counts for E2:E3BP would differ significantly depending on the model of subunit organisation as shown in Table 7.3, in particular the difference between the 48E2+12E3BP substitution model (Hiromasa, et al., 2004) and the proposed ‘variable E3BP substitution’ model (see section 7.4) based on SANS data (see section 7.3.2.1.1). It is also interesting to note that differences in predicted radiolabelling ratios for E2:E3BP over the subunit composition range of 46E2+14E3BP to 40E2+20E3BP are relatively minor (Table 7.3).

Model	Subunit stoichiometry	Ratio of LDs of E2:E3BP	Theoretical ratio of radiolabelled E2:E3BP
Addition	60E2+12E3BP	120:12	10:1
Substitution	48E2+12E3BP	<b>96:12</b>	<b>8:1</b>
Proposed (Variable E3BP substitution model)	46E2+14E3BP	<b>92:14</b>	<b>6.5:1</b>
	44E2+16E3BP	<b>88:16</b>	<b>5.5:1</b>
	42E2+18E3BP	<b>84:18</b>	<b>4.6:1</b>
	40E2+20E3BP	<b>80:20</b>	<b>4:1</b>

**Table 7.3 Core stoichiometries and their theoretical [<sup>14</sup>C] radiolabelled E2:E3BP ratios**

Interestingly, analysis of the ratio of radiolabelled E2:E3BP subunits obtained from this series of experiments suggests a marked difference in core organisation between the recombinant and bovine E2/E3BP cores and is summarized in Table 7.4. While the native bovine E2/E3BP core of PDC favours a subunit organisation of 48E2+12E3BP, recombinant E2/E3BP core contains a greater number of E3BPs (>12). Therefore, radiolabelling data suggest average core models of 48E2+12E3BP and 40E2+20E3BP for bE2/E3BP and rE2/E3BP, respectively.

A potential source of error in these estimates could stem from background labelling of cysteine residues in other PDC subunits in the presence or absence of NAD<sup>+</sup> and NADH. This was investigated by fluorography as described in the following section.

### 7.3.2.1.3 Fluorography

Substrate specific labelling of the reduced lipoyl groups of E2 and E3BP of rE2/E3BP with [<sup>14</sup>C]-NEM was confirmed by fluorography. Previous studies by Hodgson and colleagues (1986) established selective labelling of E2 and E3BP in

Samples	[ <sup>14</sup> C] counts E2	[ <sup>14</sup> C] counts E3BP	Ratio E2:E3BP
bE2/E3BP <sub>1</sub>	6025 ± 489	752 ± 31	8.03 ± 0.53:1
bE2/E3BP <sub>2</sub>	5640 ± 480	718 ± 66	7.92 ± 0.97:1
bE2/E3BP <sub>3</sub>	2038 ± 95	250 ± 20	8.20 ± 0.57:1
bE2/E3BP <sub>4</sub>	3230 ± 125	398 ± 40	8.19 ± 0.76:1
rE2/E3BP <sub>1</sub>	7802 ± 568	1908 ± 230	4.15 ± 0.33:1
rE2/E3BP <sub>2</sub>	19325 ± 1309	4558 ± 93	4.24 ± 0.09:1
rE2/E3BP <sub>3</sub>	17967 ± 267	4364 ± 11	4.12 ± 0.01:1
rE2/E3BP <sub>4</sub>	4688 ± 110	832 ± 10	5.63 ± 0.02:1

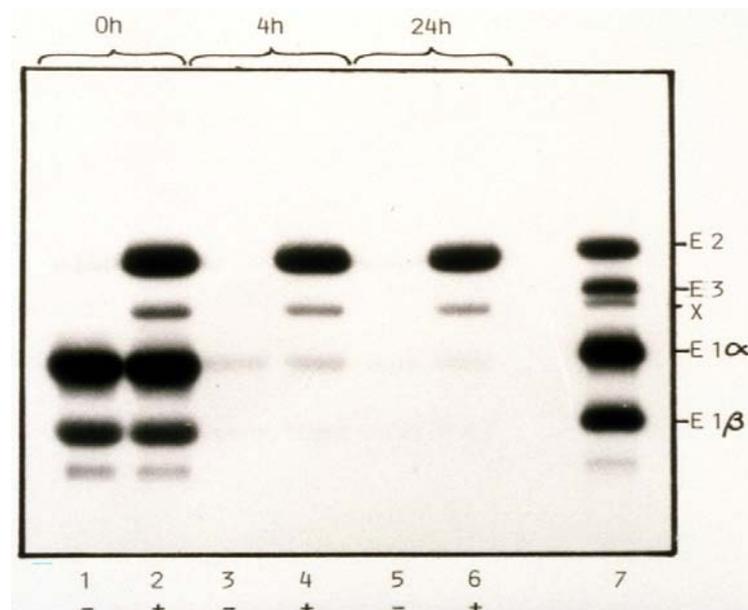
**Table 7.4 Determination of subunit composition of bovine and recombinant E2/E3BP cores via [<sup>14</sup>C]-NEM radiolabelling**

Subscripts 1 to 4 denote different protein preparations used for the experiment, and ratios of radiolabelled E2:E3BP are colour coded based on substitution (blue) or the proposed variable E3BP substitution (violet) models. Average counts of triplicate samples ± SEM (standard error of the mean) are shown.

bovine PDC. Their work also indicated that no radiolabelling of bE3 was detectable in the oxidised (NAD<sup>+</sup>) or reduced states (NADH) as shown in Fig. 7.7. In contrast E1 $\alpha$  and E1 $\beta$  that contain several accessible Cys residues are readily modified under both oxidising and reducing conditions (in the presence of NAD<sup>+</sup> and NADH).

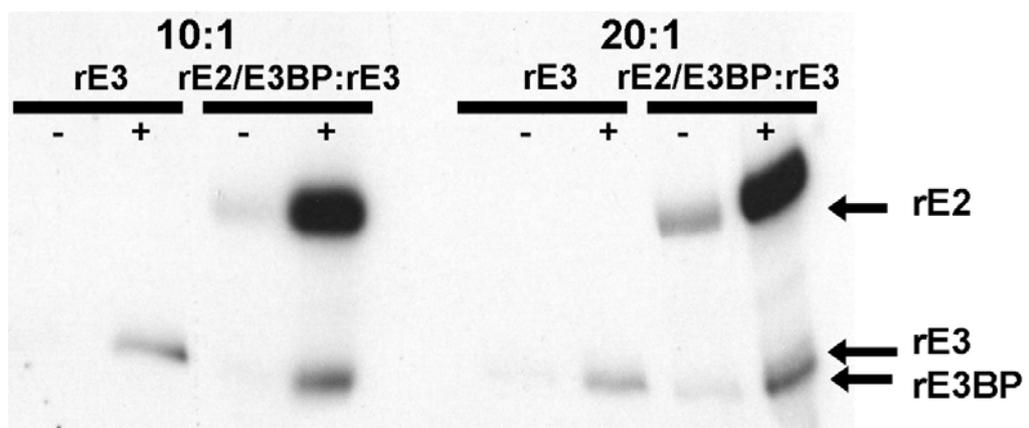
Labelling of the recombinant E2/E3BP core was performed at a 2:1 (48E2/12E3BP:6E3) stoichiometry of rE2/E3BP:rE3 to ensure maximum reduction of E2 and E3BP. However, under these conditions human rE3 labelled quite strongly. To avoid cross contamination of excised E3BP with E3, subsequent experiments were conducted employing minimal amounts of rE3 where no detectable labelling was observed. The extent of radiolabelling of E2 and E3BP was unaffected over the entire range of rE3 concentrations used.

The fluorograph obtained for rE2/E3BP:rE3 (rE2/E3BP complexed with rE3) indicates a similar result as compared with bE2/E3BP (Fig. 7.7). Selective labelling of E2 and E3BP within the complex is observed with no discernable labelling of the complexed E3 evident in the presence of NADH (Fig. 7.8). However, radiolabelling of uncomplexed rE3 is observed in the presence of



**Figure 7.7 Incorporation of [ $^{14}\text{C}$ ]-NEM into bovine heart PDC**

Fluorography confirming [ $^{14}\text{C}$ ]-NEM incorporation into the various subunits of bovine PDC is shown. Samples were pretreated over various times (0h, 4h, 24h) with radioactive NEM in the presence of  $\text{NAD}^+$  (-) or  $\text{NADH}$  (+). Native PDC in the presence of  $\text{NAD}^+$  (Lane 1) and  $\text{NADH}$  (Lane 2) is indicated. Lanes 3-6 represent radiolabelled PDC treated with non-radioactive NEM for 4h and 24 h in the presence of  $\text{NAD}^+$  (Lanes 3, 5) and  $\text{NADH}$  (Lanes 4, 6). Denatured PDC labelled with [ $^{14}\text{C}$ ]-NEM is shown in Lane 7. The various subunits of PDC are indicated. The image was taken from Hodgson, et al. (1986).



**Figure 7.8 Incorporation of [ $^{14}\text{C}$ ]-NEM into rE2/E3BP core**

Fluorography of rE2/E3BP complexed with rE3 at molar ratios of 10:1 and 20:1 was carried out in the presence of  $\text{NAD}$  (-) and  $\text{NADH}$  (+) as shown. Uncomplexed rE3 at ratios of 10:1 and 20:1 under non-reducing (-) and reducing (+) was used as control. While rE2 and rE3BP labelling is observed in the presence of  $\text{NADH}$ , radiolabelling of rE3 is evident only in the uncomplexed state and not when associated with the rE2/E3BP core.

NADH, directly in contrast to bovine E3 (Fig. 7.7), indicating exposure of thiol groups to NEM modification possibly those involved in the Cys disulphide pair that is integral to the catalytic mechanism. In addition, minimal background labelling (<5%) of rE2/E3BP was sometimes observed (Fig. 7.8) and could possibly represent spillover or contamination from adjacent tracks.

As fluorography indicates selective labeling of rE2 and rE3BP (Fig. 7.8) under reducing conditions for the rE2/E3BP:rE3 complex, the ratio of E2:E3BP counts reflect a direct measure of the subunit composition of the rE2/E3BP core without any interference from rE3 labelling.

### **7.3.2.2 E2/E3BP core composition: investigation by mathematical modelling**

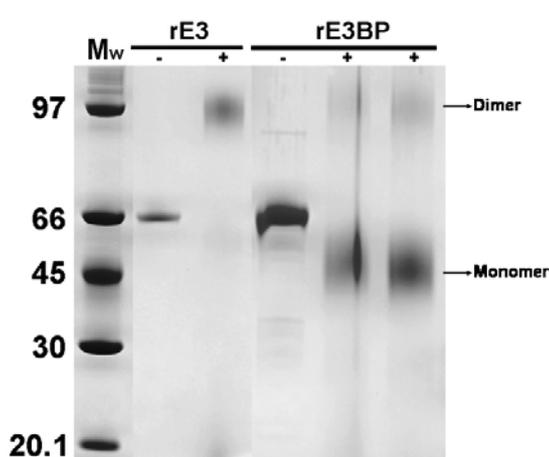
Mathematical modeling of the possible subunit organisations of the E2/E3BP core was carried out on the basis of a mixture of homotrimers (all E2 subunits) and heterotrimers (2E2+1E3BP) constituting the total of 20 trimers that formed the 60-meric core as listed in Table 7.5.

Two fundamental constraints were used in the modelling process; the first was that only one E3BP can be present per heterotrimer, and the second that E3BPs of neighbouring trimers can interact only with each other. The first constraint stems from the fact that E3BP is incapable of forming a core on its own and there has been no evidence to date to suggest incorporation of 2E3BPs per heterotrimer. In addition, work in our laboratory has shown that despite high levels of E3BP expression relative to E2, purification of rE2/E3BP always results in a core with E2 as the major component as judged by SDS-PAGE and radiolabelling. The justification for the second constraint is based on the idea that E3BP lacks the key residues involved in the hydrophobic ‘ball and socket’ connection that mediate the inter-trimer interactions within the core. Hence, as a direct interaction between E3BP of one trimer and E2 of a second trimer is not possible, it is likely that E3BP would interact weakly with itself to form a dimer on the core surface (Hiromasa, et al., 2004), although there is no experimental evidence to support this hypothesis.

Core organisation	No. of homotrimers	No. of heterotrimers
60E2 + 0E3BP	20	0
58E2 + 2E3BP	18	2
56E2 + 4E3BP	16	4
54E2 + 6E3BP	14	6
52E2 + 8E3BP	12	8
50E2 + 10E3BP	10	10
48E2 + 12E3BP	8	12
46E2 + 14E3BP	6	14
44E2 + 16E3BP	4	16
42E2 + 18E3BP	2	18
40E2 + 20E3BP	0	20

**Table 7.5** Various E2/E3BP core compositions and number of trimers

Interestingly, glutaraldehyde crosslinking studies on full-length rE3BP showed a small fraction of dimer, although the majority of the protein remained monomeric (Fig. 7.9). It is possible that in comparison to free rE3BP, geometric constraints on the surface of the core could induce higher levels of rE3BP dimerisation, resulting in the formation of a stable E2/E3BP core. This could imply that the interactions governing E3BP as part of the heterotrimer may be stronger than free E3BP, but relatively weaker than those of an E2 homotrimer.



**Figure 7.9** Chemical crosslinking of full-length rE3BP

SDS-PAGE of full-length rE3BP (-) chemically crosslinked with 2% glutaraldehyde (+) showed the presence of some rE3BP dimer. Homodimeric rE3 was used as a control. Molecular weights ( $M_w$ ) are shown in kDa.

Fig. 7.9 indicates a lower molecular weight of monomeric rE3BP in the presence of glutaraldehyde than free (or uncrosslinked) rE3BP. This is attributed to the formation of internal crosslinks within the rE3BP monomer by glutaraldehyde, resulting in a more globular (and less elongated or open) structure, thereby causing it to migrate more rapidly on SDS-PAGE.

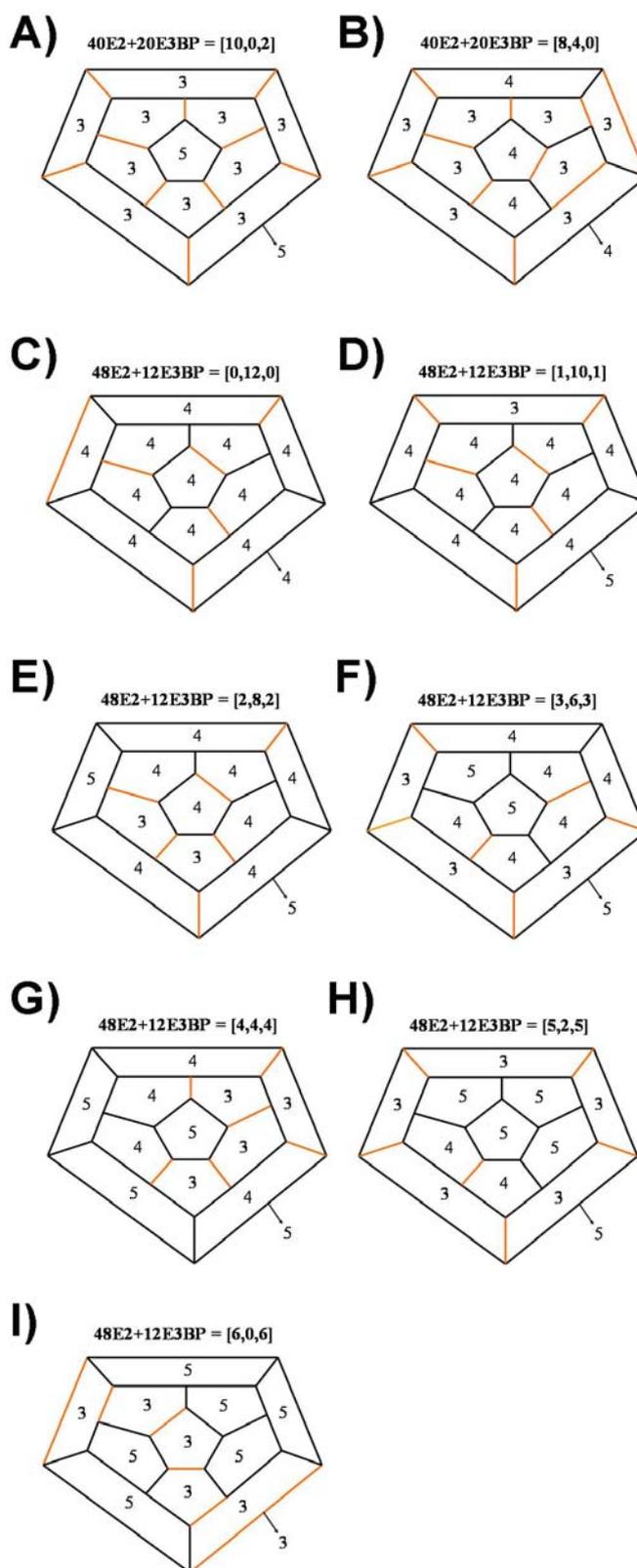
Mathematical modelling based on the above defined constraints yielded several solutions of core organisation for each model (see Table 7.5). Each of these solutions represents an arrangement of E2 and E3BP that satisfies a particular model of core composition.

Since they represent the most interesting cases with respect to this study, the different ways in which E2 and E3BP can be arranged to satisfy core compositions of 40E2+20E3BP and 48E2+12E3BP are illustrated in Fig. 7.10. Interestingly, while there are seven different solutions to obtain a complete 48E2+12E3BP core (Figs 7.10C to 7.10I), only two arrangements exist for the 40E2+20E3BP core (Figs 7.10A, 7.10B) model. The various arrangements are represented as [X,Y,Z], with X,Y and Z being the number of 3-edge, 4-edge and 5-edge E2 faces (see section 3.5.7 and 7.3.1.4 for details).

From analysis of mathematical modelling data, final states for various core models seem to follow a near Gaussian distribution as shown in Fig. 7.11. Very interestingly, while the 40E2+20E3BP model of core composition has only 2 final states (1 symmetric and 1 asymmetric), the maximum number of states peaks at the 48E2+12E3BP core model.

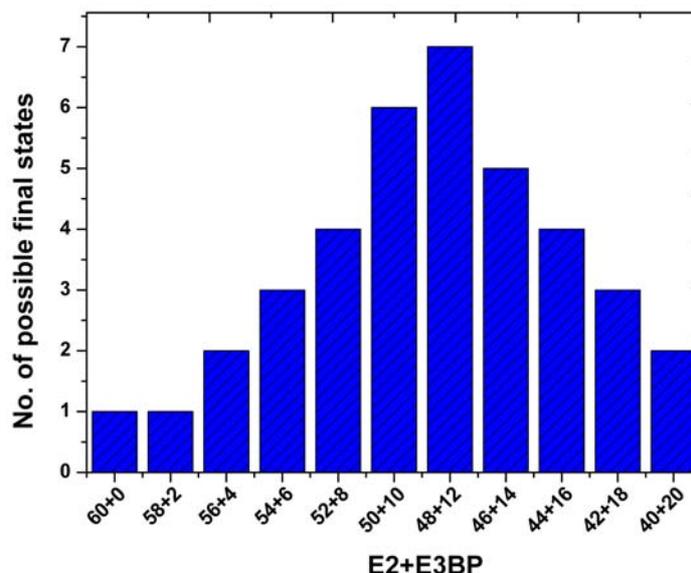
## 7.4 Discussion

The evidence presented in this chapter confirms the 2:1 binding stoichiometry of E2/E3BP:E3 on the surface of human PDC core. These findings are consistent with previous stoichiometry studies reported on the human E3BP-DD/E3 subcomplex (Smolle, et al., 2006), but are in direct contrast to the 1:1 stoichiometry observed in the crystal structures of bacterial E3/E2-SBD (Mande, et al., 1996) and human E3/E3BP-SBD (Ciszak, et al., 2006) subcomplexes. The crystal structure of the human E3/E3BP-SBD subcomplex indicates association of



**Figure 7.10 Mathematical modelling of E2/E3BP core organisation**

Various arrangements of E2 (black) and E3BP (orange) on the basis of 3-edge, 4-edge and 5-edge E2 faces,  $[X,Y,Z]$  for core models with **(A and B)**  $40E2+20E3BP$  and **(C to I)**  $48E2+12E3BP$  are shown. Integer values shown indicate the number of E2 faces (black), with the outermost face denoted by the arrow. The icosahedron schematic is drawn as a 2D diagram looking down along the 5-fold axes of symmetry.



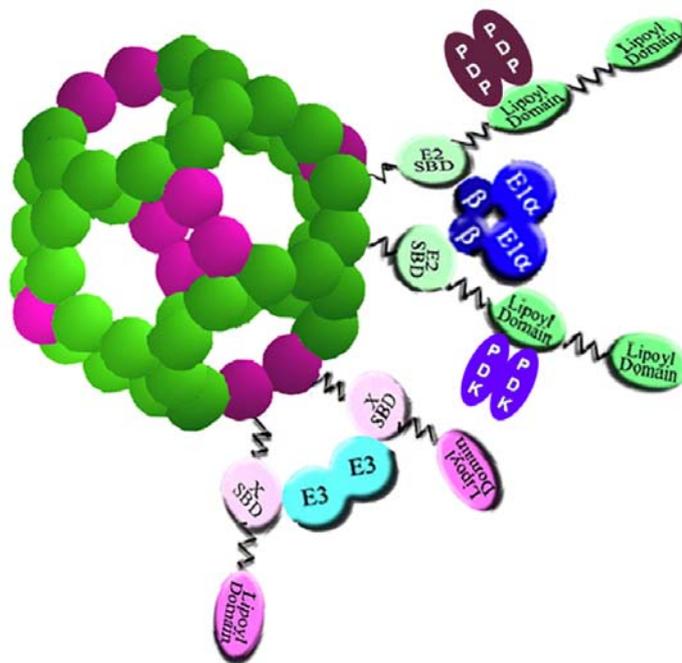
**Figure 7.11 Various core models and their number of final states**

The various final states (solutions) obtained by mathematical modelling for the various core models are shown. The trend follows a near Gaussian distribution, with a maximum at the 48E2+12E3BP core stoichiometry.

only a single SBD with E3, owing to steric hindrance of a loop in E3BP that prevents binding of a second E3BP. However as the results presented in this chapter were determined in solution, it is possible that the stoichiometry observed in the crystal structures may be a result of crystal packing effects and/or crystallisation buffer compositions. Moreover, it is possible that E3 binding could be significantly affected by geometric constraints on the surface of the E2/E3BP core; therefore the stoichiometry data reported in this chapter represent a more physiologically relevant system as compared to analysis of subunit interactions employing free E3 and monomeric di-domain constructs of E3BP.

The 2:1 binding stoichiometry of E3BP:E3 (and possibly E2:E1) in mammalian PDC suggests the possible presence of a network of ‘cross-bridges’ on the surface of the core (Fig. 7.12). Furthermore, the subunit composition of the core and its binding stoichiometry with the peripheral E1 and E3 components has significant implications for the efficient regulation of this massive complex. PDC regulation is mediated by PDK and PDP via phosphorylation and de-phosphorylation of three specific serine residues on E1. PDK phosphorylation of E1 renders the complex inactive, and only 1-3 molecules of PDK bound per complex have been reported (Yeaman, 1989). As these few molecules of PDK have to migrate around the

surface of the complex interacting with the entire population of bound E1 molecules, the presence of E1 and E3 cross-bridges on the core surface could greatly facilitate PDK migration via a hand-over hand mechanism (Liu, et al., 1995).



**Figure 7.12 Cross-bridge formation in human PDC**

Schematic representation of the cross-bridges formed between E2 (green) and E1 (blue), and E3BP (pink) and E3 (cyan) are shown. The substitution model of core organisation along with the PDC kinase (PDK, maroon) and PDC phosphatase (PDP, violet) are also depicted. Molecules are not drawn to scale.

Based on the results presented in this chapter, a new ‘variable E3BP substitution’ model of rE2/E3BP core organisation is also proposed, wherein the number of E3BPs within the core exceeds the average of 12 (in this study) and can theoretically range anywhere from 0 to a maximum of 20. With SANS data indicating an average binding of 10dE3s per rE2/E3BP core, and AUC and GFC stoichiometry studies indicating a 2:1 (rE2/E3BP:rE3) binding ratio, it can be inferred that the composition of the recombinant core is distinct from its native counterpart and comprises 40E2s and 20E3BPs at maximal occupancy. This suggests the integration of a higher number of E3BPs (>12) within the core (used in this study), contrary to the published 48E2+12E3BP substitution model of the human core (Hiromasa, et al., 2004). This proposition is further corroborated with [<sup>14</sup>C] radiolabelling data suggesting variation in core subunit composition

between bovine (favouring 48E2+12E3BP) and recombinant E2/E3BP (favouring 40E2+20E3BP) cores.

GFC purification of rE2/E3BP also yielded excess free E3BP in addition, indicating a maximum limit for integration of E3BP into the core (data not shown). However, as rE2/E3BP used in this study is produced recombinantly, over-expression of E2 and E3BP is driven to maximum capacity, in contrast to tightly coordinated expression of these subunits *in vivo*. As a consequence, it is suggested that there is a tendency to integrate higher numbers of E3BPs during the purification of rE2/E3BP, the maximum limit in this case being 20E3BPs. The justification for a maximum of 20E3BPs within the core (i.e. one E3BP per trimer) stems from the fact that E3BP is unable to form a 60-meric core on its own. Although it is theoretically possible to form trimers with 2E3BPs and 1E2, no experimental evidence for this has been reported to date. Additionally, formation of a core from this type of a heterotrimer would result in significantly large numbers of E3BP relative to E2. This would be non-optimal for efficient complex function, as E1 (that requires E2 as an anchoring site) catalyses the rate-limiting step in the overall PDC reaction.

Intriguingly, data from mathematical modelling of core organisation presented in this chapter indicates that the 48E2+12E3BP model provides the maximal flexibility in that it could potentially exist in seven possible final states differing in E3BP distribution over its surface. This is complemented by radiolabelling data which indicates that the native E2/E3BP assembly in bovine heart has a subunit composition of 48E2+12E3BP. It is likely that the relative expressions of E2 and E3BP are tightly regulated *in vivo* to generate cores only with an average capacity of about 12E3BPs per core. However, as previous EM studies on PDC cores indicate inherent heterogeneity of these massive machines (Stoops, et al., 1992; Wagenknecht, et al., 1991; Zhou, et al., 2001a), it is likely that a population of cores of several different subunit compositions are formed. Nonetheless, the probability of building complete E2/E3BP cores in the rapid folding times within the cell may be higher for the 48E2+12E3BP core model, as the maximum number of possible arrangements of the E2/E3BP core are permitted as shown in our modelling data.

Interestingly, the results in this chapter point towards an average bE2/E3BP core composition of 48E2+12E3BP, while the recombinant E2/E3BP used in this study can incorporate higher numbers of E3BP with an average core model biased towards 40E2+20E3BP. Hence, PDC may routinely comprise a range of cores of various compositions, rather than a single core with a uniquely defined composition. The variable incorporation of E3BPs resulting in different core compositions may serve to fine tune tissue-specific PDC regulation *in vivo*. The enzymes PDP and PDK control PDC regulation via E1 de-/phosphorylation. However, several isoenzymes (1-4) of PDK have been identified, and are observed to express in a tissue-specific manner displaying different specific activities (Bowker-Kinley, et al., 1998; Gudi, et al., 1995; Popov, et al., 1994; Rowles, et al., 1996). While PDK1-3 associate with the E2 inner LD, PDK4 is known to preferentially interact with the LD of E3BP (Roche, et al., 2003). Therefore, variation in the subunit composition of the core may have significant implications for the specific interactions of E2 and E3BP with these PDKs, altering the fine regulation of PDC in individual tissues in both normal and disease states.

## Chapter 8

### Conclusions

The 2-oxoacid dehydrogenase (OADC) family of multi-enzyme complexes has been extensively studied and characterised as it serves as a paradigm for analysis of protein-protein interactions, complex efficiency, enzyme cooperativity, active-site coupling and regulatory mechanisms. Despite having made considerable progress in our understanding of structural and functional aspects of these assemblies and their constituent enzymes, many questions still remain. A range of mutations, particularly in the E1 $\alpha$  component, have been identified in PDC alone. Nevertheless it has been difficult to establish a clear link between the clinical disease and complex dysfunction at a molecular level. Similarly, although several atomic structures of these constituent enzymes have been solved over the last two decades, it remains unclear as to how all the individual components come together to form these massive highly-regulated assemblies. Additionally, the various OADCs exhibit species- and complex-specific variations in their modes of interaction with constituent enzymes. For example in humans, while E1 specifically interacts with E2 in PDC, a direct association of E3 with E1 (and not E2) is observed in OGDC.

In particular, with reference to PDC, considerable effort has been directed to the characterisation of PDC from *B. stearothermophilus* and a limited number of other Gram-positive and Gram-negative bacteria. Analysis of eukaryotic PDC in contrast has progressed more slowly, although many studies over the past 30 years have revealed significant differences between the prokaryotic and eukaryotic assemblies, indicating the importance of structural evolution towards complex regulation and function of these massive machines. In addition, eukaryotic PDCs are distinct from their prokaryotic counterparts as they are transported into the mitochondria where they are assembled. However, with the discovery of E3BP as an additional subunit of eukaryotic PDC (De Marcucci and Lindsay, 1985; Jilka, et al., 1986) and the existence of specialised protein-protein interactions within the complex, determining the precise subunit organisation of the E2/E3BP core and the nature and effect of its binding with E1

and E3 on overall PDC organisation, function and regulation has been at the forefront of eukaryotic PDC research.

The research that is the subject of this thesis involves large-scale purification of individual PDC components from recombinant or native sources (Chapter 4) for subsequent use in biochemical and biophysical experiments. All proteins, with the exception of bE2/E3BP, were expressed as His-tagged products in order to aid purification. Purification of all proteins required two or three chromatography steps in order to obtain good yields of highly pure (>95%) samples, as judged by SDS-PAGE. In addition, new constructs encoding truncated CTD forms of E2 and E3BP were also designed and expressed using standard molecular cloning techniques (Chapter 4).

Detailed structural characterisation employing a diverse range of biophysical approaches was conducted on recombinant full-length rE2 and rE2/E3BP (Chapter 5) as well as the CTD tE2/tE3BP (Chapter 6) cores. Determination of the stoichiometry of binding between E2/E3BP and E3 and its consequence on E2/E3BP core subunit composition was also investigated (Chapter 7). AUC and SAXS data for rE2/E3BP support the substitution model of core organisation. Solution structures of both recombinant (rE2/E3BP) and bovine (bE2/E3BP) cores determined by *ab initio* modelling of SAXS data indicate icosahedral hollow core structures with empty faces, reflecting their high structural similarity. In addition, their dodecahedral morphology was also confirmed by negative stain EM of rE2/E3BP, bE2/E3BP and rE2 cores. The maximum diameters ( $D_{max}$ ) of rE2/E3BP and bE2/E3BP obtained from SAXS are slightly greater than previously reported values, and probably reflect the inherent size variation (heterogeneity) that is symbolic of the ‘breathing’ of these massive complexes.

The truncated E2/E3BP (tLi19/tLi30) core comprising only the CTD was designed, expressed, purified and characterised analogous to full-length rE2/E3BP (Chapter 6). Interestingly, a sedimentation coefficient similar to that for full-length rE2/E3BP was obtained from AUC studies of tE2/tE3BP. Furthermore, while the SANS solution structure and negative stain EM of tE2/tE3BP confirmed the substitution model of core organisation, homology modelling with the tE2 of *B. stearothermophilus* (Izard, et al., 1999) revealed a highly conserved icosahedral framework. Initial crystallisation trials on tE2/E3BP proved

encouraging although further work is required to produce high quality diffracting protein crystals to examine the arrangement of E2/E3BP subunits within the central core assembly.

In the case of the rE2/E3BP:rE3 complex, data obtained from AUC and GFC confirmed a 2:1 stoichiometry (Chapter 7). This is in agreement with previously published results from solution studies of the E3BP-DD:E3 sub-complex (Smolle, et al., 2006), and in contrast to the 1:1 stoichiometry observed in the crystal structure of the E3BP-XSBD:E3 sub-complex (Brautigam, et al., 2006; Ciszak, et al., 2006). However, given that the crystal structure involves a truncated construct of E3BP, it is possible that complex formation with E3 is mediated by different geometric constraints as opposed to those on the surface of the E2/E3BP core. Hence stoichiometry data obtained for the assembled E2/E3BP:E3 complex are likely to be more biologically relevant.

Interestingly, examination of a series of purified rE2/E3BP:dE3 complexes by SANS yielded 10E3 dimers bound per rE2/E3BP core at maximal occupancy, possibly indicative of a 1:1 stoichiometry. However, the 2:1 binding stoichiometry of E2/E3BP:E3 obtained using AUC and GFC along with subsequent biochemical [ $^{14}\text{C}$ ] radiolabelling studies indicated marked differences in core composition between the recombinant and bovine E2/E3BP assemblies (Chapter 7). As a result of these studies and the quantitation of E3 binding sites by SANS it is proposed that recombinant core has an overall composition approaching 40E2+20E3BP, in contrast to the 48E2+12E3BP substitution model proposed by Hiromasa and colleagues (2004). Thus evidence for variation in the composition between recombinant and bovine PDC cores has led us to propose a 'variable E3BP substitution model', wherein the number of E3BPs within the core can range from 0 to 20 depending on the availability of E3BP. While the rE2/E3BP used in this study possesses an increased number of E3BPs, approaching an average core composition of 40E2+20E3BP, the native bE2/E3BP assembly is estimated to have a stoichiometry of 48E2+12E3BP. This difference may be attributable to high levels of over-expression of E2 and E3BP in the recombinant system that drives maximum integration of E3BP, as opposed to conditions within the cell. Interestingly, mathematical models of core organisation reveal that 48E2+12E3BP cores could potentially exist in the highest number of final states differing in the distribution of E3BP around the icosahedron (Chapter 7),

further confirming the hypothesis that relative levels of expression of E2 and E3BP are tightly regulated *in vivo* (as opposed to the recombinant system), possibly in order to maximise formation of complete cores with optimal catalytic efficiency.

As the icosahedral cores are composed of trimeric building blocks, an all E2 core (e.g. in *B. stearothermophilus*) is assembled from 20 E2 trimers. However, a substitution model of E2/E3BP core organisation entails the incorporation of heterotrimers (2E2+1E3BP), their number depending on the number of E3BPs within the core. The crystal structure of the truncated E2 core indicates strong inter-E2 trimer contacts being responsible for the stability of the overall icosahedral framework (Izard, et al., 1999). Moreover, changes in these inter-trimer distances are proposed to be involved in the 'breathing' of these cores (Zhou, et al., 2001a). Analysis of the assessment of core stability of rE2/E3BP by unfolding studies using CD and fluorescence revealed interesting results with rE2/E3BP being less stable than rE2, indicative of a contribution of E3BP towards this destabilisation. Additionally, comparative analysis of recombinant full-length and truncated E2/E3BP CD profiles pointed towards a significant contribution by the N-terminal flanking arms (comprising LDs, SBD and flexible linker regions) towards core destabilisation. To our knowledge, this is the first report demonstrating the influence of E3BP integration on overall core stability. Analysis of intrinsic tryptophan fluorescence during GdmCl-induced unfolding also indicated dissociation of the overall quaternary structure into possible trimers or monomers prior to the main unfolding event, complementing previous unfolding studies carried out on bE2/E3BP (McCartney, et al., 1997).

The presence of a 2:1 stoichiometric arrangement on the core surface involving E2/E3BP interactions with E3 and possibly also E1 suggests the existence of 'cross-bridges' that would precisely position the peripherally bound E1 and E3. Such an arrangement along with the empty pentagonal faces of E2/E3BP could facilitate interaction of the lipoyl swinging arms of E2 and E3BP with the active sites during catalysis. In addition to the LD of E3BP being active, it can interact individually with both E1 and E3 (Singh, 2008). This suggests a possible shared mechanism of interaction between the LDs, thus not requiring both E2- and E3BP-LDs to visit all active sites. It is also likely that the cross-bridge formation

would introduce constraints on the mobility of the N-terminal arms that may be necessary for efficient active site coupling and enhanced rates of catalysis.

As PDC is regulated by 1-3 molecules of PDK and PDP, the cross bridge formation is proposed to enhance the migration of PDKs over the core surface via a suggested hand-over-hand mechanism (Liu, et al., 1995). Our 'variable E3BP substitution model' proposes the existence of a variable number of E3BPs ( $\geq 0$  and  $\leq 20$ ) within the E2/E3BP core. The various isoforms of PDK that regulate PDC are known to express in a tissue specific manner with differing specific activities (Bowker-Kinley, et al., 1998; Gudi, et al., 1995). While PDK1-3 associate with E2-LD (inner or outer), PDK4 specifically interacts with E3BP-LD (Roche, et al., 2003). Hence, variation in core composition could have significant implications on these PDK interactions, possibly serving to fine tune PDC regulation in a tissue-specific manner.

A number of papers have been published on eukaryotic PDC core organisation over the last 20 years, with the aim of determining the exact locations of E3BP within the core. However, this is a challenging task and requires novel approaches to this problem. Furthermore, inherent heterogeneity of the core may pose problems for protein crystallisation. Although considerable progress in human PDC research has been made by our determination of the first ever SAXS solution structure of E2/E3BP core, the E2 subunits cannot be distinguished from E3BP within the core framework. However, SANS contrast variation, single molecule studies or cryo-EM of the E2/E3BP:E3 complex present viable alternatives for future experiments. These techniques would not only aid determination of the exact positions of E3BP within the core, but will further confirm our key finding of core variation implying that the PDC represents a population of assemblies containing a range of cores with different subunit compositions rather than a unique assembly with a defined composition. In addition, these techniques will confirm the hypothesis that the spatial organisation of PDC plays a crucial role in its efficient function and regulation.

## Appendix

Amino acid sequences of the cloned truncated constructs of E2 and E3BP are given below. The His-tag is denoted in blue.

### Truncated E2 - tLi19 in pET11b

MAPAPAAVVPPTGPGMAPVPTGVFTDIPISNIRRVIAQRLMQSKQTIPHYLSIDVNMGEVLLVR  
 KELNKILEGRSKISVNDFIIKASALACLKVPEANSSWMDTVIRQNHVVDVSVAVSTPAGLITPIVF  
 NAHIKGVETIANDVVSLATKAREGKLQPHEFQGGTFTISNLGMFGIKNFSAIINPPQACILAIGAS  
 EDKLV PADNEKGF DVASMM SVTLSCDHRVVDGAVGAQWLAEFRKYLEKPTTMLL

### Truncated E2 - tLi27 in pET14b

MGSSHHHHHSSGLVPRGSHMDSFVPSKVAPAPAAVVPPTGPGMAPVPTGVFTDIPISNIRRV  
 IAQRLMQSKQTIPHYLSIDVNMGEVLLVRKELNKILEGRSKISVNDFIIKASALACLKVPEANSS  
 WMDTVIRQNHVVDVSVAVSTPAGLITPIVFNNAHIKGVETIANDVVSLATKAREGKLQPHEFQGG  
 TFTISNLGMFGIKNFSAIINPPQACILAIGASEDKLV PADNEKGF DVASMM SVTLSCDHRVVDGA  
 VGAQWLAEFRKYLEKPITMLL

### Truncated E3BP - tLi30 in pET28b

MGSSHHHHHSSGLVPRGSHMASMTGGQQMGRDPSPLQATSGPSYPRPVIPPVSTPGQPNA  
 VGTFTTEIPASNIRRVIAKRLTESKSTVPHAYATADCDLGAVLKVRQDLVKDDIKVSVNDFIIKAAA  
 VTLKQMPDVNVSWDGE GPKQLPFIDISVAVATDKGLLTPIIKDAAAKGIQEIADSVKALS K KARD  
 GKLLPEEYQGG SFSISNLGMFGIDEFTAVINPPQACILAVGRFRPV LKLTEDEEGNAKLQQRQLI  
 TVTMSSDSRVVDDELATRFLKSFKANLENPIRLA

## Bibliography

Adams, P.D., Chen, Y., Ma, K., Zagorski, M.G., Sonnichsen, F.D., McLaughlin, M.L. and Barkley, M.D. (2002) Intramolecular quenching of tryptophan fluorescence by the peptide bond in cyclic hexapeptides, *J Am Chem Soc*, **124**, 9278-9286.

Allen, A.G. and Perham, R.N. (1991) Two lipoyl domains in the dihydrolipoamide acetyltransferase chain of the pyruvate dehydrogenase multi-enzyme complex of *Streptococcus faecalis*, *FEBS Lett*, **287**, 206-210.

Allen, A.G., Perham, R.N., Allison, N., Miles, J.S. and Guest, J.R. (1989) Reductive acetylation of tandemly repeated lipoyl domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli* is random order, *J Mol Biol*, **208**, 623-633.

Allen, M.D., Broadhurst, R.W., Solomon, R.G. and Perham, R.N. (2005) Interaction of the E2 and E3 components of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. Use of a truncated protein domain in NMR spectroscopy, *FEBS J*, **272**, 259-268.

Aral, B., Benelli, C., Ait-Ghezala, G., Amessou, M., Fouque, F., Maunoury, C., Creau, N., Kamoun, P. and Marsac, C. (1997) Mutations in *PDX1*, the human lipoyl-containing component X of the pyruvate dehydrogenase-complex gene on chromosome 11p1, in congenital lactic acidosis, *Am J Hum Genet*, **61**, 1318-1326.

Arjunan, P., Nemeria, N., Brunskill, A., Chandrasekhar, K., Sax, M., Yan, Y., Jordan, F., Guest, J.R. and Furey, W. (2002) Structure of the pyruvate dehydrogenase multienzyme complex E1 component from *Escherichia coli* at 1.85 Å resolution, *Biochemistry*, **41**, 5213-5221.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics*, **22**, 195-201.

Bao, H., Kasten, S.A., Yan, X., Hiromasa, Y. and Roche, T.E. (2004a) Pyruvate dehydrogenase kinase isoform 2 activity stimulated by speeding up the rate of dissociation of ADP, *Biochemistry*, **43**, 13442-13451.

Bao, H., Kasten, S.A., Yan, X. and Roche, T.E. (2004b) Pyruvate dehydrogenase kinase isoform 2 activity limited and further inhibited by slowing down the rate of dissociation of ADP, *Biochemistry*, **43**, 13432-13441.

Beechem, J.M. (1992) Global analysis of biochemical and biophysical data, *Methods Enzymol*, **210**, 37-54.

Behal, R.H., Browning, K.S., Hall, T.B. and Reed, L.J. (1989) Cloning and nucleotide sequence of the gene for protein X from *Saccharomyces cerevisiae*, *Proc Natl Acad Sci U S A*, **86**, 8732-8736.

Behal, R.H., DeBuysere, M.S., Demeler, B., Hansen, J.C. and Olson, M.S. (1994) Pyruvate dehydrogenase multienzyme complex. Characterization of assembly intermediates by sedimentation velocity analysis, *J Biol Chem*, **269**, 31372-31377.

Behlke, J. and Ristau, O. (1997) Molecular mass determination by sedimentation velocity experiments and direct fitting of the concentration profiles, *Biophys J*, **72**, 428-434.

Berg, A., de Kok, A. and Vervoort, J. (1994) Sequential <sup>1</sup>H and <sup>15</sup>N nuclear magnetic resonance assignments and secondary structure of the N-terminal lipoyl domain of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*, *Eur J Biochem*, **221**, 87-100.

Berg, A., Smits, O., de Kok, A. and Vervoort, J. (1995) Sequential <sup>1</sup>H and <sup>15</sup>N nuclear magnetic resonance assignments and secondary structure of the lipoyl domain of the 2-oxoglutarate dehydrogenase complex from *Azotobacter vinelandii*. Evidence for high structural similarity with the lipoyl domain of the pyruvate dehydrogenase complex, *Eur J Biochem*, **234**, 148-159.

Berg, A., Vervoort, J. and de Kok, A. (1996) Solution structure of the lipoyl domain of the 2-oxoglutarate dehydrogenase complex from *Azotobacter vinelandii*, *J Mol Biol*, **261**, 432-442.

Berg, A., Vervoort, J. and de Kok, A. (1997) Three-dimensional structure in solution of the N-terminal lipoyl domain of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*, *Eur J Biochem*, **244**, 352-360.

Berg, A., Westphal, A.H., Bosma, H.J. and de Kok, A. (1998) Kinetics and specificity of reductive acylation of wild-type and mutated lipoyl domains of 2-oxo-acid dehydrogenase complexes from *Azotobacter vinelandii*, *Eur J Biochem*, **252**, 45-50.

Berova, N., Nakanishi, K. and Woody, R.W. (2000) *Circular Dichroism: Principles and Applications*, 2nd ed. Wiley-VCH, New York, pp 877.

Berry, A., Scrutton, N.S. and Perham, R.N. (1989) Switching kinetic mechanism and putative proton donor by directed mutagenesis of glutathione reductase, *Biochemistry*, **28**, 1264-1269.

Blass, J.P. and Gibson, G.E. (1991) The role of oxidative abnormalities in the pathophysiology of Alzheimer's disease, *Rev Neurol (Paris)*, **147**, 513-525.

Blass, J.P. and Gibson, G.E. (1999) Cerebrometabolic aspects of delirium in relationship to dementia, *Dement Geriatr Cogn Disord*, **10**, 335-338.

Bleile, D.M., Munk, P., Oliver, R.M. and Reed, L.J. (1979) Subunit structure of dihydrolipoyl transacetylase component of pyruvate dehydrogenase complex from *Escherichia coli*, *Proc Natl Acad Sci U S A*, **76**, 4385-4389.

Boden, G. (2001) Pathogenesis of type 2 diabetes. Insulin resistance, *Endocrinol Metab Clin North Am*, **30**, 801-815, v.

Bollag, D.M., Rozycki, M.D. and Edelstein, S.J. (1996a) Concentrating protein solutions. In *Protein Methods*. Wiley-Liss, New York, pp. 83-106.

Bollag, D.M., Rozycki, M.D. and Edelstein, S.J. (1996b) Gel electrophoresis under nondenaturing conditions. In *Protein Methods*. Wiley-Liss, New York, pp. 155-172.

Borges, A., Hawkins, C.F., Packman, L.C. and Perham, R.N. (1990) Cloning and sequence analysis of the genes encoding the dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase components of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*, *Eur J Biochem*, **194**, 95-102.

Borgnia, M.J., Shi, D., Zhang, P. and Milne, J.L. (2004) Visualization of alpha-helical features in a density map constructed using 9 molecular images of the 1.8 MDa icosahedral core of pyruvate dehydrogenase, *J Struct Biol*, **147**, 136-145.

Boulatnikov, I. and Popov, K.M. (2003) Formation of functional heterodimers by isozymes 1 and 2 of pyruvate dehydrogenase kinase, *Biochim Biophys Acta*, **1645**, 183-192.

Boulin, C.J., Kempf, R., Gabriel, A. and Koch, M.H.J. (1988) Data acquisition systems for linear and area X-ray detectors using delay-line readout, *Nucl Instr and Meth (A)*, **269**, 312-320.

Bowker-Kinley, M.M., Davis, W.I., Wu, P., Harris, R.A. and Popov, K.M. (1998) Evidence for existence of tissue-specific regulation of the mammalian pyruvate dehydrogenase complex, *Biochem J*, **329 ( Pt 1)**, 191-196.

Brautigam, C.A., Chuang, J.L., Tomchick, D.R., Machius, M. and Chuang, D.T. (2005) Crystal structure of human dihydrolipoamide dehydrogenase: NAD<sup>+</sup>/NADH binding and the structural basis of disease-causing mutations, *J Mol Biol*, **350**, 543-552.

Brautigam, C.A., Wynn, R.M., Chuang, J.L., Machius, M., Tomchick, D.R. and Chuang, D.T. (2006) Structural insight into interactions between dihydrolipoamide dehydrogenase (E3) and E3 binding protein of human pyruvate dehydrogenase complex, *Structure*, **14**, 611-621.

Brown, A.E. (2002a) Towards a recombinant model of human pyruvate dehydrogenase complex, University of Glasgow, PhD thesis.

Brown, G.K., Brown, R.M., Scholem, R.D., Kirby, D.M. and Dahl, H.H. (1989a) The clinical and biochemical spectrum of human pyruvate dehydrogenase complex deficiency, *Ann N Y Acad Sci*, **573**, 360-368.

- Brown, G.K., Otero, L.J., LeGris, M. and Brown, R.M. (1994) Pyruvate dehydrogenase deficiency, *J Med Genet*, **31**, 875-879.
- Brown, R.M., Dahl, H.H. and Brown, G.K. (1989b) X-chromosome localization of the functional gene for the E1 alpha subunit of the human pyruvate dehydrogenase complex, *Genomics*, **4**, 174-181.
- Brown, R.M., Head, R.A., Boubriak, I.I., Leonard, J.V., Thomas, N.H. and Brown, G.K. (2004) Mutations in the gene for the E1 $\beta$  subunit: a novel cause of pyruvate dehydrogenase deficiency, *Hum Genet*, **115**, 123-127.
- Brown, R.M., Head, R.A. and Brown, G.K. (2002b) Pyruvate dehydrogenase E3 binding protein deficiency, *Hum Genet*, **110**, 187-191.
- Bubber, P., Haroutunian, V., Fisch, G., Blass, J.P. and Gibson, G.E. (2005) Mitochondrial abnormalities in Alzheimer brain: mechanistic implications, *Ann Neurol*, **57**, 695-703.
- Burns, G., Sykes, P.J., Hatter, K. and Sokatch, J.R. (1989) Isolation of a third lipoamide dehydrogenase from *Pseudomonas putida*, *J Bacteriol*, **171**, 665-668.
- Byrd, D.J., Krohn, H.P., Winkler, L., Steinborn, C., Hadam, M., Brodehl, J. and Hunneman, D.H. (1989) Neonatal pyruvate dehydrogenase deficiency with lipoate responsive lactic acidemia and hyperammonaemia, *Eur J Pediatr*, **148**, 543-547.
- Byron, O. (1997) Construction of hydrodynamic bead models from high-resolution X-ray crystallographic or nuclear magnetic resonance data, *Biophys J*, **72**, 408-415.
- Byron, O. (2008) Hydrodynamic modeling: the solution conformation of macromolecules and their complexes, *Methods Cell Biol*, **84**, 327-373.
- Byron, O. and Gilbert, R.J. (2000) Neutron scattering: good news for biotechnology, *Curr Opin Biotechnol*, **11**, 72-80.
- Carothers, D.J., Pons, G. and Patel, M.S. (1989) Dihydrolipoamide dehydrogenase: functional similarities and divergent evolution of the pyridine nucleotide-disulfide oxidoreductases, *Arch Biochem Biophys*, **268**, 409-425.

Cate, R.L., Roche, T.E. and Davis, L.C. (1980) Rapid intersite transfer of acetyl groups and movement of pyruvate dehydrogenase component in the kidney pyruvate dehydrogenase complex, *J Biol Chem*, **255**, 7556-7562.

Chacón, P., Diaz, J.F., Moran, F. and Andreu, J.M. (2000) Reconstruction of protein form with X-ray solution scattering and a genetic algorithm, *J Mol Biol*, **299**, 1289-1302.

Chacón, P., Moran, F., Diaz, J.F., Pantos, E. and Andreu, J.M. (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm, *Biophys J*, **74**, 2760-2775.

Chamberlain, J.P. (1979) Fluorographic detection of radioactivity in polyacrylamide gels with the water-soluble fluor, sodium salicylate, *Anal Biochem*, **98**, 132-135.

Chatake, T., Ostermann, A., Kurihara, K., Parak, F.G. and Niimura, N. (2003) Hydration in proteins observed by high-resolution neutron crystallography, *Proteins*, **50**, 516-523.

Chen, C.C., Meng, H.C., Chang, F.Y., Yang, W.G., Ng, W.W. and Lee, S.D. (1996) Characteristics of endoscopic retrograde cholangiopancreatogram in patients with primary biliary cirrhosis, *Zhonghua Yi Xue Za Zhi (Taipei)*, **57**, 254-259.

Chen, W., Huang, X., Komuniecki, P.R. and Komuniecki, R. (1998) Molecular cloning, functional expression, and characterization of pyruvate dehydrogenase kinase from anaerobic muscle of the parasitic nematode *Ascaris suum*, *Arch Biochem Biophys*, **353**, 181-189.

Chen, W., Komuniecki, P.R. and Komuniecki, R. (1999) Nematode pyruvate dehydrogenase kinases: role of the C-terminus in binding to the dihydrolipoyl transacetylase core of the pyruvate dehydrogenase complex, *Biochem J*, **339**, 103-109.

Chen, Y. and Barkley, M.D. (1998) Toward understanding tryptophan fluorescence in proteins, *Biochemistry*, **37**, 9976-9982.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. (2003) Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Res*, **31**, 3497-3500.

Ciszak, E., Korotchkina, L.G., Hong, Y.S., Joachimiak, A. and Patel, M.S. (2001) Crystallization and initial X-ray diffraction analysis of human pyruvate dehydrogenase, *Acta Crystallogr D Biol Crystallogr*, **57**, 465-468.

Ciszak, E.M., Korotchkina, L.G., Dominiak, P.M., Sidhu, S. and Patel, M.S. (2003) Structural basis for flip-flop action of thiamin pyrophosphate-dependent enzymes revealed by human pyruvate dehydrogenase, *J Biol Chem*, **278**, 21240-21246.

Ciszak, E.M., Makal, A., Hong, Y.S., Vettaikorumakankauv, A.K., Korotchkina, L.G. and Patel, M.S. (2006) How dihydrolipoamide dehydrogenase-binding protein binds dihydrolipoamide dehydrogenase in the human pyruvate dehydrogenase complex, *J Biol Chem*, **281**, 648-655.

Claiborn, A., Ross, R.P., Ward, D., Parsonage, D. and Crane III Jr, E.J. (1994). In *Flavins and Flavoproteins*. Yagi, K. (ed), Walter de Gruyter, Berlin, pp 587-596.

Coppel, R.L., McNeilage, L.J., Surh, C.D., Van de Water, J., Spithill, T.W., Whittingham, S. and Gershwin, M.E. (1988) Primary structure of the human M2 mitochondrial autoantigen of primary biliary cirrhosis: dihydrolipoamide acetyltransferase, *Proc Natl Acad Sci U S A*, **85**, 7317-7321.

Craft, S., Asthana, S., Schellenberg, G., Baker, L., Cherrier, M., Boyt, A.A., Martins, R.N., Raskind, M., Peskind, E. and Plymate, S. (2000) Insulin effects on glucose metabolism, memory, and plasma amyloid precursor protein in Alzheimer's disease differ according to apolipoprotein-E genotype, *Ann N Y Acad Sci*, **903**, 222-228.

Dahl, H.H. (1995) Pyruvate dehydrogenase E1 alpha deficiency: males and females differ yet again, *Am J Hum Genet*, **56**, 553-557.

Dahl, H.H., Brown, R.M., Hutchison, W.M., Maragos, C. and Brown, G.K. (1990) A testis-specific form of the human pyruvate dehydrogenase E1 alpha subunit is coded for by an intronless gene on chromosome 4, *Genomics*, **8**, 225-232.

Dahl, H.H., Hansen, L.L., Brown, R.M., Danks, D.M., Rogers, J.G. and Brown, G.K. (1992) X-linked pyruvate dehydrogenase E1 alpha subunit deficiency in heterozygous females: variable manifestation of the same mutation, *J Inherit Metab Dis*, **15**, 835-847.

Dale, G.E., Oefner, C. and D'Arcy, A. (2003) The protein as a variable in protein crystallization, *J Struct Biol*, **142**, 88-97.

Damuni, Z. and Reed, L.J. (1987) Purification and characterization of a divalent cation-independent, spermine-stimulated protein phosphatase from bovine kidney mitochondria, *J Biol Chem*, **262**, 5133-5138.

Danson, M.J., Fersht, A.R. and Perham, R.N. (1978) Rapid intramolecular coupling of active sites in the pyruvate dehydrogenase complex of *Escherichia coli*: mechanism for rate enhancement in a multimeric structure, *Proc Natl Acad Sci U S A*, **75**, 5386-5390.

Dardel, F., Davis, A.L., Laue, E.D. and Perham, R.N. (1993) Three-dimensional structure of the lipoyl domain from *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex, *J Mol Biol*, **229**, 1037-1048.

Dardel, F., Laue, E.D. and Perham, R.N. (1991) Sequence-specific <sup>1</sup>H-NMR assignments and secondary structure of the lipoyl domain of the *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex, *Eur J Biochem*, **201**, 203-209.

De Marcucci, O. and Lindsay, J.G. (1985) Component X. An immunologically distinct polypeptide associated with mammalian pyruvate dehydrogenase multienzyme complex, *Eur J Biochem*, **149**, 641-648.

De Marcucci, O.G., Hodgson, J.A. and Lindsay, J.G. (1986) The Mr-50 000 polypeptide of mammalian pyruvate dehydrogenase complex participates in the acetylation reactions, *Eur J Biochem*, **158**, 587-594.

De Meirleir, L., Lissens, W., Benelli, C., Marsac, C., De Klerk, J., Scholte, J., van Diggelen, O., Kleijer, W., Seneca, S. and Liebaers, I. (1998) Pyruvate dehydrogenase complex deficiency and absence of subunit X, *J Inherit Metab Dis*, **21**, 9-16.

De Vivo, D.C. (1998) Complexities of the pyruvate dehydrogenase complex, *Neurology*, **51**, 1247-1249.

Demeler, B. (2004) UltraScan Software, University of Texas Health Science Center at San Antonio, Department of Biochemistry, [www/ultrascan.uthscsa.edu](http://www/ultrascan.uthscsa.edu).

Demeler, B. (2005) UltraScan - A comprehensive data analysis software package for analytical ultracentrifugation experiments. In *Modern Analytical Ultracentrifugation: Techniques and Methods*. Scott, D.J., Harding, S.E. and Rowe, A.J. (eds), Royal Society of Chemistry, UK, pp 210-229.

Denton, R.M., Randle, P.J. and Martin, B.R. (1972) Stimulation by calcium ions of pyruvate dehydrogenase phosphate phosphatase, *Biochem J*, **128**, 161-163.

Deonarain, M.P., Berry, A., Scrutton, N.S. and Perham, R.N. (1989) Alternative proton donors/acceptors in the catalytic mechanism of the glutathione reductase of *Escherichia coli*: the role of histidine-439 and tyrosine-99, *Biochemistry*, **28**, 9602-9607.

Dey, R., Aral, B., Abitbol, M. and Marsac, C. (2002) Pyruvate dehydrogenase deficiency as a result of splice-site mutations in the *PDX1* gene, *Mol Genet Metab*, **76**, 344-347.

Dey, R., Mine, M., Desguerre, I., Slama, A., Van Den Berghe, L., Brivet, M., Aral, B. and Marsac, C. (2003) A new case of pyruvate dehydrogenase deficiency due to a novel mutation in the *PDX1* gene, *Ann Neurol*, **53**, 273-277.

Di Mauro, S. and De Vivo, D.C. (1996) Genetic heterogeneity in Leigh syndrome, *Ann Neurol*, **40**, 5-7.

Dobson, C.M. (2004) Experimental investigation of protein folding and misfolding, *Methods*, **34**, 4-14.

Domingo, G.J., Chauhan, H.J., Lessard, I.A., Fuller, C. and Perham, R.N. (1999) Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex from *Bacillus stearothermophilus*, *Eur J Biochem*, **266**, 1136-1146.

Engelman, D.M. and Moore, P.B. (1972) A new method for the determination of biological quaternary structure by neutron scattering, *Proc Natl Acad Sci U S A*, **69**, 1997-1999.

Feigin, L.A. and Svergun, D. (1987) *Structure Analysis by Small Angle X-ray Scattering and Neutron Scattering*. Plenum Press, New York.

Fiedler, E., Thorell, S., Sandalova, T., Golbik, R., Konig, S. and Schneider, G. (2002) Snapshot of a key intermediate in enzymatic thiamin catalysis: crystal structure of the alpha-carbanion of (alpha,beta-dihydroxyethyl)-thiamin diphosphate in the active site of transketolase from *Saccharomyces cerevisiae*, *Proc Natl Acad Sci U S A*, **99**, 591-595.

Fouque, F., Marsac, C. and Benelli, C. (1998) the pyruvate dehydrogenase complex: from molecular organisation to clinical issues *Med Sci*, **14**, 1366-1374.

Frank, R.A., Pratap, J.V., Pei, X.Y., Perham, R.N. and Luisi, B.F. (2005) The molecular origins of specificity in the assembly of a multienzyme complex, *Structure*, **13**, 1119-1130.

Frank, R.A., Titman, C.M., Pratap, J.V., Luisi, B.F. and Perham, R.N. (2004) A molecular switch and proton wire synchronize the active sites in thiamine enzymes, *Science*, **306**, 872-876.

Fujita, H. (1975) *Sedimentation Equilibrium in Chemically Reacting Systems*. J. Wiley and Sons, Ltd, New York, Volume 42, Chapter 6.

Fussey, S.P., Ali, S.T., Guest, J.R., James, O.F., Bassendine, M.F. and Yeaman, S.J. (1990) Reactivity of primary biliary cirrhosis sera with *Escherichia coli* dihydrolipoamide acetyltransferase (E2p): characterization of the main immunogenic region, *Proc Natl Acad Sci U S A*, **87**, 3987-3991.

Fussey, S.P., Bassendine, M.F., Fittes, D., Turner, I.B., James, O.F. and Yeaman, S.J. (1989) The E1  $\alpha$  and  $\beta$  subunits of the pyruvate dehydrogenase complex are M2'd' and M2'e' autoantigens in primary biliary cirrhosis, *Clin Sci (Lond)*, **77**, 365-368.

Fussey, S.P., Guest, J.R., James, O.F., Bassendine, M.F. and Yeaman, S.J. (1988) Identification and analysis of the major M2 autoantigens in primary biliary cirrhosis, *Proc Natl Acad Sci U S A*, **85**, 8654-8658.

Fussey, S.P., Lindsay, J.G., Fuller, C., Perham, R.N., Dale, S., James, O.F., Bassendine, M.F. and Yeaman, S.J. (1991) Autoantibodies in primary biliary cirrhosis: analysis of reactivity against eukaryotic and prokaryotic 2-oxo acid dehydrogenase complexes, *Hepatology*, **13**, 467-474.

Gabuzda, D., Busciglio, J., Chen, L.B., Matsudaira, P. and Yankner, B.A. (1994) Inhibition of energy metabolism alters the processing of amyloid precursor protein and induces a potentially amyloidogenic derivative, *J Biol Chem*, **269**, 13623-13628.

García de la Torre, J. (2001) Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins, *Biophys Chem*, **93**, 159-170.

García de la Torre, J., Echenique Gdel, R. and Ortega, A. (2007) Improved calculation of rotational diffusion and intrinsic viscosity of bead models for macromolecules and nanoparticles, *J Phys Chem B*, **111**, 955-961.

García de la Torre, J., Navarro, S., Lopez Martinez, M.C., Diaz, F.G. and Lopez Cascales, J.J. (1994) HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules, *Biophys J*, **67**, 530-531.

Geoffroy, V., Fouque, F., Benelli, C., Poggi, F., Saudubray, J.M., Lissens, W., Meirleir, L.D., Marsac, C., Lindsay, J.G. and Sanderson, S.J. (1996) Defect in the X-lipoyl-containing component of the pyruvate dehydrogenase complex in a patient with neonatal lactic acidemia, *Pediatrics*, **97**, 267-272.

Gershwin, M.E. and Mackay, I.R. (1991) Primary biliary cirrhosis: paradigm or paradox for autoimmunity, *Gastroenterology*, **100**, 822-833.

Gershwin, M.E., Migliaccio, C.T., van De Water, J. and Coppel, R.L. (1998) Molecular considerations of primary biliary cirrhosis. In *Molecular Considerations of Primary Biliary Cirrhosis*. Lindor, K.D., Heathcote, E.J. and Poupon, R. (eds), Kluwer Academic Press, Hingham, pp 40-52.

Gibson, G.E. and Blass, J.P. (1976) Impaired synthesis of acetylcholine in brain accompanying mild hypoxia and hypoglycemia, *J Neurochem*, **27**, 37-42.

Giebeler, R. (1992) The Optima XLA: a new analytical ultracentrifuge with a novel precision absorption optical system. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Harding, S.E., Rowe, A.J. and Horton, J.C. (eds), The Royal Society of Chemistry, Cambridge, pp 16-25.

Glatter, O. (1977) New method for evaluation of small-angle scattering data, *J Appl Cryst*, **10**, 415-421.

Glatter, O. (1980) Computation of distance distribution functions and scattering functions of models for small-angle scattering experiments, *Acta Phys Austriaca*, **52**, 243-256.

Glatter, O. and Kratky, O. (1982) *Small Angle X-ray Scattering*. Academic Press, London.

Goeddel, D.V., Heyneker, H.L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D.G., Ross, M.J., Miozzari, G., Crea, R. and Seeburg, P.H. (1979a) Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone, *Nature*, **281**, 544-548.

Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A.D. (1979b) Expression in *Escherichia coli* of chemically synthesized genes for human insulin, *Proc Natl Acad Sci U S A*, **76**, 106-110.

Golberg, R.J. (1953) Sedimentation in the ultracentrifuge, *J Phys Chem Solids*, **57**, 194-202.

Gopalakrishnan, S., Rahmatullah, M., Radke, G.A., Powers-Greenwood, S. and Roche, T.E. (1989) Role of protein X in the function of the mammalian pyruvate dehydrogenase complex, *Biochem Biophys Res Commun*, **160**, 715-721.

Green, J.D., Laue, E.D., Perham, R.N., Ali, S.T. and Guest, J.R. (1995) Three-dimensional structure of a lipoyl domain from the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*, *J Mol Biol*, **248**, 328-343.

Green, J.D., Perham, R.N., Ullrich, S.J. and Appella, E. (1992) Conformational studies of the interdomain linker peptides in the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*, *J Biol Chem*, **267**, 23484-23488.

Gudi, R., Bowker-Kinley, M.M., Kedishvili, N.Y., Zhao, Y. and Popov, K.M. (1995) Diversity of the pyruvate dehydrogenase kinase gene family in humans, *J Biol Chem*, **270**, 28989-28994.

Guest, J.R., Angier, S.J. and Russell, G.C. (1989) Structure, expression, and protein engineering of the pyruvate dehydrogenase complex of *Escherichia coli*, *Ann N Y Acad Sci*, **573**, 76-99.

Guest, J.R., Lewis, H.M., Graham, L.D., Packman, L.C. and Perham, R.N. (1985) Genetic reconstruction and functional analysis of the repeating lipoyl domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*, *J Mol Biol*, **185**, 743-754.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis*, **18**, 2714-2723.

Guinier, A. and Fournet, G. (1955) *Small Angle Scattering of X-rays*. Wiley, New York.

Hakansson, A.P. and Smith, A.W. (2007) Enzymatic characterization of dihydrolipoamide dehydrogenase from *Streptococcus pneumoniae* harboring its own substrate, *J Biol Chem*, **282**, 29521-29530.

Hanlon, S., Lamers, K., Lauterbach, G., Johnson, R. and Schachman, H.K. (1962) Ultracentrifuge studies with absorption optics. I. An automatic photoelectric scanning absorption system, *Arch Biochem Biophys*, **99**, 157-174.

Harding, S.E., Horton, J.C., Jones, S., Thornton, J.M. and Winzor, D.J. (1999) COVOL: an interactive program for evaluating second virial coefficients from the triaxial shape or dimensions of rigid macromolecules, *Biophys J*, **76**, 2432-2438.

Harmych, S., Arnette, R. and Komuniecki, R. (2002) Role of dihydrolipoyl dehydrogenase (E3) and a novel E3-binding protein in the NADH sensitivity of the pyruvate dehydrogenase complex from anaerobic mitochondria of the parasitic nematode, *Ascaris suum*, *Mol Biochem Parasitol*, **125**, 135-146.

Harris, R.A., Bowker-Kinley, M.M., Huang, B. and Wu, P. (2002) Regulation of the activity of the pyruvate dehydrogenase complex, *Adv Enzyme Regul*, **42**, 249-259.

Harris, R.A., Bowker-Kinley, M.M., Wu, P., Jeng, J. and Popov, K.M. (1997) Dihydrolipoamide dehydrogenase-binding protein of the human pyruvate dehydrogenase complex. DNA-derived amino acid sequence, expression, and reconstitution of the pyruvate dehydrogenase complex, *J Biol Chem*, **272**, 19746-19751.

Harris, R.A., Huang, B. and Wu, P. (2001) Control of pyruvate dehydrogenase kinase gene expression, *Adv Enzyme Regul*, **41**, 269-288.

Hawkins, C.F., Borges, A. and Perham, R.N. (1989) A common structural motif in thiamin pyrophosphate-binding enzymes, *FEBS Lett*, **255**, 77-82.

Hayakawa, T., Kanzaki, T., Kitamura, T., Fukuyoshi, Y., Sakurai, Y., Koike, K., Suematsu, T. and Koike, M. (1969) Mammalian alpha-keto acid dehydrogenase complexes. V. Resolution and reconstitution studies of the pig heart pyruvate dehydrogenase complex, *J Biol Chem*, **244**, 3660-3670.

Head, R.A., Brown, R.M., Zolkipli, Z., Shahdadpuri, R., King, M.D., Clayton, P.T. and Brown, G.K. (2005) Clinical and genetic spectrum of pyruvate dehydrogenase deficiency: dihydrolipoamide acetyltransferase (E2) deficiency, *Ann Neurol*, **58**, 234-241.

Heller, W.T., Vigil, D., Brown, S., Blumenthal, D.K., Taylor, S.S. and Trehwella, J. (2004) C subunits binding to the protein kinase A R1 $\alpha$  dimer induce a large conformational change, *J Biol Chem*, **279**, 19084-19090.

Hengeveld, A.F., Westphal, A.H. and de Kok, A. (1997) Expression and characterisation of the homodimeric E1 component of the *Azotobacter vinelandii* pyruvate dehydrogenase complex, *Eur J Biochem*, **250**, 260-268.

Hipps, D.S., Packman, L.C., Allen, M.D., Fuller, C., Sakaguchi, K., Appella, E. and Perham, R.N. (1994) The peripheral subunit-binding domain of the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*: preparation and characterization of its binding to the dihydrolipoyl dehydrogenase component, *Biochem J*, **297** ( Pt 1), 137-143.

Hiromasa, Y., Aso, Y., Mayanagi, K., Inoue, Y., Fujisawa, T., Meno, K. and Ueki, T. (1998) Guanidine hydrochloride-induced changes of the E2 inner core of the *Bacillus stearothermophilus* pyruvate dehydrogenase complex, *J Biochem*, **123**, 564-567.

Hiromasa, Y., Fujisawa, T., Aso, Y. and Roche, T.E. (2004) Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding protein and their capacities to bind the E1 and E3 components, *J Biol Chem*, **279**, 6921-6933.

Hodgson, J.A., De Marcucci, O.G. and Lindsay, J.G. (1986) Lipoic acid is the site of substrate-dependent acetylation of component X in ox heart pyruvate dehydrogenase multienzyme complex, *Eur J Biochem*, **158**, 595-600.

Hong, Y.S., Kerr, D.S., Liu, T.C., Lusk, M., Powell, B.R. and Patel, M.S. (1997) Deficiency of dihydrolipoamide dehydrogenase due to two mutant alleles (E340K and G101del). Analysis of a family and prenatal testing, *Biochim Biophys Acta*, **1362**, 160-168.

Hong, Y.S., Korman, S.H., Lee, J., Ghoshal, P., Wu, Q., Barash, V., Kang, S., Oh, S., Kwon, M., Gutman, A., Rachmel, A. and Patel, M.S. (2003) Identification of a common mutation (Gly194Cys) in both Arab Moslem and Ashkenazi Jewish patients with dihydrolipoamide dehydrogenase (E3) deficiency: possible beneficial effect of vitamin therapy, *J Inherit Metab Dis*, **26**, 816-818.

Hope, J., Shearman, M.S., Baxter, H.C., Chong, A., Kelly, S.M. and Price, N.C. (1996) Cytotoxicity of prion protein peptide (PrP106-126) differs in mechanism from the cytotoxic activity of the Alzheimer's disease amyloid peptide, A $\beta$  25-35, *Neurodegeneration*, **5**, 1-11.

Howard, M.J., Fuller, C., Broadhurst, R.W., Perham, R.N., Tang, J.G., Quinn, J., Diamond, A.G. and Yeaman, S.J. (1998) Three-dimensional structure of the major autoantigen in primary biliary cirrhosis, *Gastroenterology*, **115**, 139-146.

Hoyer, S. (2004) Causes and consequences of disturbances of cerebral glucose metabolism in sporadic Alzheimer disease: therapeutic implications, *Adv Exp Med Biol*, **541**, 135-152.

Huang, B., Gudi, R., Wu, P., Harris, R.A., Hamilton, J. and Popov, K.M. (1998) Isoenzymes of pyruvate dehydrogenase phosphatase. DNA-derived amino acid sequences, expression, and regulation, *J Biol Chem*, **273**, 17680-17688.

Huang, B., Wu, P., Bowker-Kinley, M.M. and Harris, R.A. (2002) Regulation of pyruvate dehydrogenase kinase expression by peroxisome proliferator-activated receptor-alpha ligands, glucocorticoids, and insulin, *Diabetes*, **51**, 276-283.

Huang, Y.J. and Komuniecki, R. (1997) Cloning and characterization of a putative testis-specific pyruvate dehydrogenase beta subunit from the parasitic nematode, *Ascaris suum*, *Mol Biochem Parasitol*, **90**, 391-394.

Huang, Y.J., Walker, D., Chen, W., Klingbeil, M. and Komuniecki, R. (1998) Expression of pyruvate dehydrogenase isoforms during the aerobic/anaerobic transition in the development of the parasitic nematode *Ascaris suum*: altered stoichiometry of phosphorylation/inactivation, *Arch Biochem Biophys*, **352**, 263-270.

Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph*, **14**, 33-38.

Ibel, K. (1975) Comparison of neutron and X-ray scattering of dilute myoglobin solutions, *J Mol Biol*, **93**, 255-265.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F. and Boyer, H.W. (1977) Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin, *Science*, **198**, 1056-1063.

Iwayama, T., Leung, P.S., Coppel, R.L., Roche, T.E., Patel, M.S., Mizushima, Y., Nakagawa, T., Dickson, R. and Gershwin, M.E. (1991) Specific reactivity of recombinant human PDC-E1 alpha in primary biliary cirrhosis, *J Autoimmun*, **4**, 769-778.

Iwayama, T., Leung, P.S., Rowley, M., Munoz, S., Nishioka, M., Nakagawa, T., Dickson, E.R., Coppel, R.L., Mackay, I.R. and Gershwin, M.E. (1992) Comparative immunoreactive profiles of Japanese and American patients with primary biliary cirrhosis against mitochondrial autoantigens, *Int Arch Allergy Immunol*, **99**, 28-33.

Izard, T., Aevarsson, A., Allen, M.D., Westphal, A.H., Perham, R.N., de Kok, A. and Hol, W.G. (1999) Principles of quasi-equivalence and euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes, *Proc Natl Acad Sci U S A*, **96**, 1240-1245.

Jacrot, B. (1976) The study of biological structures by neutron scattering from solution, *Rep Prog Phys*, **39**, 911-953.

Jemth, P., Gianni, S., Day, R., Li, B., Johnson, C.M., Daggett, V. and Fersht, A.R. (2004) Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation, *Proc Natl Acad Sci U S A*, **101**, 6450-6455.

Jilka, J.M., Rahmatullah, M., Kazemi, M. and Roche, T.E. (1986) Properties of a newly characterized protein of the bovine kidney pyruvate dehydrogenase complex, *J Biol Chem*, **261**, 1858-1867.

Jones, D.D., Horne, H.J., Reche, P.A. and Perham, R.N. (2000a) Structural determinants of post-translational modification and catalytic specificity for the lipoyl domains of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*, *J Mol Biol*, **295**, 289-306.

Jones, D.D., Stott, K.M., Howard, M.J. and Perham, R.N. (2000b) Restricted motion of the lipoyl-lysine swinging arm in the pyruvate dehydrogenase complex of *Escherichia coli*, *Biochemistry*, **39**, 8448-8459.

Joplin, R. and Gershwin, M.E. (1997) Ductular expression of autoantigens in primary biliary cirrhosis, *Semin Liver Dis*, **17**, 97-103.

Joplin, R.E., Wallace, L.L., Lindsay, J.G., Palmer, J.M., Yeaman, S.J. and Neuberger, J.M. (1997) The human biliary epithelial cell plasma membrane antigen in primary biliary cirrhosis: pyruvate dehydrogenase X?, *Gastroenterology*, **113**, 1727-1733.

Jung, H.I., Bowden, S.J., Cooper, A. and Perham, R.N. (2002a) Thermodynamic analysis of the binding of component enzymes in the assembly of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*, *Protein Sci*, **11**, 1091-1100.

Jung, H.I., Cooper, A. and Perham, R.N. (2002b) Identification of key amino acid residues in the assembly of enzymes into the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*: a kinetic and thermodynamic analysis, *Biochemistry*, **41**, 10446-10453.

Junger, E. and Reinauer, H. (1971) Structure of pyruvate dehydrogenase complex of pig heart muscle, *Biochim Biophys Acta*, **250**, 478-490.

Kalia, Y.N., Brocklehurst, S.M., Hipps, D.S., Appella, E., Sakaguchi, K. and Perham, R.N. (1993) The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*, *J Mol Biol*, **230**, 323-341.

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

Kern, D., Kern, G., Neef, H., Tittmann, K., Killenberg-Jabs, M., Wikner, C., Schneider, G. and Hubner, G. (1997) How thiamine diphosphate is activated in enzymes, *Science*, **275**, 67-70.

Khailova, L.S. and Korochkina, L.G. (1985) Half-of-the-site reactivity of the decarboxylating component of the pyruvate dehydrogenase complex from pigeon breast muscle with respect to 2-hydroxyethyl thiamine pyrophosphate, *Biochem Int*, **11**, 509-516.

Khailova, L.S., Korochkina, L.G. and Severin, S.E. (1990). In *Biochemistry and Physiology of TDP Enzymes* Bisswanger, H. and Ullrich, J. (eds), VCH Weinheim, Blaubeuren, Germany, pp 251-265.

Kim, H. and Patel, M.S. (1992) Characterization of two site-specifically mutated human dihydrolipoamide dehydrogenases (His-452----Gln and Glu-457----Gln), *J Biol Chem*, **267**, 5128-5132.

Klingbeil, M.M., Walker, D.J., Arnette, R., Sidawy, E., Hayton, K., Komuniecki, P.R. and Komuniecki, R. (1996) Identification of a novel dihydrolipoyl dehydrogenase-binding protein in the pyruvate dehydrogenase complex of the anaerobic parasitic nematode, *Ascaris suum*, *J Biol Chem*, **271**, 5451-5457.

Klivenyi, P., Starkov, A.A., Calingasan, N.Y., Gardian, G., Browne, S.E., Yang, L., Bubber, P., Gibson, G.E., Patel, M.S. and Beal, M.F. (2004) Mice deficient in dihydrolipoamide dehydrogenase show increased vulnerability to MPTP, malonate and 3-nitropropionic acid neurotoxicity, *J Neurochem*, **88**, 1352-1360.

Knutson, J.R., Walbridge, D.G. and Brand, L. (1982) Decay-associated fluorescence spectra and the heterogeneous emission of alcohol dehydrogenase, *Biochemistry*, **21**, 4671-4679.

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

Kolobova, E., Tuganova, A., Boulatnikov, I. and Popov, K.M. (2001) Regulation of pyruvate dehydrogenase activity through phosphorylation at multiple sites, *Biochem J*, **358**, 69-77.

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

Kong, Y., Ming, D., Wu, Y., Stoops, J.K., Zhou, Z.H. and Ma, J. (2003) Conformational flexibility of pyruvate dehydrogenase complexes: a computational analysis by quantized elastic deformational model, *J Mol Biol*, **330**, 129-135.

Korotchkina, L.G. and Patel, M.S. (1995) Mutagenesis studies of the phosphorylation sites of recombinant human pyruvate dehydrogenase. Site-specific regulation, *J Biol Chem*, **270**, 14297-14304.

Korotchkina, L.G. and Patel, M.S. (2001) Site specificity of four pyruvate dehydrogenase kinase isoenzymes toward the three phosphorylation sites of human pyruvate dehydrogenase, *J Biol Chem*, **276**, 37223-37229.

Kovina, M.V. and Kochetov, G.A. (1998) Cooperativity and flexibility of active sites in homodimeric transketolase, *FEBS Lett*, **440**, 81-84.

Kozin, M.B. and Svergun, D.I. (2001) Automated matching of high- and low-resolution structural models., *J Appl Cryst*, **34**, 33-41.

Kratky, O. and Pilz, I. (1972) Recent advances and applications of diffuse X-ray small-angle scattering on biopolymers in dilute solutions, *Q Rev Biophys*, **5**, 481-537.

Kuntz, I.D. and Kauzmann, W. (1974) Hydration of proteins and polypeptides. In *Advances in Protein Chemistry*. Anfinsen, C.B., Edsall, J.T. and Richards, F.M. (eds), Academic Press, London, pp 239-345.

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

Lamm, O. (1929) Die differentialgleichung der ultrazentrifugierung, *Ark Mat Astr Fys*, **21B**, 1-4.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics*, **23**, 2947-2948.

Lattman, E.E. (1989) Rapid calculation of the solution scattering profile from a macromolecule of known structure, *Proteins*, **5**, 149-155.

Laue, T.M. (1994) An on-line interferometer for the XL-A ultracentrifuge, *Prog Coll Polym Sci*, **94**, 74-81.

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

Lawson, J.E., Behal, R.H. and Reed, L.J. (1991a) Disruption and mutagenesis of the *Saccharomyces cerevisiae* *PDX1* gene encoding the protein X component of the pyruvate dehydrogenase complex, *Biochemistry*, **30**, 2834-2839.

Lawson, J.E., Niu, X.D. and Reed, L.J. (1991b) Functional analysis of the domains of dihydrolipoamide acetyltransferase from *Saccharomyces cerevisiae*, *Biochemistry*, **30**, 11249-11254.

Lebowitz, J., Lewis, M.S. and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review, *Protein Sci*, **11**, 2067-2079.

Lee, K.K., Gan, L., Tsuruta, H., Hendrix, R.W., Duda, R.L. and Johnson, J.E. (2004) Evidence that a local refolding event triggers maturation of HK97 bacteriophage capsid, *J Mol Biol*, **340**, 419-433.

Lemonnier, M., Fourme, R., Rousseaux, F. and Kahn, R. (1978) X-ray curved-crystal monochromator system at the storage ring DCI, *Nucl Instr and Meth*, **152**, 173-177.

Lessard, I.A., Domingo, G.J., Borges, A. and Perham, R.N. (1998) Expression of genes encoding the E2 and E3 components of the *Bacillus stearothermophilus* pyruvate dehydrogenase complex and the stoichiometry of subunit interaction in assembly *in vitro*, *Eur J Biochem*, **258**, 491-501.

Lessard, I.A. and Perham, R.N. (1994) Expression in *Escherichia coli* of genes encoding the E1 $\alpha$  and E1 $\beta$  subunits of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus* and assembly of a functional E1 component ( $\alpha_2\beta_2$ ) *in vitro*, *J Biol Chem*, **269**, 10378-10383.

Lessard, I.A. and Perham, R.N. (1995) Interaction of component enzymes with the peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*: stoichiometry and specificity in self-assembly, *Biochem J*, **306 ( Pt 3)**, 727-733.

Leung, P.S., Iwayama, T., Coppel, R.L. and Gershwin, M.E. (1990) Site-directed mutagenesis of lysine within the immunodominant autoepitope of PDC-E2, *Hepatology*, **12**, 1321-1328.

Lindsay, H., Beaumont, E., Richards, S.D., Kelly, S.M., Sanderson, S.J., Price, N.C. and Lindsay, J.G. (2000) FAD insertion is essential for attaining the assembly competence of the dihydrolipoamide dehydrogenase (E3) monomer from *Escherichia coli*, *J Biol Chem*, **275**, 36665-36670.

Ling, M., McEachern, G., Seyda, A., MacKay, N., Scherer, S.W., Bratinova, S., Beatty, B., Giovannucci-Uzielli, M.L. and Robinson, B.H. (1998) Detection of a homozygous four base pair deletion in the protein X gene in a case of pyruvate dehydrogenase complex deficiency, *Hum Mol Genet*, **7**, 501-505.

Linn, T.C., Pettit, F.H., Hucho, F. and Reed, L.J. (1969) Alpha-keto acid dehydrogenase complexes. XI. Comparative studies of regulatory properties of the pyruvate dehydrogenase complexes from kidney, heart, and liver mitochondria, *Proc Natl Acad Sci U S A*, **64**, 227-234.

Linn, T.C., Pettit, F.H. and Reed, L.J. (1969) Alpha-keto acid dehydrogenase complexes. X. Regulation of the activity of the pyruvate dehydrogenase complex from beef kidney mitochondria by phosphorylation and dephosphorylation, *Proc Natl Acad Sci U S A*, **62**, 234-241.

Lissens, W., De Meirleir, L., Seneca, S., Liebaers, I., Brown, G.K., Brown, R.M., Ito, M., Naito, E., Kuroda, Y., Kerr, D.S., Wexler, I.D., Patel, M.S., Robinson, B.H. and Seyda, A. (2000) Mutations in the X-linked pyruvate dehydrogenase (E1) alpha subunit gene (*PDHA1*) in patients with a pyruvate dehydrogenase complex deficiency, *Hum Mutat*, **15**, 209-219.

Liu, S., Baker, J.C. and Roche, T.E. (1995) Binding of the pyruvate dehydrogenase kinase to recombinant constructs containing the inner lipoyl domain of the dihydrolipoyl acetyltransferase component, *J Biol Chem*, **270**, 793-800.

Lynch, M. and Walsh, B. (1997) *Genetics and Analysis of Quantitative Traits*. Sinauer Association, USA.

Machius, M., Chuang, J.L., Wynn, R.M., Tomchick, D.R. and Chuang, D.T. (2001) Structure of rat BCKD kinase: nucleotide-induced domain communication in a mitochondrial protein kinase, *Proc Natl Acad Sci U S A*, **98**, 11218-11223.

Mackay, I.R., Whittingham, S., Fida, S., Myers, M., Ikuno, N., Gershwin, M.E. and Rowley, M.J. (2000) The peculiar autoimmunity of primary biliary cirrhosis, *Immunol Rev*, **174**, 226-237.

Maeng, C.Y., Yazdi, M.A., Niu, X.D., Lee, H.Y. and Reed, L.J. (1994) Expression, purification, and characterization of the dihydrolipoamide dehydrogenase-binding protein of the pyruvate dehydrogenase complex from *Saccharomyces cerevisiae*, *Biochemistry*, **33**, 13801-13807.

Maeng, C.Y., Yazdi, M.A. and Reed, L.J. (1996) Stoichiometry of binding of mature and truncated forms of the dihydrolipoamide dehydrogenase-binding protein to the dihydrolipoamide acetyltransferase core of the pyruvate dehydrogenase complex from *Saccharomyces cerevisiae*, *Biochemistry*, **35**, 5879-5882.

Mande, S.S., Sarfaty, S., Allen, M.D., Perham, R.N. and Hol, W.G. (1996) Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase, *Structure*, **4**, 277-286.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1987a) Gel Electrophoresis. In *Molecular Cloning - a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 149-186.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1987b) Introduction of plasmid and bacteriophage DNA into *Escherichia coli*. In *Molecular cloning - a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 247-268.

Marsac, C., Stansbie, D., Bonne, G., Cousin, J., Jehenson, P., Benelli, C., Leroux, J.P. and Lindsay, J.G. (1993) Defect in the lipoyl-bearing protein X subunit of the pyruvate dehydrogenase complex in two patients with encephalomyelopathy, *J Pediatr*, **123**, 915-920.

Mattevi, A., Obmolova, G., Kalk, K.H., Teplyakov, A. and Hol, W.G. (1993c) Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p), *Biochemistry*, **32**, 3887-3901.

Mattevi, A., Obmolova, G., Kalk, K.H., van Berkel, W.J. and Hol, W.G. (1993a) Three-dimensional structure of lipoamide dehydrogenase from *Pseudomonas fluorescens* at 2.8 Å resolution. Analysis of redox and thermostability properties, *J Mol Biol*, **230**, 1200-1215.

Mattevi, A., Obmolova, G., Kalk, K.H., Westphal, A.H., de Kok, A. and Hol, W.G. (1993b) Refined crystal structure of the catalytic domain of dihydrolipoyl transacetylase (E2p) from *Azotobacter vinelandii* at 2.6 Å resolution, *J Mol Biol*, **230**, 1183-1199.

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K.H., Westphal, A.H., de Kok, A. and Hol, W.G. (1992a) Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex, *Science*, **255**, 1544-1550.

Mattevi, A., Obmolova, G., Sokatch, J.R., Betzel, C. and Hol, W.G. (1992b) The refined crystal structure of *Pseudomonas putida* lipoamide dehydrogenase complexed with NAD<sup>+</sup> at 2.45 Å resolution, *Proteins*, **13**, 336-351.

Mattevi, A., Schierbeek, A.J. and Hol, W.G. (1991) Refined crystal structure of lipoamide dehydrogenase from *Azotobacter vinelandii* at 2.2 Å resolution. A comparison with the structure of glutathione reductase, *J Mol Biol*, **220**, 975-994.

Mayers, R.M., Butlin, R.J., Kilgour, E., Leighton, B., Martin, D., Myatt, J., Orme, J.P. and Holloway, B.R. (2003) AZD7545, a novel inhibitor of pyruvate dehydrogenase kinase 2 (PDHK2), activates pyruvate dehydrogenase *in vivo* and improves blood glucose control in obese (fa/fa) Zucker rats, *Biochem Soc Trans*, **31**, 1165-1167.

Mayers, R.M., Leighton, B. and Kilgour, E. (2005) PDH kinase inhibitors: a novel therapy for Type II diabetes?, *Biochem Soc Trans*, **33**, 367-370.

McCartney, R. (1998) Folding and assembly studies on the composition of mammalian PDC and OGDC, University of Glasgow, PhD thesis.

McCartney, R.G., Sanderson, S.J. and Lindsay, J.G. (1997) Refolding and reconstitution studies on the transacetylase-protein X (E2/X) subcomplex of the mammalian pyruvate dehydrogenase complex: evidence for specific binding of the dihydrolipoamide dehydrogenase component to sites on reassembled E2, *Biochemistry*, **36**, 6819-6826.

Miles, J.S., Guest, J.R., Radford, S.E. and Perham, R.N. (1988) Investigation of the mechanism of active site coupling in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli* by protein engineering, *J Mol Biol*, **202**, 97-106.

Milne, J.L., Shi, D., Rosenthal, P.B., Sunshine, J.S., Domingo, G.J., Wu, X., Brooks, B.R., Perham, R.N., Henderson, R. and Subramaniam, S. (2002) Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine, *EMBO J*, **21**, 5587-5598.

Morrell, J.A., Orme, J., Butlin, R.J., Roche, T.E., Mayers, R.M. and Kilgour, E. (2003) AZD7545 is a selective inhibitor of pyruvate dehydrogenase kinase 2, *Biochem Soc Trans*, **31**, 1168-1170.

Morten, K.J., Beattie, P., Brown, G.K. and Matthews, P.M. (1999) Dichloroacetate stabilizes the mutant E1 $\alpha$  subunit in pyruvate dehydrogenase deficiency, *Neurology*, **53**, 612-616.

Muller, Y.A., Lindqvist, Y., Furey, W., Schulz, G.E., Jordan, F. and Schneider, G. (1993) A thiamin diphosphate binding fold revealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase, *Structure*, **1**, 95-103.

Naito, E., Ito, M., Yokota, I., Saijo, T., Matsuda, J. and Kuroda, Y. (1998) Thiamine-responsive lactic acidemia: role of pyruvate dehydrogenase complex, *Eur J Pediatr*, **157**, 648-652.

Naito, E., Ito, M., Yokota, I., Saijo, T., Matsuda, J., Ogawa, Y., Kitamura, S., Takada, E., Horii, Y. and Kuroda, Y. (2002) Thiamine-responsive pyruvate dehydrogenase deficiency in two patients caused by a point mutation (F205L and L216F) within the thiamine pyrophosphate binding region, *Biochim Biophys Acta*, **1588**, 79-84.

Nakanuma, Y. and Ohta, G. (1979) Histometric and serial section observations of the intrahepatic bile ducts in primary biliary cirrhosis, *Gastroenterology*, **76**, 1326-1332.

Nakanuma, Y., Tsuneyama, K., Gershwin, M.E. and Yasoshima, M. (1995) Pathology and immunopathology of primary biliary cirrhosis with emphasis on bile duct lesions: recent progress, *Semin Liver Dis*, **15**, 313-328.

Neagle, J., De Marcucci, O., Dunbar, B. and Lindsay, J.G. (1989) Component X of mammalian pyruvate dehydrogenase complex: structural and functional relationship to the lipoate acetyltransferase (E2) component, *FEBS Lett*, **253**, 11-15.

Neagle, J.C. and Lindsay, J.G. (1991) Selective proteolysis of the protein X subunit of the bovine heart pyruvate dehydrogenase complex. Effects on dihydrolipoamide dehydrogenase (E3) affinity and enzymic properties of the complex, *Biochem J*, **278 ( Pt 2)**, 423-427.

Odievre, M.H., Chretien, D., Munnich, A., Robinson, B.H., Dumoulin, R., Masmoudi, S., Kadhom, N., Rotig, A., Rustin, P. and Bonnefont, J.P. (2005) A novel mutation in the dihydrolipoamide dehydrogenase E3 subunit gene (DLD) resulting in an atypical form of alpha-ketoglutarate dehydrogenase deficiency, *Hum Mutat*, **25**, 323-324.

Olijve, W., de Boer, W., Mulders, J.W. and van Wezenbeek, P.M. (1996) Molecular biology and biochemistry of human recombinant follicle stimulating hormone (Puregon), *Mol Hum Reprod*, **2**, 371-382.

Oliver, R.M. and Reed, L.J. (1982). In *Electron Microscopy of Proteins*. Harris, R. (ed), Academic Press London, Vol. 2, pp 1-48.

Packman, L.C. and Perham, R.N. (1986) Chain folding in the dihydrolipoyl acyltransferase components of the 2-oxo-acid dehydrogenase complexes from *Escherichia coli*. Identification of a segment involved in binding the E3 subunit, *FEBS Lett*, **206**, 193-198.

Pan, K. and Jordan, F. (1998) D,L-S-methylipoic acid methyl ester, a kinetically viable model for S-protonated lipoic acid as the oxidizing agent in reductive acyl transfers catalyzed by the 2-oxoacid dehydrogenase multienzyme complexes, *Biochemistry*, **37**, 1357-1364.

Pandya, M.J., Cerasoli, E., Joseph, A., Stoneman, R.G., Waite, E. and Woolfson, D.N. (2004) Sequence and structural duality: designing peptides to adopt two stable conformations, *J Am Chem Soc*, **126**, 17016-17024.

Patel, M.S. and Harris, R.A. (1995) Alpha-keto acid dehydrogenase complexes: nutrient control, gene regulation and genetic defects. Overview, *J Nutr*, **125**, 1744S-1745S.

Patel, M.S. and Harris, R.A. (1995) Mammalian alpha-keto acid dehydrogenase complexes: gene regulation and genetic defects, *FASEB J*, **9**, 1164-1172.

Patel, M.S. and Korotchkina, L.G. (2001) Regulation of mammalian pyruvate dehydrogenase complex by phosphorylation: complexity of multiple phosphorylation sites and kinases, *Exp Mol Med*, **33**, 191-197.

Patel, M.S. and Korotchkina, L.G. (2003) The biochemistry of the pyruvate dehydrogenase complex, *Biochem Mol Biol Education*, **31**, 5-15.

Patel, M.S. and Roche, T.E. (1990) Molecular biology and biochemistry of pyruvate dehydrogenase complexes, *FASEB J*, **4**, 3224-3233.

Peitsch, M.C., Schwede, T. and Guex, N. (2000) Automated protein modelling-the proteome in 3D, *Pharmacogenomics*, **1**, 257-266.

Perham, R.N. (1991) Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein, *Biochemistry*, **30**, 8501-8512.

- Perham, R.N. (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions, *Annu Rev Biochem*, **69**, 961-1004.
- Perham, R.N., Duckworth, H.W. and Roberts, G.C. (1981) Mobility of polypeptide chain in the pyruvate dehydrogenase complex revealed by proton NMR, *Nature*, **292**, 474-477.
- Perkins, S.J. (1988) Structural studies of proteins by high-flux X-ray and neutron solution scattering, *Biochem J*, **254**, 313-327.
- Perkins, S.J. (1988) X-ray and neutron solution scattering. In *Modern Physical Methods in Biochemistry, Part B*. Neuberger, A. and Deenen, L.L.M.V. (eds), Elsevier Science Publishers, London.
- Perrin, J. (1936) Mouvement Brownien d'un ellipsoïde. II. Rotation libre et depolarisation des fluorescences. Translation et diffusion de molécules ellipsoïdales, *J Phys Radium*, **7**, 1-11.
- Petoukhov, M.V. and Svergun, D.I. (2003) New methods of domain structure determination of proteins from solution scattering data, *J Appl Cryst*, **36**, 540-544.
- Pettit, F.H., Roche, T.E. and Reed, L.J. (1972) Function of calcium ions in pyruvate dehydrogenase phosphatase activity, *Biochem Biophys Res Commun*, **49**, 563-571.
- Planel, E., Miyasaka, T., Launey, T., Chui, D.H., Tanemura, K., Sato, S., Murayama, O., Ishiguro, K., Tatebayashi, Y. and Takashima, A. (2004) Alterations in glucose metabolism induce hypothermia leading to tau hyperphosphorylation through differential inhibition of kinase and phosphatase activities: implications for Alzheimer's disease, *J Neurosci*, **24**, 2401-2411.
- Plum, F. and Posner, J.B. (1980) *The Diagnosis of Stupor and Coma*. Davis, F.A. (eds), Philadelphia.
- Popov, K.M., Kedishvili, N.Y., Zhao, Y., Gudi, R. and Harris, R.A. (1994) Molecular cloning of the p45 subunit of pyruvate dehydrogenase kinase, *J Biol Chem*, **269**, 29720-29724.

Popov, K.M., Kedishvili, N.Y., Zhao, Y., Shimomura, Y., Crabb, D.W. and Harris, R.A. (1993) Primary structure of pyruvate dehydrogenase kinase establishes a new family of eukaryotic protein kinases, *J Biol Chem*, **268**, 26602-26606.

Popov, K.M., Zhao, Y., Shimomura, Y., Kuntz, M.J. and Harris, R.A. (1992) Branched-chain alpha-ketoacid dehydrogenase kinase. Molecular cloning, expression, and sequence similarity with histidine protein kinases, *J Biol Chem*, **267**, 13127-13130.

Powers-Greenwood, S.L., Rahmatullah, M., Radke, G.A. and Roche, T.E. (1989) Separation of protein X from the dihydrolipoyl transacetylase component of the mammalian pyruvate dehydrogenase complex and function of protein X, *J Biol Chem*, **264**, 3655-3657.

Quinn, J., Diamond, A.G., Masters, A.K., Brookfield, D.E., Wallis, N.G. and Yeaman, S.J. (1993) Expression and lipoylation in *Escherichia coli* of the inner lipoyl domain of the E2 component of the human pyruvate dehydrogenase complex, *Biochem J*, **289**, 81-85.

Quinn, J., Diamond, A.G., Palmer, J.M., Bassendine, M.F., James, O.F. and Yeaman, S.J. (1993) Lipoylated and unlipoylated domains of human PDC-E2 as autoantigens in primary biliary cirrhosis: significance of lipoate attachment, *Hepatology*, **18**, 1384-1391.

Radford, S.E. (2000) Protein folding: progress made and promises ahead, *Trends Biochem Sci*, **25**, 611-618.

Radford, S.E., Laue, E.D., Perham, R.N., Martin, S.R. and Appella, E. (1989a) Conformational flexibility and folding of synthetic peptides representing an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*, *J Biol Chem*, **264**, 767-775.

Radford, S.E., Laue, E.D., Perham, R.N., Miles, J.S. and Guest, J.R. (1987) Segmental structure and protein domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Genetic reconstruction *in vitro* and <sup>1</sup>H NMR spectroscopy, *Biochem J*, **247**, 641-649.

Rahmatullah, M., Gopalakrishnan, S., Andrews, P.C., Chang, C.L., Radke, G.A. and Roche, T.E. (1989a) Subunit associations in the mammalian pyruvate dehydrogenase complex. Structure and role of protein X and the pyruvate dehydrogenase component binding domain of the dihydrolipoyl transacetylase component, *J Biol Chem*, **264**, 2221-2227.

Rahmatullah, M., Gopalakrishnan, S., Radke, G.A. and Roche, T.E. (1989b) Domain structures of the dihydrolipoyl transacetylase and the protein X components of mammalian pyruvate dehydrogenase complex. Selective cleavage by protease Arg C, *J Biol Chem*, **264**, 1245-1251.

Rahmatullah, M., Radke, G.A., Andrews, P.C. and Roche, T.E. (1990) Changes in the core of the mammalian-pyruvate dehydrogenase complex upon selective removal of the lipoyl domain from the transacetylase component but not from the protein X component, *J Biol Chem*, **265**, 14512-14517.

Rahmatullah, M. and Roche, T.E. (1987) The catalytic requirements for reduction and acetylation of protein X and the related regulation of various forms of resolved pyruvate dehydrogenase kinase, *J Biol Chem*, **262**, 10265-10271.

Rai, N., Nöllmann, M., Spotorno, B., Tassara, G., Byron, O. and Rocco, M. (2005) SOMO (SOlution MOdeler) : differences between x-ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics, *Structure*, **13**, 723-734.

Ralston, G. (1993) *Introduction to Analytical Ultracentrifugation*. Beckman, Fullerton, USA.

Ramadan, D.G., Head, R.A., Al-Tawari, A., Habeeb, Y., Zaki, M., Al-Ruqum, F., Besley, G.T., Wraith, J.E., Brown, R.M. and Brown, G.K. (2004) Lactic acidosis and developmental delay due to deficiency of E3 binding protein (protein X) of the pyruvate dehydrogenase complex, *J Inherit Metab Dis*, **27**, 477-485.

Randle, P.J. (1986) Fuel selection in animals, *Biochem Soc Trans*, **14**, 799-806.

Randle, P.J., Priestman, D.A., Mistry, S. and Halsall, A. (1994) Mechanisms modifying glucose oxidation in diabetes mellitus, *Diabetologia*, **37 Suppl 2**, S155-161.

Reed, L.J. (2001) A trail of research from lipoic acid to alpha-keto acid dehydrogenase complexes, *J Biol Chem*, **276**, 38329-38336.

Reed, L.J. and Hackert, M.L. (1990) Structure-function relationships in dihydrolipoamide acyltransferases, *J Biol Chem*, **265**, 8971-8974.

Reed, L.J., Lawson, J.E., Niu, X.D. and Yan, J. (1996). In  *$\alpha$ -Keto Acid Dehydrogenase Complexes*. Patel, M.S., Roche, T.E. and Harris, R.A. (eds), Birkhäuser Verlag, Basel, Switzerland, pp 131-138.

Reed, L.J. and Oliver, R.M. (1968) The multienzyme alpha-keto acid dehydrogenase complexes, *Brookhaven Symp Biol*, **21**, 397-412.

Reed, L.J. and Oliver, R.M. (1982) Structure-function relationships in pyruvate and alpha-ketoglutarate dehydrogenase complexes, *Adv Exp Med Biol*, **148**, 231-241.

Reiman, E.M., Caselli, R.J., Yun, L.S., Chen, K., Bandy, D., Minoshima, S., Thibodeau, S.N. and Osborne, D. (1996) Preclinical evidence of Alzheimer's disease in persons homozygous for the epsilon 4 allele for apolipoprotein E, *N Engl J Med*, **334**, 752-758.

Ricaud, P.M., Howard, M.J., Roberts, E.L., Broadhurst, R.W. and Perham, R.N. (1996) Three-dimensional structure of the lipoyl domain from the dihydrolipoyl succinyltransferase component of the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*, *J Mol Biol*, **264**, 179-190.

Richards, S.D. (1999) Protein-protein interactions within the 2-oxoacid dehydrogenase complexes, University of Glasgow, PhD thesis.

Robinson, B.H. (1995) Lactic acidemia - disorders of pyruvate carboxylase, pyruvate dehydrogenase. In *The Metabolic and Molecular Basis of Inherited Disease*. Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. (eds), McGraw Hill, New York, pp 1479-1499.

Robinson, B.H., MacMillan, H., Petrova-Benedict, R. and Sherwood, W.G. (1987) Variable clinical presentation in patients with defective E1 component of pyruvate dehydrogenase complex, *J Pediatr*, **111**, 525-533.

Roche, T.E., Hiromasa, Y., Turkan, A., Gong, X., Peng, T., Yan, X., Kasten, S.A., Bao, H. and Dong, J. (2003) Essential roles of lipoyl domains in the activated function and control of pyruvate dehydrogenase kinases and phosphatase isoform 1, *Eur J Biochem*, **270**, 1050-1056.

Roche, T.E., Powers-Greenwood, S.L., Shi, W.F., Zhang, W.B., Ren, S.Z., Roche, E.D., Cox, D.J. and Sorensen, C.M. (1993) Sizing of bovine heart and kidney pyruvate dehydrogenase complex and dihydrolipoyl transacetylase core by quasielastic light scattering, *Biochemistry*, **32**, 5629-5637.

Rowles, J., Scherer, S.W., Xi, T., Majer, M., Nickle, D.C., Rommens, J.M., Popov, K.M., Harris, R.A., Riebow, N.L., Xia, J., Tsui, L.C., Bogardus, C. and Prochazka, M. (1996) Cloning and characterization of PDK4 on 7q21.3 encoding a fourth pyruvate dehydrogenase kinase isoenzyme in human, *J Biol Chem*, **271**, 22376-22382.

Royer, C.A. (2006) Probing protein folding and conformational transitions with fluorescence, *Chem Rev*, **106**, 1769-1784.

Sanderson, S.J., Khan, S.S., McCartney, R.G., Miller, C. and Lindsay, J.G. (1996a) Reconstitution of mammalian pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase complexes: analysis of protein X involvement and interaction of homologous and heterologous dihydrolipoamide dehydrogenases, *Biochem J*, **319 (Pt 1)**, 109-116.

Sanderson, S.J., Miller, C. and Lindsay, J.G. (1996b) Stoichiometry, organisation and catalytic function of protein X of the pyruvate dehydrogenase complex from bovine heart, *Eur J Biochem*, **236**, 68-77.

Sayle, R.A. and Milner-White, E.J. (1995) RASMOL: biomolecular graphics for all, *Trends Biochem Sci*, **20**, 374.

Schachman, H.K. (1959) *Ultracentrifugation in Biochemistry* Academic Press, New York.

Schachman, H.K., Gropper, L., Hanlon, S. and Putney, F. (1962) Ultracentrifuge studies with absorption optics. II. Incorporation of a monochromator and its application to the study of proteins and interacting systems, *Arch Biochem Biophys*, **99**, 175-190.

Schellenberger, A. (1998) Sixty years of thiamin diphosphate biochemistry, *Biochim Biophys Acta*, **1385**, 177-186.

Schierbeek, A.J., Swarte, M.B., Dijkstra, B.W., Vriend, G., Read, R.J., Hol, W.G., Drenth, J. and Betzel, C. (1989) X-ray structure of lipoamide dehydrogenase from *Azotobacter vinelandii* determined by a combination of molecular and isomorphous replacement techniques, *J Mol Biol*, **206**, 365-379.

Schmidt, B. and Riesner, D. (1992) A fluorescence detection system for the analytical ultracentrifuge and its application to proteins, nucleic acids, viroids and viruses. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Harding, S.E., Rowe, A.J. and Horton, J.C. (eds), The Royal Society of Chemistry, Cambridge, pp 176-207.

Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling, *Biophys J*, **78**, 1606-1619.

Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation, *Anal Biochem*, **320**, 104-124.

Schuck, P. and Demeler, B. (1999) Direct sedimentation analysis of interference optical data in analytical ultracentrifugation, *Biophys J*, **76**, 2288-2296.

Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J. and Schubert, D. (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems, *Biophys J*, **82**, 1096-1111.

Schulze, E., Westphal, A.H., Boumans, H. and de Kok, A. (1991a) Site-directed mutagenesis of the dihydrolipoyl transacetylase component (E2p) of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Binding of the peripheral components E1p and E3, *Eur J Biochem*, **202**, 841-848.

Schulze, E., Westphal, A.H., Obmolova, G., Mattevi, A., Hol, W.G. and de Kok, A. (1991b) The catalytic domain of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii* and *Escherichia coli*. Expression, purification, properties and preliminary X-ray analysis, *Eur J Biochem*, **201**, 561-568.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003) SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res*, **31**, 3381-3385.

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

Serdyuk, I.N., Zaccai, N.R. and Zaccai, J.G. (2007) *Methods in Molecular Biophysics: Structure, Dynamics, Function*. Cambridge University Press, UK.

Sergienko, E.A. and Jordan, F. (2002) Yeast pyruvate decarboxylase tetramers can dissociate into dimers along two interfaces. Hybrids of low-activity D28A (or D28N) and E477Q variants, with substitution of adjacent active center acidic groups from different subunits, display restored activity, *Biochemistry*, **41**, 6164-6169.

Sergienko, E.A., Wang, J., Polovnikova, L., Hasson, M.S., McLeish, M.J., Kenyon, G.L. and Jordan, F. (2000) Spectroscopic detection of transient thiamin diphosphate-bound intermediates on benzoylformate decarboxylase, *Biochemistry*, **39**, 13862-13869.

Seyda, A., McEachern, G., Haas, R. and Robinson, B.H. (2000) Sequential deletion of C-terminal amino acids of the E(1) $\alpha$  component of the pyruvate dehydrogenase (PDH) complex leads to reduced steady-state levels of functional E(1) $\alpha$ (2) $\beta$ (2) tetramers: implications for patients with PDH deficiency, *Hum Mol Genet*, **9**, 1041-1048.

Shany, E., Saada, A., Landau, D., Shaag, A., Hershkovitz, E. and Elpeleg, O.N. (1999) Lipoamide dehydrogenase deficiency due to a novel mutation in the interface domain, *Biochem Biophys Res Commun*, **262**, 163-166.

Sheu, K.F. and Blass, J.P. (1999) The alpha-ketoglutarate dehydrogenase complex, *Ann N Y Acad Sci*, **893**, 61-78.

Sheu, K.F., Cooper, A.J., Koike, K., Koike, M., Lindsay, J.G. and Blass, J.P. (1994) Abnormality of the alpha-ketoglutarate dehydrogenase complex in fibroblasts from familial Alzheimer's disease, *Ann Neurol*, **35**, 312-318.

Shi, Q., Chen, H.L., Xu, H. and Gibson, G.E. (2005) Reduction in the E2k subunit of the alpha-ketoglutarate dehydrogenase complex has effects independent of complex activity, *J Biol Chem*, **280**, 10888-10896.

Shoffner, J.M. (1997) Oxidative phosphorylation defects and Alzheimer's disease, *Neurogenetics*, **1**, 13-19.

Shu, F., Ramakrishnan, V. and Schoenborn, B.P. (2000) Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin, *Proc Natl Acad Sci U S A*, **97**, 3872-3877.

Sims, N.R., Anderson, M.F., Hobbs, L.M., Kong, J.Y., Phillips, S., Powell, J.A. and Zaidan, E. (2000) Impairment of brain mitochondrial function by hydrogen peroxide, *Brain Res Mol Brain Res*, **77**, 176-184.

Singh, G. (2008) Analysis of genetic mutations using a recombinant model of the mammalian pyruvate dehydrogenase complex, University of Glasgow, PhD thesis.

Small, G.W., Mazziotta, J.C., Collins, M.T., Baxter, L.R., Phelps, M.E., Mandelkern, M.A., Kaplan, A., La Rue, A., Adamson, C.F., Chang, L. and et al. (1995) Apolipoprotein E type 4 allele and cerebral glucose metabolism in relatives at risk for familial Alzheimer disease, *JAMA*, **273**, 942-947.

Smolle, M. (2005) Protein-protein interactions within the human pyruvate dehydrogenase complex, University of Glasgow, PhD thesis.

Smolle, M., Prior, A.E., Brown, A.E., Cooper, A., Byron, O. and Lindsay, J.G. (2006) A new level of architectural complexity in the human pyruvate dehydrogenase complex, *J Biol Chem*, **281**, 19772-19780.

Snijder, H.J., Timmins, P.A., Kalk, K.H. and Dijkstra, B.W. (2003) Detergent organisation in crystals of monomeric outer membrane phospholipase A, *J Struct Biol*, **141**, 122-131.

Sousa, R. (1995) Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization, *Acta Crystallogr D Biol Crystallogr*, **51**, 271-277.

Spencer, M.E., Darlison, M.G., Stephens, P.E., Duckenfield, I.K. and Guest, J.R. (1984) Nucleotide sequence of the *sucB* gene encoding the dihydrolipoamide succinyltransferase of *Escherichia coli* K12 and homology with the corresponding acetyltransferase, *Eur J Biochem*, **141**, 361-374.

Spotorno, B., Piccinini, L., Tassara, G., Ruggiero, C., Nardini, M., Molina, F. and Rocco, M. (1997) BEAMS (BEAdS Modelling System): a set of computer programs for the generation, the visualization and the computation of the hydrodynamic and conformational properties of bead models of proteins, *Eur Biophys J*, **25**, 373-384.

Squire, P.G. and Himmel, M.E. (1979) Hydrodynamics and protein hydration, *Arch Biochem Biophys*, **196**, 165-177.

Stafford, W.F. (1992) Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile, *Anal Biochem*, **203**, 295-301.

Stafford, W.F. (1994) Boundary analysis in sedimentation velocity experiments, *Methods Enzymol*, **240**, 478-501.

Stafford, W.F. and Sherwood, P.J. (2004) Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants, *Biophys Chem*, **108**, 231-243.

Stanley, C.J. and Perham, R.N. (1980) Purification of 2-oxo acid dehydrogenase multienzyme complexes from ox heart by a new method, *Biochem J*, **191**, 147-154.

Steussy, C.N., Popov, K.M., Bowker-Kinley, M.M., Sloan, R.B., Jr., Harris, R.A. and Hamilton, J.A. (2001) Structure of pyruvate dehydrogenase kinase. Novel folding pattern for a serine protein kinase, *J Biol Chem*, **276**, 37443-37450.

Stoops, J.K., Baker, T.S., Schroeter, J.P., Kolodziej, S.J., Niu, X.D. and Reed, L.J. (1992) Three-dimensional structure of the truncated core of the *Saccharomyces cerevisiae* pyruvate dehydrogenase complex determined from negative stain and cryoelectron microscopy images, *J Biol Chem*, **267**, 24769-24775.

Stoops, J.K., Cheng, R.H., Yazdi, M.A., Maeng, C.Y., Schroeter, J.P., Klueppelberg, U., Kolodziej, S.J., Baker, T.S. and Reed, L.J. (1997) On the unique structural organization of the *Saccharomyces cerevisiae* pyruvate dehydrogenase complex, *J Biol Chem*, **272**, 5757-5764.

Strumilo, S. (2005) Short-term regulation of the mammalian pyruvate dehydrogenase complex, *Acta Biochim Pol*, **52**, 759-764.

Stuhrmann, H.B. (1970) New method for determination of surface form and internal structure of dissolved globular proteins from small angle x-ray measurements, *Zeitschr Phys Chem*, **72**, 177-198.

Stuhrmann, H.B., Scharpf, O., Krumpolc, M., Niinikoski, T.O., Rieubland, M. and Rijllart, A. (1986) Dynamic nuclear polarisation of biological matter, *Eur Biophys J*, **14**, 1-6.

Sugden, M.C. and Holness, M.J. (2003) Recent advances in mechanisms regulating glucose oxidation at the level of the pyruvate dehydrogenase complex by PDKs, *Am J Physiol Endocrinol Metab*, **284**, E855-862.

Surh, C.D., Ahmed-Ansari, A. and Gershwin, M.E. (1990) Comparative epitope mapping of murine monoclonal and human autoantibodies to human PDH-E2, the major mitochondrial autoantigen of primary biliary cirrhosis, *J Immunol*, **144**, 2647-2652.

Surh, C.D., Roche, T.E., Danner, D.J., Ansari, A., Coppel, R.L., Prindiville, T., Dickson, E.R. and Gershwin, M.E. (1989) Antimitochondrial autoantibodies in primary biliary cirrhosis recognize cross-reactive epitope(s) on protein X and dihydrolipoamide acetyltransferase of pyruvate dehydrogenase complex, *Hepatology*, **10**, 127-133.

Svedberg, T. and Pedersen, K.O. (1940) *The Ultracentrifuge*. Oxford University Press, London and New York.

Svergun, D. and Koch, M.H.J. (2003) Small-angle scattering studies of biological macromolecules in solution, *Rep Prog Phys*, **66**, 1735-1782.

Svergun, D., Semenyuk, A.V. and Feigin, L.A. (1988) Small angle scattering data treatment by the regularization method, *Acta Cryst*, **A44**, 244-250.

Svergun, D. and Stuhrmann, H.B. (1991) New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations, *Acta Cryst*, **A47**, 736-744.

Svergun, D., Volkov, V.V., Kozin, M.B. and Stuhrmann, H.B. (1996) New developments in direct shape determination from small-angle scattering. 2. Uniqueness, *Acta Cryst*, **A52**, 419-426.

Svergun, D.I. (1992) Determination of the regularization parameter in indirect transform methods using perceptual criteria., *J Appl Cryst*, **25**, 495-503.

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

Svergun, D.I., Barberato, C. and Koch, M.H.J. (1995) CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates, *J Appl Cryst*, **28**, 768-773.

Svergun, D.I., Burkhardt, N., Pedersen, J.S., Koch, M.H., Volkov, V.V., Kozin, M.B., Meerwink, W., Stuhrmann, H.B., Diedrich, G. and Nierhaus, K.H. (1997) Solution scattering structural analysis of the 70 S *Escherichia coli* ribosome by contrast variation. I. Invariants and validation of electron microscopy models, *J Mol Biol*, **271**, 588-601.

Svergun, D.I. and Koch, M.H. (2002) Advances in structure analysis using small-angle scattering in solution, *Curr Opin Struct Biol*, **12**, 654-660.

Svergun, D.I., Petoukhov, M.V. and Koch, M.H. (2001) Determination of domain structure of proteins from X-ray solution scattering, *Biophys J*, **80**, 2946-2953.

Svergun, D.I., Richard, S., Koch, M.H., Sayers, Z., Kuprin, S. and Zaccai, G. (1998) Protein hydration in solution: experimental observation by x-ray and neutron scattering, *Proc Natl Acad Sci U S A*, **95**, 2267-2272.

Szabo, P., Sheu, K.F., Robinson, R.M., Grzeschik, K.H. and Blass, J.P. (1990) The gene for the alpha polypeptide of pyruvate dehydrogenase is X-linked in humans, *Am J Hum Genet*, **46**, 874-878.

Tanford, C. (1961) *Physical Chemistry of Macromolecules*. John Wiley and Sons, Inc, New York and London.

Teller, D.C. (1973) Characterisation of proteins by sedimentation equilibrium in the analytical ultracentrifuge. In *Methods in Enzymology*. Hirs, C.H.W. and Timasheff, S.N. (eds), Academic Press, New York and London.

Texter, F.L., Radford, S.E., Laue, E.D., Perham, R.N., Miles, J.S. and Guest, J.R. (1988) Site-directed mutagenesis and <sup>1</sup>H NMR spectroscopy of an interdomain segment in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*, *Biochemistry*, **27**, 289-296.

Titman, C.M. (2005) The inter-domain regions in the *Bacillus stearothermophilus* pyruvate dehydrogenase complex, University of Cambridge, PhD thesis.

Toyoda, T., Kobayashi, R., Sekiguchi, T., Koike, K., Koike, M. and Takenaka, A. (1998a) Crystallization and preliminary X-ray analysis of pig E3, lipoamide dehydrogenase, *Acta Crystallogr D Biol Crystallogr*, **54**, 982-985.

Toyoda, T., Suzuki, K., Sekiguchi, T., Reed, L.J. and Takenaka, A. (1998b) Crystal structure of eukaryotic E3, lipoamide dehydrogenase from yeast, *J Biochem*, **123**, 668-674.

Tuganova, A., Yoder, M.D. and Popov, K.M. (2001) An essential role of Glu-243 and His-239 in the phosphotransfer reaction catalyzed by pyruvate dehydrogenase kinase, *J Biol Chem*, **276**, 17994-17999.

Vachette, P. and Svergun, D. (2000) Small-angle x-ray scattering by solutions of biological macromolecules. In *Structure and Dynamics of Biomolecules*. Franchon, E., Geissler, E., Hodeau, J.L., Regnard, J.R. and Timmins, P.A. (eds), Oxford University Press, Oxford, pp 199-235.

Van de Water, J., Fregeau, D., Davis, P., Ansari, A., Danner, D., Leung, P., Coppel, R. and Gershwin, M.E. (1988a) Autoantibodies of primary biliary cirrhosis recognize dihydrolipoamide acetyltransferase and inhibit enzyme function, *J Immunol*, **141**, 2321-2324.

Van de Water, J., Gershwin, M.E., Leung, P., Ansari, A. and Coppel, R.L. (1988b) The autoepitope of the 74 kDa mitochondrial autoantigen of primary biliary cirrhosis corresponds to the functional site of dihydrolipoamide acetyltransferase, *J Exp Med*, **167**, 1791-1799.

van Holde, K.E. (1971) *Physical Biochemistry*. Prentice Hall, Inc, Englewoods Cliffs, New Jersey.

Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R.A., Rouault, T.A. and Schuck, P. (2004) Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition, *Anal Biochem*, **326**, 234-256.

Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering, *J Appl Cryst*, **36**, 860-864.

Wagenknecht, T., Grassucci, R., Radke, G.A. and Roche, T.E. (1991) Cryoelectron microscopy of mammalian pyruvate dehydrogenase complex, *J Biol Chem*, **266**, 24650-24656.

Wagenknecht, T., Grassucci, R. and Schaak, D. (1990) Cryoelectron microscopy of frozen-hydrated alpha-ketoacid dehydrogenase complexes from *Escherichia coli*, *J Biol Chem*, **265**, 22402-22408.

Wall, M.E., Gallagher, S.C. and Trehwella, J. (2000) Large-scale shape changes in proteins and macromolecular complexes, *Annu Rev Phys Chem*, **51**, 355-380.

Wallis, N.G., Allen, M.D., Broadhurst, R.W., Lessard, I.A. and Perham, R.N. (1996) Recognition of a surface loop of the lipoyl domain underlies substrate channelling in the pyruvate dehydrogenase multienzyme complex, *J Mol Biol*, **263**, 463-474.

Wallis, N.G. and Perham, R.N. (1994) Structural dependence of post-translational modification and reductive acetylation of the lipoyl domain of the pyruvate dehydrogenase multienzyme complex, *J Mol Biol*, **236**, 209-216.

Westphal, A.H., Fabisz-Kijowska, A., Kester, H., Obels, P.P. and de Kok, A. (1995) The interaction between lipoamide dehydrogenase and the peripheral-component-binding domain from the *Azotobacter vinelandii* pyruvate dehydrogenase complex, *Eur J Biochem*, **234**, 861-870.

Wexler, I.D., Hemalatha, S.G., McConnell, J., Buist, N.R., Dahl, H.H., Berry, S.A., Cederbaum, S.D., Patel, M.S. and Kerr, D.S. (1997) Outcome of pyruvate dehydrogenase deficiency treated with ketogenic diets. Studies in patients with identical mutations, *Neurology*, **49**, 1655-1661.

Wieland, O.H. (1983) The mammalian pyruvate dehydrogenase complex: structure and regulation, *Rev Physiol Biochem Pharmacol*, **96**, 123-170.

Wijburg, F.A., Barth, P.G., Bindoff, L.A., Birch-Machin, M.A., van der Blij, J.F., Ruitenbeek, W., Turnbull, D.M. and Schutgens, R.B. (1992) Leigh syndrome associated with a deficiency of the pyruvate dehydrogenase complex: results of treatment with a ketogenic diet, *Neuropediatrics*, **23**, 147-152.

Wilkinson, K.D. and Williams, C.H., Jr. (1981) NADH inhibition and NAD activation of *Escherichia coli* lipoamide dehydrogenase catalyzing the NADH-lipoamide reaction, *J Biol Chem*, **256**, 2307-2314.

Williams, C.H. (1992). In *Chemistry and Biochemistry of Flavoenzymes*. Muller, F. (ed), CRC Boca Raton, FL, pp 121-211.

Williams, C.H., Jr. (1965) Studies on lipoyl dehydrogenase from *Escherichia coli*, *J Biol Chem*, **240**, 4793-4800.

Williams, C.H., Jr., Allison, N., Russell, G.C., Prongay, A.J., Arcsott, L.D., Datta, S., Sahlman, L. and Guest, J.R. (1989) Properties of lipoamide dehydrogenase and thioredoxin reductase from *Escherichia coli* altered by site-directed mutagenesis, *Ann N Y Acad Sci*, **573**, 55-65.

Yeaman, S.J. (1989) The 2-oxo acid dehydrogenase complexes: recent advances, *Biochem J*, **257**, 625-632.

Yeaman, S.J., Fussey, S.P., Danner, D.J., James, O.F., Mutimer, D.J. and Bassendine, M.F. (1988) Primary biliary cirrhosis: identification of two major M2 mitochondrial autoantigens, *Lancet*, **1**, 1067-1070.

Yeaman, S.J., Hutcheson, E.T., Roche, T.E., Pettit, F.H., Brown, J.R., Reed, L.J., Watson, D.C. and Dixon, G.H. (1978) Sites of phosphorylation on pyruvate dehydrogenase from bovine kidney and heart, *Biochemistry*, **17**, 2364-2370.

Yi, J., Nemeria, N., McNally, A., Jordan, F., Machado, R.S. and Guest, J.R. (1996) Effect of substitutions in the thiamin diphosphate-magnesium fold on the activation of the pyruvate dehydrogenase complex from *Escherichia coli* by cofactors and substrate, *J Biol Chem*, **271**, 33192-33200.

Yphantis, D.A., Lary, J.W., Stafford, W.F., Liu, S., Olsen, P.H., Hayes, D.B., Moody, T.P., Ridgeway, T.M., Lyons, D.A. and Laue, T.M. (1994) Online data acquisition for the Rayleigh interference optical system of the analytical ultracentrifuge. In *Modern Analytical Ultracentrifugation*. Schuster, T.M. and Laue, T.M. (eds), Birkhäuser, Boston, pp 209-226.

Yu, X., Hiromasa, Y., Tsen, H., Stoops, J.K., Roche, T.E. and Zhou, Z.H. (2008) Structures of the human pyruvate dehydrogenase complex cores: a highly conserved catalytic center with flexible N-terminal domains, *Structure*, **16**, 104-114.

Zaccai, G. and Jacrot, B. (1983) Small angle neutron scattering, *Annu Rev Biophys Bioeng*, **12**, 139-157.

Zhang, Y., Taiming, L. and Liu, J. (2003) Low temperature and glucose enhanced T7 RNA polymerase-based plasmid stability for increasing expression of glucagon-like peptide-2 in *Escherichia coli*, *Protein Expr Purif*, **29**, 132-139.

Zhou, Z.H., Liao, W., Cheng, R.H., Lawson, J.E., McCarthy, D.B., Reed, L.J. and Stoops, J.K. (2001a) Direct evidence for the size and conformational variability of the pyruvate dehydrogenase complex revealed by three-dimensional electron microscopy. The "breathing" core and its functional relationship to protein dynamics, *J Biol Chem*, **276**, 21704-21713.

Zhou, Z.H., McCarthy, D.B., O'Connor, C.M., Reed, L.J. and Stoops, J.K. (2001b) The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes, *Proc Natl Acad Sci U S A*, **98**, 14802-14807.