Small-Angle X-ray Scattering Studies of Bone Morphogenetic Protein Regulators

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD) in the Faculty of Biology, Medicine and Health

2017

Michael Paul Lockhart-Cairns

Division of Cell Matrix Biology & Regenerative Medicine

School of Biological Sciences

"It's amazing how far someone can get by learning the times tables on the walk to school."

John 'Iain' Mollison Cairns

1934-2017

Contents

Word count: 45660

Figures

Supplementary Figures

Tables

Supplementary Tables

Abstract

The implementation of small-angle X-ray scattering (SAXS) in combination with size exclusion chromatography (SEC) has provided a useful tool for increasing the purity of samples during SAXS data collection. With the use of SEC-SAXS, samples become diluted and so methods to enhance the scattering signal of the samples at lower concentrations have been applied to beamline B21 using a modified sample cell, a variety of window materials and longer exposure times. SEC-SAXS at B21 provided a method of collecting data for biological macromolecules where aggregation and low concentration issues are a major problem. This method of data collection is shown with a variety of proteins, including matrix proteins, which are less amenable to SAXS data collection.

Bone morphogenetic proteins (BMP) are essential signalling molecules involved in developmental and pathological processes and are regulated in the matrix by secreted glycoproteins. One such regulator is BMP-binding endothelial cell precursor-derived regulator (BMPER) which can both inhibit and enhance BMP signalling in a context and concentration-dependant manner. Twisted gastrulation (Tsg) also has pro- and anti-BMP properties but it is unclear whether Tsg and BMPER directly interact and whether they act synergistically on BMP signalling. Here, we show that BMPER binds to Tsg with high-affinity through the N-terminal region of BMPER. Cell-based inhibition assays show that the N-terminal region of BMPER is a better inhibitor of BMP-4 signalling than full-length BMPER. Furthermore, BMPER and Tsg cooperatively inhibit BMP-4 signalling suggesting they act in concert. Full-length BMPER, but not the N-terminal region, binds heparan sulphate (HS) proteoglycans at the cell surface. A disease-causing BMPER point mutation, P370L, which is found in the acid-catalysed internal cleavage site, was introduced to BMPER. The mutation caused a change to the location of the cleavage site, decreased HS-binding capability and increased BMP inhibition. These data suggest that localisation of BMPER at the cell surface reduces BMPER activity and consequently BMP inhibitory potential, which may be ameliorated by the mutation. To interrogate how these regions within BMPER are arranged, small-angle X-ray scattering and electron microscopy were used to show that BMPER is elongated which spaces the N-terminal BMP-binding and C-terminal cell-interactive regions.

Declaration

That no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

I.   The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

II.   Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

III.   The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of Copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV.   Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Acknowledgements

Firstly I would like to take this opportunity to thank my supervisors Professor Clair Baldock and Dr Katsuaki Inoue for their support and guidance throughout my PhD at both the University of Manchester and Diamond Light Source Ltd.

From the University of Manchester I would like to extend my thanks to the members of the Baldock lab, past and present, for any and all questions answered and advice given. I would like to thank Dr Stuart Cain of the Kimber lab for his advice and use of his inducible BMP constructs, and Dr Michael Leverentz for the use of his pCDH lentiviral constructs. I would like to thank my advisor Dr Steve Prince for his advice.

From Diamond Light Source I would like to acknowledge the staff of beamline B21, past and present, especially Dr Robert Rambo, for their advice and help, and to Professor Nick Terrill of beamline I22. I would also like to acknowledge any and all useful conversations with users of B21 while I was helping with user support. The exposure to groups working on a variety of projects has been very insightful for my own work.

I would like to thank members of all facilities, beamline and collaborators throughout my degree including: Biomolecular analysis facility, electron microscopy facility, mass spectrometry facility, beamline BM29 at the European Synchrotron Radiation Facility, beamline P12 at Petra III storage ring and Professor Christian Siebold and his lab at the Division of Structural Biology at the University of Oxford (STRUBI).

I would like to acknowledge the funding supplied by both the University of Manchester and Diamond Light Source and offer them my thanks.

I would like to thank my family and friends for their support throughout my degree. To my friends, I would like to thank their support for my work and equally for their support in taking my mind off work, from trips away to Saturday rugby with the lads at Broughton Park FC. I would like to thank my family for their unquestionable support throughout my PhD.

Finally I would like to offer my warmest and most sincere thanks to Alisa Vowell for your love, support and encouragement through my PhD, but also your understanding and patience when additional hours and days were required in the lab.

Abbreviation List

Abbreviations	Verbose
% (v/v)	Percentage of volume per volume
% (w/v)	Percentage of weight per volume
1D	One dimensional
A_{280}	Absorbance at 280 nm
A-domain	Adenylation domain
AUC	Analytical ultracentrifugation
BCA	Bicinchoninic acid
BioSAXS	Biological SAXS
BMP	Bone morphogenetic protein
BMPER	Bone morphogenetic protein binding endothelial cell precursor-derived regulator
BMPR	Bone morphogenetic protein receptor
BSA	Bovine serum albumin
CAR	Carboxylic acid reductase
CCD	Charge coupled device
CD	Circular dichroism
CHRD	Chordin domains
CMOS	Complementary metal-oxide-semiconductor
CORAL	Complexes with random loops
CPU	Central processing unit
CTF	Contrast transfer function
CV	Column volume
CV-2	Crossveinless-2
Da	Dalton
DAM	Dummy atom model
DAWN	Data Analysis WorkbeNch
DDD	Direct electron detector
DESY	Deutsches Elektronen-Synchrotron
DLP	Digital light processing
D_{max}	Maximum dimension of electron density
DMEM	Dulbecco's modified eagle medium
DMM	Double-multilayer monochromator
DPP	Decapentaplegic
dRI	Differential refractive index

DSD	Diaphanospondylodysostosis
e^-	Electron
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ERK	Extracellular signal-regulated kinase
ESRF	European Synchrotron Radiation Facility
$F(q)$	Form factor
FACS	Fluorescence activated cell sorting
FBS	Foetal bovine serum
FMDV	Foot-and-Mouth disease virus
FSC	Fourier shell correlation
g	gram
GA	Guinier approximation
GAGS	Glycosaminoglycans
GDA	Generic Data Acquisition
GDF	Growth differentiation factor
GFP	Green fluorescent protein
GI	Glucose isomerase
GIFT	Generalized indirect Fourier transform
GPA	Guinier peak analysis
HEK	Human embryonic kidney
HEK293-EBNA	Human embryonic kidney containing Epstein-barr virus nuclear antigen-1
HEK293-T	Human embryonic kidney containing SV40 large T antigen
HESC	Human embryonic stem cell
HPLC	High performance liquid chromatography
HS	Heparan sulphate
HSPG	Heparan sulphate proteoglycans
HSV-1	Herpes simplex virus-1
$I(0)$	Scattering intensity at the zero angle
$I(q)$	Scattering intensity
ICP4	Infected cell protein 4
ICP4N	N-terminus of the ICP4
IDR	Intrinsically disordered protein
IE	Immediate-early
IE3_19mer	19 base pair IE3

17

IFT	Indirect Fourier transform
IP	Immunoprecipitations
ISD	Ischiospinal dysostosis
JNK	Jun N-terminal kinase
k_a	Association rate constant
KCP	Kielin/chordin-like protein
K_D	Binding constant
K_d	Dissociation rate constant
L	Litre
LLC	Large latent complex
Lm-GpsB	Listeria monocytogenes GpsB
LTBP-1	Latent TGF-β binding protein 1
m	Meter
M	Molar
MMPs	Matrix metalloproteases
MUC2	Mucin-2
MUC5A	Mucin-5AC
MWCO	Molecular weight cut-off
N-BMPER	N-terminal cleavage product of BMPER
NFkB	Nuclear factor kappa beta
NMR	Nuclear magnetic resonance
NS	Nonsense mutation
NSD	Normalised spatial discrepancy
OPPF	Oxford protein production facility
$P(r)$	Pair-distance distribution function
PBS	Phosphate buffered saline
PBS-500	Phosphate buffered saline with 500 mM NaCl
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PDDF	Pair distance distribution function
PEI	Polyethylenimine
pSMAD	Phosphorylated SMADs
QCMD	Quartz crystal microbalance with dissipation monitoring
q_{max}	Maximum q range of X-ray scattering
R-domain	Reductase termination domain
R_g	Radius of gyration

RGM	Repulsive guidance molecule
R_H	Hydrodynamic radius
RIPA	Radioimmunoprecipitation assay
RONN	Regional order neural network
rpm	Revolutions per minute
RU	Response units
$S(q)$	Structure factor
S200	Superdex 200
S200i	Superdex 200 increase
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
Scw	Screw
SD	Standard deviation
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC-MALS	Size exclusion chromatography multi-angle light scattering
SEC-SAXS	Size exclusion chromatography small-angle X-ray scattering
SiN	Silicon nitride
SLA	Stereolithography
SMI	Super matrix inversion
SOC	Super Optimal broth with Catabolite repression
SOMO	Solution modeller
SPR	Surface plasmon resonance
STRUBI	Division of Structural Biology at the University of Oxford
TAE	Tris base, acetic acid and EDTA
TBS	Tris buffered saline
TBS-T	Tris buffered saline with Tween-20
TEM	Transmission electron microscopy
tGFP	Turbo green fluorescent protein
TGF-β	Transforming growth factor beta
TIL	Trypsin inhibitor-like domain
Tsg	Twisted gastrulation
V	Volts
V_c	Volume of correlation
vWF	von Willebrand factor
wp	Well plate

x g	Number of times g-force
XFEL	X-ray free-electron lasers
ΔN-BMPER	Volume of BMPER minus N-BMPER
$\Delta\rho(r)$	Difference of average electron density of solute and bulk solvent
$S_{20,w}$	Sedimentation coefficient
dn/dc	Refractive index increment
f/f_0	Frictional ratio
$\gamma(r)$	Gamma function
λ	Wavelength
$\rho(r)$	Average electron density of solute and bulk solvent
$\rho(s)$	Average electron density of bulk solvent

1 Introduction

1.1 Extracellular Matrix

The extracellular matrix (ECM) is composed of matrix molecules secreted by cells that provide mechanical support to the surrounding cells whilst eliciting biochemical responses by the incorporated molecules. The matrix has been defined as being composed of a 'core matrisome' containing approximately 300 proteins which can be categorised into: proteoglycans; collagens; modifiers of matrix structure and function; and matrix bound growth factors and secreted factors (Hynes and Naba, 2012). The matrix is difficult to study using biochemical methods due to its inherent insolubility conferred by its composition of large proteins that are often cross-linked, causing higher order oligomerisation (Hynes and Naba, 2012).

Proteoglycans are modified with large glycosaminoglycans (GAGs), such as heparan sulphate, keratan sulphate and chondroitin sulphate. These GAGs are linear polysaccharide chains that comprise of repeating disaccharide subunits (Caligur, 2008). GAGs are connected to core proteins such as perlecan or decorin, and their interactions with various growth factors facilitates signalling (Esko et al., 2009).

Collagens are very abundant and are the main structural protein in the matrix, providing tensile strength. Although there are 28 types of collagen, collagen I is the most common (Bella and Hulmes, 2017). Collagens are essential to many biological processes such as cell attachment, tissue scaffolding and tissue repair (Kadler et al., 2007). Fibrillar collagens require N- and C-terminal proteolytic processing of propeptides for assembly by members of the tolloid and a disintegrin and metalloproteinase (ADAM) metalloproteinase families (Sharma et al., 2017). Tolloids, ADAMs and matrix metalloproteases (MMPs) are grouped into the category of modifiers of matrix structure and function. Where tolloid and ADAM proteases cleave proteins to promote assembly, MMPs having a degradative function. They are responsible for processing the matrix, but also in the release of signalling molecules by cleavage of antagonists (Larrain et al., 2001).

There is a large group of matrix-bound signalling molecules and secreted factors (Hynes and Naba, 2012) including members of the transforming growth factor beta (TGF-β) superfamily. BMPs are the largest sub-group of the TGF-β superfamily with more than 20 members (Bragdon et al., 2011), and are inhibited by secreted glycoproteins such as BMPER, Tsg and chordin (Mulloy and Rider, 2015; Rider and Mulloy, 2010).

BMPER, the focus of this thesis, is a large, multi-domain glycoprotein secreted into the matrix. It is composed of five von Willebrand Factor (vWF) C domains, a vWFD domain and a trypsin inhibitor-like domain (TIL)(Figure 1.3-A). BMPER directly interacts with BMPs through the first vWFC domain (Zhang et al., 2007a; Zhang et al., 2008), and with the cell surface through the vWFD domain (Rentzsch et al., 2006). BMPER has been shown to have concentration dependant pro- and anti-BMP signalling functions. Mutations within BMPER cause a rare recessive skeletal disorder known as diaphanospondylodysostosis (DSD) (Funari et al., 2010), and an attenuated version ischiospinal dysostosis (ISD) (Kuchinskaya et al., 2016). DSD is often embryonic lethal (Gonzales et al., 2005) where ISD is less severe (Legare et al., 2017). This thesis will focus on the structure of BMPER, its interactions with extracellular regulators and the effect of a BMPER-P370L mutation.

1.2 Bone Morphogenetic Proteins

BMPs are secreted matrix binding signalling molecules, first discovered in the extract of bone (Urist, 1965) and able to regulate many processes throughout an organism. The BMP pathway is highly phylogenically conserved dating back 1.2 – 1.4 billion years (Salazar et al., 2016). Indeed, BMP signalling molecules are highly conserved across a range of species; Drosophila contains signalling molecules similar to BMPs -2 & -4, decapentaplegic (Dpp), and -5, -6, -7 & -8, screw (Scw) (Arora et al., 1994; Shimmi et al., 2005). Dpp and screw are shown to work together in *Drosophila* to determine dorsal cell fates in embryogenesis. Although initially identified in the formation of bone and cartilage, BMPs are essential for a wide-range of functions throughout embryogenesis (Eldar et al., 2002) and into adult life where BMP misregulation has been linked to diseases such as cancer (Ikushima and Miyazono, 2010) and vascular disease (Cai et al., 2012).

1.2.1   Expression and Secretion

BMPs are synthesised as large precursor proteins, of around 400 – 500 residues, consisting of a signal peptide, a pro-domain and the BMP growth factor (Figure 1.1-A). The signal peptide is required for secretion and the pro-domain is required for the mature BMP growth factor to fold and dimerise correctly (Figure 1.1-B). The BMP growth factor comprises a cysteine knot (Allendorph et al., 2007; Scheufler et al., 1999). Cysteine knots are a structural motif present in BMPs and other molecules that lack a globular core, and provide a rigid centre (Daly and Craik, 2011); a hydrophobic interface then stabilises the dimers (Scheufler et al., 1999). Each BMP monomer contains seven cysteine residues which form three intramolecular disulphide bonds with the final cysteine disulphide bonding to another BMP molecule to form the small, ~28 kDa BMP dimer (Figure 1.1-C).

Under certain circumstances the pro-domains of BMPs remain associated to the growth factors when secreted, but the BMPs are still biochemically active, as in the cases of BMP-7 (Sengle et al., 2008) and BMP-9 (Brown et al., 2005). Fibrillin-1 was shown to provide latency to BMP-7 by interacting with the pro-domain (Wohl et al., 2016).
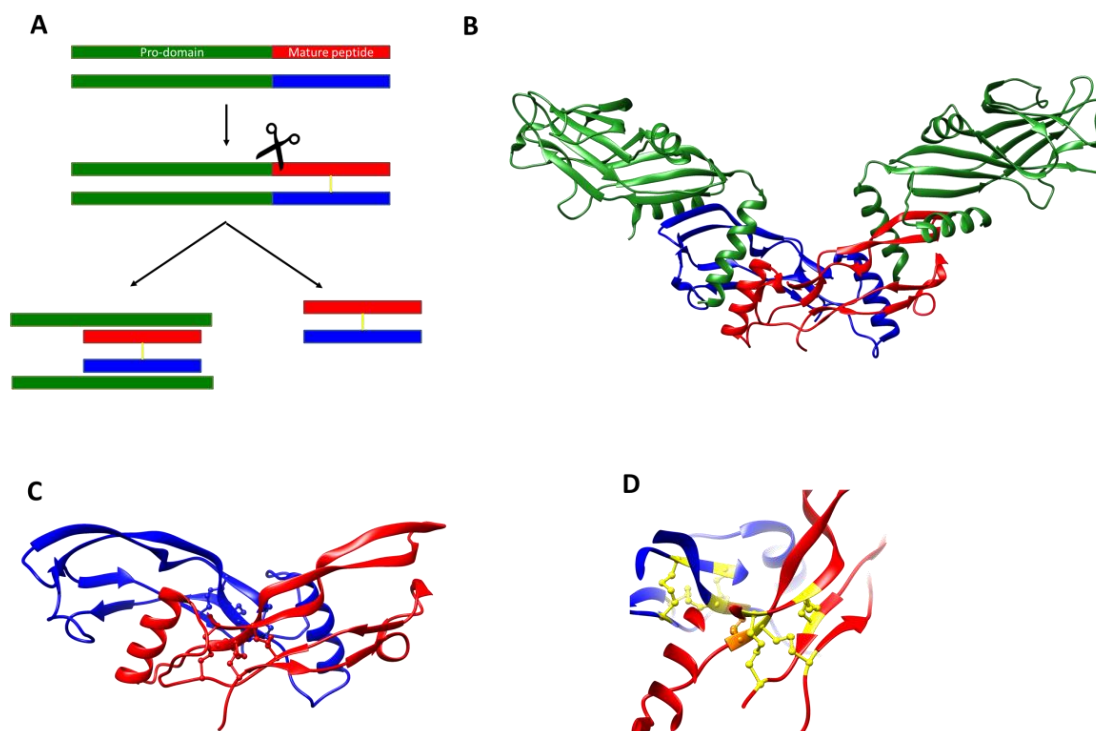
Figure 1.1 – BMP Synthesis and Processing

A) BMPs are synthesised as a polypeptide with a signal peptide, a pro-domain (green) and a mature peptide (blue and red). After post-translational modifications in the Golgi apparatus, the mature peptides form a disulphide bond and the pro-domains are cleaved by a serine protease (often furin). The resulting mature peptide is then secreted from the cell either in association with the cleaved pro-domains, or as a single mature peptide. B) The crystal structure of a mature peptide of BMP-9 associated with its pro-domain (PDB ID 4YCG (Mi et al., 2015)) and of the BMP-2 dimer (C) (PDB ID 3BMP (Scheufler et al., 1999)) with one chain of the dimer shown in red and the other shown in blue. D) A magnified image of BMP-2 where the three disulphide bonds making up a cysteine knot coloured in yellow and the intermolecular disulphide bond is coloured orange. Images were produced using UCSF Chimera (Pettersen et al., 2004).

Cleavage of the pro-domain occurs at a conserved serine protease site and furin is responsible for this cleavage (Heng et al., 2010), which recognises the 'RXXR' cleavage motif and cleaves after the final arginine. Heng et al. showed that knocking down BMP-2 or furin had the same effect on decidualization of human embryonic stem cells (HESCs) demonstrating that BMP-2 is critically dependant on furin cleavage for activation.

1.2.2 Signalling

BMP signalling is initiated by members of the BMP family binding to the BMP receptors (BMPR), which are members of the serine/threonine kinase family of receptors. The receptors belong to two different subfamilies, types I and II, where BMPR-I have five members: ALK1, ALK2, ALK3, ALK4, ALK6; and BMPR-II have three members: BRII, ActRIIa and ActRIIb (Attisano et al., 1993; Kingsley, 1994). The BMP ligands signal by binding to BMPR-I and -II on the cell surface causing the receptors to oligomerise. BMPs bind to receptors sequentially, where binding to BMPR-I occurs before binding to BMPR-II due to higher binding affinities to the type I receptor (Groppe et al., 2008; Weber et al., 2007). Allendorph et al. showed that each of the BMPRs have separate binding sites on the BMP molecule, which facilitates hetero-oligomerisation (Allendorph et al., 2006; Yadin et al., 2016). The receptors span the cell membrane and hetero-oligomerisation of the BMPRs and BMPs activates the intracellular serine/threonine kinase of BMPR-II (Figure 1.2) (Allendorph et al., 2006). Once BMPR-II is activated it can phosphorylate BMPR-I, which in turn phosphorylates the intracellular signalling molecules, SMAD-1/-5/-8, responsible for gene expression. There are eight SMAD molecules, classified into three subgroups based on their function. SMAD-1/-5/-8 become transiently associated with BMPR-I and become phosphorylated. In order for the phosphorylated SMADs to translocate through the nuclear membrane they require cooperative interactions with SMAD-4.
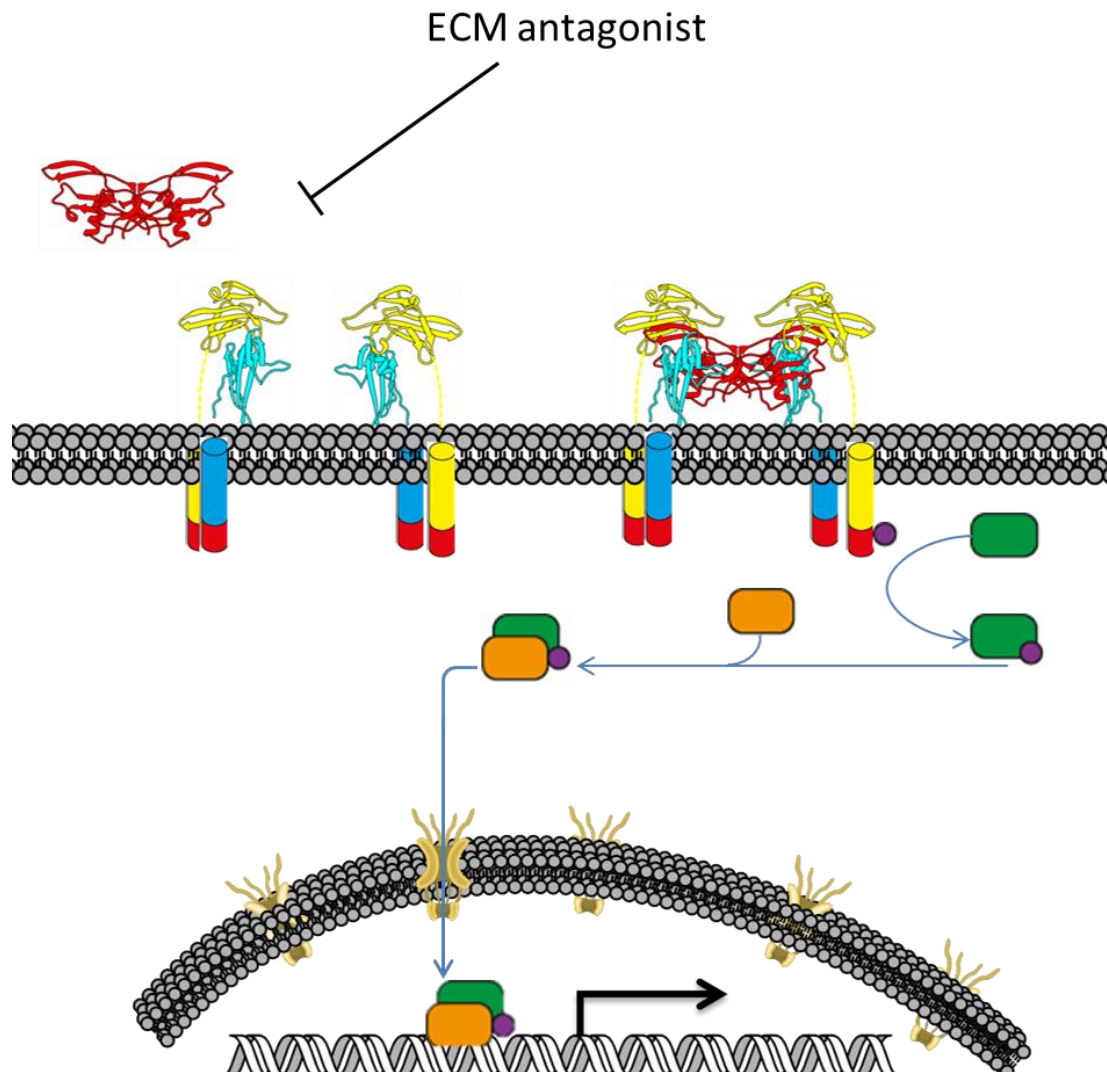
Figure 1.2 – BMP Signalling Pathway

BMPs dimers (red) in the matrix bind cooperatively to the ecto domains of BMP receptors type I (cyan) and II (yellow) causing BMPR-II to become an active kinase, able to phosphorylate BMPR-I. Phosphorylation (purple sphere) of BMPR-I leads to an intracellular signalling cascade where the SMAD proteins 1, 5, and 8 are phosphorylated (pSMAD, Green), and can translocate through the nuclear membrane after recruitment of SMAD-4 (orange).. BMP signalling can be modulated extracellularly though BMP antagonists, preventing binding to the BMPRs. The crystal structures are taken from PDB ID 2goo (Allendorph et al., 2006) and the images were created in UCSF Chimera (Pettersen et al., 2004).

Although BMP signalling can occur through the SMAD mediated pathway it is also mediated through SMAD independent pathways. These pathways include the extracellular signal-regulated kinase (ERK), MAP kinase p38, C-jun N-terminal kinase (JNK), ERK, and nuclear factor kappa beta (NFkB) (Bragdon et al., 2011). Through the initiation of these pathways, BMPs are able to elicit their effects on cell survival, apoptosis, migration, and differentiation (Bragdon et al., 2011).

1.2.3 Extracellular Regulation

In the matrix, BMPs are regulated by antagonists, consisting mainly of secreted multi-domain glycoproteins. Many of the regulators directly bind to BMPs and prevent them binding to their respective receptors (Groppe et al., 2002; Zhang et al., 2008). This introduction will focus on the extracellular regulation of BMP through the morphogenetic pathway which contains BMPs, chordin, Tsg, BMP-binding endothelial cell precursor-derived regulator (BMPER) and tolloid (De Robertis and Moriyama, 2016). Chordin, Tsg and BMPER are able to directly bind to BMP (Kišonaitė et al., 2016; Troilo et al., 2016a; Troilo et al., 2014; Zhang et al., 2007a), tolloid is a metalloprotease, of which chordin is a substrate (Berry et al., 2009; Berry et al., 2010; Blader et al., 1997; Piccolo et al., 1997).

1.3 BMP Endothelial Cell Precursor-Derived Regulator

The homologue of human BMPER, crossveinless-2 (CV-2), was initially discovered in the development of cross veins in *Drosophila*; CV-2 was shown to regulate the signalling of BMP homologues: Dpp and Gbb (Conley et al., 2000). Similar to the *Drosophila* homologue of chordin, Sog, the vWFC domains in CV-2 were thought to play a role in growth factor regulation (Conley et al., 2000). Murine BMPER was cloned and shown to have a similar BMP binding capacity to *Drosophila* CV-2 (Coffinier et al., 2002; Moser and Patterson, 2003). Human BMPER was expressed to provide a protein with similar BMP binding capacity (Binnerts et al., 2004). In this thesis BMPER will refer to human BMPER.

1.3.1 BMPER Expression and Secretion

The discovery of CV-2 and BMPER showed that there was strong sequence conservation across different organisms (Coffinier et al., 2002; Conley et al., 2000; Moser et al., 2003). The mammalian BMPERs contain an additional trypsin inhibitor-like (TIL) domain not present in CV2; the function of this domain is not known (Figure 1.3-A) (Coffinier et al., 2002; Conley et al., 2000; Moser et al., 2003). BMPER is a multi-domain secreted glycoprotein with five predicted N-linked glycosylation sites, four in the vWFC domains and one in the vWFD domain, although only the vWFC domains appear to be glycosylated (Figure 1.3-A) (Kamimura et al., 2004b). When BMPER is secreted it is found to be approximately 80 kDa by Western blot under non-reducing conditions, although under reducing conditions a band at 40 kDa appears

(Figure 1.3-B) (Binnerts et al., 2004). This indicates that there is cleavage within BMPER with the two cleavage fragments stabilised by a disulphide bond. Cleavage occurs in the 'GDPH' motif, between the aspartic acid and proline residues (Rentzsch et al., 2006). This cleavage sequence is conserved across many proteins (Figure 1.3-C). There has been much dispute over the cleavage mechanism of BMPER; initially a protease was thought to be the primary mechanism (Binnerts et al., 2004; Rentzsch et al., 2006; Serpe et al., 2008). However, it was shown that BMPER was not cleaved in a proteolytic manner, but cleavage was due to a decrease in the pH of the secretary system (Ambrosio et al., 2008). Similarly, work carried out on mucins showed that neutralising the secretory pathway prevents cleavage at the 'GDPH' motif (Lidell and Hansson, 2006; Lidell et al., 2003) building upon chemical studies into aspartyl-proline peptide bonds (Piszkiewicz et al., 1970).
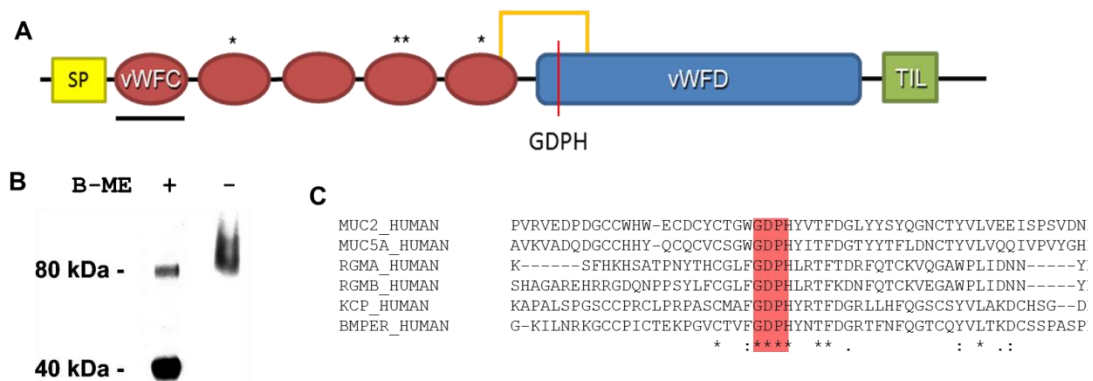
Figure 1.3 – Schematic Diagram of BMPER and its Cleavage Reaction

A) BMPER is made up of a signal peptide (yellow), five vWFC domains (red) with asterisks denoting N-linked glycosylation sites, a vWFD domain (blue) linked to the final vWFC domain by a disulphide bond and containing an acid catalysed 'GDPH' motif, and a trypsin inhibitor like domain (green). The first domain is underlined to show the location of direct BMP interaction. B) Western blot analysis of BMPER after recombinant expression and purification. A cleavage product is shown under reducing conditions showing that the cleavage fragments are disulphide linked. Image adapted from Binnerts et al (Binnerts et al., 2004). C) Sequence alignment of human proteins: mucins -2 and -5AC (MUC2_HUMAN and MUC5A_HUMAN), repulsive guidance molecules A and B (RGMA_HUMAN and RGMB_human), kielin/chordin-like protein (KCP_HUMAN) and BMPER (BMPER_HUMAN). Sequences were aligned with Clustal Omega (Sievers and Higgins, 2014a, b). The red highlighting shows the conservation of the 'GDPH' motif across proteins, the '*' represents a fully conserved residue, ':' represents a conserved residue with highly similar properties and '.' represents a conserved residue with loosely similar properties.

1.3.2   Sites of BMPER Expression

In humans, *bmper* was shown to be expressed throughout the body, with brain and lung being the most prevalent, especially in foetal lung (Figure 1.4-A) (Binnerts et al.,

2004). As the name BMPER suggests, it is not unexpected for BMPER to be expressed in these regions due to the high number of endothelial cells in the brain and lung (Crapo et al., 1982; Lippmann et al., 2012).

Although BMPER developmental studies have been carried out in multiple organisms, here I will focus on mammalian models. The *bmper*$^{-/-}$ mice have a perinatal lethal phenotype (Ikeya et al., 2006). Additional defects were observed throughout the mouse including skeletal effects, decreased body size (Figure 1.4-A,-B), eyes, kidneys and at low penetrance, exencephaly was also observed (Figure 1.4-B) (Ikeya et al., 2006).
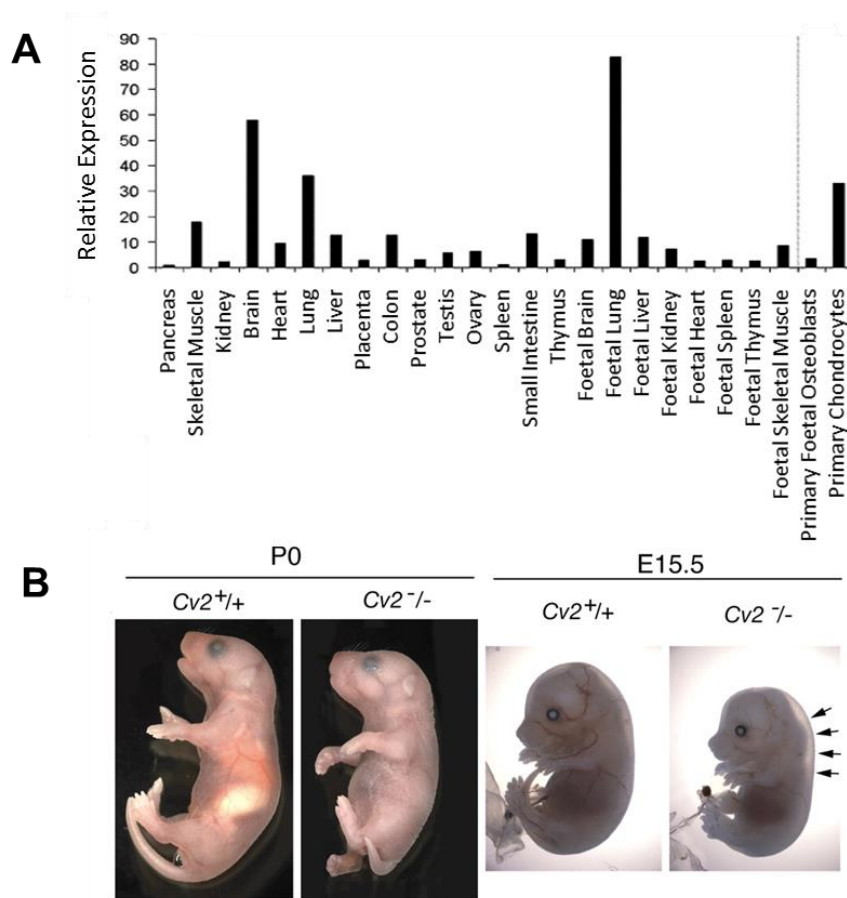
Figure 1.4 – Expression Levels and Location of BMPER Expression and Mouse Knockouts

A) Expression of BMPER mRNA in human foetal tissues and primary osteoblasts and chondrocytes (shown after the dashed line). mRNA expression levels are relative and detected by qPCR. Image adapted from Binnerts et al (Binnerts et al., 2004). B) Mouse knockout of BMPER (CV-2-/-) compared to wildtype mice (CV-2+/+) at postnatal day 0 (P0) and embryonic day 15.5 (E15.5). Arrows indicate exencephaly. Images are taken from Ikeya et al (Ikeya et al., 2006).

1.3.3   BMPER Inhibits BMP Signalling by Directly Binding BMPs

BMPER has been shown to interact with BMPs -2, -4, -6 and -7 as shown by immunoprecipitation experiments (IP) (Binnerts et al., 2004; Coffinier et al., 2002; Moser et al., 2003). Due to the similarity with chordin-BMP binding, it was expected that the vWFC domains would bind BMPs and so surface plasmon resonance (SPR) experiments were carried out to determine which domains were responsible (Zhang et al., 2007a). Zhang et al. showed that the first vWFC domain of BMPER interacts directly with BMP-2 and -7 and growth differentiation factor (GDF) -5 with high affinity (K_D of 2.4, 7.0 and 34.0 nM respectively). Zhang et al. also showed that vWFC domains 2-5 were unable to bind to BMPs or GDF5 through IPs. To determine the

interaction site of BMPER with BMP-2, BMP-2 mutations were generated guided by the crystal structures of BMP-2, BMPR-IA and ActR-IIB (PDB accession code 2GOO) (Allendorph et al., 2006). Mutations of both L100K and N102D in BMP2 appeared to disrupt the binding of vWFC1 to BMP-2 by two orders of magnitude (from 2.4 nM to 660 nM). Interestingly the L100P mutation did not disrupt binding indicating that BMPER binds to BMP-2 by a mainly hydrophobic interaction. Subsequently the BMPER-BMP-2 complex crystal structure confirmed that the predicted sites on BMP-2 did interact with BMPER (Figure 1.5-A) (Zhang et al., 2008). Interestingly the vWFC1 domain of BMPER interacts with BMP-2 through a tripartite interaction including an N-terminal extension, termed the clip-region. Figure 1.5-B shows a magnified version of how the clip region interacts with each of the BMP-2 monomers. With this knowledge, further SPR experiments to determine the required portion of the vWFC1 domain were carried out. By segmenting the domain into the clip, sub-domain (SD) 1 and SD2, it was shown that the clip and SD1 are sufficient for binding to BMP-2 (Zhang et al., 2008). Furthermore, the crystal structure of BMPER and BMP-2 has shown directly how BMPER can inhibit BMP signalling by interaction of SD1 with BMP-2 at the same region responsible for binding the ActR-IIb ecto-domain while the clip region simultaneously interferes with the BMPR-IA binding region (Figure 1.5-C).
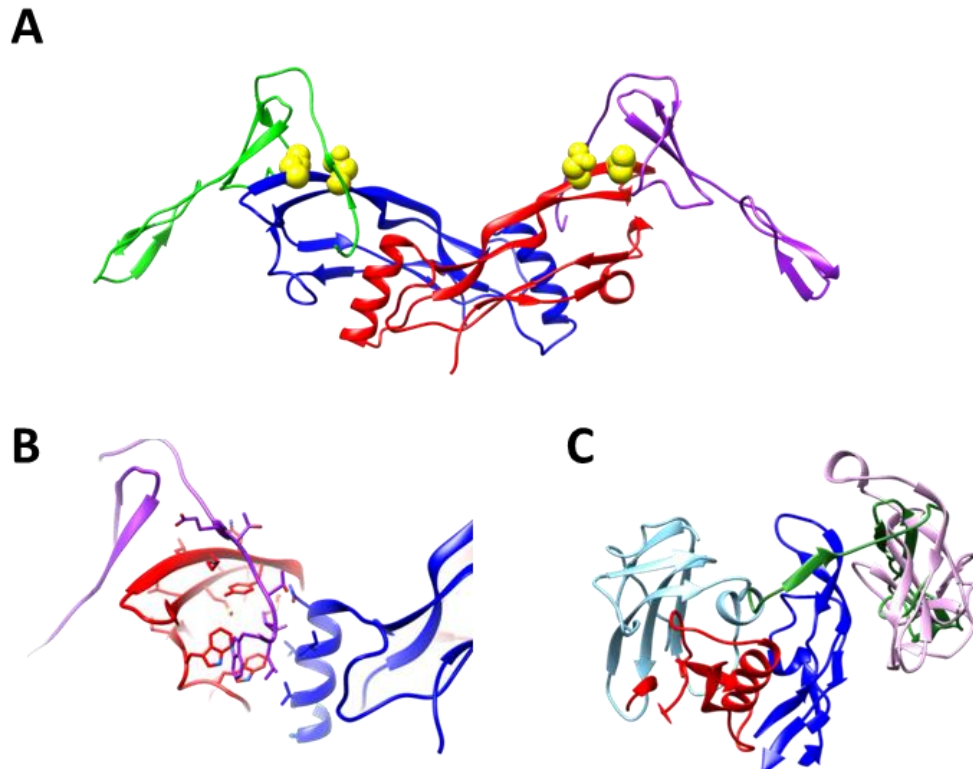
Figure 1.5 – Structural Information of BMPER-BMP Binding

A) Crystal structure of BMP-2 (red and blue) with the first vWFC domain of BMPER binding to each side of the molecule (green and purple) (PDB ID 3BK3) (Zhang et al., 2008). The yellow residues are locations of the L100K and N102D mutations on BMP which greatly reduced binding. B) A magnified view of Figure 1.5-A showing the location and interactions of the residues of the clip region. C) Superimposed crystal structures of BMPs showing how BMPER overlaps with the binding sites of the BMPRs (PDB ID 2GOO) (Allendorph et al., 2007). The BMP-2 dimer is represented by the blue and red chains, the vWFC1 of BMPER is shown in green with the BMPRs shown in pink (type I) and pale blue (type II).

In *Drosophila* it was shown that CV-2 was able to bind both Dpp and the human BMPR-I (Figure 1.6-B) (Serpe et al., 2008) which was specific for the type I receptor. It was also shown that Dpp could simultaneously co-IP both CV-2 and BMPR-I suggesting that they may form a tripartite complex. Whether this would be via the N- or C- terminus of BMPER is unknown. Furthermore, BMPER prevented the binding of BMPs to the BMPR-IB receptor (Figure 1.6-F) (Ambrosio et al., 2008).

1.3.4 Interaction of BMPER with Heparan Sulphate

BMPER has also been shown to bind to additional matrix components such as heparan sulphate proteoglycans (HSPGs) (Figure 1.6-A) (Rentzsch et al., 2006) and BMPR-IB shown by co-localisation (Figure 1.6-B) (Serpe et al., 2008). BMPER is

32

also involved in interaction with other BMP regulators, namely chordin, as shown by IPs and SPR (Figure 1.6-C) (Ambrosio et al., 2008; Rentzsch et al., 2006; Zhang et al., 2010a; Zhang et al., 2007b). Whereas there is conflicting literature surrounding an interaction with Tsg where a direct interaction was not observed by SPR (Zhang et al., 2007a) but immunoprecipitation detected an interaction (Figure 1.6-D) (Ambrosio et al., 2008), however this could be indirectly mediated.

Work completed by Rentzsch et al, showed that zebrafish BMPER, which shares 66% sequence identity with human BMPER, was able to bind to heparin beads through the vWFD domain (Figure 1.6-A) (Rentzsch et al., 2006). This was validated with a deletion mutation of positively charged residues, '393-396', in the vWFD domain which showed a reduction in heparin binding (Figure 1.6-A). Furthermore, the N-terminus of BMPER (N-BMPER) did not appear to bind to heparin beads, and so localising the binding to the C-terminus (Rentzsch et al., 2006).

1.3.5   Interaction of BMPER with other BMP Regulators

The pro-BMP activity of BMPER is mediated via interactions with other BMP regulators such as chordin and Tsg. The interactions of BMPER and chordin have been well characterised by biochemical and biophysical methods (Ambrosio et al., 2008; Rentzsch et al., 2006; Zhang et al., 2010a; Zhang et al., 2007b). SPR analysis of truncations of BMPER and chordin were completed to show that SD2 of vWFC1 to vWFC4 of BMPER were required to bind vWFC2 of chordin (Zhang et al., 2010a). Ambrosio et al showed that the ability of BMPER to bind BMP-4 was increased in the presence of chordin (Figure 1.6-C) (Ambrosio et al., 2008). A mechanism of action for BMPERs pro-BMP function though chordin interaction was hypothesised; the BMP:chordin complex binds to the surface tethered BMPER, increasing the concentration of the complex, and allowing tolloid to cleave chordin, releasing the BMPs for signalling. Surprisingly, Ambrosio et al also showed that Tsg and BMPER interact (Figure 1.6-E). Tsg binding increased the affinity of BMPER-BMP-4 binding (Figure 1.6-E) and a BMPER-Tsg-BMP-4 ternary complex was formed, in a similar way to chordin (Ambrosio et al., 2008). Furthermore CV-2 and Tsg additively block the binding of BMP-4 to BMPR-IB (Figure 1.6-E).
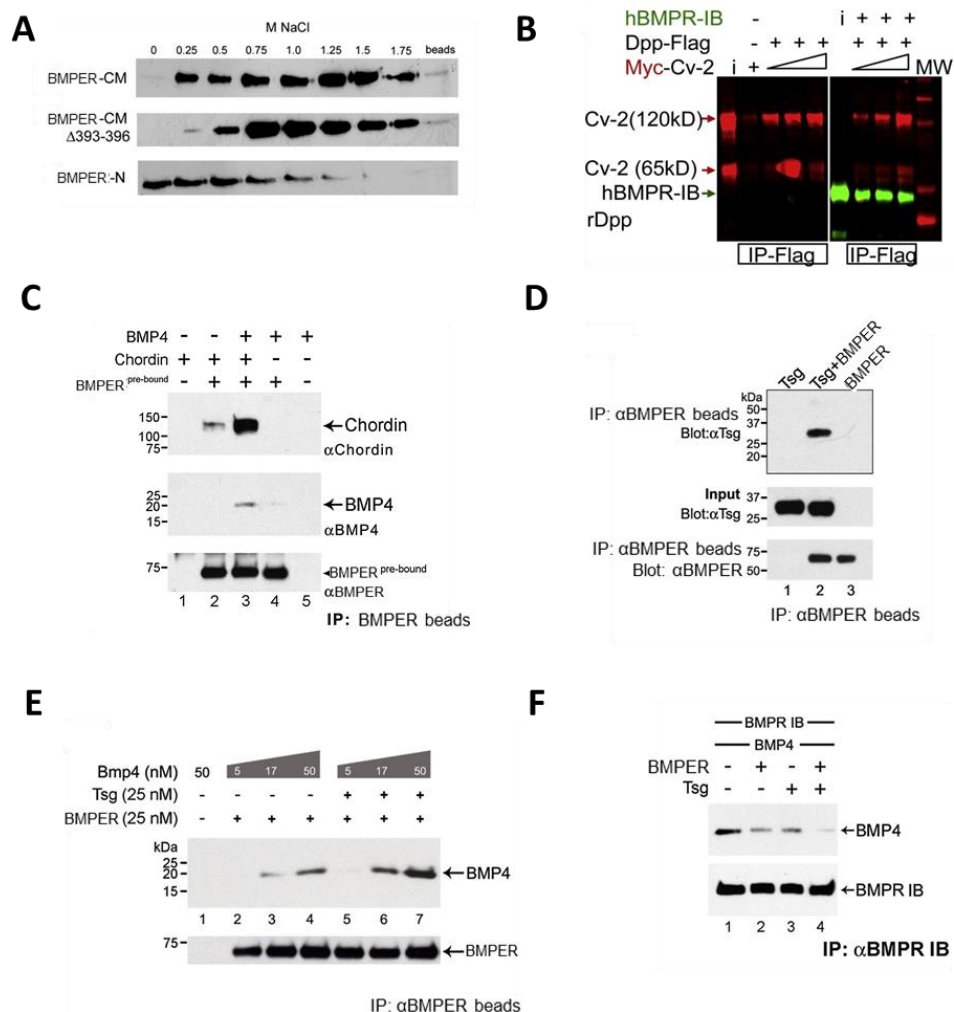
Figure 1.6 – BMPER Interactions with Matrix Components

A) Purified Zebrafish BMPER, including FL-BMPER, a BMPER deletion of residues 393-396 and N-BMPER, was bound to heparin beads and washed with increasing concentrations of NaCl. FL-BMPER bound most tightly, followed by the deletion mutant whereas N-BMPER bound much less. This localised the HSPG binding to the vWFD domain. This image was taken from Rentzsch et al (Rentzsch et al., 2006). B) Immunoprecipitation of Drosophila CV-2 and human BMPR-IB with Dpp-FLAG. Red bands represent CV-2 detected by a Myc-tag and green bands represent BMPR-IB (i = input). This image was taken from Serpe et al (Serpe et al., 2008). C) Co-IPs of BMPER, chordin and BMP-4 using BMPER conjugated to beads showing that BMPER can bind to chordin and BMP-4 but to a greater extent when all are present. D) Immunoprecipitation showing that BMPER interacts with Tsg. E) In the presence of Tsg, BMPER is able to bind to BMP-4 more efficiently. F) Immunoprecipitation showing that BMPER and Tsg inhibit BMP-4 binding to BMPR-IB. Images from C-F were modified from Ambrosio et al (Ambrosio et al., 2008).

Chordin is a member of the cysteine rich chordin family of proteins. It is a 120 kDa secreted modular glycoprotein made up of four vWFC and four chordin specific domains (CHRD) which are located between the first and second vWFC domain. The

vWFC domains of chordin can interact directly with BMPs -2, -4 and -7 and prevent the interaction of chordin with BMPRs (Figure 1.7-B) (Piccolo et al., 1996). Chordin is cleaved by tolloid at two locations, between vWFC1 and CHRD 1 and also between vWFC3 and -4 to produce ΔN- and ΔC-chordin, as shown in Figure 1.7-A (Larraín et al., 2001; Piccolo et al., 1997). The low resolution structure of chordin was determined by negative stain transmission electron microscopy (TEM) resulting in a horseshoe shaped structure (Figure 1.7-C) (Troilo et al., 2014). vWFC domains and the SAXS model of the CHRD domains were fitted within the TEM model with the BMP-2 crystal structure docked into the area that would represent binding to the vWFC1 and vWFC3 domains.

Twisted gastrulation is a small ~32 kDa secreted glycoprotein that acts as both an agonist and antagonist of BMP signalling (Troilo et al., 2016b). The SAXS model of Tsg reveals an elongated molecule with glycans clustered on one face of the molecule (Figure 1.7-C) (Troilo et al., 2016a). Tsg acts as a BMP antagonist by binding to the C-terminal vWFC domains of chordin (Troilo et al., 2016a) and enhancing formation of the BMP:chordin complex, forming a tripartite complex (Chang et al., 2001; Ross et al., 2001; Scott et al., 2001). The ternary complex is involved in the sequestration of BMP-2 and allows for the dorsal-ventral diffusion gradient to form as the complex moves through the matrix by facilitated diffusion. BMP-2 is released from the complex through the cleavage of chordin by tolloid proteases (Ambrosio et al., 2008). Tsg also works as an agonist of BMP signalling by enhancing the ability of tolloid to cleave chordin and release the BMP signalling molecule (Larrain et al., 2001). The mechanism of how Tsg enhances chordin cleavage is not well understood but it is thought that chordin is required to undergo a conformational change in Tsg:BMP-2:chordin complex, allowing tolloid to cleave chordin (Figure 1.7-D).
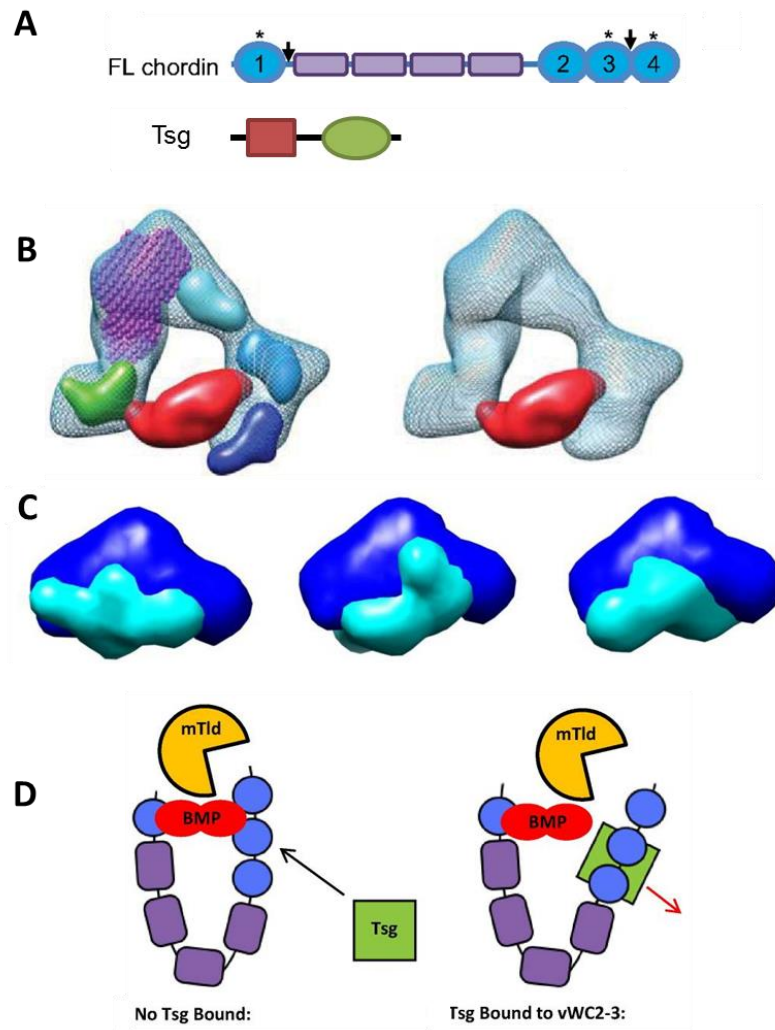
Figure 1.7 – Structure and Interaction of Tsg and Chordin with BMPs

A) Schematic diagram of the chordin domain structure, where the blue circles represent vWFC domains and the purple rectangles represent CHRD domains, arrows represent the cleavage sites recognised by tolloid and the '*' represent the domains that bind BMPs. Tsg with the red square representing the cysteine rich domain and the green oval representing the Tsg domain. Chordin schematics taken from Troilo et al. (Troilo et al., 2014). B) TEM structure of chordin with BMP-2 modelled into the density. Images take from Troilo et al. (Troilo et al., 2014). C) SAXS modelling of Tsg with glycans mapped through multiphase SAXS modelling. Image taken from Troilo et al (Troilo et al., 2016a). D) Mode of action of tolloid on the BMP:Tsg:chordin complex to release BMP. Image taken from Troilo et al (Troilo et al., 2016a).

Expression of recombinant, functional Tsg has been shown to be sensitive to the expression system used in its production. Previous work has showed that production of Tsg from mammalian, insect and bacterial cells resulting in markedly different binding to BMPs, with the difference being glycosylation (Billington Jr et al., 2011). This interesting difference in the functional properties of protein binding may account for the conflicting results shown for the binding of BMPER to Tsg, where no binding

was shown by SPR using Tsg produced in SF9 cells, producing truncated glycans (Zhang et al., 2007b), but interactions were shown though co-IPs of mammalian produced protein (Ambrosio et al., 2008).

Tsg homozygous null mice have skeletal defects where the mouse is smaller than its wildtype counterpart, and has a kinked tail as a result of vertebrate abnormalities (Sun et al., 2010). This is observed in the surviving mice as the study showed that there was a high mortality rate for Tsg null mice, where few survive to breeding age.

1.3.6 Concentration-Dependent Effects of BMPER on BMP signalling

How BMPER is able to regulate BMP signalling has been a point of contention in the literature as both a pro- (Conley et al., 2000; Ikeya et al., 2006; Kamimura et al., 2004a; Rentzsch et al., 2006) and anti- (Binnerts et al., 2004; Moser et al., 2003; Rentzsch et al., 2006; Zhang et al., 2007b) BMP signalling mechanisms have been described. Kelley et al showed that BMPER effects on BMP signalling occured in a concentration dependant manner (Kelley et al., 2009). This was shown by monitoring the levels of pSMADs in mouse endothelial cells (Figure 1.8-A). When BMPER was at a concentration lower than BMP-4, there was increased pSMAD, reflecting an increase in BMP signalling. However, when BMPER exceeded the concentration of BMP-4 the level of pSMAD decreased. When repeating the experiment with another BMP inhibitor, noggin, pSMAD levels decreased in a concentration dependent manner (Figure 1.8-A) (Kelley et al., 2009). BMP inhibition at higher BMPER concentrations is consistent with a 2:1 molar stoichiometry which was shown by Zhang et al (Zhang et al., 2007b).

The concentration dependence of BMPERs pro- and anti-BMP function was further tested in a biologically relevant model using mouse lung embryonic fibroblasts (MEFs) treated with BMP-4 and varying concentrations of BMPER (Figure 1.8-B). The level of apoptosis was detected by cleavage of caspase-3, where BMP-4 protects cells from apoptosis. When cells were treated with BMP-4 and a lower concentration of BMPER the cells were further protected from apoptosis, whereas a higher concentration of BMPER resulted in increased apoptosis(Figure 1.8-B) (Kelley et al., 2009).
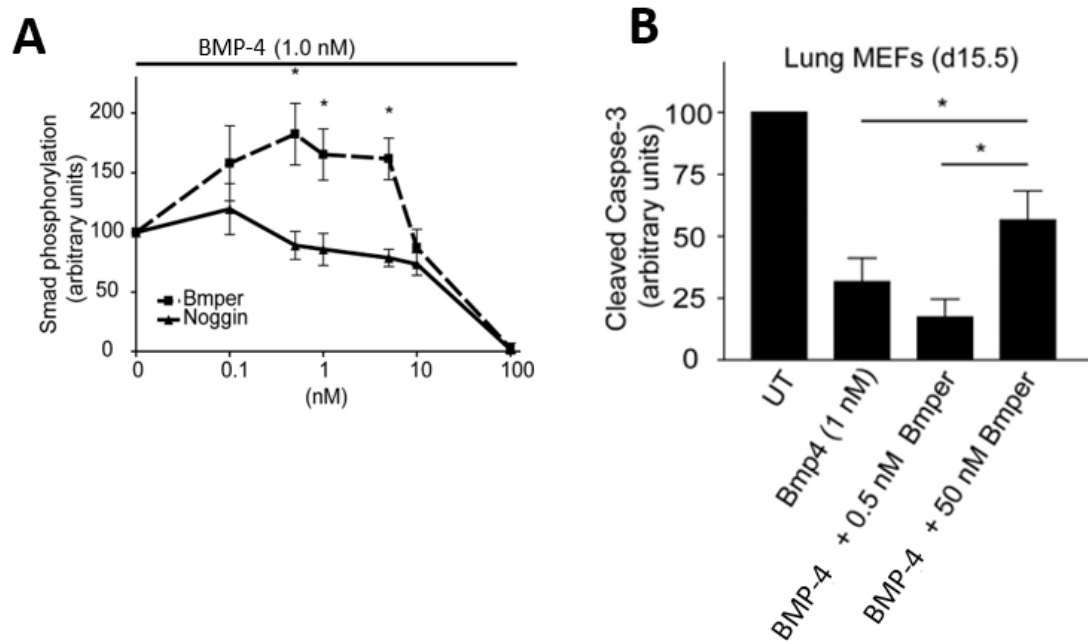
Figure 1.8 – BMPERs Pro- and Anti- BMP Signalling Effects

A) Analysis of SMAD phosphorylation when BMP-4 is added to mouse endothelial cells with varying concentrations of BMPER and noggin. BMPER enhances BMP-4 signalling at concentrations lower than BMP-4 and inhibits it at concentrations higher than BMP-4. Whereas noggin only inhibits BMP-4 signalling. B) Quantification of caspase-3 cleavage in lung mouse embryonic fibroblasts showed that BMPER can enhance the effect of BMP-4 at concentrations lower than BMP-4 and enhance caspase-3 cleavage by the inhibition of BMP-4 when at concentrations higher than BMP-4. Images taken from Kelley et al (Kelley et al., 2009).

1.3.7 Disease Causing Mutations in BMPER

Mutations in BMPER have been shown to cause a range of pathologies which have been characterised as two diseases: DSD (Ben-Neriah et al., 2011; Funari et al., 2010; Kuchinskaya et al., 2016; Legare et al., 2017; Scottoline et al., 2012; Zong et al., 2015) and ISD (Almasri et al., 2017; Kuchinskaya et al., 2016). Of the mutations recorded for DSD, most appear to be nonsense (NS) mutations (Ben-Neriah et al., 2011; Funari et al., 2010; Tasian et al., 2012; Vatanavicharn et al., 2007) while one results in a missense mutation which coincides with the 'GDPH' cleavage motif of BMPER (Ben-Neriah et al., 2011). For ISD, the mutations are both nonsense and missense mutations and are equally distributed over the length of BMPER (Kuchinskaya et al., 2016; Zong et al., 2015).

Number	cDNA	Protein	In-silico analysis	Diagnosis	Reference
1	c.925C>T	p.Gln309*	NS	DSD	(Funari et al., 2010)
2	c.26_35del10ins14	p.Ala9Glufs*4	NS		(Funari et al., 2010, Vatanavicharn et al., 2007, Ben-Neriah et al., 2011, Tasian et al., 2012)
3	c.1032+5G>A		Splice		
4	c.514C>T	p.Gln172*	NS		(Funari et al., 2010)
5	c.1109C>T	p.Pro370Leu	Missense		
6	c.1638T>A	p.Cys546*	NS		
7	c.310C>T	p.Gln104*	NS		(Ben-Neriah et al., 2011,)
8	c.251G>T	p.Cys84Phe	Missense	ISD	(Zong et al., 2015)
9	c.1078+5G>A		Splice		
10	c.416C>G	p.Thr139Arg	Missense		(Kuchinskaya et al., 2016)
11	c.942G>A	p.Trp314*	NS		
12	c.1672C>T	p.Arg558*	NS		
13	c.1988G>A	p.Cys663Tyr	Missense		
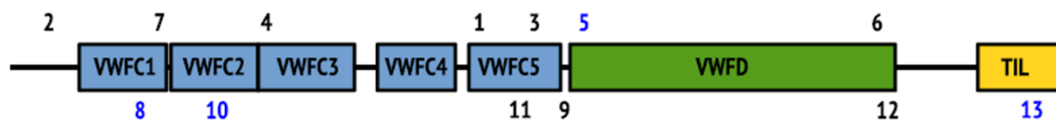
Figure 1.9 – Disease Causing Mutations of BMPER

Documented mutations involved in DSD and ISD with their locations mapped onto BMPER where the in silico predictions were carried out using SIFT (http://sift-dna.org), AlignGVGD (http://agvgd.iarc.fr/ and MutationTaster (http://www.mutationtaster.org/). The schematic shows the locations of the mutations with DSD on the top and ISD on the bottom. The black numbers are truncating mutations and the blue are missense mutations. This table and image was adapted from Kuchinskaya et al (Kuchinskaya et al., 2016).

Both DSD and ISD are rare, distinct skeletal dysplasias despite their similarities as shown in Table 1. Although DSD was not identified until 2005 (Gonzales et al., 2005), several cases reported earlier describe this disorder (Nisbet et al., 1999; Prefumo et al., 2003). Whilst in the majority of cases DSD is lethal, one patient survived to nine years of age (Legare et al., 2017). Although the patient is the longest lived of DSD sufferers, they display all of the symptoms recorded in Table 1, and required ventilation for the first 4.5 years of life although now only requires ventilation at night (Legare et al., 2017).

Diaphanospondylodysostosis	Ischiospinal dysostosis
Life expectancy	
Lethal or sublethal	Not lethal
Respiratory	
Marked respiratory insufficiency	Thoracic hypoplasia
Bell-shaped thorax	
Spine	
Low vertebral and sacral ossification; no ossification below T12	Delay in spinal ossification
Short neck/cervical spine	Hypoplasia of lumbosacral spine
Vertebral segmentation defects	Vertebral segmentation defects
	Kyphoscoliosis
Ribs	
Decreased number of ribs	Decreased number of ribs
Thin ribbon-like ribs	Floating 12th rib
Posterior rib gaps	Posterior rib gaps, fused ribs
Pelvis	
Unusual downward tilt of ischiopubic rami	Ischial hypoplasia
Upward widening of the iliac wings	Narrow iliac bones
Renal	
Cystic kidney with or without nephroblastoma	Cystic kidney with or without nephroblastoma
Height	
Short spine and short stature	Short spine and short stature

Table 1 – Phenotypes of DSD and ISD

The table was adapted from Legare et al (Legare et al., 2017).

Although there are several clinical reports on the state of DSD and ISD, to date there have been no biochemical studies to determine the cause of the phenotypes.

1.4 **Small-Angle X-ray Scattering**

The use of X-rays in the study of materials dates back to the Bragg formulation of X-ray diffraction, where it was observed that crystalline solids have a predictable diffraction pattern (Equation 1-1) (Bragg and Bragg, 1913). From this formula a whole field of X-ray studies has evolved.

$$n\lambda = 2dsin\theta$$

Equation 1-1 – Bragg's Law

Where n is a positive integer, λ is the wavelength of the incident wave, d is the distance between atomic layers and θ is the angle of incidence.

1.4.1   *History of Small-Angle X-ray Scattering*

Small-angle X-ray scattering (SAXS) was a concept developed by André Guinier (Guinier, 1939). Guinier made use of monochromatic X-ray radiation scattering to discern the size of Al-Cu grains, from which the Guinier Approximation was derived (section 1.5.2). This formula holds true for both X-ray and neutron scattering (SANS).

The work completed by Guinier, utilising X-ray scattering, led to the use of SAXS to study a variety of materials. Kratky, best known for his work in polymer physics and the study of worm-like chains (Kratky and Porod, 1949b), developed the Kratky plot to observe the flexibility of systems (section 1.5.4). Porod used SAXS to describe the dense arrangements of colloids, micelles and fibres (Kratky and Porod, 1949a). Porod's law describes how the signal of a SAXS curve decays at a rate of q^{-4} at high angles of scatter. Porod's law was later derived by Peter Debye (Debye et al., 1957) and their findings have been used to calculate the volumes of particles obeying these laws (Rambo and Tainer, 2011).

Glatter represented reciprocal space scattering in real space by way of an Indirect Fourier Transform (IFT) (Glatter, 1977a, b). With real space or pair distance distribution functions (PDDF), or $P(r)$, it was possible to achieve greater understanding of studies samples (section 1.5.3).

With the advances in the field as a whole, a more targeted development was occurring in the field of biological SAXS (BioSAXS) by Svergun. The reconstruction of biological sample shapes from SAXS curves was introduced with a suite of programs known as ATSAS (Konarev et al., 2006). Further developments have allowed the utilisation of atomic structures of proteins to generate accurate quasi-atomic protein models from SAXS data (Franke et al., 2017; Petoukhov et al., 2012; Petoukhov et al., 2007).

Along with software advances there have been advances in data collection hardware, moving from in-house X-ray sources to 3rd generation synchrotrons, capable of high flux (photons per second) and high data collection from complementary metal-oxide-semiconductor (CMOS) detectors, with high signal to noise efficiency. Furthermore, the ability to produce high levels of protein by recombinant methods rather than from source organisms has also allowed researchers to improve their SAXS studies.

1.5 Principles of SAXS

1.5.1 X-ray Scattering

Unlike X-ray crystallography, SAXS is inherently a contrast based method, where the signal is derived from the difference in the contrast of the average electron density ($\Delta\rho(\boldsymbol{r})$) of the solute ($\rho(r)$) (0.44 e$^-$/Å^3) and of the bulk solvent ($\rho(s)$) (0.33 e$^-$/Å^3) as shown in Equation 1-2.

$$\Delta\rho(\boldsymbol{r}) = \rho(r) - \rho(s)$$

Equation 1-2 – SAXS Contrast Differences

Where $\Delta\rho(\boldsymbol{r})$ is the difference in the average electron density, $\rho(r)$ is the electron density of the protein in solution and $\rho(s)$ is the electron density of the bulk solvent.

SAXS relies on the elastic scattering of photons, that is, the kinetic energy of a particle is retained after interacting with a material, but the direction of the photon is modified. In this case the direction is the angle of scatter and can described by Equation 1-3, where the values denoting the momentum transfer vector are: $q = s = h = k = Q$, and can be used interchangeably, although q and s are most often used in BioSAXS.

$$q = \frac{4\pi}{\lambda} . \sin(\theta)$$

Equation 1-3 – Momentum Transfer Equation

Where q is the momentum transfer, 2θ is the angle between the incident and scattered X-rays, and λ is the wavelength of light.

Whilst q is represented by Equation 1-3, the scattering intensity ($I(q)$) can also be represented as a product of the structure factor ($S(q)$) and the form factor ($F(q)$) as shown in Equation 1-4. The form factor is described as the scattering contribution of an isolated molecule, and is used to measure internal distances. The form factor provides characteristic scattering curves for the shape of the molecule. The structure factor is the contribution in densely packed systems where the intermolecular distances come within the same magnitude as those of the intramolecular distances and so the interference patterns of the neighbour contribute to the overall scattering.

$$I(q) = F(q) . S(q)$$

Equation 1-4 – Scattering Intensity as a Product of the Form and Structure Factor

Where $I(q)$ is the scattering intensity, $F(q)$ is the form factor and $S(q)$ is the structure factor.

Most SAXS studies measure only the form factor of proteins, and wish to reduce contributions from the structure factor. Under these conditions the structure factor has a value of 1. In a concentrated solution, where proteins become aggregated, the structure factor has a value greater than 1, and so the low q regions have an increased scattering contribution (Figure 1.10-A). In studies where there is active repulsion of molecules within the solution, the value of the structure factor decreases below 1 and there is a decreased scattering contribution at low q (Figure 1.10-B, C).
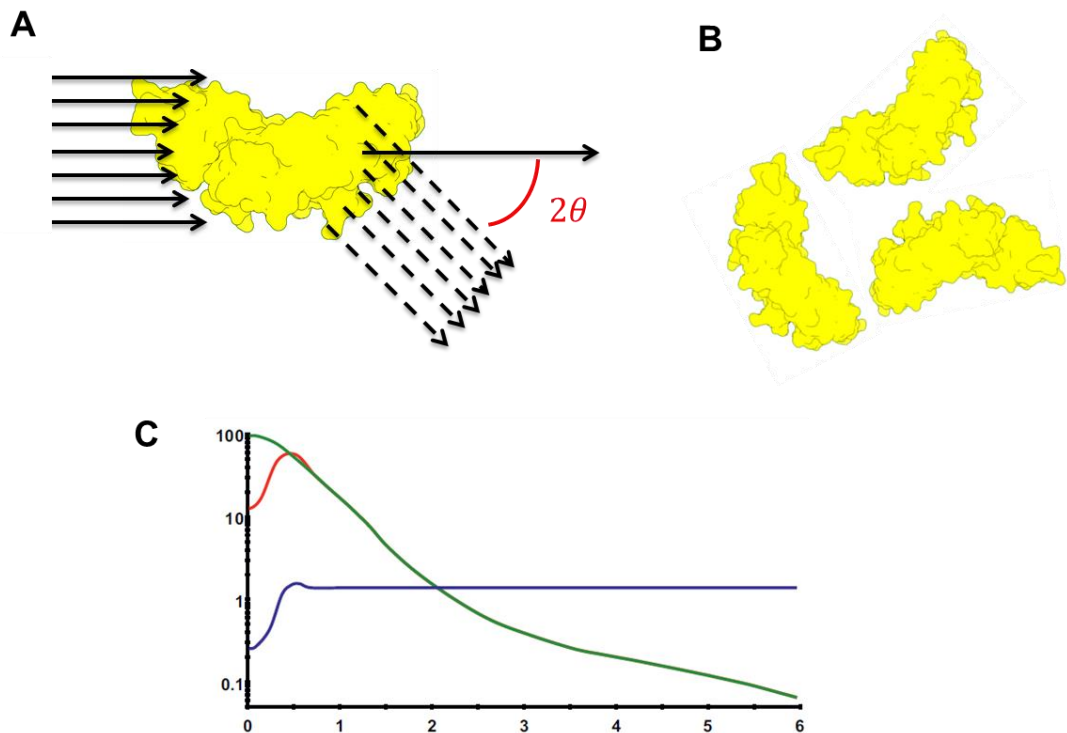
Figure 1.10 – Depiction of Form and Structure Factors

Schematic diagram of the form factor contributions from a single particle (A) and the structure factor contributions from multiple molecules at intermolecular distances in the same magnitude as the intramolecular distances, i.e., less that the D_{max} (B). The structure is generated from PDB 3BK3 (Zhang et al., 2008). C) A scattering curve (red) shown with the form factor (green) and structure factor contributions (blue).

The resolution of SAXS is lower than other structural techniques in the range of 10 – 30 Å. At high q_{max} values the random tumbling effect of protein in solution means that the scattering of all proteins begin to converge (Figure 1.11) (Svergun et al., 2001). This prevents the technique from being used for high resolution studies. As shown in Equation 1-5, the highest resolution achievable is inversely proportional to the q_{max}. Similarly the largest possible distance measured by SAXS is inversely proportional to the q_{min} and thus the maximum dimension of the protein (D_{max}) must be smaller than this.
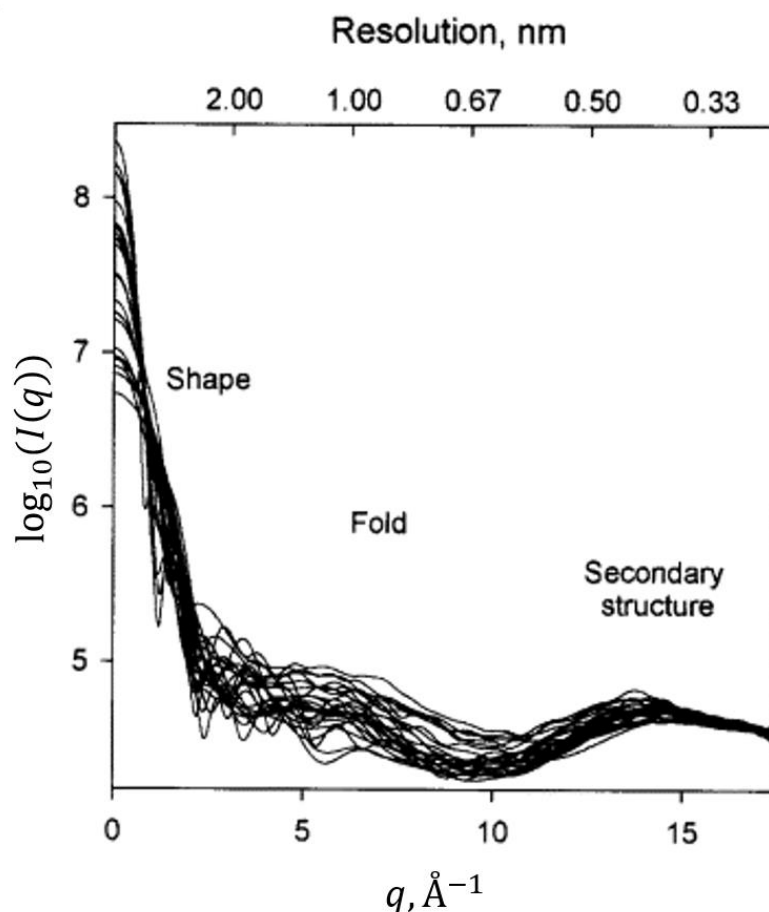
Figure 1.11 – Resolution Limit of SAXS

The scattering pattern of multiple protein structures showing a convergence at very high q (Svergun et al., 2001). The scattering profile shows descriptors of what value of q are required to see the overall shape, the protein folds and the secondary structure of the protein. The resolution shown across the top of the curve is calculated using Equation 1-5.

$$D_{min} = \frac{\pi}{q_{max}}$$

Equation 1-5 – Arbitrary Resolution Calculation

Where q_{max} is the highest scattering angle recorded and D_{min} is the corresponding smallest distance that can be measured at that angle, and can thus be assumed to be resolution.

SAXS is a differential technique where the scattering pattern of the bulk solvent is subtracted from the scattering of the protein sample in buffer to provide a SAXS profile of the protein (Figure 1.12-A). This requires the collection of a protein sample in solution and subtraction of the corresponding buffer (Figure 1.12-B).
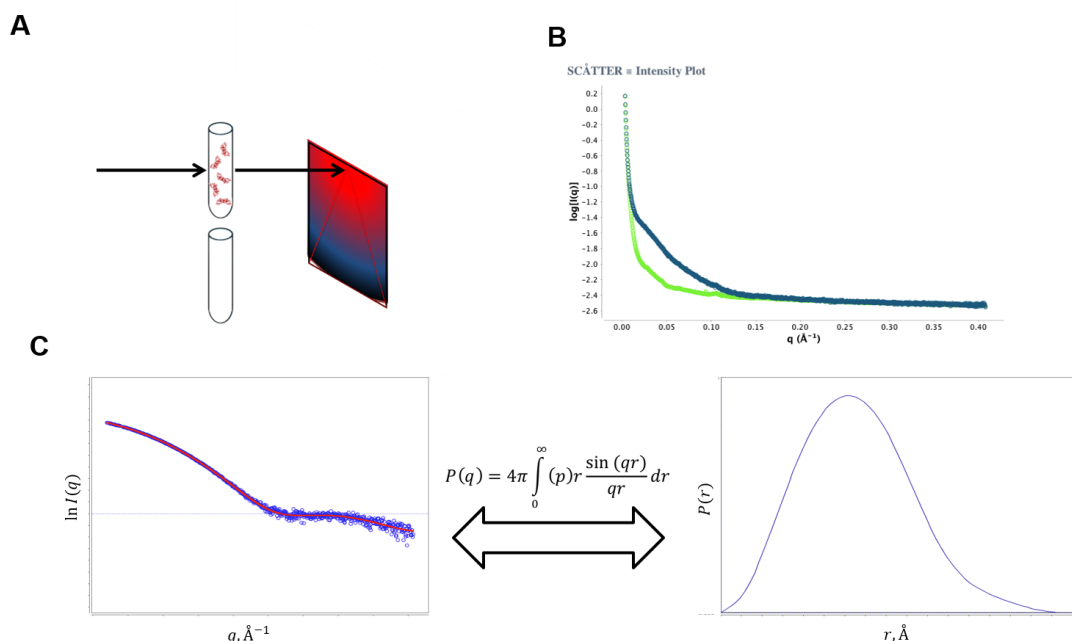
Figure 1.12 – SAXS Collection and Subtraction

A) Sample collection of protein samples in the top tube, and corresponding buffer in the lower tube, with a mock intensity on a detector. PDB structure in the top tube is from PDB ID 3BK3 (Zhang et al., 2008). B) Corresponding protein (blue) and buffer (green) intensity curves, taken from Tunnicliffe et al. (Tunnicliffe et al., 2017) from Section 6.3. C) An example of a real IFT with scattering data shown as blue circles and the red line showing the IFT of the PDDF on the right. Data were taken from (Tunnicliffe et al., 2017)

1.5.2 Guinier Approximation

While the shape of the scattering curve may provide us with some information regarding the overall shape of the molecule, Guinier derived an approximation to assign the radius of gyration through the Guinier Approximation (GA) (Equation 1-6).

$$I(q) = I(0)e^{-\frac{q^2 . R_g^2}{3}}$$

Equation 1-6 – Guinier Approximation

Where $I(q)$ is the scattering intensity, q is the angle of scatter, $I(0)$ is the scattering intensity when $q = 0$ and R_g is the radius of gyration.

The GA provides a simultaneous determination of the radius of gyration (R_g) and the $I(q)$ at a zero angle of scatter ($I(0)$) through extrapolating the scattering curve by conversion to a straight line (Equation 1-7).

$$I(q) = I(0)e^{-\frac{q^2 \cdot R_g^2}{3}} \qquad (1)$$

$$lnI(q) = lnI(0) - \frac{R_g^2}{3} \cdot q^2 \qquad (2)$$

$$y = c + m \cdot x \qquad (3)$$

Equation 1-7 – Guinier Derivation

Where (1) is the Guinier Approximation, (2) represents taking the natural logarithm of the Guinier Approximation and (3) represents a straight line, where $-\frac{R_g^2}{3}$ is the gradient of the line and the R_g can be calculated.

The R_g is defined as the distribution of mass around a particle's centre of inertia and is independent of the particle. It is particularly useful when comparing conformational changes of proteins. The $I(0)$ is directly proportional to the particle volume. The Guinier Approximation has been shown to be valid for q values of $R_g \leq 1.3$ (Feigin and Svergun, 1987) due to 20 – 30% deviations from the line, and when $R_g \leq 2$ the GA breaks down entirely (Feigin and Svergun, 1987).

A further iteration of the GA has been developed; the Guinier Peak Analysis (GPA) provides an easy method of estimating parameter errors from problematic low q scattering (Putnam, 2016). Although this method has not been popularised, it could be well suited for beamline pipelines.

1.5.3 Pair Distance Distribution Function

Another useful parameter from the scattering curve is the intramolecular distances of a molecule, as represented by the gamma function ($\gamma(r)$). The gamma function is not often used to represent data, but the PDDF Equation 1-8.

$$P(r) = r^2 \cdot \gamma(r)$$

Equation 1-8 – Pair Distance Distribution Function from Gamma Function

Where $P(r)$ is the PDDF, r^2 is the distance squared, and $\gamma(r)$ is the probability of finding the given distance within the particle.

The PDDF describes the probability of distances occurring between atoms within a protein molecule. As mentioned in Section 1.4.1, Glatter presented a new method of determining the PDDF by way of IFT (Glatter, 1977a, b) as presented in Equation 1-9. Although the PDDF is not easily calculated from the $I(q)$, as it is not possible to collect scattering data to infinite angles, it is possible to accurately reconstruct a scattering curve from the PDDF by integrating between the 0 and the D_{max} (Equation 1-9).

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty \frac{I(q).\sin(q.r)}{q.r} dq \qquad (1)$$

$$I(q) = 4\pi \int_0^{D_{max}} P(r).\frac{\sin(q.r)}{q.r} dr \qquad (2)$$

Equation 1-9 – Indirect Fourier Transform

Where $P(r)$ is the pair distance distribution function (PDDF), r^2 is the distance squared, q is the scattering angle, $I(q)$ is the scattering intensity and D_{max} is the maximum dimension of electron density within the molecule.

Experimentally the correct PDDF is determined by estimating the PDDF and completing an IFT to obtain the scattering curve. This scattering curve is then compared to the scattering data, and the most representative scatter curve is selected. An amendment to this rule is that the shape of the PDDF can take many forms (Figure 1.13-A), but must connect to the domain axis with a smooth curve.
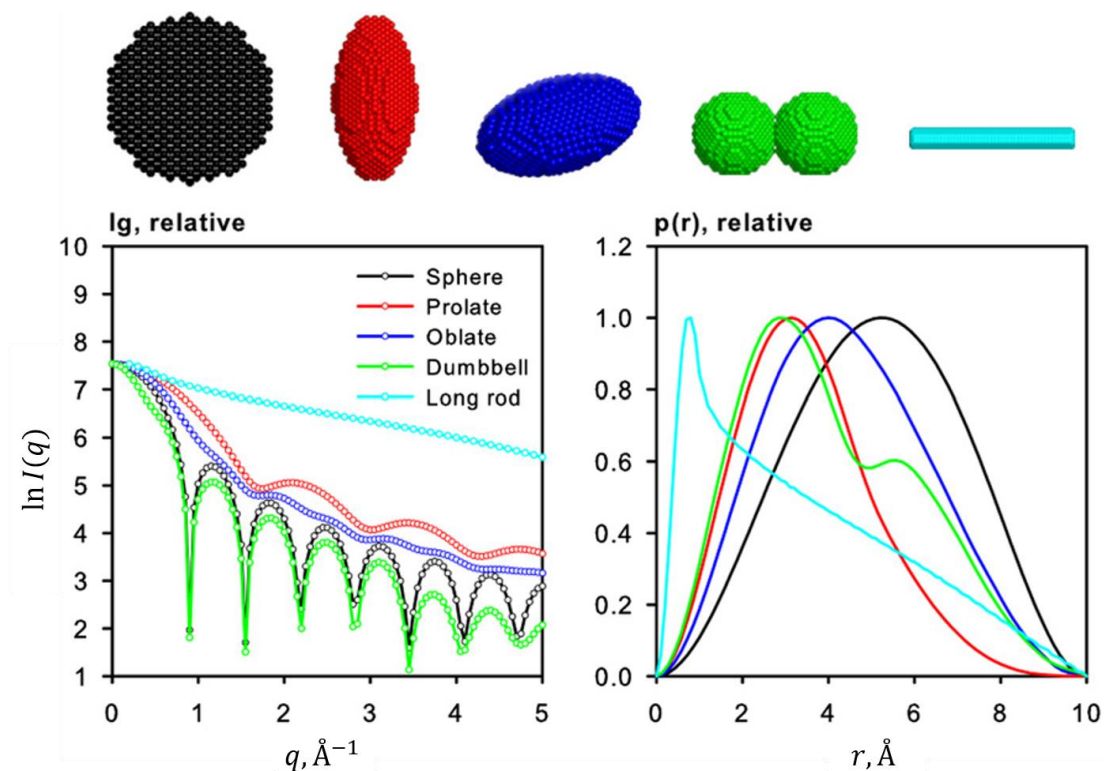
Figure 1.13 – PDDF of Model Systems

Various model shapes were constructed with beads and the scattering patterns of these models are shown. Pair distance distribution functions (PDDFs) as calculated from bead models, producing indicative scattering curves of model systems using Indirect Fourier Transforms (IFT). Taken from Mertens and Svergun (Mertens and Svergun, 2010). B).

From the PDDF it is possible to calculate the R_g in real space. This has advantages over the GA as the R_g is calculated from the entire scattering curve and so is much less sensitive to any problematic data collection in the Guinier region, at low q. This is also particularly useful for larger molecules that would have smaller Guinier regions.

$$R_g{}^2 = \frac{\int_0^{D_{max}} r^2 p(r) dr}{2 \int_0^{D_{max}} p(r) dr}$$

Equation 1-10 – Real Space Radius of Gyration

Where R_g is the radius of gyration, r^2 is the distance squared, $P(r)$ is the pair distance distribution function (PDDF) and D_{max} is the maximal dimension of electron density.

1.5.4 Determining Flexibility

The implementation of the Kratky Plot for use in polymer systems, as outlined by Kratky and Porod (Kratky and Porod, 1949b), has had a huge implication in determining the folded state and flexibility of proteins (Figure 1.14-A). The Kratky Plot

is not specifically useful for the comparison of proteins of different sizes, and so the dimensionless Kratky Plot was used (Durand et al., 2010; Receveur-Brechot and Durand, 2012). The dimensionless Kratky Plot allows for the scaling of protein size as it takes into account the $I(0)$ and the R_g of proteins (Figure 1.14-B). The crosshair shown in Figure 1.14-B demonstrates a perfect sphere and is derived in Equation 9-1.
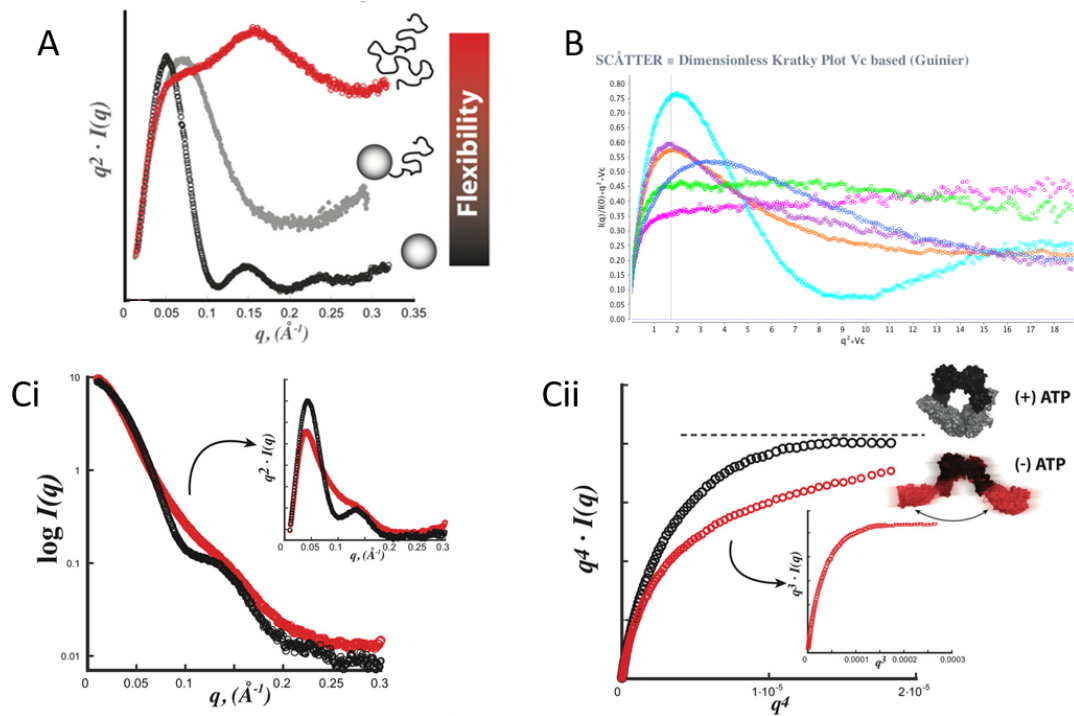
Figure 1.14 – Protein Flexibility

A) Kratky Plot showing unfolded (red), partially unfolded (grey) and folded proteins (black), as taken from (Rambo and Tainer, 2011). B) Dimensionless Kratky Plot showing possible conformational outcomes for a well folded protein: xylanase (cyan), for larger asymmetric molecules P4-P6 RNA domain (orange) and SAM-1 riboswitch (purple), an intrinsically disordered protein: RAD51-AP1 (magenta), and an intrinsically disordered protein attached to a globular protein: RAD51-AP1 fused to maltose binding protein (blue), as taken from www.bioisis.net. C) Determination of the flexibility of Mre11-Rad50 complex in the presence (black) and absence (red) of ATP from Rambo and Tainer (Rambo and Tainer, 2011). Ci) The scattering profiles of each are shown, with an inset showing the Kratky Plot which does not accurately represent the flexibility of the protein with and without ATP. Cii) Porod-Debye Plot showing the plateau of the Mre11-Rad50 complex in the presence of ATP (black) and q^3 Plot of the Mre11-Rad50 complex in the absence of ATP (red) with a clearer depiction of flexibility.

Although both Kratky Plots have the ability to show the change in flexibility it is not very sensitive and not applicable in all circumstances (Figure 1.14-Ci). Rambo and Tainer (Rambo and Tainer, 2011) devised the q^3 plot which can be used in conjunction with the Porod-Debye (q^4) plot to determine the degree of flexibility semi-

qualitatively by examining the point where the scatter plateaus (Figure 1.14-Cii). For compact proteins, $I(q)$ will decay as q^{-4}, whereas the scattering intensity of a flexible Gaussian chain will decay as q^{-2} or slower (Bernado and Svergun, 2012), the q^3 plot makes use of the middle ground and is used to identify semi-flexible proteins.

1.5.5 Molecular Mass Determination

SAXS has been used in a variety of ways to determine the molecular mass of proteins. The most common way at a beamline has been to calibrate the beam with a standard, predominantly bovine serum albumin (BSA), and use this to approximate the mass of the sample. Additionally the Porod Volume of the sample can be calculated from the measured scattering curve (Glatter and Kratky, 1982). As a rule of thumb the volume is approximated to the molecular mass by dividing the volume of the protein by 1.6.

A more robust method of calculation was determined by Rambo and Tainer using the volume of correlation (V_c) (Rambo and Tainer, 2013). The V_c is defined as the ratio of the $I(0)$ to the $I(q)$ of a molecule (Equation 1-11-1) (Rambo and Tainer, 2013). As such this is very sensitive to conformational changes and can also be used in the dimensionless Kratky Plot replacing R_g.

$$V_c = \frac{I(0)}{\int q.I(q)dq} \qquad\qquad (1)$$

$$Q_R = \left(\frac{V_c^2}{R_g}\right) \qquad\qquad (2)$$

$$mass = \left(\frac{Q_R}{e^c}\right)^{\frac{1}{k}} \qquad\qquad (3)$$

$$\ln(Q_R) = k.\ln(mass) + c \qquad (4)$$

Equation 1-11 – Volume of Correlation in Molecular Mass Estimation

Where V_c is the volume of correlation, $I(0)$ is the scattering intensity, $I(q)$ is the scattering intensity, q is the angle of scatter, R_g is the radius of gyration and k and c are determined empirically.

Using the ratio Q_R of R_g and V_c (Equation 1-11-2) it is possible to determine the mass of the protein though the transformation of Equation 1-11-3 to Equation 1-11-4, which utilises the power law to calculate the mass (Figure 1.15).
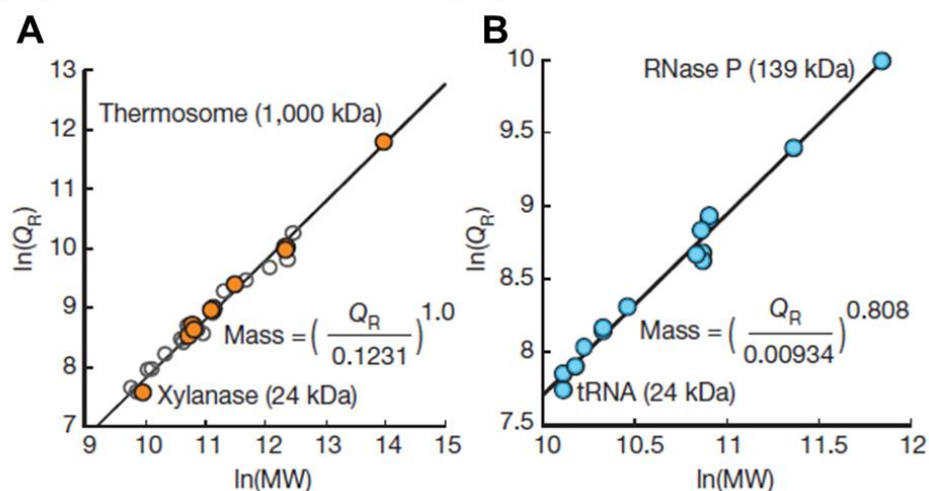
Figure 1.15 – Volume of Correlation Mass Estimates of Protein, DNA and RNA

The estimation of the mass of protein (A) and DNA and RNA (B) is shown for proteins spanning several magnitudes of size. The scattering of DNA and RNA is different to that of protein and so the values of k and c differ slightly in Equation 1-11-3. This figure is take from Rambo and Tainer (Rambo and Tainer, 2013).

1.6 Collecting SAXS Data

Modern SAXS data studies are conducted at synchrotrons as synchrotron radiation has the advantage of being higher energy meaning that experiments can be conducted with a higher signal to noise ratio at a faster speed. Synchrotrons operate at high energies and allow users to utilise X-rays at end stations of beamlines. Typically there are two types of beamlines for BioSAXS collections and are characterised by the way in which X-rays are produced: undulator beamlines (e.g. P12 (DESY, Germany)) (Blanchet et al., 2015) and bending magnet beamlines (e.g. B21 (Diamond Light Source, UK)), where undulator beamlines produce higher flux.

1.7 SAXS Modelling

Determination of the qualitative SAXS parameters provides limited information but in most cases a model of the protein is often useful. This area has been mostly dominated by the Svergun group at EMBL, Hamburg through the development of the ATSAS package (Franke et al., 2017; Konarev et al., 2006; Petoukhov et al., 2012; Petoukhov et al., 2007). Through software refinements, introduction of new algorithms and more powerful hardware, this package contains modelling software for all types of data.

1.7.1 Bead Modelling

Bead modelling, or dummy atom modelling (DAM), was the first implementation of SAXS modelling that provided a more accurate shape of the protein. DAMMIN (Svergun, 1999) and DAMMIF (Franke and Svergun, 2009) are an example of iteration and advancement in the algorithms that allowed a bead model to be reconstructed from 1D SAXS curve. Both programs use the Monte Carlo simulated annealing method but implemented it in very different ways. Where DAMMIN started with a sphere of beads with a radius of half the D_{max}, DAMMIF is able to break this constraint and has an unlimited search volume (Franke and Svergun, 2009). Another difference in the bead modelling is in the selection of beads. DAMMIN begins with a mixture of beads denoted as solvent and as sample, where DAMMIF used an approach of sample but then the solvent has three different phases: a solvent, a possible solvent and a search volume, where the possible solvent can easily change between a sample and solvent. Finally, the DAMMIF beads are all linked, and not independent like the DAMMIN beads. This has the added advantage of moving beads with fewer penalties (Franke and Svergun, 2009). A further modelling method from the ATSAS suite, which can arguably obtain higher resolution models is GASBOR (Svergun et al., 2001). This method uses simulated annealing, and takes advantage of the fact that at a certain q range (~ 5 Å resolution), all residues are indistinguishable. GASBOR uses the number of dummy residues is equal to the number of residues of the protein; the down side to the constrained dummy residue number is that this process cannot be used to model proteins with post translational modifications.

1.7.2 Bead Model Averaging

When 10 – 20 bead models have been generated they are compared using the DAMAVER suite from within ATSAS (Volkov and Svergun, 2003). The DAMAVER suite (Volkov and Svergun, 2003) processes the DAMs, in order to average and filter the models, resulting in a final model. Within the DAMAVER suite are five individual programs: DAMSEL, DAMSUP, DAMAVER, DAMFILT and DAMSTART. DAMSEL compares each of the DAMs using a program SUPCOMB (Kozin and Svergun, 2001). SUPCOMB selects the most probable model and calculates the outliers, returning the normalised spatial discrepancy (NSD) as a metric for selecting DAMs that will contribute to the final model. Values of the NSD are determined by principle axis

alignment by aligning the x-, y- and z-axis, followed by a local grid search. DAMSEL uses the mean and standard deviation of the NSD to remove any DAMs with an NSD greater than two standard deviations over the mean. DAMSUP then reads the models selected by DAMSEL and aligns each DAM with the reference, considered the most probable representation of the final DAM. DAMAVER then takes the aligned models and averages them. The method of completing an average DAM uses a frequency probability map which is calculated for use with DAMFILT. DAMFILT then filters the bead model from DAMAVER given a specific cut-off volume. With the cut-off volume and the frequency map, low probability beads and beads with high loose-ness are removed from the final model. In most cases this is considered the final model. DAMSTART is the final program within the DAMAVER suite and works in a similar manner to DAMFILT, by producing a DAM with beads at a higher specified volume. DAMSTART models are generated from a small number of DAMMIN models generated in fast mode to enable an input search model for DAMMIN (Svergun, 1999) as an initial constraint for faster DAM generation. DAMMIF has been optimised for accuracy and speed so that this method is rarely used. SASRES is a new interpretation of model variance that has been used to calculate the resolution for SAXS models in a similar fashion to electron microscopy (EM) with a Fourier shell correlation (FSC). Resolution determined this way used a FSC cut-off at 0.5 (Tuukkanen et al., 2016).

Bead models can be generated from either SAXS or SANS models with very similar results of the same protein (Figure 1.16-A) (Baldock et al., 2011). This method has also been shown to correlate well with negative stain electron microscopy (Figure 1.16-B) (Troilo et al., 2014) and thus a strong corroborative technique.
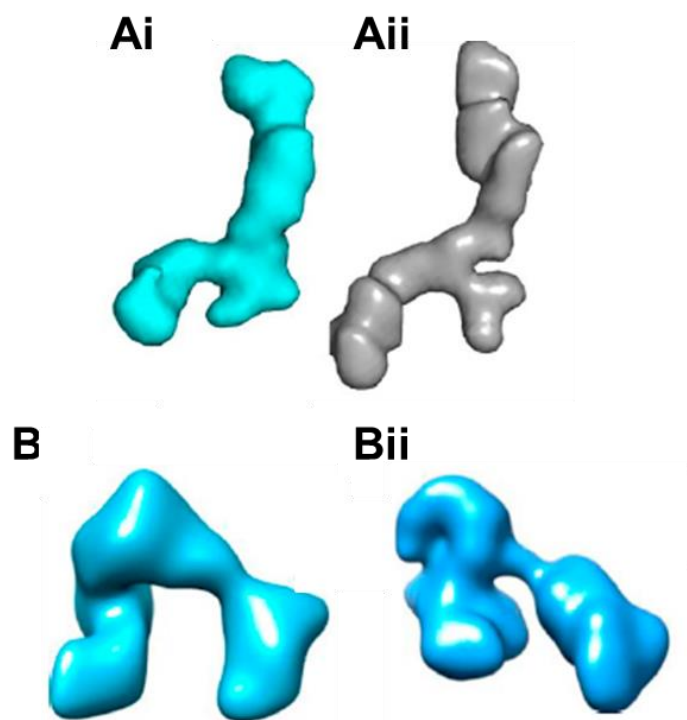
Figure 1.16 – Bead Modelling Accuracy

Reconstruction of SAXS (Ai) and SANS (Aii) profiles collected on tropoelastin, generated by dummy atom modelling using GASBOR (Baldock et al., 2011). Comparison of a 3D reconstruction of chordin from particles identified by negative stain microscopy (Bi) with a DAMMIN bead model (Bii) generated from SAXS data collected on chordin (Troilo et al., 2014).

1.7.3 Atomistic Modelling

SAXS is very useful as part of a combinatorial technique with other structural methods. Where crystallographers would have just used SAXS to validate X-ray structures by comparing the scattering intensity to the calculated intensity of the crystal structure (Grudinin et al., 2017; Putnam et al., 2013; Schneidman-Duhovny et al., 2013; Svergun et al., 1995), SAXS data can be used to construct more complex biological systems that may previously not have been possible. Using high resolution structures of domains it is possible to obtain SAXS information of the entire molecule, and use a variety of molecular dynamic methods to determine how those domains together fit the SAXS curve. This has been implemented in a number of ways, taking into account flexible systems using an ensemble of models to reconstruct a selection of contributing models (Schneidman-Duhovny et al., 2016; Tria et al., 2015), multi-domain modelling of proteins and complex interactions (Pelikan et al., 2009; Perkins et al., 2016; Petoukhov et al., 2012; Petoukhov and Svergun, 2005; Schneidman-Duhovny et al., 2016).
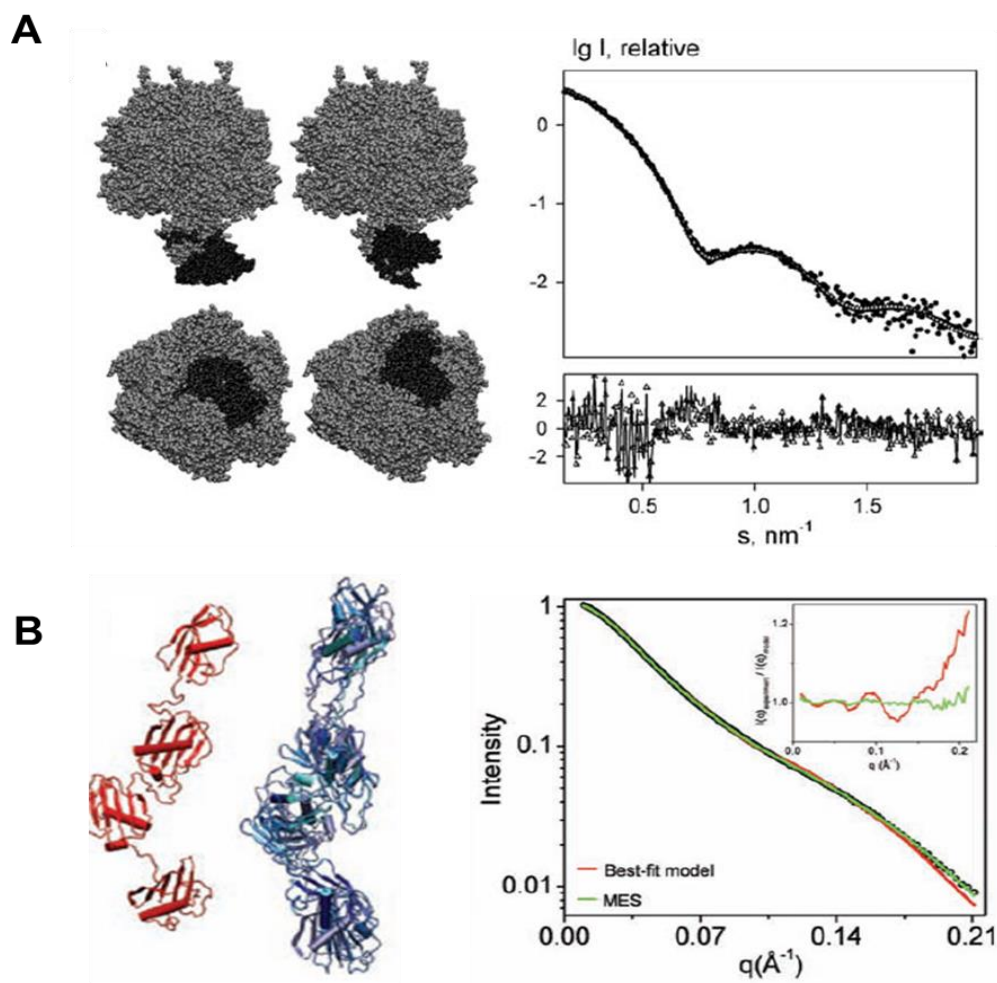
Figure 1.17 – Atomistic Modelling of Proteins

Modelling of a single F1 ATPase using a DIMFOM, a precursor to BUNCH, and shows accurate modelling of the structure, shown by the line through the scattering pattern, with low residual divergence, shown below the scattering graph (A) (Petoukhov and Svergun, 2005). Ensemble modelling shown using BILBOMD (B) where a single 'best model' of extracellular adherence protein (EAP), shown by the red protein, does not accurately represent the scatter curve, shown as the red line, with a high divergence at higher q in the residual plot. An ensemble built form five conformers, shown by blue models and represented by the green line on the graph fits the scattering much better, with lower residual plot divergence (Pelikan et al., 2009).

1.7.4   Combined Techniques

As mentioned above it is possible to combine and compare SAXS with negative-stain EM (Troilo et al., 2014), and with X-ray crystallography through model validation (Grudinin et al., 2017; Putnam et al., 2013; Schneidman-Duhovny et al., 2013; Svergun et al., 1995) and through modelling of domains and complexes (Guttman et al., 2013; Petoukhov et al., 2012; Petoukhov and Svergun, 2005; Schneidman-

Duhovny et al., 2010; Schneidman-Duhovny et al., 2016). In addition to those methods mentioned above, SAXS has been used in conjunction with nuclear magnetic resonance (NMR) spectroscopy to facilitate a comprehensive characterisation of biological molecules in solution.

Additional solution based biophysical techniques, including multi-angle light scattering (MALS) and analytical ultracentrifugation (AUC) can be used to assess the validity of SAXS models by determining their hydrodynamic parameters. One such example is the use of SOlution MOdeller (SOMO) (Brookes et al., 2010a; Brookes et al., 2010b). This program is able to calculate the hydrodynamic radius (R_H) and the frictional ratio (f/f_0) from DAMs to be compared to experimental AUC data.

1.8 Future of SAXS

Future advances in SAXS include advances in detectors collecting data with an increase in the signal to noise ratio (Dectris, 2017), upgrades to beamlines improving the flux (Blanchet et al., 2015) and improving on sample environments. The way in which these advances will improve data collecting capacities will vary on beamlines. The higher flux may be used for higher throughput of samples collected using robots, where it may conversely be used in collecting data at much lower concentrations from difficult to express proteins after experiencing dilution on SEC-SAXS. Furthermore, conformational changes over time may be detected using a higher flux and a faster, more sensitive detector.

Although not strictly SAXS, it would be remiss of me not to mention the possible applications of high energy X-rays generated by X-ray free-electron lasers (XFEL) and the possibility of observing individual molecules (Donatelli et al., 2017; Fratalocchi and Ruocco, 2011). In this case fluctuating X-ray scattering would use X-ray pulses in order to collect data of randomly oriented molecules in times shorter than the tumbling time of the molecules (Donatelli et al., 2015). This snapshot of protein molecules would be akin to the way in which cryo-EM images are observed. The reconstruction of the molecular shape from this data would pose a challenge, although the deconvolution of the data through a flexible iterative method by Donatelli et al would allow rapid determination of the molecular structure (Donatelli et al., 2015).

1.9 **Aims**

The work here is carried out as part of a joint student-ship between the University of Manchester and Diamond Light Source, the aims of this project were designed to reflect this. Across both sites, the structural and functional properties of matrix regulators associated with BMP signalling were studied. Protein production, purification and biophysical characterisation were mainly carried out at the University of Manchester. At Diamond Light Source, the main focus was on the implementation of new hardware on beamline B21.

Currently little is known about the structure of BMPER other than the first domain. Here I intend to use SAXS, EM and a number of biophysical techniques to structurally characterise BMPER to provide an insight into the relationship of structure and function. The binding of BMPER with Tsg, another BMP regulator was to be further investigated as there is conflict in the literature as to whether BMPER and Tsg directly interact. A mutation in the BMPER cleavage site (P370L) has been linked to the disease DSD. Currently, only clinical studies have been published on this mutation and a biochemical study into this mutation will allow us to determine whether the protein is secreted, if the mutation effects cleavage and BMP signalling. Therefore, adopting a multifaceted approach the biological study will have the following objectives:

1. Recombinant production of BMPER constructs in a mammalian expression system in order to produce enough protein for structural and biochemical studies.

2. To complete hydrodynamic and structural analyses on BMPER constructs to investigate their structure.

3. Determine the binding capacity of human BMPER for Tsg and subsequently investigate their effect on BMP signalling as assessed by inhibition of alkaline phosphatase production.

4. Generate and conduct the first biochemical studies on the BMPER-P370L mutation, responsible for the disease DSD.

Beamline B21 is a relatively new installation of a high throughput solution state BioSAXS beamline at Diamond Light Source, where commissioning must be completed on the beamline and periphery equipment used in experimental setups and so in order to help with the optimisation of the beamlines experimental setup, the aims reflect this. The aims of the Diamond Light Source bound project are to:

1. Aid in the implementation and optimisation of HPLC-SAXS at the beamline

2. Use 3D printing to make prototype sample cells for testing new window materials, and for implementation of HPLC-SAXS stop flow systems.

3. Apply BioSAXS data collection and analysis to address key biological questions.

2 Materials and Methods

2.1 Bacterial Strains

XL10-Gold Ultracompetent cells and XL1-Blue Competent cells (Agilent).

2.2 Mammalian Cell Lines

HEK (human embryonic kidney) 293 cells containing the Epstein-barr virus nuclear antigen-1 (EBNA) and the SV40 large T antigen (T) cells were available in the lab, and were used for all mammalian cell culture.

2.3 Vectors

The vectors used were the pCEP-Pu/Ac7, a derivative of pCEP4 vector (Kohfeldt et al., 1997); the pCEP4 vector was adapted to be puromycin resistant and have a BM40 signal peptide for increased protein expression. A variant of the pCDH vector designed by Dr Michael Leverentz (University of Manchester); the pCDH vector was modified to add a His$_6$ tag and a V5 tag before the T2A peptide and the turbo green fluorescent protein (GFP). T2A is a 'self-cleaving' peptide of the Foot-and-Mouth Disease Virus (FMDV) allowing for turbo GFP to be expressed but not fused to the gene of interest (Radcliffe and Mitrophanous, 2004). The pHLsec vector was also used (Aricescu et al., 2006); the pHLsec vector contains a γ-secretase signal peptide and a β-actin promoter and is designed as a transient transfection expression vector. Vector maps can be found in Appendix C.

2.4 Molecular Biology Methods

2.4.1 Construct Generation

The full length human BMPER construct was generated by GeneArt gene synthesis (Thermo Fischer) based on accession number NM_133468.4. The additional BMPER constructs derived from this were generated by PCR. A summary of the primers used to generate constructs through PCR (2.4.2) for recombinant protein expression can be found in Table 2-1. Additional pHLsec primers can be found in Appendix B.

Construct	Forward Primer (5' – 3')	Restriction endonuclease	Reverse Primer (5' – 3')	Restriction endonuclease
BMPER – pCEP-Pu/Ac7	**TGGCAGCCCCG** <u>CTAG</u>CGTCCTCC TTCTTGACAGGT TCTGT	NheI	CCTTGCCGGC <u>CTCGAG</u>ACTA ATGGTGGTGG TGATGATGGG AG	XhoI
N-BMPER – pCEP-Pu/Ac7	**TGGCAGCCCCG** <u>CTAG</u>CGTCCTCC TTCTTGACAGGT TCTGT	NheI	**CCTTGCCGGC** <u>CTCGA</u>GCTAA TGATGGTGGT GATGATGATC TCCA	XhoI
BMPER/B MPER-P370L – pCDH	**CTACTCTAGAGC** <u>TAG</u>CGAATTCGA AAGCACCATGCT CTGGTTCTCCGG CGT	NheI	**TAGGCTTACC** <u>GGATC</u>CCCG CTGGGGACAA AGGAC	BamHI
P370L mutation primers	C<u>T</u>CCACTACAAC ACTTTTGACGGT CGG	N/A	ATCTCCAAAC ACCGTGCAAA CGCCG	N/A

Table 2-1 – Table of Primers for PCR Amplification

Table showing the primers designed for each construct and their corresponding vector and quick change mutagenesis primers. Black text denotes the overlap of the primer with the gene of interest, text underlined in red denotes the endonuclease recognition sequence, bold blue text denotes the In-Fusion overlapping regions with the desired vector and green underlining denotes the region of mutation in the mutagenesis primers.

2.4.2 Polymerase Chain Reaction (PCR) Methods

PCR reactions were carried out using Kapa HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems) or Phusion Flash High-Fidelity PCR Master Mix (Thermo Fischer) for amplification and quick-change mutagenesis as both kits and methods followed the same principles. For the Kapa kit the manufacturer's instructions were followed, where the Oxford Protein Production Facility (OPPF) protocol was followed for the Phusion Flash reaction mixture. Below, the respective composition (Table 2-2) and amplification protocols (Table 2-3) for each polymerase kit is shown.

	Kapa HiFi	Phusion Flash
Template DNA (ng/reaction)	80 – 100	20 – 40
Forward Primer from 10 µM stock (µL)	1.5	3.0
Reverse Primer from 10 µM stock (µL)	1.5	3.0
2 x Ready Mix (µL)	25.0	25.0
Sterile Water (µL)	Up to 50	Up to 50

Table 2-2. Table of PCR Composition.

A table showing the composition of PCR reaction mixtures for each of Kapa HiFi HotStart ReadyMix PCR Kit and Phusion Flash High-Fidelity PCR Master Mix reaction mixes.

	Kapa HiFi			Phusion Flash		
	Temperature (°C)	Time (s)	Cycles	Temperature (°C)	Time (s)	Cycl
Initial denaturation	95	180	1	98	10	1
Denaturation	98	20	25	98	1	29
Annealing	60	15		60	5	
Extension	72	30 s/kb		72	15 s/kb	
Final extension	72	60 s/kb	1	72	120	1
Hold	4	∞	1	4	∞	1

Table 2-3. Table of PCR Steps.

A table showing the temperature, duration and number of cycles for each step of the reaction for both the KAPA HiFi HotStart ReadyMix PCR Kit and the Phusion Flash High-Fidelity PCR Master Mix.

2.4.3   Agarose Gel Electrophoresis

1 %(v/w) agarose gels were prepared to separate DNA by size for comparison to a DNA standard ladder by adding agarose (Bioline) to TAE buffer (40 mM Tris (pH 7.6), 20 mM acetic acid and 1 mM EDTA) at 1% (w/v). This was heated for 60 seconds then cooled on ice for 5 minutes before adding SafeView (NBS Biologicals) DNA Gel Stain in a 1:10,000 dilution. The gel was cast with an appropriate toothed comb to make wells, and allowed to set. Once set, the gel was immersed in TAE buffer. Samples were prepared by adding Purple Gel Loading Dye (6X) (New England BioLabs) to the desired amount of DNA and loading into the wells of the gel with one lane reserved for HyperLadder 1kb (Bioline) as a DNA molecular weight marker. Gels were electrophoresed for 30 minutes at 100 V and imaged using a Dark Reader transilluminator (Clare Chemical Research) to avoid DNA crosslinking.

2.4.4   Purification of DNA

2.4.4.1   PCR Clean Up

DNA from a PCR reaction was treated with endonuclease DpnI (New England Biolabs) for one hour in a 50 µL reaction at 37 °C. The entire reaction volume was used in the PCR clean up. DNA purification was completed with either the QIAquick PCR Purification Kit (Qiagen) or the NucleoSpin Plasmid Kit (Macherey-Nagel), and following the protocol outlined by the manufacturer. Briefly, binding buffer was added to the DNA solution in a 5:1 ratio and added to a spin-column. The binding column and mixture was centrifuged 16,100 $\times g$ for 60 seconds to remove the excess material. The bound DNA was washed with wash buffer before centrifuging for a further 60 seconds. DNA was eluted from the column using 30 – 50 µL of sterile water, which was incubated on the membrane of the column for five minutes before centrifuging the samples for 60 seconds. If samples were not immediately used they were stored at -20 °C.

2.4.4.2   Agarose Gel Extraction

Once the correct band had been identified on an agarose gel, by comparison with DNA standards, it was excised with a sterile scalpel and purified using either the QIAquick Gel Extraction Kit (Qiagen) or the NucleoSpin Plasmid Kit (Macherey-Nagel). Briefly, two to three times the gel volume (1 g – 1 mL) of solubilisation and binding buffer was added to the excised gel band, and the mixture heated at 55 °C and mixed by vortex for 10 minutes, or until the gel was dissolved. For the Qiagen kit, isopropanol was added to the mixture at one gel volume. For both methods the mixture was added to a binding column and centrifuged at 16,100 $\times g$ for 60 seconds. A washing buffer was then applied to the column before centrifugation for 60 seconds at 16,100 $\times g$. A further centrifugation step was required to dry the membrane. DNA was eluted from the column using 30 – 50 µL of sterile water, which was incubated on the membrane of the column for five minutes before centrifuging the samples for 60 seconds at 16,100 $\times g$. If samples were not immediately used they were stored at -20 °C.

2.4.5   Preparation of PCR Products for Ligation

Ligation was carried out by one of two methods, either by cohesive-end ligation with a Quick Ligation Kit (New England Biolabs) or using the In-Fusion HD Cloning Kit (Takara). Cohesive-end ligation was predominantly used where possible and In-

Fusion cloning was used when two of the same cloning sites were in the construct. Blunt end ligation was performed for mutagenesis in a method akin to QuikChange Site Directed Mutagenesis (Agilent). Recognition sires for the restriction endonucleases used to digest the vector were included in the design of the primers and added by PCR to the constructs (Table 2-1).

2.4.5.1 Cohesive-end Preparation

PCR products were digested with the corresponding endonucleases, as per the manufacturer's instructions. The resulting PCR product was purified as described in section 2.4.4 to produce a product with complementary base-pair overhangs.

2.4.5.2 Vector Linearization

The digestion of the vector was carried out with 3 – 5 µg of DNA, 5 µL of 10X CutSmart Buffer (New England Biolabs) and 2 µL of each of the desired restriction endonucleases. The mixture was incubated at 37 °C for one hour before purification by agarose gel extraction, section 2.4.4.2.

2.4.6 Ligation

2.4.6.1 Cohesive-end Ligation

Cohesive-end ligation was completed with a Quick Ligation Kit (New England Biolabs) by following the manufacturer's instructions. Briefly, 50 ng of linearised vector was mixed with digested purified insert in a 1:3 ratio (vector to insert) and the volume adjusted to 10 µL. 10 µL of 2X Quick Ligation Buffer was added to the DNA mixture followed by 1 µL of Quick T4 DNA Ligase and incubated at room temperature for 5 – 10 minutes. The sample was then transformed into *E. coli*, see section 2.4.7, or frozen at -20 °C.

2.4.6.2 In-Fusion Ligation

In-Fusion ligation could be completed with a purified PCR product without further treatment. Linearised vector (50 µg) was mixed with PCR product in a 1:2 ratio, and 2 µL of 5X In-Fusion HD Enzyme Mix was added and made up to 10 µL with sterile water. The mixture was incubated at 50 °C for 15 minutes before being placed on ice. The sample was then transformed into *E. coli*, see section 2.4.7, or frozen at -20 °C.

2.4.6.3 Blunt End Ligation

As a result of a modified QuikChange Site Directed Mutagenesis, the PCR product would be blunt ended. The same protocol from section 2.4.6.1 was used, where a total of 50 µg of DNA was used. The sample was then transformed into *E. coli*, see section 2.4.7, or frozen at -20 °C.

2.4.7 Transformation

3 - 10 µL of ligation reaction was mixed with 50 µL of competent XL10-Gold Ultracompetent cells or XL1-Blue cells. The mixture was incubated on ice for 45 minutes before heat shocking the mixture in a 42 °C water bath for 45 seconds. Cells were immediately placed back on ice and incubated for 5 minutes. Transformations were mixed with 500 µL of warmed Super Optimal broth with Catabolite repression (SOC) medium (Thermo Fischer Scientific) and incubated at 37 °C for 1 hour. The mixture was centrifuged slowly, ~ 200 × g, for 1 minute to obtain a loose pellet of cells. 250 µL of SOC medium was removed and discarded before the cells were re-suspended in the 250 µL remaining media. 150 µL of cells were then plated onto an agar plate containing either ampicillin (100 µg/mL) or kanamycin (50 µg/mL), depending on the resistance cassette of the vector, and incubated at 37 °C overnight.

2.4.8 Overnight Cultures

Colonies resulting from transformations were picked from agar plates using a sterile loop and used to inoculate a 50 mL centrifuge tube (Corning) containing 10 mL of LB broth, with either ampicillin (100 µg/mL) or kanamycin (50 µg/mL). The cultures were maintained at 37 °C overnight whilst shaking (225 rpm).

2.4.9 DNA Extraction by Miniprep

Overnight cell cultures were pelleted by centrifugation at 2,260 × g for 10 minutes and the supernatant removed. DNA was extracted from the pellet using QIAprep Spin Miniprep Kit (Qiagen). Briefly, pelleted cells were re-suspended in 250 µL of re-suspension buffer and transferred into a 1.5 mL centrifuge tube. 250 µL of lysis buffer was added to the mix, gently mixed and allowed to stand for five minutes. 350 µL of neutralisation buffer was used to neutralise the reaction and precipitate proteins and chromosomal DNA. The precipitant was pelleted by centrifugation at 16,100 × g. The clarified buffer, containing the soluble vector DNA, was loaded onto a DNA-binding spin column. The spin column and mixture was centrifuged 16,100 × g for 60 seconds

to remove the excess material. The bound DNA was washed with wash buffer before centrifuging for a further 60 seconds. DNA was eluted from the column using 30 – 50 µL of sterile water, which was incubated on the membrane of the column for five minutes before centrifuging the samples for 60 seconds. If samples were not immediately used they were stored at -20 °C.

2.4.10 Construct Validation

2.4.10.1 PCR Verification

PCR verification was carried out in the same manner as those mentioned in section 2.4.2. The forward sequencing primer was used in conjunction with the reverse PCR primer for a PCR reaction in order amplify a fragment containing both vector and construct. Results were analysed by agarose gel electrophoresis (section 2.4.3).

2.4.10.2 DNA Sequencing

DNA sequencing reactions were prepared in a total volume of 10 µL with 5 µM of primer and up to 500 ng of DNA. Sequencing primers were selected depending on the vector, unless internal primers were required. Primer sequences can be found in Table 2-4. Sequencing samples were sent to GATC Biotech (Cologne, Germany) using the LIGHTrun service for sequencing.

Name	Forward (5' – 3')	Reverse (5' – 3')
pMK-RQ	GTAAAACGACGGCCAGT	AACAGCTATGACCATG
pCEP-Pu/AC7	AGCAGAGCTCGTTTAGTGAACCG	CCAGGATGGAGAGATGTGGTCCTCTATCAATTG (BMPER internal forward primer)
pCDH	CTCCACGCTTTGCCTGACCCT	CCAGGATGGAGAGATGTGGTCCTCTATCAATTG (BMPER internal forward primer)
pHLsec	GCTGGTTGTTGTGCTGTCTCATC	CACCAGCCACCACCTTCTGATAG

Table 2-4 – Sequencing Primers for Each Vector

Sequencing primers for each of the vectors used in expression or amplification of genes. The pCEP-Pu/AC7 and pCDH vectors required an internal BMPER forward primer to sequence them.

2.5 Protein Expression & Purification

2.5.1 HEK293 Cell Culture

HEK293-EBNA and –T cells were cultured on tissue culture plastic in flasks with either 25 or 75 cm^2 surface area (Corning), and 5 or 10 mL of growth media (Dulbecco's Modified Eagles Media, DMEM) supplemented with foetal bovine serum (FBS) (10% v/v, Gibco) and penicillin streptomycin (1% v/v, Sigma). Cells were incubated at 37 °C and 5% CO_2, until confluent. Cells were passaged by removing the growth media and washing the cells with 10 mL of phosphate buffered saline (PBS, Sigma). Cells were then detached by incubating the cells with 1 or 2 mL of trypsin-EDTA (Sigma), for 5 minutes at room temperature. Trypsin was inhibited by the addition of growth media to suspend the cells. Cells were then passaged at a 1:5 dilution for further growth.

2.5.2 HEK293 Transfection

A confluent T75 flask with HEK293-EBNA or –T cells was diluted to 1:40 in growth media. For a six well plate transfection, 1 mL of cell suspension was added to the plate 24 hours before transfection to achieve 80% confluency. This was scaled for any larger flask. In a six well plate, 5 μg of DNA was used per transfection. Transfections were completed as specified below.

2.5.2.1 Transient transfection

DNA was diluted into 250 µL of expression media (a 1:1 mixture of DMEM and HAM's F-12 nutrient mixture (Sigma) with penicillin streptomycin (1% v/v)). A 1 mg/mL solution of polyethylenimine (PEI) was added to the DNA in a ratio of 2:1 and incubated for 10 minutes before adding to cells and increasing the volume of media to 1 mL. The media was changed after 24 hours. Cells were then allowed to incubate for 48 hours before harvesting conditioned media or cell lysate.

2.5.2.2 Stable transfection

DNA was mixed with P3000 Reagent and 250 µL of serum free DMEM (Sigma) while Lipofectamine 3000 Reagent (Thermo Fischer Scientific) was also mixed with 250 µL of serum free DMEM. The mixtures were combined and allowed to mix for 10 minutes. This mixture was then added to the cells and allowed to incubate. After 24 hours, cells were passaged to a 25 cm^2 flask and allowed to reach confluency. Media was exchanged with growth media supplemented with 2 µg/mL of puromycin (Gibco) to select transfected cells. Cells were selected for a period of two weeks.

2.5.2.3 Lentiviral preparation and transduction

In the case of lentivirus transfections, the pCDH vector was transfected along with the packaging vectors (pMD2.G and psPAX2) using PEI transfection reagent in a 1:1.5 ratio (DNA to PEI). After 24 hours, the media was removed and replaced with growth media supplemented with 10 mM sodium phenyl butyrate, a histone deacetylase inhibitor (Sigma). After a further 4 - 8 hours the media was replaced with growth media and allowed to incubate for a further 24 hours.

The media was removed from cells transfected with lentiviral vectors and supplanted onto the target cells. The cells were then incubated for 48 hours before sorting for the corresponding fluorescent protein encoded in the vector by fluorescence activated cell sorting (FACS). Cells were sorted using a FACS Aria Fusion (BD Biosciences). Turbo GFP (tGFP) expressing cells were sorted using an excitation wavelength of 482 nm and an emission wavelength of 502 nm. Blue fluorescent protein was sorted using an excitation wavelength of 402 nm and an emission wavelength of 457 nm

2.5.3 Cell Storage

2.5.3.1 Freezing Cells

HEK293-T and -EBNA cells lines were stored for future use by freezing. Cells at 90% confluency were treated with trypsin (section 2.5.1), re-suspended in DMEM and transferred to a 15 mL centrifuge tube. Cells were centrifuged at 16,100 $\times g$ for 5 minutes and the supernatant removed. Cells were then re-suspended in freezing media composed of DMEM (50% v/v), FBS (40% v/v) and dimethyl sulfoxide (DMSO) (10% v/v). Aliquots (1 mL) were frozen to -80 °C at a rate of 1 °C per minute.

2.5.3.2 Raising Cells

Cells were raised from -80 °C by thawing them quickly at 37 °C and transferring them to a 75 cm^2 flask with 10 mL of DMEM with penicillin (100 units) and streptomycin (0.1 mg/mL) (Sigma), L-glutamine (2 mM) (Sigma) and FBS (10% v/v), known as growth media. After 24 hours the media was replaced with growth media, and growth was carried out as described in section 2.5.1.

2.5.4 Growing cells for protein expression

Stably selected, or sorted, cell lines were grown and passaged up to 225 cm^2 flasks, from which, cells were transferred to HYPERflasks (Corning) or roller bottles (Grenier Bio One). Upon reaching confluency the media was exchanged to expression media and the cells were incubated at 37 °C. Roller bottles did not require 5% CO_2.

2.5.5 Harvesting conditioned and clarifying media

For both HYPERflasks and roller bottles, media was harvested after one week. The media was poured from the flasks and replaced with fresh media. This could be repeated for four weeks before the cells began to detach. The media was then passed through grade 1 filtration paper (Whatman) under vacuum to remove cell debris before passing the media through 0.65 µm filter (Millipore) to remove finer particulate matter. Media was then ready for purification.

2.5.6 Immobilized Metal Affinity Chromatography

BMPER constructs were purified by utilising their His$_6$ tags using HisTrap Excel chromatography columns (GE Healthcare). Media was loaded onto the buffer equilibrated columns using a peristaltic pump at 4 °C. N-BMPER was purified in PBS,

and BMPER in PBS with 500 mM NaCl (PBS-500). Once loaded, columns were washed for 20 column volumes (CVs) with their respective buffers with the addition of 40 mM imidazole to wash off any non-specifically bound contaminants. Protein was then eluted from the columns with the addition of 250-500 mM imidazole. Eluted fractions were monitored by SDS-PAGE, as described in section 2.6.1. All fractions containing protein were concentrated to 500 µL by centrifugal concentration using Vivaspin columns (Sartorius), with a molecular weight cut off (MWCO) of 30 kDa for BMPER and 10 kDa for N-BMPER.

2.5.7   Size Exclusion Chromatography

Size exclusion chromatography was performed on a Superdex 200 Increase (S200i) 10/300 GL column (GE Healthcare). The S200i column was equilibrated in either PBS or PBS-500 on an ÄKTA Purifier (GE Healthcare). 500 µL samples were loaded onto the column and injected at a rate of 0.75 mL/min. The eluted volume was collected in 0.5 mL fractions with protein elution monitored using A_{280}.

2.6   Protein Characterisation

2.6.1   Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein samples were mixed with LDS NuPAGE buffer in 20 µL (4X, Thermo Fischer). This was completed in either the presence or absence of 2-mercaptoethanol (5%). Samples were then boiled at 100 °C for 10 minutes. Samples were then loaded onto a 4-12% NuPAGE SDS-PAGE gel (Thermo Fischer) and electrophoresed at 200 V for 50 minutes in a MOPS buffer (Thermo Fischer) in an X-Cell Surelock Mini-cell Electrophoresis System (Thermo Fischer). Molecular mass estimations were carried out using either the SeeBlue Plus2 Pre-stained Protein Standard (Thermo Fischer) or Precision Plus Protein Standards (Bio-Rad). The protein was visualised in the SDS-PAGE gels using InstantBlue (Expedeon), as per the manufacturer's instructions.

2.6.2   Western Blotting

Following SDS-PAGE, the proteins were transferred to a nitrocellulose membrane (GE Healthcare) at 35 V for one hour using an X-Cell Surelock Mini-cell Electrophoresis System (Thermo Fischer). The transfer buffer was composed of 96mM Tris-HCL, 780mM glycine, 0.075% (v/v) SDS with 20% (v/v) methanol. Immediately after transfer the membrane was blocked for an hour with either a milk

71

solution (5% w/v) or BSA solution (5% w/v) in Tris-buffered saline (10 mM Tris pH 7.4, 150 mM NaCl) with Tween-20 (0.05%) (TBS-T). The blocking solution was then removed and replaced with the blocking solution in the presence of a primary antibody (Table 2-5). Primary antibodies were left on the membrane overnight at 4 °C whilst rocking. The primary antibody was then removed and the membrane washed with TBS-T three times for 10 minutes per wash. If the primary antibody was conjugated to horseradish peroxidase then it was imaged directly, if not a secondary antibody was used. The secondary antibody in blocking buffer was added to the membrane at room temperature for one hour. The membranes were then washed with TBS-T, as above, and stored in TBS-T until imaged. The blots were visualised using UptiLight HRP Blot chemiluminescent substrate (Interchim) on a ChemiDoc Imaging System (BioRad).

Name	Dilution	Source	Product Code	Company
Anti-His$_6$	1:10,000	Mouse	MAB050	R&D Systems
Anti-V5	1:5,000	Mouse	MCA1360GA	Bio-Rad
Anti-pSMAD-1/5/8	1:500	Rabbit	AB3848-I	Merck Millipore
Anti SMAD1	1:1,000	Rabbit	D59D7	Cell Signalling Technology
Anti-mouse	1:4,000	Goat	HAF007	R&D Systems
Anti-Rabbit	1:2,000	Goat	P0448	Dako

Table 2-5 – Table of Western Blot Antibodies

List of primary and secondary antibodies, the dilutions used, their source, and the product code and company they were purchased from.

2.6.3 UV Absorbance Spectroscopy

Protein concentrations were estimated by applying the Beer-Lambert Law (Equation 2-1) to the absorbance measure from a Nanodrop-1000 spectrophotometer (Nanodrop) at A_{280}.

$$A = \varepsilon C L$$

Equation 2-1 – Beer-Lambert law

Where 'A' is the A_{280}, 'C' is molar concentration, 'L' is length in centimetres and 'ε' is the molar extinction coefficient of the protein, per centimetre. Here 'L' is set at 1 cm.

The extinction coefficients were determined through the Expasy ProtParam web service (http://web.expasy.org/protparam/), and are listed in Table 2-6.

Protein	Extinction coefficient $(M^{-1}cm^{-1})$
BMPER	74985
N-BMPER	21450
Tsg	23960
Chordin vWFC2-3	20730

Table 2-6 – Table of Extinction Coefficients.

Extinction coefficients were determined by submitting the protein sequence, without the signal peptide, to the online ProtParam web server (http://web.expasy.org/protparam/), where the additive effects of each residues contribution to the extinction coefficient was calculated.

2.6.4   Mass Spectrometry

Samples were submitted to the University of Manchester Mass Spectrometry Core Facility (Michael Smith Building, University of Manchester, Oxford Road, M13 9PL). Protein bands were excised from SDS-PAGE gels, trypsin digested and analysed using the HCT Ultra mass spectrometer (Bruker Daltonics). Peptide fragments were then analysed using the programme Scaffold (Proteome Software).

2.7   **Cell based assays**

2.7.1   Heparan Sulphate Binding Assays

Cells were seeded into a six well plate as described above (section 2.5.2) and allowed to become confluent overnight. The media was then removed and the cells washed with PBS. Fresh expression media was added to the cells containing 0, 1 or 10 mg/mL of unfractionated heparin (Iduron). The cells were allowed to incubate for 48 hours before the media was analysed by Western blot (section 2.6.2)

2.7.2   Alkaline Phosphatase Assays

Alkaline phosphatase assays were carried out in the Sengle Lab (University of Cologne, Germany) with recombinant protein sent to them. Full details can be found in Troilo et al (Troilo et al., 2014).

2.7.3 Phospho-SMAD assays

Stably transduced HEK293-T cells were seeded into a 6 well plate in a 1:40 dilution from a confluent T75. After 24 hours the growth media was replaced and the cells were incubated for a further 48 hours. Media was removed and the cells washed with 1 mL of ice cold PBS before being lysed with 100 µL of radioimmunoprecipitation assay (RIPA) buffer containing Protease Inhibitor Cocktail (Sigma) (1 µL per 100 µL of RIPA buffer). Cell debris was then transferred to a centrifuge tube in the RIPA buffer and mixed at 4 °C for 1 hour before centrifugation at 4 °C for 10 minutes. The supernatant was transferred to new tubes and stored at -80 °C. When required, the total protein concentration of the lysate was measured using a bicinchoninic acid (BCA) assay as per the manufacturer's instructions (Thermo Fischer Scientific). 50 µg of lysate protein was loaded onto an SDS-PAGE under reducing conditions, see section 2.6.1, and analysed by Western blot, see section 2.6.2, using anti-pSMAD-1/5/8. Secondary antibody, anti-mouse conjugated to HRP, was used with ECL reagent to image the bands. The blotting membranes were stripped by incubating with 5 mL of Restore Western Blot Stripping Buffer (Thermo Fischer Scientific) at 37 °C for 30 – 60 minutes. The membrane was then blocked and re-probed with anti-SMAD1. Blots were imaged and analysed using Bio-Rad instruments and software, see section 2.6.2. Image intensity analysis was performed using GraphPad Prism version 7.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

2.8 Biophysical Methods

2.8.1 Multi-angle Light Scattering

Samples of BMPER, (0.1 - 0.5 mg/mL), were loaded onto a S200 10/300GL column (GE Healthcare) at a flow rate of 0.75 mL/min, equilibrated in a buffer compatible for the desired protein. Samples eluted from the column were flowed through a DAWN HELEOS II (Wyatt), coupled with an Optilab rEX differential refractive index detector (dRI) (Wyatt). The DAWN HELEOS II contains an 18-angle laser photometer, where one of the detectors was replaced with a QELS detector (Wyatt) to determine the hydrodynamic radii (R_H) of particles. The molecular mass moments, R_H, polydispersity and concentration of the samples were analysed using the software Astra 6.1 (Wyatt).

2.8.2 Analytical Ultracentrifugation

Samples of BMPER (0.1 - 0.5 mg/mL) were analysed using an XL-A centrifuge (Beckman). Samples, in a compatible buffer, were centrifuged at 163296 $\times g$ in an An60Ti-4 Hole rotor at a temperature of 20 °C with a buffer blank. Every 180 seconds the sedimenting boundary was monitored at a wavelength of 230 nm for 250 scans. The resulting data were analysed using the Sedfit software suite where a continuous model-based distribution was evaluated by the Lamm equation (Schuck, 2000). The resulting sedimentation coefficients were then corrected for solvent conditions using Sednterp (Biomolecular Interaction Technologies Centre), producing a more accurate R_H and f/f_0 values for each construct.

2.8.3 Surface Plasmon Resonance

SPR was carried out on the ProteOn XPR36 Protein Interaction Array System (Bio-Rad), using the ProteOn GLS Sensor Chip (Bio-Rad) primed in PBS with Tween (0.005% v/v, PBS-T). Tsg or chordin vWFC2-3 was immobilised onto the sensor chip at approximately 300 response units (RU) using NHS-EDC amine coupling and the reference lane blocked with ethanolamine. Analytes were diluted in PBS-T to their desired concentration for kinetic analysis. Analytes were flowed over the chip at 50 µL/min and the response curves recorded. Data from N-BMPER binding to Tsg or chordin vWFC2-3 were fitted to a 1:1 Langmuir model, where binding data for BMPER binding to Tsg or chordin vWFC2-3 were fitted to an equilibrium model.

2.9 Structural Methods

2.9.1 Transmission Electron Microscopy

2.9.1.1 Data Collection

Negative-stain electron microscopy was carried out using carbon coated 400 mesh copper grids (Electron Microscopy Sciences). Grids were glow discharged at 25 mA for 25 seconds using a K100X Glow Discharger (EMITECH). 3 µL of recombinant BMPER (0.02 mg/mL) was added to the grid and allowed to adsorb to the surface for 60 seconds before being washed with 60 µL of 2% uranyl acetate (w/v) and wicked dry with blotting paper. Samples were loaded into a Tecnai 12 Twin microscope (FEI) and images recorded onto a TVIPS charge coupled device (CCD) camera (Tietz) fitted to the microscope. The microscope operated at 120 KV with electrons produced from a LaB_6 electron source. Samples were imaged with a sampling of 2.8 Å/pixel.

2.9.1.2 *Data Processing and model reconstruction*

Micrographs were analysed using EMAN 2.0 software suite (Ludtke, 2016). Particles were picked using the swarm picking method, using a box size of 128 pixels (358 Å). The particles were then contrast transfer function (CTF) corrected in using e2ctf.py and a particle set of 3013 particles generated using e2boxer.py in the EMAN 2.1 suite (Tang et al., 2007). The particles underwent a reference-free alignment and averaging to produce 40 classes of different orientation particle averages. An initial model was generated using e2initialmodel.py and was followed by the e2refine workflow, generating model refined class averages (Tang et al., 2007). These underwent eight rounds of 3D refinement to produce a final model of BMPER. The final resolution for each map refined in RELION 1.4 (Scheres, 2012b), using e2refinetorelion3d.py (Ludtke, 2016), was determined using the FSC at the 0.143 criterion (Scheres, 2012a).

2.9.2 *Small-angle X-ray Scattering*

2.9.2.1 *Data Collection*

SAXS data were collected using size-exclusion chromatography SAXS (SEC-SAXS), although some data shown in Chapters 4 and 6 were collected using a robot sample changer (Round et al., 2015; Round et al., 2008) in batch mode.

At Diamond Light source, SAXS studies are carried out at beamline B21 which, making use of the bending magnet, can produce a flux of 1×10^{11} photons per second. At the sample, the beam has a cross section of 5 mm by 1 mm (width by height) which allows for a low flux density. The beam has a wavelength (λ) of 1 Å, and the scattering data are collected on a Pilatus 2M detector (Dectris) set at 4.09 m from the sample. The detector was configured to measure a scattering vector (q) range from 0.0022 to 0.42 Å^{-1}.

At B21 data were collected using SEC-SAXS on an Agilent HPLC system and the 3.2/300 Superdex Increase (S200i) series columns (GE Healthcare) with a 2.4 mL bed volume. Samples were loaded into the HPLC using 96-well plates and the samples were loaded onto the desired column using the Agilent injection system. The sample cell was a custom sample cell, described in Chapter 4. When using SEC-SAXS there are two different methods for collecting data, the first is to collect each

frame over the length of the elution at three second exposures, and the second is to pause the flow of the SEC elution, where the protein peak with highest concentration of desired protein is in the path of the beam. HPLC flow rate was 0.05 mL/min for S200i columns.

SAXS data were collected and images corrected for beam current variance through Diamond's Generic Data Acquisition (GDA) software (www.opengda.org) and each frame was processed by Data Analysis WorkbeNch (DAWN) (Basham et al., 2015) through the B21 pipeline to produce reduced one dimensional (1D) scatter curves for each frame collected. The reduced frames are outputted as '.dat' files for processing.

2.9.2.2 Data Processing

For each SEC-SAXS run, the processed 1D frames were analysed in ScÅtter (www.bioisis.net). ScÅtter was used to monitor the change of sample intensity by subtracting a predicted background buffer from each frame. This trace is akin to the A_{280} trace monitored during HPLC of the sample. Using the trace, the background is selected from a region where the buffer is flat and the peak is then selected. The background is subtracted from each of the selected peak frames and the R_g is estimated and plotted across the integral of ratio to background per frame. Regions of a flat, and stable R_g are selected for subtraction. The frames selected for subtraction are first scaled together, and averaged before the scaled and averaged buffer is subtracted to give a final scatter curve. All data processing was carried out in Ångstroms.

From the 1D reciprocal scatter plot it is possible to transform the data into real space though an IFT (Equation 1-9) resulting in a PDDF. ScÅtter required the use of DATGNOM (Petoukhov et al., 2012) in order to produce '.out' files, compatible with bead modelling software, described below.

2.9.2.3 Bead Modelling

2.1.1.1.1   DAMMIF

DAMMIF (Franke and Svergun, 2009), part of the ATSAS suite, was run using the ScÅtter GUI. Due to this, only the default settings were used, with the option for 'Slow' or 'Fast' specified. 23 DAMs were generated in ScÅtter, run in slow mode, and were allocated to eight central processing unit (CPU) threads in order to decrease the run

times and prevent overloading the CPU; ScÅtter allows parallelism for DAMMIF, a single threaded program. These models read data from the '.out' files which were generated in DATGNOM. The DAMAVER suite (Volkov and Svergun, 2003) was then run to process the DAMs, in order to average and filter the models, resulting in a final model.

2.9.2.3.1 Multi-Phase Modelling

Multi-phase bead modelling was carried out using MONSA (Svergun, 1999). As data were collected for both N-BMPER and BMPER, two phases are required for bead modelling. The refined '.dat' files were produced and the R_g, D_{max} and volume for each sample were calculated in ScÅtter. In order to run MONSA, both a master ('.mst') and a control ('.con') were created. The master file contains the volumes, the R_g of the constant phase, and the phase information. The volume was determined by fitting a straight line to the Porod-Debye plot. The master file also contains the names of the two control files. Within each of the control files are the names of the '.dat' files used in the calculation. MONSA requires a sphere of beads to be generated using DAMESV, where the lengths in x-, y- and z-axes are entered. From the D_{max} of BMPER, a DAMESV sphere of 200 Å × 200 Å × 200 Å was generated of two different phase beads and randomised. 20 MONSA models were generated independently on a desktop computer with a maximum number of steps of 500, and an annealing schedule factor of 0.95. Each of the completed MONSA phases were split into directories determined by their phases using a Ruby script (Rambo, 2015). The primary phase models, representing N-BMPER, were averaged together using the DAMAVER suite (Volkov and Svergun, 2003). The models of the secondary phase were then rotated according to the transformation applied to the matrix of the final DAMFILT model using a further Ruby Script (Rambo, 2015). The DAMFILT model and the phase-2 models were viewed in UCSF Chimera (Pettersen et al., 2004) in order to manually select the models with the best fit.

2.9.2.4 Rigid Body Modelling

2.9.2.4.1 CORAL

CORAL (Complexes with Random Loops) is a program for generating rigid-body models of individual domains (Petoukhov et al., 2012) by implementing a combined approach of SASREF and BUNCH (Petoukhov and Svergun, 2005). Homology models of vWFC domains one to five of BMPER were generated using the SWISS-

MODEL online server (Biasini et al., 2014), which were the rigid components of the modelling. A control file for CORAL was written specifying the linker regions between each of the domains and the correct sequence of domains. This was then used to assemble a protein chain where domains were connected in order but not fixed in space, and no domains were paired. CORAL was repeated 10 times and the outputs were compared to SAXS parameters measured from the data.

2.9.2.5   UCSF Chimera Commands for SAXS Models

UCSF Chimera (Pettersen et al., 2004) was used to render a surface onto bead models at varying resolutions. In order to represent DAMs with a surface the "molmap" function was used with Chimera Command 1.

```
molmap #x value
```

Chimera Command 1 – Where 'x' is the model to have the surface applied and 'value' is the resolution at what the surface is applied.

Comparisons between two volumes, either volumes rendered with Chimera Command 1 or from EM, were made using the "Fit in map" function in UCSF Chimera using Chimera Command 2.

```
fit #w #x search y res z
```

Chimera Command 2 - Where 'w' is the model to be fitted into volume 'x', which are searched 'y' times and 'z' as the resolution of the surface rendered onto model to be fitted.

These commands, and variations thereof, were used throughout the comparisons of structural models.

2.9.3   Computational methods

2.9.3.1   Calculating hydrodynamic parameters from structure

SOMO has been developed as a way of calculating the hydrodynamic parameters of both atomic models and low resolution bead models.

HYDROMIC (Garcia de la Torre et al., 2001), fills the volume of an EM map with beads given a specific threshold. The hydrodynamic parameters were calculated from

the resulting bead model. EM2DAM, within the ATSAS suite (Petoukhov et al., 2012), also fills the volume of an EM map with beads given a specific threshold, similar to HYDROMIC. From a bead model generated from the EM map it was possible to obtain the SAXS parameters calculated with CRYSOL.

2.9.3.2 *Calculating SAXS Curves from high resolution structures*

Crysol3 (Svergun et al., 1995) and FoXS (Schneidman-Duhovny et al., 2010; Schneidman-Duhovny et al., 2013, 2016) were used independently calculate the scattering curves of atomic structures. Both were used with default settings, with the exception that the scattered intensity for each of the models was cut to the q_{max} of the N-BMPER data using either the '-sm' argument for CRYSOL3, or '-q' for FoXS.

2.10 3D Printing

3D printing of sample cells was completed using a MiiCraft+ 3D printer (http://www.miicraft.com/) utilising stereolithography (SLA) based digital light processing (DLP). Using the MiiCraft interface, models were printed using the slow setting, printing ~100 µm in 14 seconds. The completed 3D print was then cured for 300 seconds. Clear BV-003 resin was used to print the sample cells. All sample cells were designed by beamline B21 staff and generated as stereolithography files, '.stl'. These were then converted to MiiCraft files, '.mii', in the MiiCraft software suite for printing. Once printing and curing was completed the sample cells were submerged in isopropanol and cleaned by sonication. The scaffold was then removed using a sharp scalpel and smoothed with sand paper. Each end of the sample cells were tapped to allow room for screws by the engineering team in the soft condensed matter village at Diamond Light Source.

3   Results Chapter 1: Nanoscale Structure of BMPER and N-BMPER

To investigate the hydrodynamic and structural parameters of full length BMPER and the vWFC domains of BMPER (N-BMPER), purified, recombinant proteins were required. HEK293-EBNA cells were stably transfected with expression vectors encoding BMPER and N-BMPER. Transient expression of BMPER fragments was also attempted. The overexpressed proteins were purified, and their hydrodynamic parameters were measured SEC-MALS and AUC. Structural measurements were carried out using solution SAXS and negative-stain TEM with single particle analysis. Proteins were freshly expressed and purified for each experiment throughout this thesis.

3.1   Expression and Purification of BMPER and N-BMPER

The study of BMPER and N-BMPER (Figure 3.1A) commenced with the generation of a BMPER construct through GeneArt Gene Synthesis (Thermo Fisher Scientific) based on NCBI Reference Sequence: NM_133468.4. N-BMPER was amplified from the synthetic gene by PCR (section 2.4.2) using primers shown in Table 2-1. Constructs were stably transfected into HEK293-EBNA cells (section 2.5.2.2) in T75 flasks, and expanded to larger flasks (section 2.5.1) and expression media was harvested (section 2.5.5) for purification (section 2.5.6). C-terminal His_6 protein purification tags provided a method of immobilised metal affinity chromatography (section 2.5.6) and for specific detection of the constructs using Western blot analysis using anti-His_6 antibodies (R & D Systems) (section 2.6.2). This can be seen by the expression tests of BMPER in Figure 3.1-D. When Western blotting for BMPER in conditioned media under non-reducing conditions there is a band representing the full-length of BMPER at approximately 75 kDa (compared to 77.2 kDa by sequence) and a C-terminal cleavage product, containing a His_6 tag at approximately 38 kDa. Under reducing conditions, the 75 kDa band decreased in intensity and the 38 kDa (compared to 37.1 kDa by sequence) band increases (Binnerts et al., 2004; Moser et al., 2003). This was caused through the reduction of the disulphide bond linking the acid-catalysed cleavage products. Affinity purified BMPER and N-BMPER were further purified by size exclusion chromatography (SEC) using S200 or S200i resin (GE Healthcare) (section 2.5.7). SEC resulted in monodisperse proteins shown by cleaner bands on an SDS-PAGE (section 2.6.1), migrating at estimated molecular

masses of above 62 kDa for BMPER and slightly below 42 kDa for N-BMPER when compared to a SeeBlue® Plus2 Pre-Stained Standard (Thermo Fischer Scientific) (Figure 3.1-E). BMPER maintained a faint band at approximately 38 kDa which likely represents the C-terminal cleavage product. Using different protein markers resulted in different mass estimates from an SDS-PAGE gel.

By analysing the sequence of BMPER and N-BMPER, and from work from Kamimura *et al.* (Kamimura et al., 2004b), it was shown that the N-terminal fragment of BMPER, but not the C-terminal fragment, contains N-linked glycosylation sites (Figure 3.1-A). The PNGase-F assay conducted by Kamimura *et al.,* was repeated using recombinant N-BMPER. From an SDS-PAGE analysis, visualised by Coomassie staining, it appears that approximately 12 kDa of mass was removed from N-BMPER through deglycosylation (Figure 3.1-E).
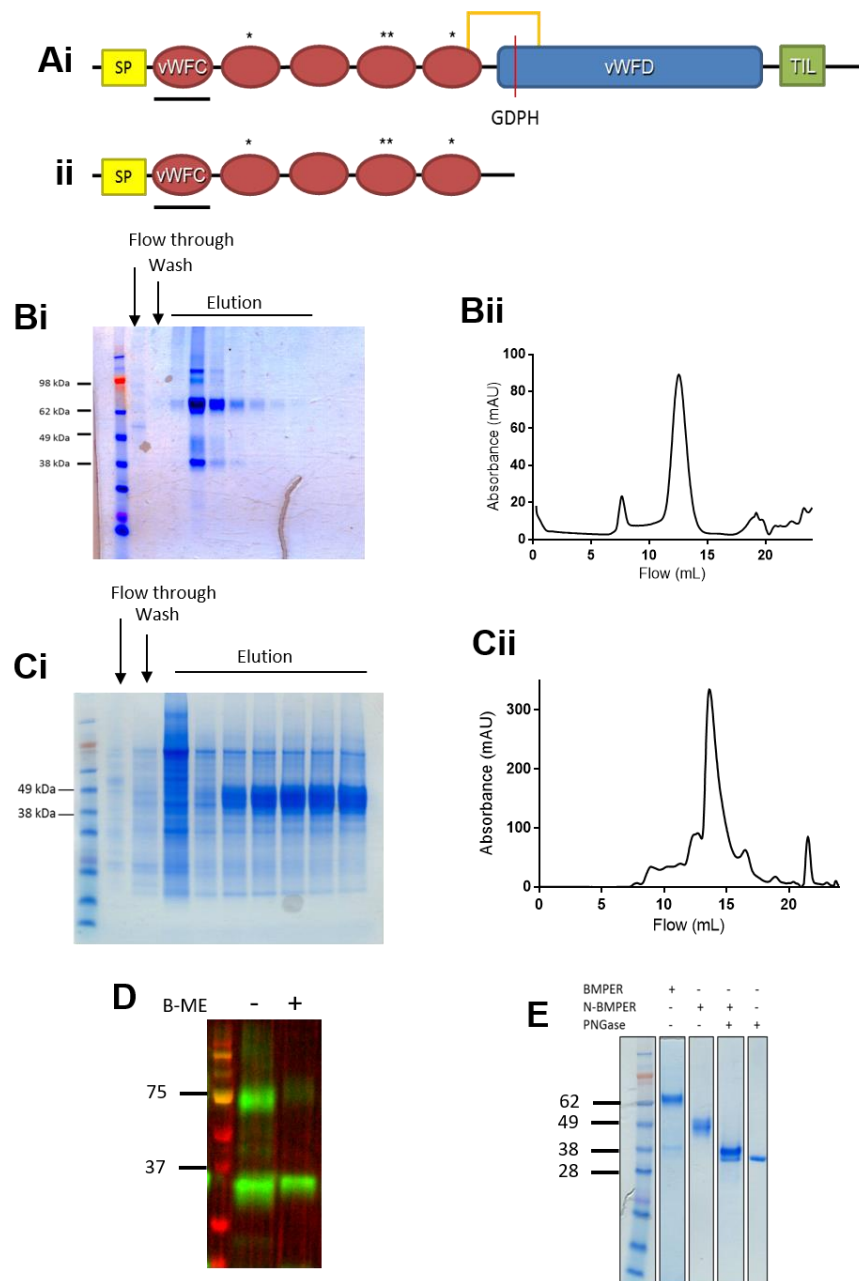
Figure 3.1 – Expression and Purification of BMPER and N-BMPER.

Schematic representations of BMPER (Ai) and N-BMPER (Aii), showing the domain type and arrangement (vWF = von Willebrand Factor, TIL = Trypsin inhibitor like domain), the BMP binding domain (vWFC1, underlined in black), glycosylation sites (*), inter-domain disulphide bond (between vWFC5 and vWFD) and the location of the 'GDPH' cleavage site (red line). Bi) Coomassie-stained SDS-PAGE Gel of Ni-NTA purification of recombinant BMPER showing the flow through of the column, the wash step and the elution fractions. Bii) SEC trace of the most concentrated HisTrap elution fraction where BMPER is the largest peak on the trace shown by A_{280}. Ci) Coomassie-stained SDS-PAGE Gel of Ni-NTA purification of recombinant N-BMPER showing the flow through of the column, the wash step and the elution fractions. Cii) SEC trace of the largest elution fraction where BMPER is shown to be the largest peak on the trace. D) An expression and secretion test of BMPER was carried out using reducing (β-ME +) and non-reducing conditions (β-ME -) on media secreted from BMPER. E) Coomassie-stained SDS-PAGE gel of SEC purified BMPER and N-BMPER, along with a PNGase F deglycosylation assay.

In addition to stable cell lines producing BMPER constructs (section 2.5.2.2), 25 additional BMPER constructs were designed for use in the pHLsec vector with either a His_6 or a 1D4 purification tag (Aricescu et al., 2006). This was completed in combination with The Division of Structural Biology at the University of Oxford (STRUBI). The constructs were designed to be used in conjunction with each other in the crystallisation of domain groups. Four constructs were also designed for chordin and two for Tsg as these would also contribute to projects within the lab. Each of the constructs' domain boundaries can be seen in Table 3-1. The pHLsec vectors contain a β-actin promoter, have a cleavable γ-phosphatase signal peptide and have been used widely in the expression of recombinant proteins at high level (Aricescu et al., 2006; Bell et al., 2013; Chang et al., 2007; Healey et al., 2015; Zhao et al., 2011).

Construct Boundaries	Labels	Construct Boundaries	Labels
BMPER S40 – G226	1	BMPER I298 – L496	16
BMPER S40 – G236	2	BMPER I298 – H503	17
BMPER S40 – L291	3	BMPER I298 – S530	18
BMPER S40 – V300	4	BMPER I298 – E621	19
BMPER S40 – T358	5	BMPER I298 – R687	20
BMPER S40 – L496	6	BMPER T358 – L496	21
BMPER S40 – H503	7	BMPER T358 – H503	22
BMPER S40 – S530	8	BMPER T358 – S530	23
BMPER S40 – E619	9	BMPER T358 – E619	24
BMPER S40 – R687	10	BMPER T358 – R687	25
BMPER V230 – L496	11	Chordin A27 – A866	26
BMPER V230 – H503	12	Chordin A27 – S955	27
BMPER V230 – S530	13	Chordin D154 – A866	28
BMPER V230 – E621	14	Chordin D154 – S955	29
BMPER V230 – R687	15	Tsg C26 – F223	30
		Tsg N27 – F 223	31

Table 3-1 – BMPER, Chordin and Tsg constructs.

25 BMPER constructs were generated by PCR and ligated into pHLsec vectors with a C-terminal His6 tag or a 1D4 antibody epitope tag. BMPER constructs highlighted in red represent N-BMPER (construct 5) and BMPER (construct 10). Four chordin constructs were designed around the cleavage sites (constructs 26 to 29) and two Tsg constructs with the presence or absence of the first cysteine (constructs 30 and 31). Label numbers correspond to construct boundaries in Table 3-2.

Construct Labels	Construct Boundaries
1	VWC VWC VWC
2	VWC VWC VWC
3	VWC VWC VWC VWC
4	VWC VWC VWC VWC
5	VWC VWC VWC VWC VWC
6	VWC VWC VWC VWC VWC VWD
7	VWC VWC VWC VWC VWC VWD
8	VWC VWC VWC VWC VWC VWD
9	VWC VWC VWC VWC VWC VWD
10	VWC VWC VWC VWC VWC VWD TIL
11	VWC VWC VWD
12	VWC VWC VWD
13	VWC VWC VWD
14	VWC VWC VWD
15	VWC VWC VWD TIL
16	VWC VWD
17	VWC VWD
18	VWC VWD
19	VWC VWD
20	VWC VWD TIL
21	VWD
22	VWD
23	VWD
24	VWD
25	VWD TIL
26	VWC CHRD CHRD CHRD CHRD VWC VWC
27	VWC CHRD CHRD CHRD CHRD VWC VWC VWC
28	CHRD CHRD CHRD CHRD VWC VWC
29	CHRD CHRD CHRD CHRD VWC VWC VWC
30	C Tsg
21	Tsg

Table 3-2 – BMPER, chordin and Tsg construct compositions

The domain compositions of the constructs listed in Table 3-1 are shown.

The pHLsec vectors are transient mammalian expression vectors. Each construct was transfected into HEK293-T cells in six well plates and allowed to incubate for 48 hours (section 2.5.2.1). Western blots were performed on both the conditioned media and the cell lysate ('M' and 'L' respectively) of cells transfected with constructs with both the C-terminal His_6 and 1D4 purification tags (Figure 3.2). The transfections of all constructs were completed with the exception of construct eight for the 1D4 tagged proteins due to a failure in its generation. Conditioned media from stably transfected BMPER HEK293-EBNA cells was also Western blotted to compare expression, shown in Figure 3.2-A and labelled 'S'. Both His_6 and 1D4 tags were used with a positive control (+ve), using a protein characterised in STRUBI, and a negative control

(-ve), untransfected HEK293-T. Clearly, the stable expression system shows no band due to the low concentration of BMPER in the conditioned media. While this may indicate that the expression of BMPER from a transient system may be no better than the expression of BMPER from a stable system at producing protein, under these circumstances the stable cell lines are not creating a bottle neck, relative to the transient system, which the pHLsec vector was designed to remove (Aricescu et al., 2006).

Visualising the Western blots probed with the His$_6$ antibody (Figure 3.2-A) it is possible to see very few bands were present. Of the bands that were present it appeared that the majority of the bands were seen in the cell lysate ('L'). Although this was predominantly the case, some constructs were secreted, including: 6, 7, 27, 28 and 29, as well as the positive control. These secreted bands included two BMPER constructs (six and seven) and three chordin constructs (27, 28, and 29). Visualisation of the Western Blot probed with the 1D4 antibody (Figure 3.2-B) showed similar results to the His$_6$ Western blot in that most of the samples were shown to be expressed but not secreted. Exceptions to this were seen by constructs three and four where they appeared to be secreted at the appropriate molecular mass. Further exceptions were seen for the chordin and Tsg constructs, lanes 26 to 31. As these constructs were secreted in high quantities they were pursued, although this is not within the scope of this thesis. As only two of the BMPER constructs were secreted, the HEK293-EBNA system was pursued.
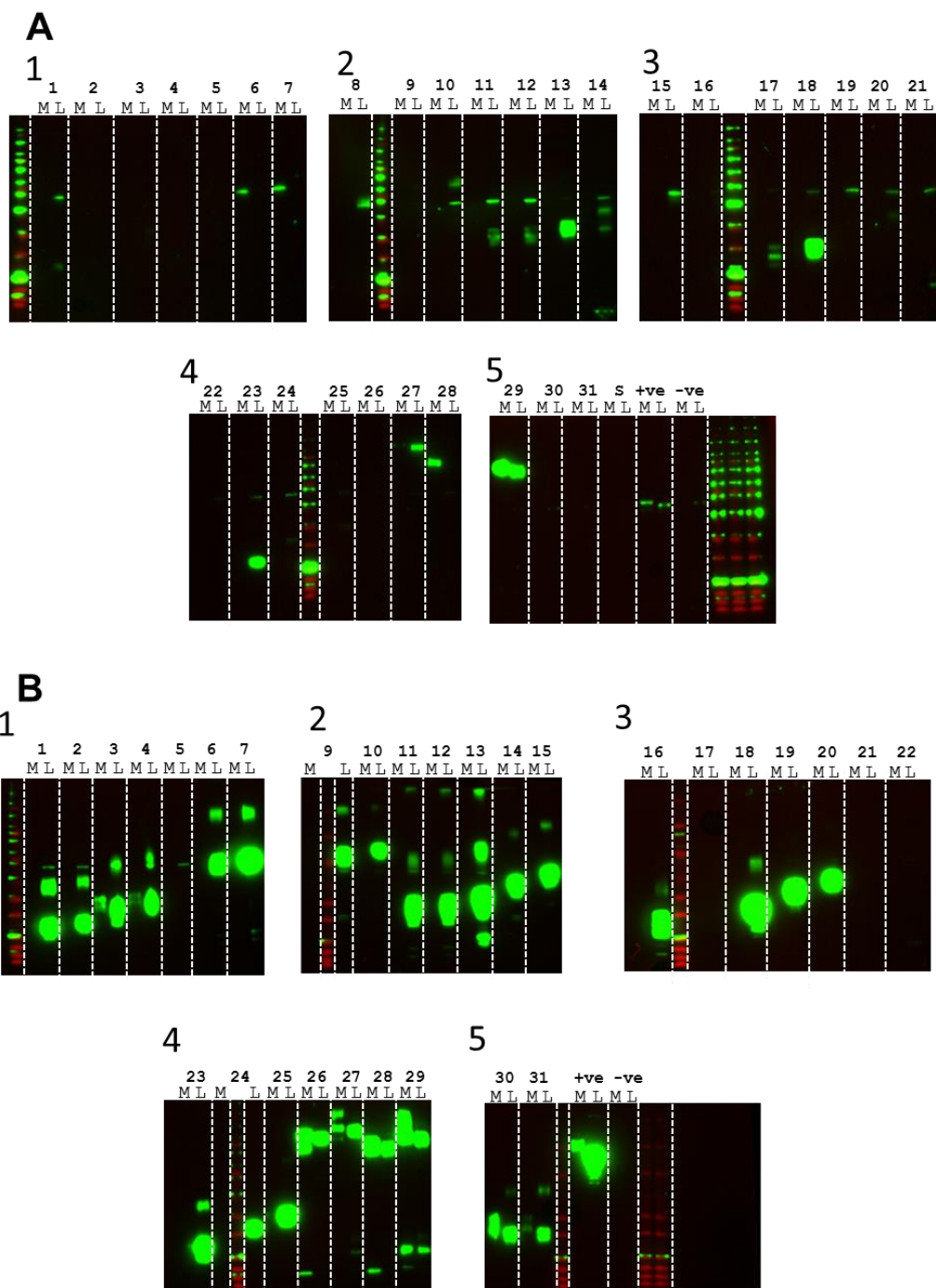
Figure 3.2 – Results of Transient Expression Trials

Expression trials of BMPER, chordin and Tsg transiently expressed in HEK293-T cells with a His_6 or 1D4 tag. A) Anti-His_6 Western blot showing the conditioned media ,'M', and lysate, 'L', of HEK293-T cells transfected with 31 constructs (25 BMPER, 4 chordin and 2 Tsg). Conditioned media and lysate of HEK293-EBNA cells stably transfected with BMPER was also shown under 'S'. Positive and negative controls were also shown by '+ve' and '-ve'. B) Anti-1D4 Western blot showing the conditioned media ,'M', and lysate, 'L', of HEK293-T cells transfected with 30 constructs (24 BMPER, 4 chordin and 2 Tsg). Positive and negative controls were also shown by '+ve' and '-ve'.

3.2 **Solution Hydrodynamic Measurements of BMPER and N-BMPER**

3.2.1 Hydrodynamic Measurements and Analysis of BMPER and N-BMPER

SEC-MALS (section 2.8.1) and AUC (section 2.8.2) are excellent methods for looking at the hydrodynamic properties of samples in their solution state. SEC-MALS shows that BMPER (blue) and N-BMPER (red) have single peaks measured by dRI (Figure 3.3-A). From the measurements it was possible to determine the estimated molecular mass of BMPER and N-BMPER as 100.4 kDa and 48.3 kDa. The hydrodynamic radii (R_H) measured by both SEC-MALS and velocity AUC were consistent for BMPER (51 and 53 Å) and N-BMPER (40 and 41 Å). Velocity AUC was also able to determine the frictional ratio (f/f_0) of 1.73 and 1.74 for BMPER and N-BMPER respectively, showing that although there is a difference in mass, they are elongated at the same proportional ratio. Comparisons of hydrodynamic measurements to structural model can be seen in **Error! Reference source not found.**.
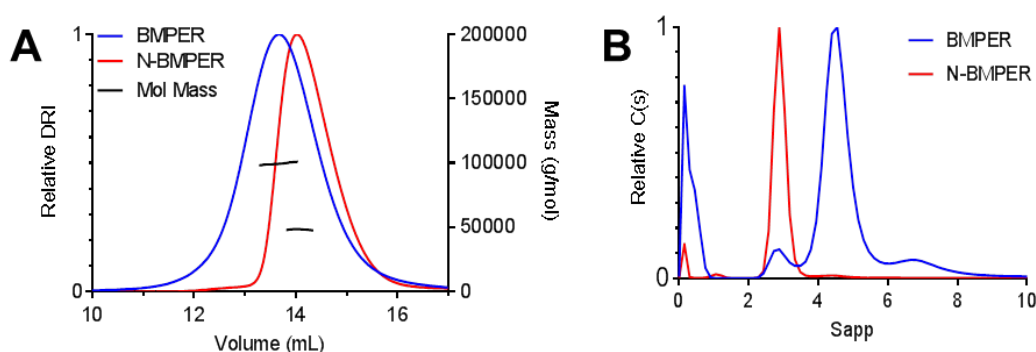
Figure 3.3 – Hydrodynamic Measurements of BMPER and N-BMPER.

A) SEC-MALS was used to show the monodispersity of the sample and measure the molecular mass of BMPER (blue line) and N-BMPER (red line) by dRI. The black line corresponds to the molecular mass of the protein. B) Velocity AUC of BMPER (blue, 500 mM NaCl) and N-BMPER (red, 150 mM NaCl) was used to measure the stoichiometry, R_H, $S_{20,w}$ and the f/f_0 of the samples. All parameters can be seen in **Error! Reference source not found.**.

3.3 Structural Measurements and Model generation of BMPER and N-BMPER

3.3.1 SAXS measurements of BMPER and N-BMPER

Solution SAXS data for both BMPER and N-BMPER were collected using SEC-SAXS to remove any trace aggregates (section 2.9.2.1). This allowed the purified sample to flow directly into the beamline. Advantages of this were that sample frames are averaged over the peak, encompassing a range of concentrations. HPLC-SAXS also allows a buffer exchange and so the sample frames have the same buffer that the column was equilibrated in. The data were reduced at Diamond Light Source using DAWN (Basham et al., 2015) and averaged using ScÅtter (http://www.bioisis.net). All further processing was completed with ScÅtter (section 2.9.2.2). The 1D scatter intensity data of BMPER and N-BMPER are represented in reciprocal space as a function of q (Å^{-1}) (Figure 3.4-A). The Guinier plots show all trace aggregates were removed by displaying a straight line (Figure 3.4-Ci &-Di) with corresponding residual plot confirming this (Figure 3.4-Cii &-Dii). The GA of BMPER and N-BMPER calculated the R_g at 51 Å and 34 Å respectively. An IFT produced a PDDF determining the real space R_g, 48 Å and 33 Å (**Error! Reference source not found.**). The shape of the PDDF is also indicative of elongated proteins (Figure 3.3-B), corroborating the elongation of both BMPER and N-BMPER. The D_{max} of BMPER and N-BMPER was 160 Å and 111 Å respectively.
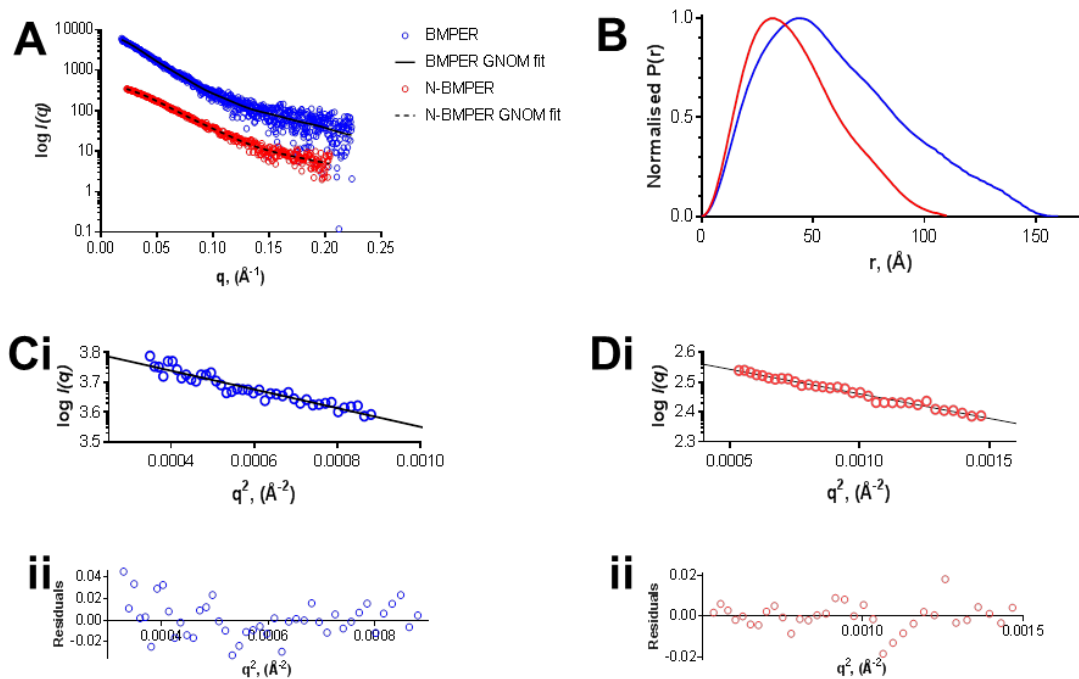
Figure 3.4 – SAXS Processing of BMPER and N-BMPER.

A) One-dimensional scattering profiles showing the log to the base 10 of BMPER (blue) and N-BMPER (red) scattering vectors as a function of q ($\log_{10} I(q)$ vs q). The results of processing the data are shown with the fitting of the IFT from the pair distance distribution function (PDDF) as a solid black line (BMPER) and a dashed line (N-BMPER). B) The normalised PDDF of BMPER (blue) and N-BMPER (red), representing the real space distance probabilities of each is presented with the band limited D_{max} of 160 Å and 111 Å respectively. The Guinier plots ($\log_{10} I(q)$ vs q^2) for BMPER and N-BMPER are represented in Ci and Di with their corresponding residual plots ($E_{obs} - E_{pred}$ vs q) in Cii and Dii. The respective reciprocal radius of gyration (R_g) and $I(0)$ were calculated at 51 Å and 7.92×10^{-3} for BMPER and 34Å and 21.23 for N-BMPER.

3.3.2 Structural Flexibility of BMPER and N-BMPER

SAXS is a powerful tool in assessing the 'folded-ness' of a protein as well as its flexibility. The dimensionless Kratky plot provides a semi-quantitative assessment of proteins regardless of size as the R_g and $I(0)$ are taken into account in the normalisation process (Durand et al., 2010; Receveur-Brechot and Durand, 2012). Examination of the dimensionless Kratky plot shows that for both BMPER and N-BMPER the peaks are not coincident with the cross-hair ($\sqrt{3}$, $3e^{-1}$) on both axis, signifying folded, elongated proteins (Figure 3.5-A). Both BMPER and N-BMPER are proportionally elongated, as was described by the frictional ratio by AUC (Figure 3.3). To assess the flexibility of proteins it is possible to exploit a power-law relationship that exist between compact and flexible proteins (Rambo and Tainer, 2011). As the dimensionless Kratky plot shows no intrinsic disorder, or high levels of flexibility, the

90

Kratky-Debye plot was omitted from flexibility analysis. Comparison of the SIBYLS Plot and the Porod-Debye Plot allows determination of whether BMPER and N-BMPER are flexible (Figure 3.5-B). Both proteins plateau in the SIBYLS plot showing that there is a degree of flexibility, between intrinsic disorder and complete rigidity (Rambo and Tainer, 2011).
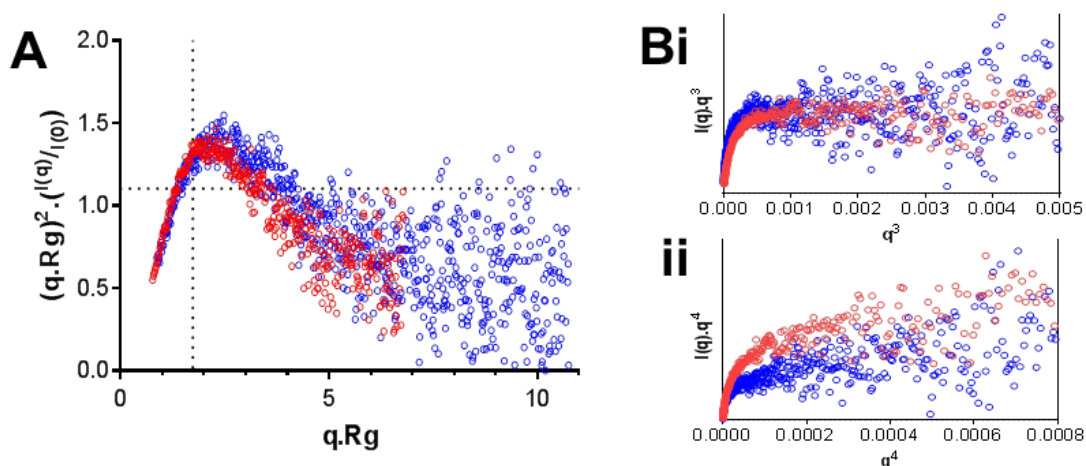
Figure 3.5 – Analysis of Flexibility of BMPER and N-BMPER.

A) Dimensionless Kratky plot ($(q.R_g)^2.(I(q)/I(0))$ vs $q.R_g$) showing that both BMPER (blue) and N-BMPER (red) are folded to a similar degree. B) The degree of flexibility of BMPER and N-BMPER can be seen by comparing the SIBYLS plot ($I(q).q^3$ vs q^3)(Bi) and the Porod-Debye plot ($I(q).q^4$ vs q^4) (Bii) to observe the first region to plateau.

3.3.3 ab initio Shape Reconstruction of BMPER

From SAXS data, low resolution DAMs can be generated for shape determination (section 2.9.2.3). 23 *ab initio* models were generated using DAMMIF to best represent the scatter profile of BMPER by DAMs (Franke and Svergun, 2009). The DAMAVER suite (Volkov and Svergun, 2003) produced a unique finalised model. For the 23 models generated, the mean NSD was 0.672 with a standard deviation of 0.035. Only one model was removed from the selection as the NSD had a difference of greater than 2σ. This consistency within the shape of the models shows a unique solution to the data. Each model was also checked to make sure that the bead models did indeed fit with the data. An example model has been used in Figure 3.6 showing the bead model fitting to the curve with a χ^2 of 1.126, and the inset showing a residual plot ($residuals = experimental\ values - expected\ values$) with a good fit. For BMPER, the resolution of the models is 44 Å (± 3 Å) with the FSC curve shown in Figure 3.6. The final filtered model appeared to be consistent with the data provided by SAXS and AUC with regards to an elongated structure (Figure 3.6). Further comparison of

BMPER to hydrodynamic data is carried out in section 3.4 and values are represented in **Error! Reference source not found.**.
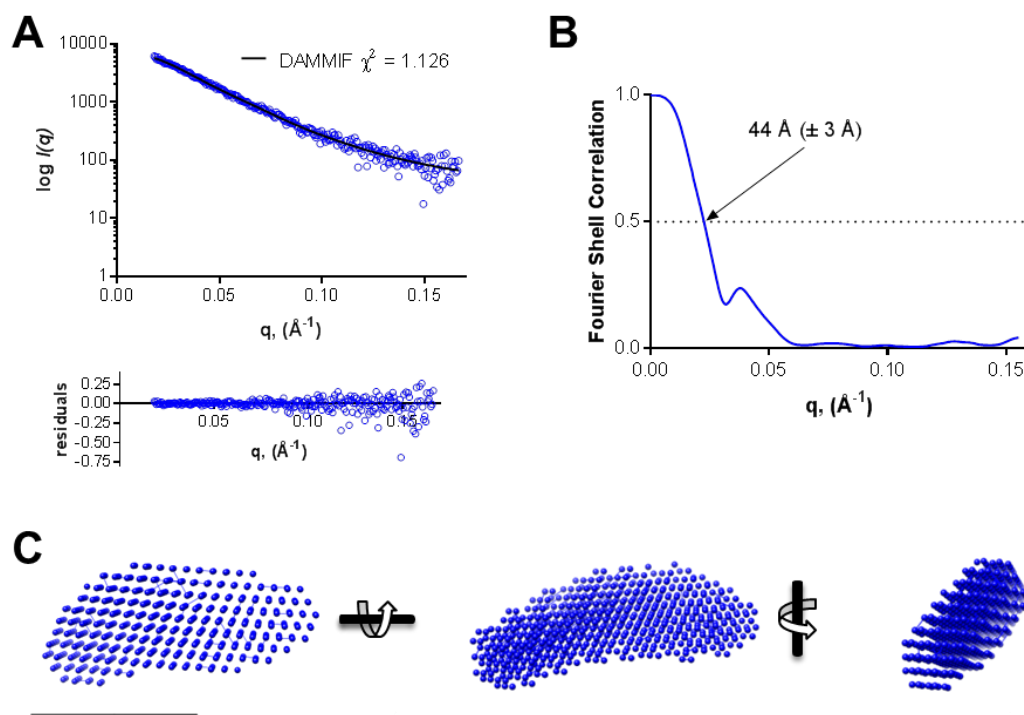
Figure 3.6 – *ab initio* Modelling of BMPER.

A) 1D scattering profile showing the log to the base 10 of BMPER scattering vectors as a function of q ($\log_{10} I(q)$ vs q) with an overlay of the bead model simulated scatter with a χ^2 of 1.126. B). The resolution of the most likely *ab initio* DAM is 44 Å (± 3 Å) and is plotted as the FSC as a function of q (FSC vs q). Resolution was determined by a 0.5 cut off on the FSC giving a resolution of 44 Å (± 3 Å). C) The final DAMFILT model of BMPER is shown with 90 ° rotations in the x- and y-axis. The scale bars represent 100 Å.

3.3.4   ab initio Shape Reconstruction of N-BMPER

DAM models were generated for N-BMPER, again using 23 models. The mean NSD of N-BMPER was calculated at 0.674 with a standard deviation of 0.064; DAMAVER excluded a single model, again showing consistency of the models giving a unique result to the data. An example of the data is represented in Figure 3.7-A, with a χ^2 of 0.213. The resolution of the resulting models generated for N-BMPER is 35 Å (± 3 Å) and represented in Figure 3.7-B by the FSC curve with a cut-off of 0.5. The resulting filtered model is shown to be elongated. Further comparison of N-BMPER to hydrodynamic data will be represented in section 3.4 and values are represented in **Error! Reference source not found.**.
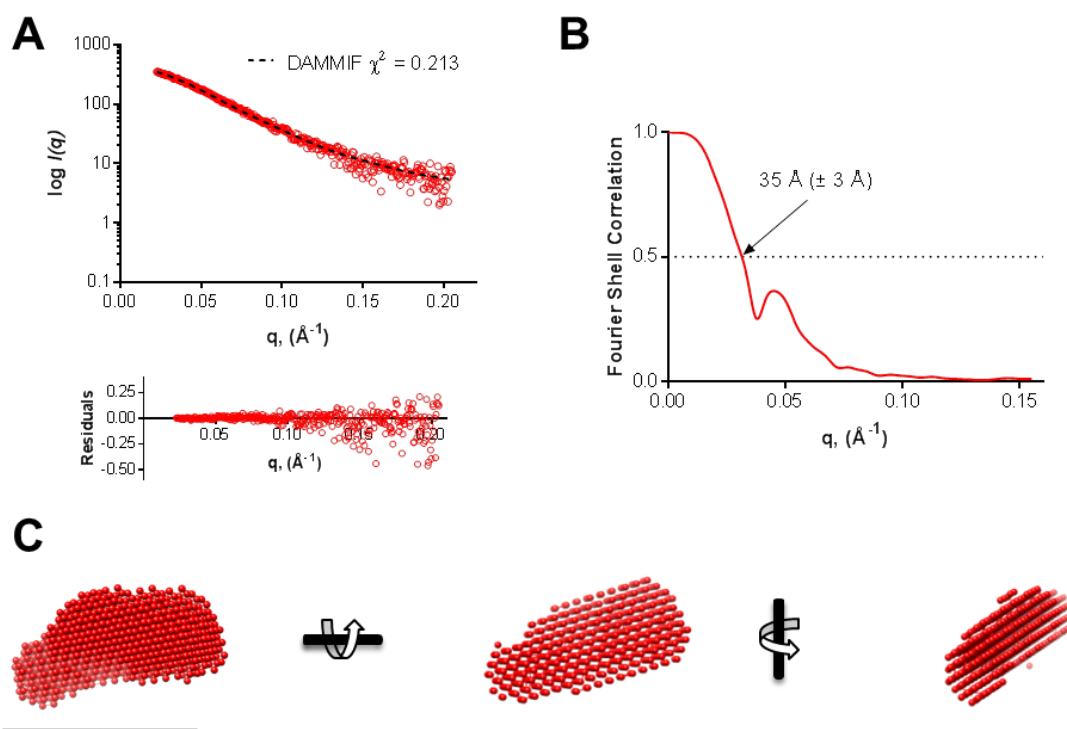
Figure 3.7 – *ab initio* Modelling of N-BMPER.

A) 1D scattering profile showing the log to the base 10 of N-BMPER scattering vectors as a function of q ($\log_{10} I(q)$ vs q) with an overlay of bead model simulated scatter with a χ^2 of 0.213. B) The resolution of the most likely *ab initio* DAM is 35 Å (± 3 Å) and is plotted as the FSC as a function of q (FSC vs q). Resolution was determined by a 0.5 cut off on the FSC giving a resolution of 35 Å (± 3 Å). C) The final DAMFILT model of BMPER is shown with 90° rotations in the x- and y-axis. The scale bars represent 100 Å.

3.3.5 *Quasi-atomic Resolution Model of N-BMPER using Rigid Body Modelling*

The crystal structure (Zhang et al., 2008) and solution NMR structure of the first vWFC domain of BMPER (Fiebig et al., 2013) provided a structure for homology models of vWFC domains to be generated. Ten rigid body models of N-BMPER were generated using the program CORAL (Petoukhov et al., 2012) (section 2.9.2.4.1). Filtering of the models was required to obtain the model that best fitted the data. The ten CORAL models fit the data with a χ^2 of around 0.2 (data not shown). With SAXS the aim is to fit the data with a χ^2 close to 1. Here the data appeared to be over fitted but with equally distributed residuals and so another mode of selection was required.

Firstly, the D_{max} for each of the 10 CORAL models was measured using ScÅtter (Table 3-3). In Figure 3.8, it is possible to see that some models have a D_{max} much

93

larger than that of N-BMPER, whereas others are within a 10% margin: models 1, 2, 5 and 7 (Figure 3.8-B).
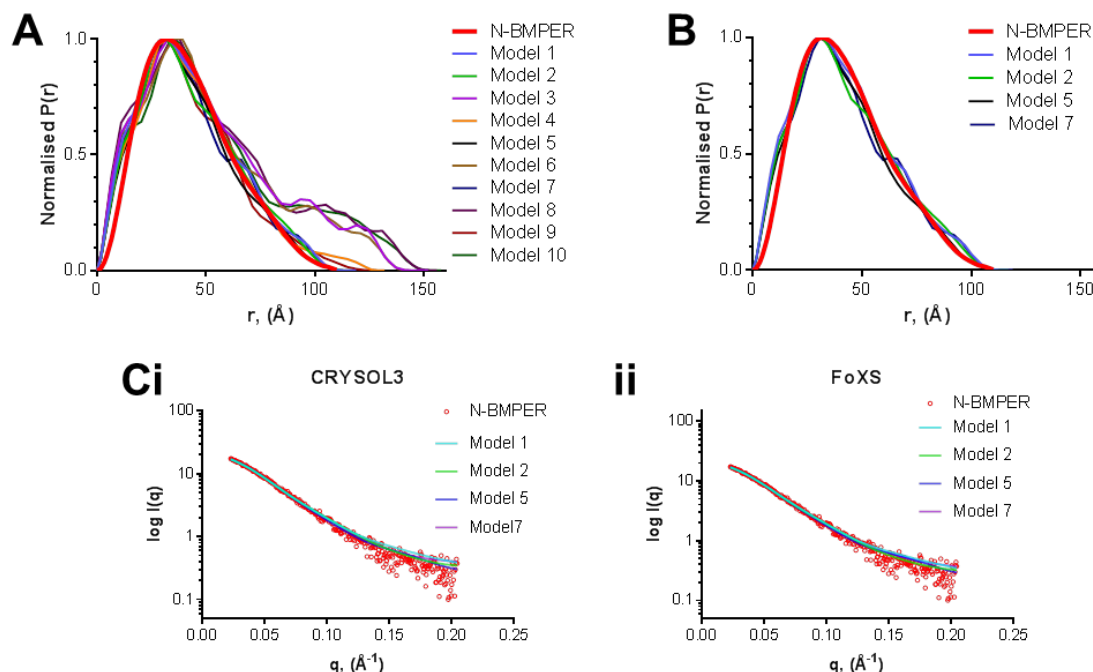
Figure 3.8 – Selecting CORAL models.

Ai) The normalised P(r) for each of the 10 CORAL models, with the real data overlaid in red. B) Normalised P(r) for each of the CORAL models (1, 2, 5, 7) that have a D_{max} that consistent with the measured data. C) The resulting scattering curves for the four models were simulated using CRYSOL 3 (Ci) and FoXS (Cii) to recalculate the fitting of the models to the scatter. D_{max} and χ^2 values can be seen in Table 3-3.

Secondly, the CORAL models generated were then compared to the N-BMPER data by calculating the 1D scatter curves using CRYSOL 3 (Svergun et al., 1995) and FoXS (Schneidman-Duhovny et al., 2010; Schneidman-Duhovny et al., 2016) (Figure 3.8Ci & -Cii), FoXS gives its results in χ rather than χ^2 so these values were squared to match CRYSOL for ease of reading. From this further comparison of the models to the SAXS curve it is possible to see that the original models that were selected based on D_{max} (1, 2, 5 & 7) were again the models with the best fitting to the data (Figure 3.8, Table 3-3). Exceptions can be seen with models four and nine with χ^2 of 0.32/0.25 and 0.41/0.34, calculated using CRYSOL 3 and FoXS respectively, but these models had a D_{max} larger than N-BMPER.

Model	D_{max}	χ^2		$R_g(\text{Å})$	
		Crysol 3	FoXS	Crysol 3	FoXS
N-BMPER	109	N/A		33.10	
1	119	0.44	0.32	34.30	33.52
2	117	0.33	0.24	33.93	33.63
3	152	0.87	0.90	43.08	42.50
4	132	0.32	0.25	35.37	34.71
5	113	0.24	0.24	33.68	32.67
6	144	0.98	1.14	42.77	41.96
7	113	0.37	0.28	34.14	33.46
8	156	1.29	1.42	44.80	44.55
9	128	0.41	0.34	34.76	33.64
10	158	0.82	0.98	44.78	44.59

Table 3-3 – Model Comparison Table.

This table shows all of the parameters that were calculated for each of the 10 CORAL models using ScÅtter, CRYSOL and FoXS.

With models narrowed down from ten to four, the remaining models were cDSompared to the DAM model of N-BMPER rendered to 35 Å resolution. With the resolution defined for the CORAL models there can be a correlation calculation for the fit of the CORAL models within the N-BMPER DAM volume, which is more widely reported as a 'goodness of fit'. For the CORAL models, the search was completed ten times and the resolution was set at 35 Å, the same as the DAM model resolution and surface. Figure 3.9 shows each of the models and their relative correlation to the SAXS model. Visually we can see that model 2 is too long and should be removed from consideration. Other than model 2, the correlation of fit for the atomistic models to the DAM models is high, greater than 0.9, where 1 is perfect. With this modelling information and the quantitative results above, CORAL model 5 fits the data the best and will be used moving forward in describing BMPER and N-BMPER in this thesis, although models 1 and 7 are very close overall.
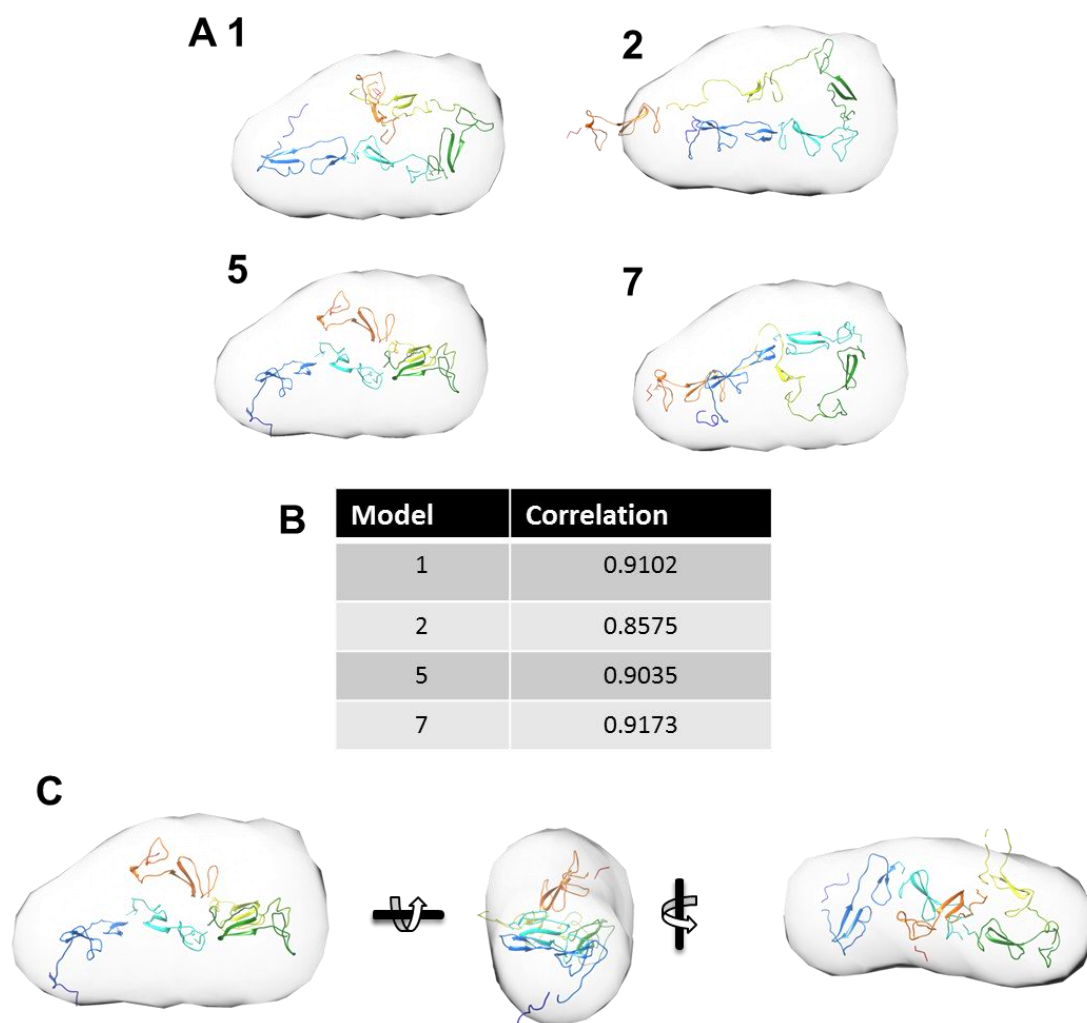
Figure 3.9 – CORAL Models Fitted into the N-BMPER DAMMIF Model.

A) Representation of all models fitted into the N-BMPER surface representation of a DAMMIF model using the UCSF Chimera function 'fit-in map' with a surface map representing a resolution of 35 Å. B) A table showing the correlation of models to the surface map at 35 Å. C) CORAL model 5 shown at three rotations about its axis.

3.3.6 Determination of the Locations of the N- and C-Termini of BMPER by Multiphase *ab initio* Modelling

Although SAXS is a powerful tool for looking at the structure of proteins, it isn't able to provide us with any information regarding the orientation of the proteins from a single-phase *ab initio* model. Volumes of BMPER and N-BMPER were measured from the SAXS data using ScÅtter at 320,000 and 140,000 Å^3 respectively. Having two scattering curves meant that two phases could be used to construct a multi-phase model (section 2.9.2.3.1). The first phase was N-BMPER with a volume of 140,000 Å^3 and the second phase was ΔN-BMPER, the remainder of BMPER representing the C-terminus, with a volume of 180,000 Å^3. 20 multi-phase *ab initio* models were generated using MONSA. The phases were split and phase-1 was averaged with the

DAMAVER suite and showed a distinct similarity to the DAM model of N-BMPER (Figure 3.10-B). Currently there is not a way to average the second phase and so a representative model can be found in Figure 3.10-B. Figure 3.10-C represents the fitting of the N-BMPER CORAL model 5 into phase-1 of the MONSA model.
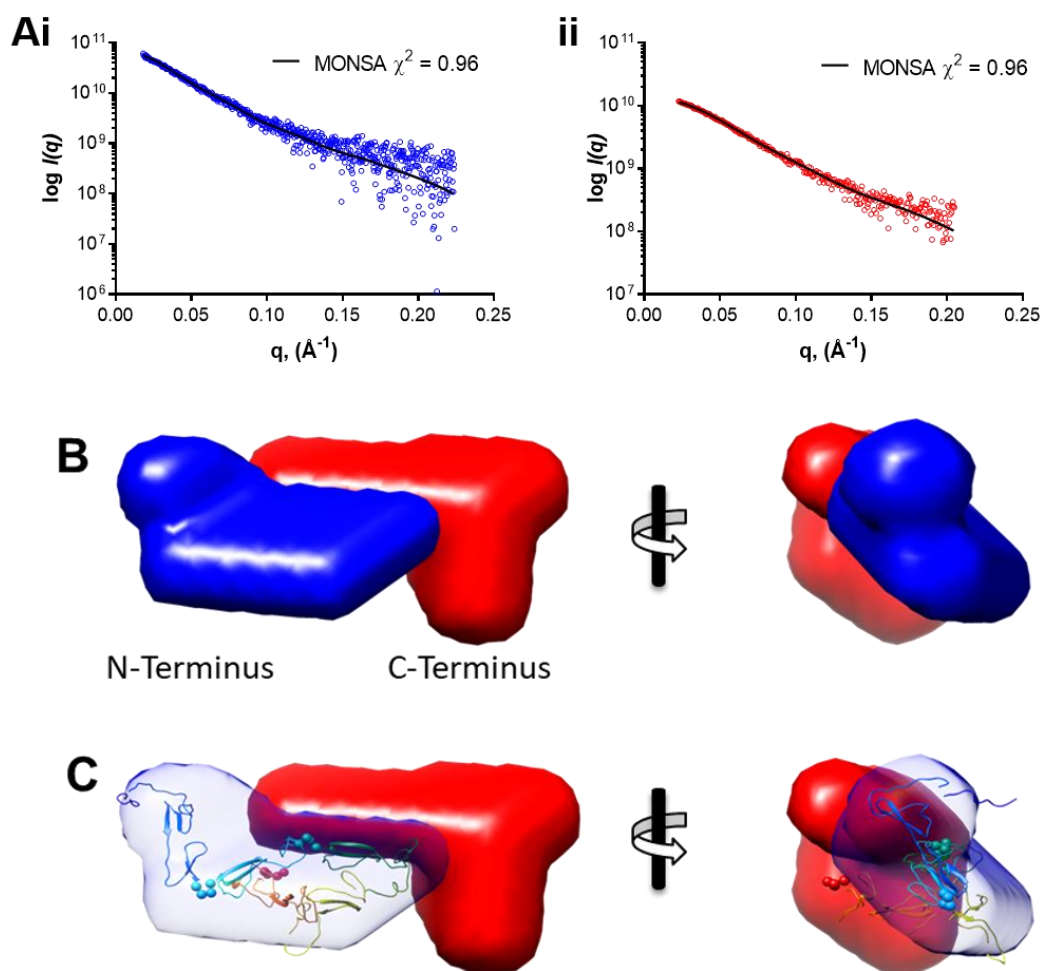
Figure 3.10 – Multiphase *ab initio* Modelling of BMPER.

A) The χ^2 values of BMPER (Ai, blue) and N-BMPER (Aii, red) are 0.96 and 0.23 respectively. B) A representative model from 20 MONSA runs shows the N-terminus of BMPER in red and the C-terminus (ΔN-BMPER) in blue with a 20 Å resolution shell over the beads. The models are presented with a 90 ° rotation in the y-axis. C) The N-BMPER CORAL model is fitted into the N-BMPER phase of the MONSA model using UCSF Chimera with a correlation of 0.8 at a resolution of 20 Å.

3.3.7 Validating the N-terminal Models of N-BMPER

It is possible to validate the phase-1 model using the single phase DAMMIF model representing N-BMPER (Figure 3.7-C) as they are generated from the same data set. Phase-1 was run through the DAMAVER suite to generate a final model (Figure 3.11-B) with an NSD value of 0.6 (SD of 0.032 and one model excluded). The final MONSA

model was compared to the DAM from DAMMIF in two ways. The first was using DAMSEL to align the models and compare them. This resulted in an NSD of 0.526, showing good agreement, but also using 'fit-in map' at a resolution of 35 Å providing a correlation of 0.98. Both methods showed high levels of agreement through independent model generation.
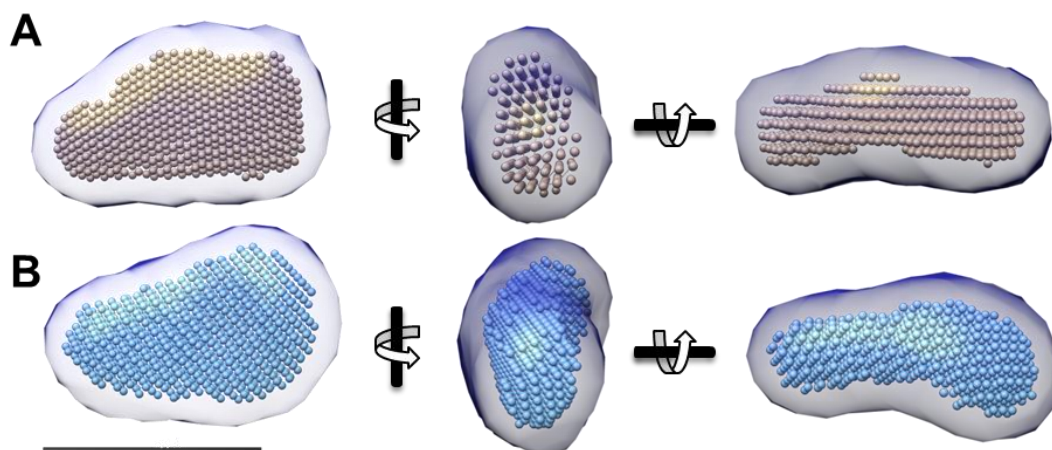
Figure 3.11 – Comparison of N-BMPER DAMs from Single and Multiphase *ab initio* Modelling.

A) DAMFILT model of N-BMPER generated from 22 models, shown as gold beads with a blue surface at a resolution of 35 Å and rotation in the x- and y- axes. B) DAMFILT models of phase-1 generated from 19 models, shown as blue beads with a blue surface at a resolution of 35 Å and rotated in the x- and y- axes. All surfaces were generated using 'molmap' Chimera Command 1. The scale bar represents 100 Å.

3.3.8 3D reconstruction of BMPER Using Negative Stain Transmission Electron Microscopy

Negative-stain TEM was used to further validate the structural information of BMPER. A dilute, monodisperse sample of BMPER was added to a carbon coated copper grid and subjected to negative staining. Discreet negatively stained particles could be imaged, as represented in Figure 3.12-A, with an inset showing particles at a magnification of 1.5x (section 2.9.1.1). The micrographs revealed that individual particles of BMPER are elongated. 3013 particles were picked using the EM software EMAN 2.1 (Tang et al., 2007) (section 2.9.1.2). Eight rounds of reference-free class averaging was undertaken in order to remove any 'bad' particles. The reference-free class averages are then generated and revealed BMPER as an elongated molecule with a bend in the centre (Figure 3.12-B). No symmetry was applied throughout the EM processing.
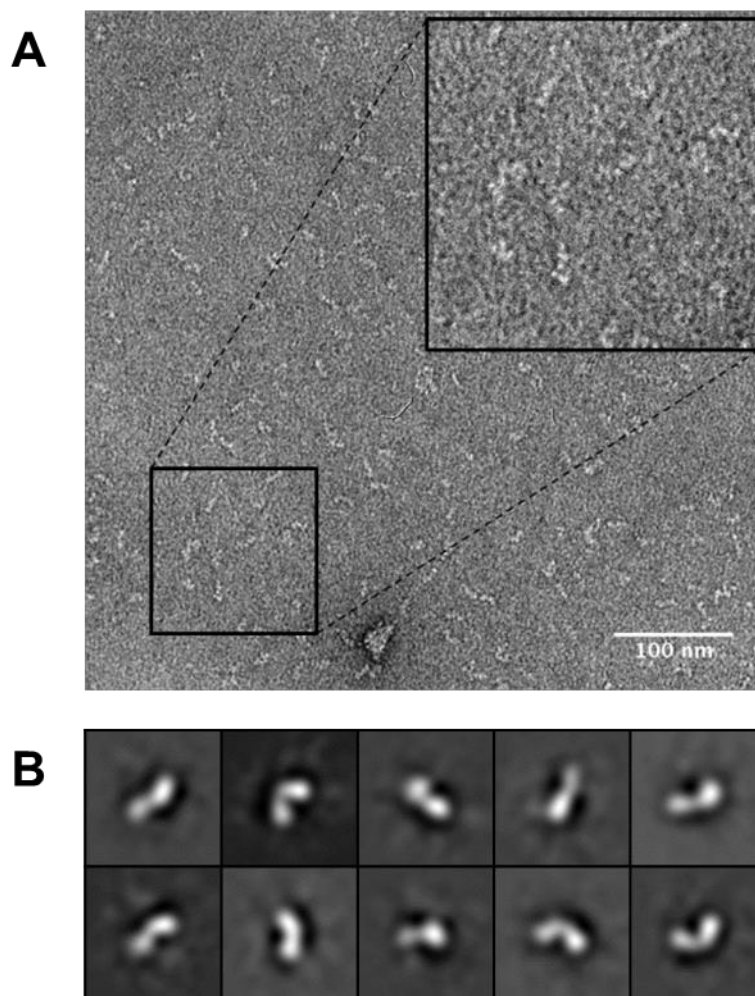
Figure 3.12 – Negative-stain Transmission Electron Microscopy of Purified BMPER

A) Representative image of negatively stained BMPER particles. Scale bars represent 100 nm. B) Representative reference-free class averages generated from the picked particle set using EMAN2.1. Class averages are elongated with a defined bend in the centre of the molecule. A box size of 128 pixels (258 Å) was used.

Utilising the information from the reference free class averages, it was possible to construct an initial 3D model of BMPER. The initial 3D model was projected in 2D, which was then compared to the particles that were used to make the 3D model. Particles with similar orientations are then aligned and averaged in 2D, which are then used to construct a refined 3D model. This model is then used as a projection for the next round of refinement. Eight iterations were completed before a final model was constructed (Figure 3.13). The final model was analogous to the BMPER bead model with regards to scale and overall shape. Overall shape comparisons are evaluated in section 3.3.9.
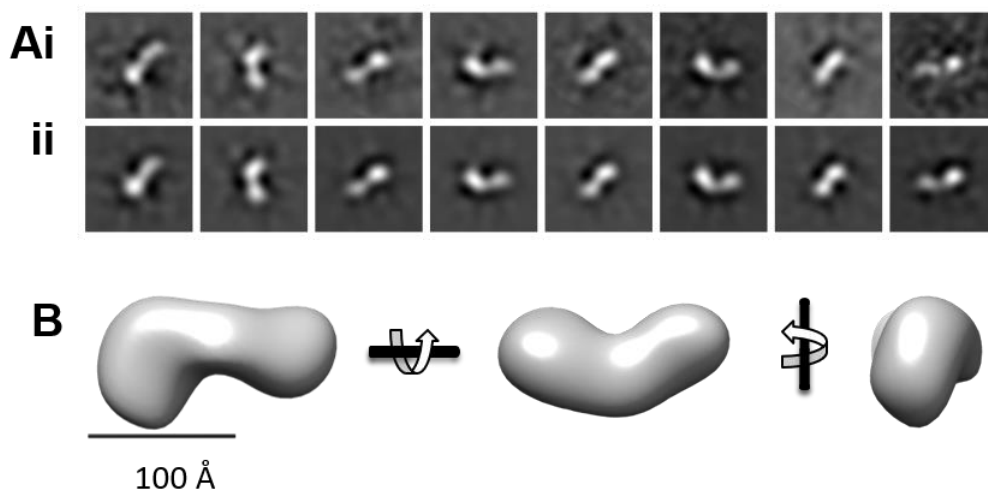
Figure 3.13 – 3D Reconstruction of BMPER from 2D Class Averages.

Ai) Model refined class averages of BMPER, with corresponding back projections (Aii) from the 3D reconstruction of BMPER. A box size of 128 pixels (258 Å) was used. B) 3D reconstruction of BMPER from iterative model refined class averages is shown with 90° rotations in the x- and y-axis. The model appears elongated with a bend in the centre of the structure. Scale bar represents 100 Å.

The final model was represented at a resolution of 35.8 Å, as determined by the Gold Standard resolution calculation (Figure 3.14-A). The particles that were used to reconstruct the 3D model of BMPER sampled the space well, providing a good distribution of angles, shown by the Euler angles (Figure 3.14-B).
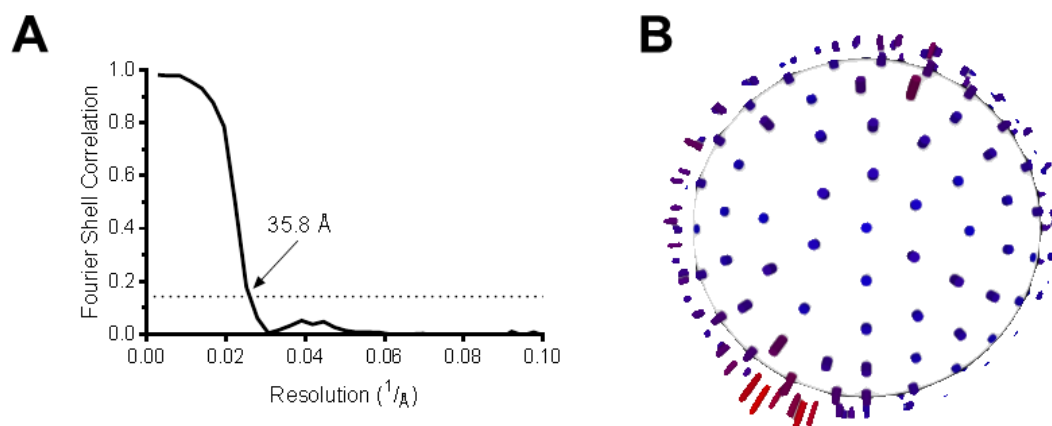
Figure 3.14 – Resolution Assessment of the BMPER Model.

A) The Gold-standard FSC of BMPER is calculated with a 0.143 FSC criteria, shown with a dashed line representing a final resolution of 35.8 Å. B) Euler angle distribution of all particles contributing to the final 3D reconstruction at a given angle. Red columns signify more particle contributions at designated angles relative to the blue columns.

3.3.9   SAXS and EM Model Comparisons

In order to generate a final model that represents BMPER as accurately as possible the final models from SAXS and from EM were compared with each other (Figure 3.15) (section 2.9.2.5). In Figure 3.15-A the model of BMPER from DAMMIF modelling was shown to be of a good overall fit to that of the 3D reconstruction from negative stain EM. Looking to a more complete model of BMPER the comparison of the negative stain EM model to the multi-phase *ab initio* model from MONSA was undertaken (Figure 3.15-B). The overall model fits within the volume of the mesh of the 3D reconstruction. Using multi-phase *ab initio* modelling it is possible to represent the N- and C-termini of BMPER within the mesh. Finally, in Figure 3.15-C, it is possible to use CORAL model 5, as was shown in Figure 3.10-C, to represent the N-terminus of BMPER at an all-atom level.
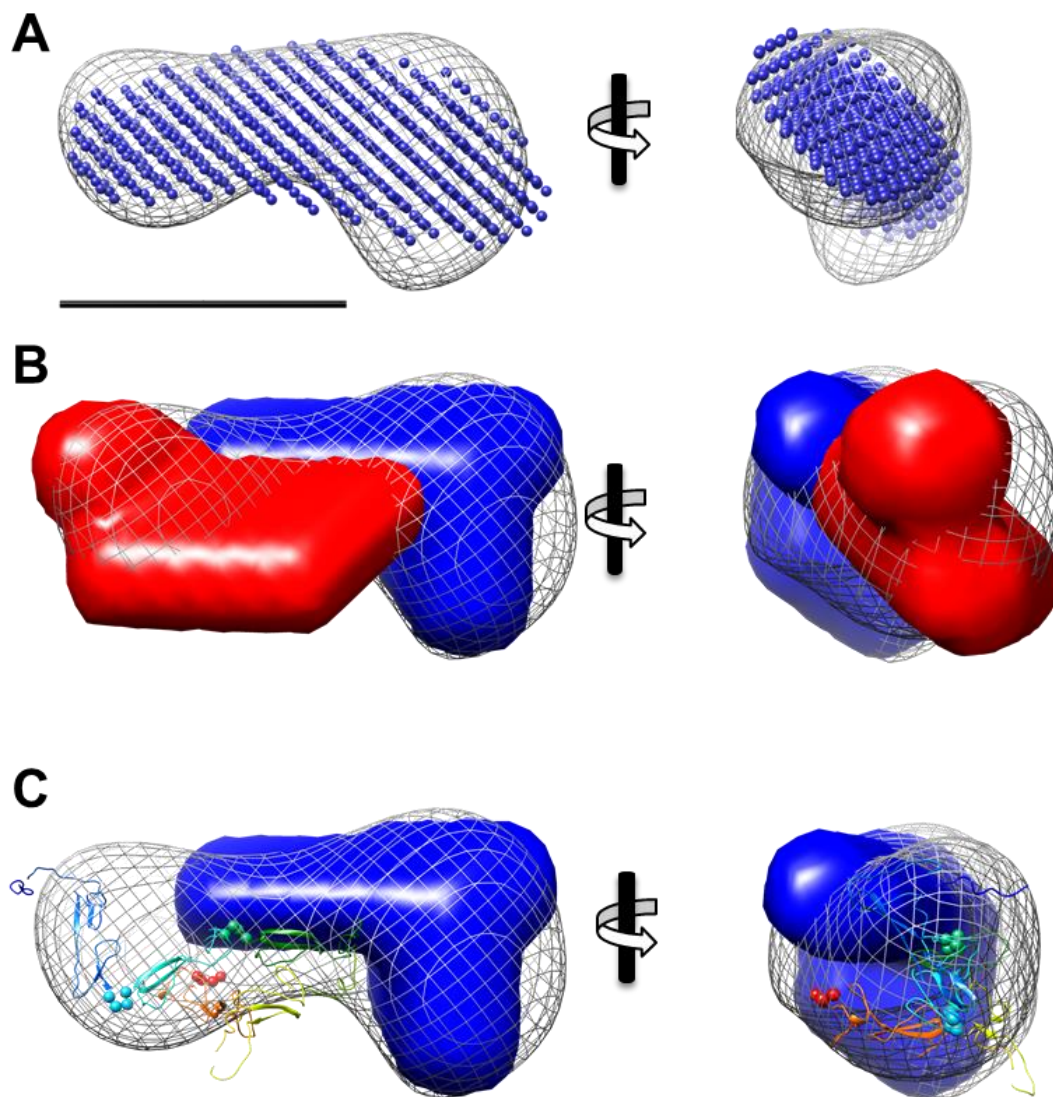
Figure 3.15 – Comparison of Negative Stain EM Reconstruction and SAXS Models.

A) Single phase *ab initio* BMPER bead models (blue beads) are docked into the volume of the 3D reconstruction of BMPER obtained by negative stain EM, rotated along the y-axis. B) Multi- phase *ab initio* BMPER bead models (N-terminal segment is red, C-terminal segment is blue) are docked into the volume of the 3D reconstruction of BMPER obtained by negative stain EM, rotated along the y-axis. C) CORAL model 5 and phase-2 (ΔN-BMPER) *ab initio* BMPER bead model is docked into the volume of the 3D reconstruction of BMPER obtained by negative stain EM, rotated along the y-axis.

3.4 **Comparison of Hydrodynamic and Structural Parameters**

Hydrodynamic parameters were calculated using the super matrix inversion (SMI) method, built into SOMO, and the Zeno method (Kang et al., 2004) which is piped through SOMO. These methods were used for BMPER and N-BMPER DAM models and CORAL model 5 (section 2.9.3.2). These results can be seen in **Error! Reference source not found.**.

Hydrodynamic parameters for EM models were calculated using HYDROMIC (Garcia de la Torre et al., 2001) as seen in **Error! Reference source not found.** (section 2.9.3.1). SAXS parameters can also be simulated from EM models using EM2DAM and by running CRYSOL on the resulting bead models (**Error! Reference source not found.**) (section 2.9.3.2). From here it was possible to directly compare the results of measured parameters for each method in the structural study.

3.5 Discussion

In summary, wild-type BMPER and N-BMPER were expressed and secreted from stably transfected HEK293-EBNA cells. BMPER showed the characteristic cleavage products under reducing conditions that have been shown previously (Ambrosio et al., 2008) at the 'GDPH' cleavage motif (Rentzsch et al., 2006). Each construct was then purified from the media using IMAC in the form of HisTrap nickel affinity columns. Further purification using a S200 or S200i SEC column yielded predominantly monodisperse protein (Figure 3.1-C). From the analysis of the sequence, and by reproducing the work of Kamimura *et al.*, (Kamimura et al., 2004b), it was shown that N-BMPER was glycosylated with approximately 12 kDa of glycosylation, consistent with four glycans.

Although BMPER and N-BMPER were being produced at levels for structural and hydrodynamic analysis, an opportunity to use the STRUBI mammalian expression vector pHLsec, with His$_6$ and 1D4 purification tags, arose with the vision to crystallise BMPER. 25 BMPER constructs were generated along with four chordin and two Tsg constructs (Table 3-1). The domain boundaries were designed to extend to the end of the predicted domains and to minimise the flexibility of the linkers using RONN (Yang et al., 2005). In most cases construct boundaries ended at either the predicted site, or the region of most order, predicted by RONN. Domain boundaries for chordin were determined by tolloid cleavage sites. For Tsg, the final residue remained the same for both of the constructs, but the first residue was varied. The first residue in the mature peptide, after the signal peptide, is a cysteine this was omitted from one construct and retained in the other as it was thought that this may either hinder or help folding of the mature peptide. Although the cloning was successful for all constructs, with the exception of construct 8 in the 1D4 vector, and the constructs

appeared to be expressed, few BMPER constructs were secreted and so these were not taken forward for further study (Figure 3.2-A & B). The chordin and Tsg constructs with a 1D4 tag were highly expressed, but their use is not within the scope of this thesis. It may be possible that due to the complexity of BMPER, with respect to the high number of domains, disulphide bonds and glycans, and the strong promoter (β-actin promoter) and signal peptide (μ-phosphatase), BMPER may not be folding correctly which in turn would increase endoplasmic reticulum stress and activate the unfolded protein response. A method to compensate for this may be to grow the cells under hypothermic conditions at temperatures of 30 – 35 °C. This has been shown to reduce the growth rate of cells whilst increasing the expression of protein by ~1.5 times which may increase the expression and secretion of BMPER fragments (Lin et al., 2015).

Although BMPER constructs had been expressed by several groups in the literature (Ambrosio et al., 2008; Fiebig et al., 2013; Kamimura et al., 2004b; Moser et al., 2003; Rentzsch et al., 2006; Serpe et al., 2008; Zakin et al., 2010; Zhang et al., 2007a; Zhang et al., 2010a; Zhang et al., 2008), only the first vWFC domain was structurally characterised. This left a gap in the literature for a structural model of full length of BMPER, and corresponding hydrodynamic data. Hydrodynamic characterisation of BMPER and N-BMPER was carried out using SEC-MALS (Figure 3.3-A) and AUC (Figure 3.3-B). MALS provided information on the molecular mass, revealing BMPER to be 100 kDa and N-BMPER to be 48 kDa showing both samples to be monomeric. The measurements of mass were not entirely accurate as the specific refractive index increment (dn/dc) was not calibrated for glycoproteins as glycans scatter light more than peptides. The mass measured by MALS was ~26 kDa larger than the mass prediction by sequence which is through a combination of N-linked glycans and the differences in dn/dc calibration. AUC provided information on elongation via the frictional ratio, and interestingly both BMPER and N-BMPER appeared elongated by the same proportion, although their mass is significantly different. Both SEC-MALS and AUC provided measurements of the R_H for BMPER and N-BMPER. SEC-MALS measured R_Hs of 51 Å and 40 Å for BMPER and N-BMPER while AUC measured values of 53 Å and 41 Å. These are very consistent across the two techniques and provide a consensus for the final hydrodynamic parameters using two independent approaches. A summary of all hydrodynamic parameters can be seen in **Error! Reference source not found.**.

Measured		BMPER	N-BMPER
MALS	R_h (Å)	51	40
	MW Measure/by sequence (kDa)	100.4/73.3	48.3/36.6
AUC	R_h (Å)	53	41
	f/f$_0$	1.73	1.74
	$S_{w,20}$	4.59	2.92
SAXS	R_g (Å) Reciprocal/Real	51/48	34/33
	I(0) Reciprocal/Real	$7.04 \times 10^{-3}/7.92 \times 10^{-3}$	21.23/18.55
	D_{max} (Å)	160	111
	Volume (Å^3)	320,000	140,000
	Model resolution (Å)	44 (± 3)	35 (± 3)
Calculated from:			
EM	R_h (Å)	54.79	N/A
	$S_{w,20}$	4.09	N/A
	R_g (Å)	49.39	N/A
	D_{max} (Å)	166.3	N/A
SAXS – DAM	R_h (Å)	48.8 (4.81)	36.5 (33.9)
	f/f$_0$	1.59 (1.62)	1.52 (1.54)
	$S_{w,20}$	4.89 (5.74)	3.14 (3.09)
SAXS - CORAL	R_h (Å)	N/A	35.4 (34.8)
	f/f$_0$	N/A	1.66 (1.64)
	$S_{w,20}$	N/A	2.47 (2.52)

Table 3-4 – Comparison of Measured Structural and Hydrodynamic Parameters

The table shows measured and calculated structural parameters. The values calculated from structural models are directly compared to the values of the measured data. Calculated values for the SAXS DAM and CORAL models are shows with super matrix inversion (SMI) or Zeno approaches through SoMo (Brookes et al., 2010a; Brookes et al., 2010b).

SEC-SAXS was employed to determine the overall shape of BMPER through DAMs, but also from the 1D scatter curve. This also provided information regarding the shape, folded-ness and flexibility of BMPER and N-BMPER. An IFT was then applied to the 1D scattering curves to transform the data to real space from reciprocal space resulting in a D_{max} of 160 Å and 111 Å for BMPER and N-BMPER respectively. The values obtained from the gradient of the GA (Figure 3.4-C & D), R_g of 51 Å and 34 Å for BMPER and N-BMPER respectively, were very close to the real space values of BMPER and N-BMPER (48 Å and 33 Å), shown in Table 3-4**Error! Reference source**

not found.. The $I(0)$ values were also shown to be close for BMPER at 7.04×10^{-3} and 7.92×10^{-3}, and for N-BMPER at 21.23 and 18.55, in reciprocal and real space respectively, also shown in **Error! Reference source not found.**. As the values from the GA and the PDDF are close, it shows that the processing is most likely correct.

The flexibility of BMPER and N-BMPER was determined through transformations of the domain and range axes of the 1D curves. The normalised Kratky plot allowed the semi-quantitative comparison of the BMPER constructs regardless of their size (Figure 3.5-A). Similar to the frictional ratio, as described in section 3.2.1, the normalised Kratky plot showed that the BMPER constructs were roughly proportionally elongated by the shift of their peaks away from the cross-hair. The PDDF (Figure 3.4-B) also showed elongation for each construct due to the tail on the right of the peak. Hydrodynamic measurements, reported by the frictional ratio (section 3.2.1) also showed that BMPER and N-BMPER were elongated. As the normalised Kratky plot did not show any form of intrinsic disorder or high flexibility, the flexibility of BMPER and N-BMPER were assessed by the SIBYLS and the Porod-Debye plots. As both BMPER constructs showed a plateau in the SIBYLS plot, the protein could not be a single rigid system. They were therefore defined as a set of rigid domains connected with flexible linkers, accounting for a degree of flexibility. The SIBYLS plot has been used to describe partially flexible multi-domain proteins (Troilo et al., 2016c).

The single phase DAMs generated from the BMPER (Figure 3.6-C) and N-BMPER (Figure 3.7-C) scattering confirmed the molecules are elongated. 23 DAMs were generated using DAMMIF. The fitting of the representative BMPER DAM (Figure 3.6-A) is close to an ideal (χ^2 of 1.126 versus 1), but for N-BMPER the χ^2 was lower (0.213). Although this would normally be considered over fitting, the residual plot inset into the scatter plot (Figure 3.7-A) shows a close match to the data. In explaining this, the problem may be due to the errors reported on the data from the sample detector due to a low concentration of sample. The 23 DAMs were analysed with the DAMAVER suite with the NSD values for BMPER as 0.672 and for N-BMPER 0.674, with a single model excluded from each pool of models. With values less than '0.7' it is clear that these are DAMs unique to the corresponding scattering profiles (Volkov and Svergun, 2003). Resolution of the models could also be calculated using

SASRES and was shown to be 44 Å (± 3 Å) and 35 Å (± 3 Å). These values were used when generating surface maps or fitting models within a volume using UCSF Chimera 'molmap' and 'fit-in map' commands. Hydrodynamic parameters of the DAMs were calculated using SOMO by both SMI and ZENO methods of calculation. Here the values are close enough as to say the models generated from SAXS agree with two other independent methods for measuring hydrodynamic parameters, SEC-MALS and AUC (**Error! Reference source not found.**). It can be understood that slight variations in the final values come from errors with the measurements and the imperfect nature of model calculations matching real data.

From the work of Zhang *et al.,* (Zhang et al., 2008) and Fiebig *et al., (Fiebig et al., 2013)* in solving the structure of the first vWFC of BMPER, homology models were generated in order to generate a rigid body model for N-BMPER using CORAL. As shown above, in Figure 3.4-A and Figure 3.7-A, when N-BMPER DAM reconstructions are compared to the data the χ^2 is low, around 0.2. This would usually be considered over fitting, but, as mentioned, the residual plots show no divergence from the data. Of the CORAL models generated, only models with a D_{max} close to that of N-BMPER were kept for further analysis (Figure 3.8-A). Interestingly, as the D_{max} of the models increased further from that measured for N-BMPER, the χ^2 of the CORAL output increased also (Table 3-3). As the fit of the CORAL models to the scattering curves (Figure 3.8-B) and the correlation of the CORAL models to the DAM volume could not definitively select a single model to represent N-BMPER, CORAL model 5 was selected as the best overall fit, including hydrodynamic values calculated from SOMO (Table 3-4**Error! Reference source not found.**).

Yet, further information can be obtained from the SAXS data. As the single-phase DAM of BMPER cannot inform us of the orientation, MONSA was used to construct 20 multi-phase *ab initio* models to elucidate the terminal fragments (Figure 3.10). The fitting of the models for both phase-1, N-BMPER, and the combined phase-1 and -2, N-BMPER and ΔN-BMPER, were good (χ^2 values at 0.23 and 0.96). As mentioned there are no specific methods of generating a final model of MONSA models, but the scripts by Rambo (Rambo, 2015) allowed the splitting of the phases allowing averaged DAMs of phase-1 to be generated. This averaged model was similar to the DAM of N-BMPER generated from DAMMIF, which is understandable as phase-1

was generated from N-BMPER scattering data (Figure 3.11). A consistent location for the N- and C-terminus was found throughout the models (Figure 3.10-B). CORAL model 5 fitted well within the phase-1 volume as a representative quasi-atomic N- and C-terminus model (Figure 3.10-C).

The hydrodynamic information coupled with the SAXS data has provided corroborating information regarding the elongation of BMPER and N-BMPER. Hydrodynamic data supplied information regarding the elongation of proteins through the frictional ratio while SAXS provided this though the dimensionless Kratky plot and the PDDF. DAMs generated through *ab initio* modelling also provided a visual representation of the elongation of the molecule. Negative stain electron microscopy was employed as a final corroborative technique to independently determine the structure of BMPER. This was not possible for N-BMPER as it is too small. From the micrographs, represented by Figure 3.12-A, approximately 4,000 particles were picked using the EMAN2.1 software suite. Through reference free classification 1,000 bad or damaged particles were removed from the data set by eliminating bad classes. Examples of classes can be seen in Figure 3.12-B. An initial 3D model was generated from the reference free classes and allowed for a refinement to achieve a final 3D reconstruction (Figure 3.13-A). The final 3D reconstruction showed a slight bending in the structure of BMPER. The model refined class averages (Figure 3.13-Bi) match the projections from the 3D reconstruction (Figure 3.13-Bii) and so gives rise to a final model of BMPER by negative stain EM with a resolution of 35.8 Å (Figure 3.14-A). The contribution of particles to the final model can be assessed visually by way if the Euler angles (Figure 3.14); Euler angles show a good distribution of particles contributing to the final model.

As mentioned above, this information regarding the measurements of hydrodynamic and structural parameters, as well as the generation of models goes a long way to filling gaps in the literature regarding the structure of BMPER.

4 Results Chapter 2: Investigating BMPERs Inhibition of BMPs and the Effect of the Diaphanospondylodysostosis Causing Mutation

BMPER is a large, multi-domain glycoprotein from the chordin family. Chordin contains four vWFC domains that have been shown to facilitate binding to different ECM ligands and regulators, such as Tsg (Troilo, 2013; Troilo et al., 2016a; Troilo et al., 2014). As the vWFC domains of human BMPER share between 23% and 38% sequence identity with vWFC domains of human chordin it could be expected that BMPER may also bind Tsg.

4.1 BMPER Binds to ECM Components

A variety of binding assays were used in order to determine the binding of BMPER to various ECM proteins and components. Inhibition assays were also employed to determine how these interactions effect BMP signalling.

4.1.1 BMPER Binds to Heparan Sulphate

As it was shown that BMPER from zebrafish can bind to cell surfaces through HSPGs (Rentzsch et al., 2006), validation was required for the human homologue of BMPER. Using a different approach from the heparin beads, utilised by Rentzsch *et al.* (Rentzsch et al., 2006), cells expressing either BMPER or N-BMPER were dosed with unfractionated heparin, a more sulphated form of HS, to compete the protein from cell surface interactions (section 2.7.1). Figure 4.1 shows both BMPER and N-BMPER being dosed with unfractionated heparin, and compared with conditioned media from non-dosed cells and cells washed with PBS. In lane 1 of Figure 4.1-A, full length BMPER is not detected in the conditioned media due to low protein concentrations. Lane 2 shows that when BMPER expressing cells were washed with PBS no additional protein was observed. However, when BMPER cells were incubated with 1 µg/mL and 10 µg/mL of heparin (lanes 4 and 5), there was a clear increase in the intensity of the bands. When looking at the corresponding N-BMPER lanes there were no overall increase in the band intensity Figure 4.1-B. This shows that the C-terminus of BMPER is the region binding to HS.
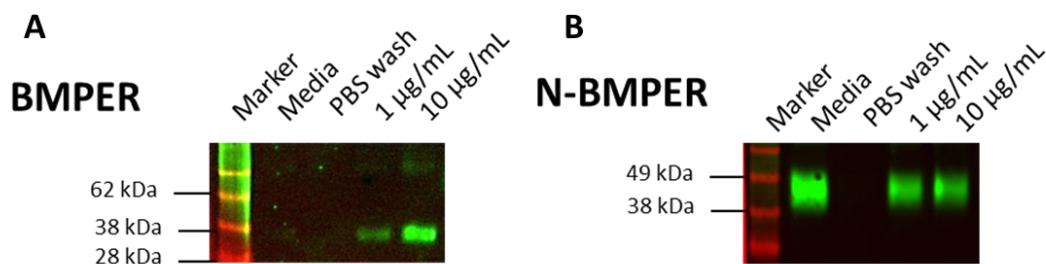
Figure 4.1 – Heparin Competition Assay.

Conditioned media from cells incubated in the absence of heparin, cells washed with PBS, cells incubated with 1 µg/mL and 10 µg/mL of unfractionated heparin were analysed by Western blotting using an anti-His$_6$ antibody. A) Western blot analysis of the media and wash collected from the BMPER cell line under different culturing conditions, including 1 µg/mL and 10 µg/mL of heparin. B) Western blot analysis of the media and wash collected from the N-BMPER cell line under different culturing conditions, including 1 µg/mL and 10 µg/mL of heparin..

4.1.2 BMPER Binds to Tsg

The ability of BMPER to bind to Tsg has been reported within the literature with differing results. It has been shown that BMPER is able to bind to Tsg through crosslinking studies using the mouse homologues of BMPER and Tsg (Ambrosio et al., 2008) but not through SPR using the zebrafish homologue of BMPER and the mouse homologue of Tsg (Zhang et al., 2007a). To provide a conclusive result we re-examined the interactions of human BMPER and Tsg by SPR studies (section 2.8.3). Recombinant BMPER and N-BMPER and recombinant Tsg, as purified previously (Troilo et al., 2016a), was used to determine binding affinities. Tsg was immobilised to the GLC sensor chip (Bio-Rad) through amine coupling, as had previously been reported (Troilo et al., 2016a). Recombinant BMPER and N-BMPER were injected over the chip in concentrations of 0 – 150 nM and 0 – 50 nM, respectively. The resulting sensorgrams showed that both BMPER and N-BMPER were able to bind to Tsg (Figure 4.2). It was not possible to fit the sensorgram data from BMPER interactions to a 1:1 Langmuir binding model (Figure 4.2-Ai) so an equilibrium analysis was performed using a non-linear regression using GraphPad Prism version 7.00 for Windows (GraphPad Software); this provided a binding constant (K_D) of 68.37 nM (Figure 4.2-Aii). With N-BMPER it was possible to fit the data to a 1:1 binding model to obtain a K_D of 0.55 nM (Figure 4.2-C). These data showed that BMPER is able to bind to Tsg with a high binding affinity, and that the binding is localised to the vWFC regions in N-BMPER. A complete table of results can be seen in Figure 4.2-C.
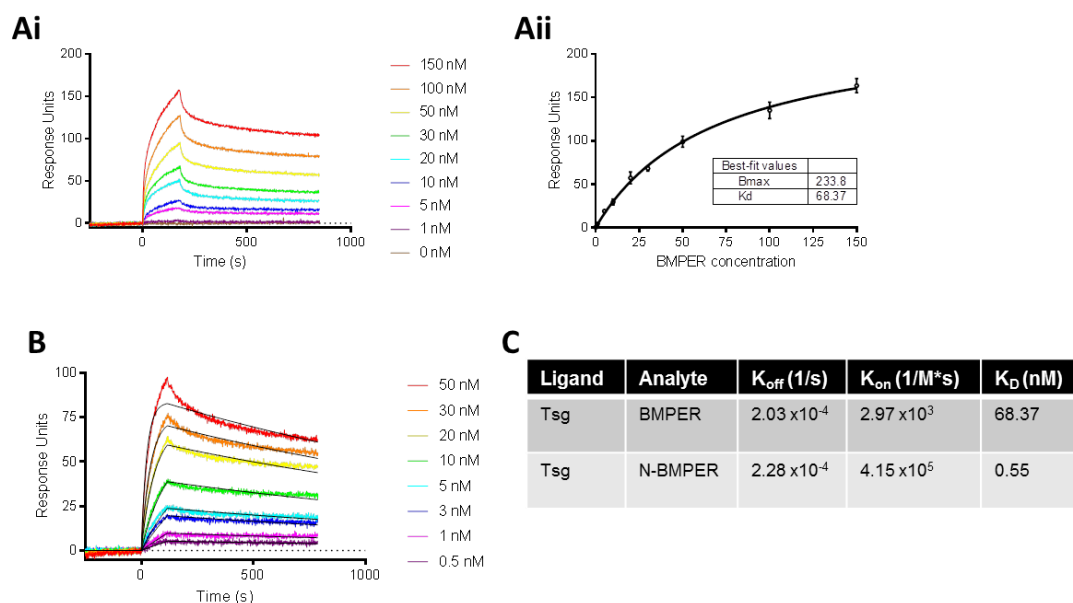
Figure 4.2 – BMPER and N-BMPER Binds to Tsg by SPR.

Ai) SPR sensorgram of the analyte, BMPER, at varying concentrations (0 – 150 nM) binding to the ligand, Tsg, with the binding analysis conducted using equilibrium analysis to give a K_D of 68.37 nM (Aii). B) Sensorgram of the analyte, N-BMPER, at varying concentrations (0 – 50nM) binding to the ligand, Tsg, with a Langmuir 1:1 binding model showing a K_D of 0.55 nM. C) A table showing the binding parameters of BMPER constructs and Tsg. All experiments were carried out in triplicate.

As the binding affinities of BMPER and N-BMPER differ by two orders of magnitude, further analysis was conducted in order to determine why this may be. Looking at the dissociation constants, k_d, of BMPER and N-BMPER, values were very close at $2.03 \times 10^{-4} s^{-1}$ and $2.28 \times 10^{-4} s^{-1}$, respectively (Figure 4.3). This indicates that it is the association constant, k_a, that differs for the analytes, because when the constructs are bound to the ligand they dissociate at the same rate. The data provided here then leads us to believe that both of the BMPER constructs interact in a similar fashion with Tsg. The k_a value may be higher for BMPER due to a conformational change which occurs in the N-terminus of BMPER when the C-terminus is present.

111
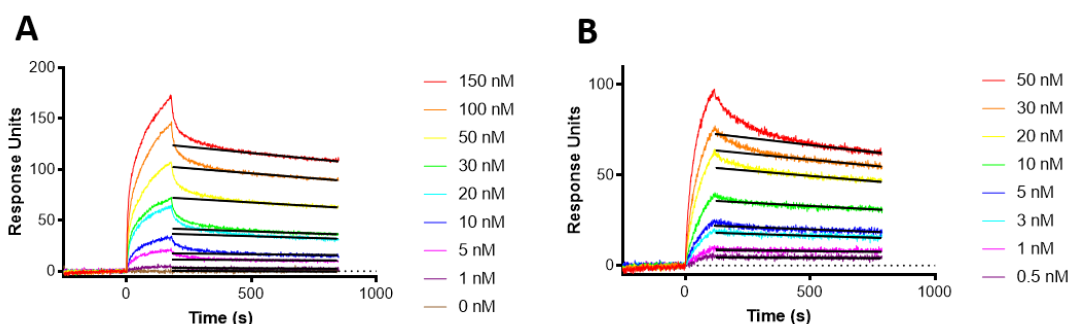
Figure 4.3 –Dissociation Constants of BMPER and N-BMPER from Tsg.

SPR sensorgrams showing the dissociation constant of BMPER (A) from Tsg, represented by a black line and the dissociation constant of N-BMPER (B) from Tsg, represented by a black line.

4.1.3 BMPER Binds to Chordin

Although it has been reported elsewhere that BMPER is able to bind to chordin, with K_Ds of 1.4 nM (Ambrosio et al., 2008), 25 nM (Rentzsch et al., 2006) and 175 nM (Zhang et al., 2010a), the study was repeated using recombinant human N-BMPER and the chordin fragment composed of the vWFC domains 2-3, that had previously been expressed in the lab (Troilo et al., 2016a) (section 2.8.3). Chordin fragment vWFC 2-3 was chosen as it has been shown that BMPER only binds to vWFC2 of chordin (Zhang et al., 2010b). The chordin fragment was immobilised to a GLC sensor chip (Bio-Rad). N-BMPER was used as the analyte and injected over the ligand at concentrations of 0 – 100 nM (Figure 4.4). From the sensorgrams it is possible to see an interaction between the ligand and the analyte and it was possible to fit the data to a 1:1 Langmuir binding model with a K_D of 0.20 nM. This allowed the corroboration of the binding site of the chordin fragment to N-BMPER.

A

B

Ligand	Analyte	K_{off} (1/s)	K_{on} (1/M*s)	K_D (nM)
Chordin vWFC2-3	N-BMPER	8.22×10^{-5}	4.04×10^{5}	0.20
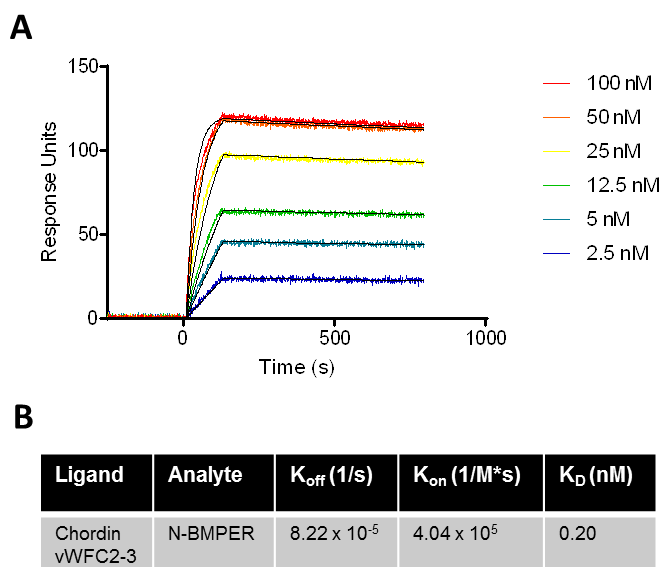
Figure 4.4 – vWFC2-3 of Chordin Binding to N-BMPER

A) SPR sensorgram of the analyte, N-BMPER, at varying concentrations (0 – 100nM) binding to the ligand, fitted to a Langmuir 1:1 binding model showing a K_D of 0.20 nM. B) A table showing the binding parameters of BMPER constructs and chordin vWFC2-3.

4.2 BMPER Works in Concert with Tsg to Inhibit BMP Signalling

As BMPER and Tsg have been shown to interact with one another, we wished to pursue the effect that the combination of proteins would have on BMP signalling. Previous studies looking at the effect of ECM regulators on BMP signalling have been monitored using alkaline phosphatase (ALP) assays (Troilo et al., 2016a; Troilo et al., 2014). ALP assays were carried out using both BMPER constructs to assess the impact of the constructs on BMP-4 signalling in both the presence and absence of Tsg (section 2.7.2). As Tsg also inhibits BMP-4 signalling to 20% at 28× molar excess (Troilo et al., 2016a), this concentration was used for the ALP assays.

The ALP assays were carried out and analysed in collaboration with the Sengle lab (University of Cologne, Germany) with recombinant proteins purified in the Baldock lab. BMPER and Tsg constructs were purified as described above (section 4.1.2) and used in ALP assays to assess their efficacy in BMP-4 inhibition. Figure 4.5-A shows that BMPER inhibits BMP-4 by approximately 5%, as shown by the red line. When Tsg was added to BMPER, it was possible to see a downward trend in the activity of BMP-4, shown by the blue line. The inhibition was such that it was possible to determine an IC_{50} of 2.6× molar excess of BMPER to BMP-4. Comparing the percentage inhibition of BMP-4 at a BMPER excess of 2.6 fold in the absence and presence of Tsg, the activity of BMP-4 decreases from 5% to 50%. Although Tsg has been added, it can only provide 20% inhibition at this concentration showing BMPER and Tsg work in concert to inhibit BMP-4 signalling (Troilo et al., 2016a).
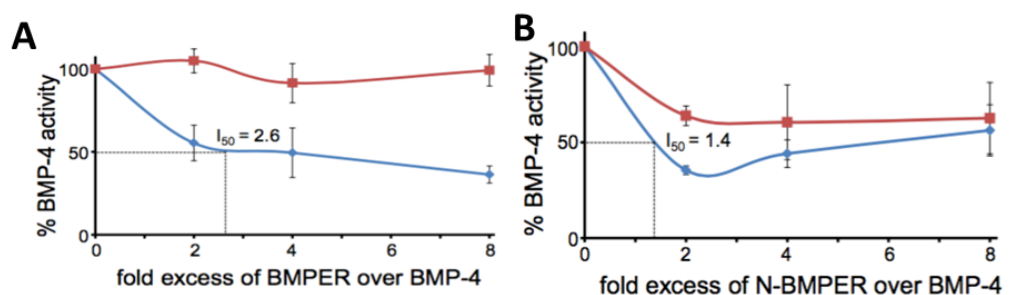
Figure 4.5 – Alkaline Phosphatase Assay Showing Inhibition of BMP-4 by BMPER and N-BMPER.

For all inhibition assays BMP-4 was used at 30 ng/mL (1.15 nM) and the concentrations of BMPER constructs used at two, four and eight times the concentration of BMP-4. Tsg was used at 28× molar excess of BMP-4 (32.3 µM), as per previous studies (Troilo et al., 2016a). A) BMPER shows little effect (5% inhibition at 2.6 × molar excess) in inhibiting

BMP signalling (red). When Tsg was added (blue) the inhibition was increased and an IC_{50} of 2.6 × molar excess of BMPER was observed. B) N-BMPER shows a greater effect (20% inhibition at 1.4 × molar excess) in inhibiting BMP signalling (red). When Tsg was added (blue) the inhibition was increased and an IC_{50} of 1.4 × molar excess of N-BMPER was observed. Results can be seen in Table 4-1.

Figure 4.5-B provides us with the same experiment using N-BMPER to inhibit BMP-4 signalling activity. When N-BMPER was used in the absence of Tsg, it showed a distinct decrease in the percentage of BMP-4 activity (red line). When Tsg was combined with the N-BMPER titrations (blue line), it was possible to see a further decrease, providing an IC_{50} of 1.4× molar excess of BMPER to BMP-4. Comparing the inhibition of N-BMPER in the absence of Tsg, at 1.4 fold molar excess, N-BMPER provides 20% inhibition of BMP-4 activity, and in the presence of Tsg provides 50% inhibition of BMP-4 signalling. This shows further synergistic effects of BMPER and Tsg. A full comparison of results at their percentage inhibition can be seen in Table 4-1.

Samples	% inhibition
Tsg	20
BMPER	5
BMPER + Tsg	50
N-BMPER	20
N-BMPER + Tsg	50

Table 4-1 – Table Showing the Percentage Inhibition of BMP-4 Signalling.

Inhibition of BMP-4 signalling by BMPER and N-BMPER concentrations in line with their corresponding percentage inhibition, with and without Tsg at 28× molar excess to BMP-4.

4.3 Investigating the Diaphanospondylodysostosis Causing Mutation

Clinical studies into the cause of the rare, recessively inherited, perinatally lethal skeletal disorder DSD identified 13 mutations in BMPER (Funari et al., 2010). Differing severities of DSD have also been described by Kuchinskaya *et al.* (Kuchinskaya et al., 2016), as well as a less severe ISD. Although some of the mutations in BMPER were shown to influence the severity of DSD, studies into the mutation of proline 370 to leucine (BMPER-P370L) were pursued, as this was in the 'GDPH' acid catalysed cleavage motif of BMPER. As no biochemical studies have

been conducted on DSD mutants and so secretion studies of BMPER and BMP inhibition assays were performed.

4.3.1 Generating Clones of BMPER and BMPER-P370L

BMPER-P370L was generated by site directed mutagenesis of BMPER by PCR and cloned into modified pCDH vectors (sections 2.4.6 and 11.3), along with BMPER, with the native BMPER signal sequence.

4.3.2 Generating HEK293-T Cell Lines with Lentiviral Constructs

Lentivirus containing BMPER and BMPER-P370L were generated in HEK293-T cells by transfecting BMPER containing pCDH vectors with packaging vectors into them (section 2.5.2.3). HEK293-T cells were transduced with BMPER containing lentivirus transplanted from transfected HEK293-T cells and sorted for tGFP with an excitation wavelength of 482 nm and an emission wavelength of 502 nm. These were compared to a HEK293T control, and cells containing tGFP were selected for experimental use. Figure 4.6-A shows the control of HEK293-T cells, where Figure 4.6-B and –C shows that there are emissions in the tGFP wavelength and these are selected in the range designated by 'P3'. Figure 4.6-D, shown the percentage of tGFP expressing cells that were selected from each cell sort for BMPER and BMPER-P370L. Sorted cells were cultured in a similar way to standard HEK293-T cells.
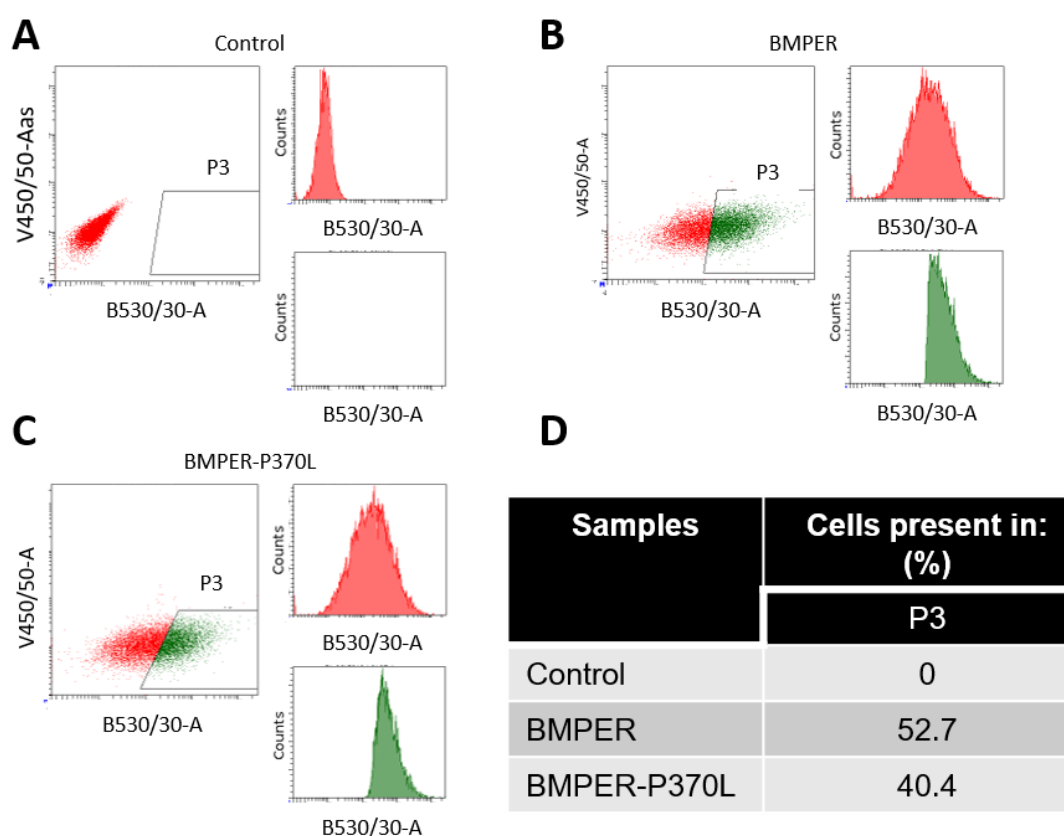
Figure 4.6 – Sorting for Cells Expressing tGFP

BMPER and BMPER-P370L transduced HEK293-T cells were sorted. P3 shows the cells that express tGFP, A) Control HEK293T cells with no tGFP expression, B) Cells selected for tGFP and also expressing BMPER. C) Cells selected for tGFP and also expressing BMPER-P370L. D) Table showing the populations of cells presenting tGFP.

In addition to expression and secretion studies, stable cell lines were developed to test the inhibition of BMP signalling of BMPER and BMPER-P370L. An additional, adapted pCDH vector was used to generate cells lines that have inducible BMP-2 expression, provided by Dr Stuart Cain (*University of Manchester, UK*). BMP-2 expression was induced by doxycycline and was co-expressed with a blue fluorescent protein (BFP). Viruses were generated for the BMP-2 and the BMPER constructs; cell-lines transduced with BMP-2, BMP-2 and BMPER and BMP-2 and BMPER-P370L were produced. Cells were sorted (Figure 4.7) to select for only those expressing both tGFP and BFP. The control (Figure 4.7-A) shows cells without the expression of tGFP or BFP. BFP can be detected by excitation at a wavelength of 402 nm and emission at a wavelength of 457 nm. In Figure 4.7, P3 shows the cells that express only tGFP; P4 shows cells that express only BFP and P5 shows cells that express both tGFP and BFP. Figure 4.7-B shows that 69.3% of the cells sorted expressed BFP, corresponding to BMP-2. Figure 4.7-C and –D showed that 9.7%

and 7.0% of cells sorted contained both BMPER & BMP-2 and BMPER-P370L & BMP-2 respectively by selecting based on tGFP and BFP.
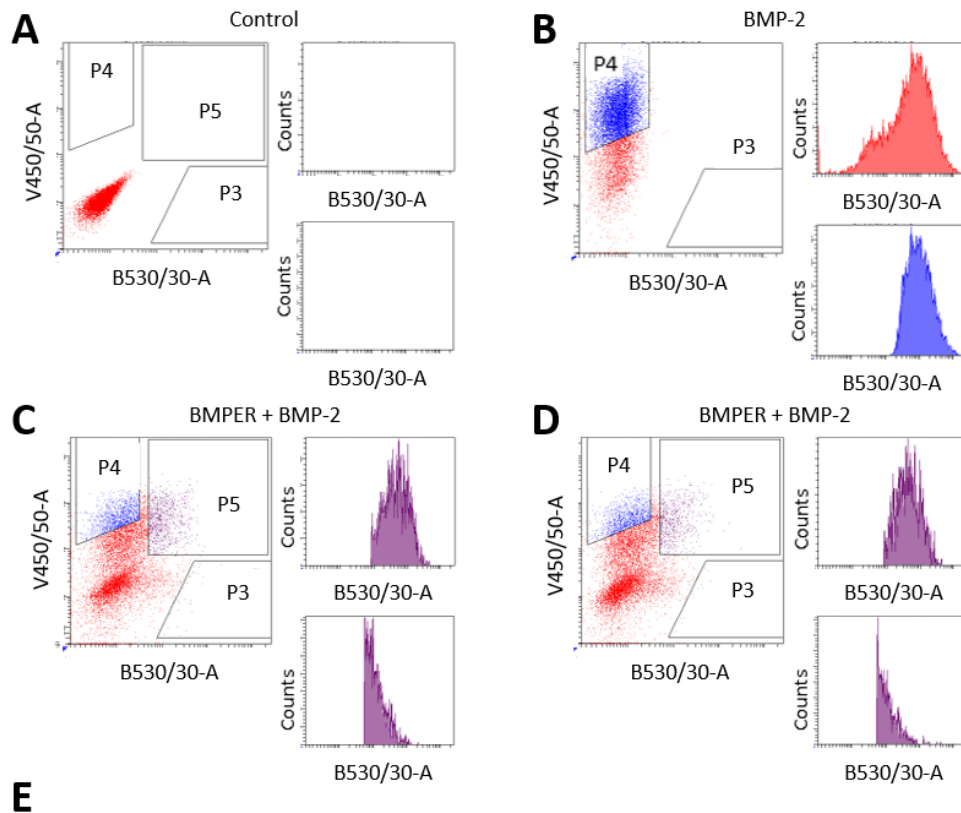
Figure 4.7 – Sorting for Cells Expressing tGFP or BFP

BMP-2, BMP-2 & BMPER and BMP-2 & BMPER-P370L transduced HEK293-T cells were sorted. P3 shows the cells that express only tGFP, P4 shows cells that express only BFP and P5 shows cells that express both tGFP and BFP. A) Control HEK293T cells, B) Cells selected for BMP-2 containing cells. C) Cells selected for BMP-2 and BMPER containing cells. D) Cells selected for BMP-2 and BMPER-P370L containing cells.

4.3.3   BMPER-P370L is Secreted and Cleaved from HEK293-T Cells

Media from BMPER and BMPER-P370L cell lines were analysed by Western blot, under reducing (Figure 4.8-A) and non-reducing (Figure 4.8-B) conditions using a V5 antibody. BMPER and BMPER-P370L were both shown to be secreted. BMPER appears to be expressed and secreted in a characteristic, two band fashion, as shown in (Figure 3.1-D) under the reducing and non-reducing conditions. Interestingly, BMPER-P370L is secreted as a high molecular weight species at the same mass as

118

BMPER. More surprisingly, BMPER-P370L has a cleavage product that can be seen at a smaller molecular mass than the BMPER cleavage product.

To further investigate the origin of the cleavage product shown by the BMPER-P370L cell-lines, the cell lysates were Western blotted under reducing conditions using the V5 antibody (Figure 4.8-C). Figure 4.8-C clearly shows that BMPER is cleaved within the cell where BMPER-P370L shows no signs of cleavage. This points to the cleavage of the mutant occurring in the ECM, leading to the conclusion that cleavage is most likely completed by an ECM protease rather than by an acid catalysed mechanism for BMPER.
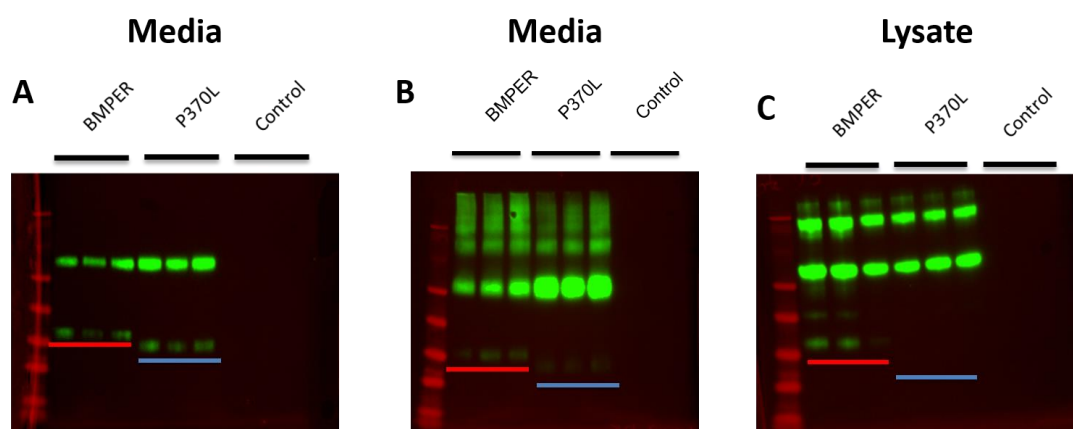
Figure 4.8 – BMPER-P370L Secretion Tests

The conditioned media and cell lysates of BMPER and BMPER-P370L cell-lines and HEK293T control cells were probed by Western blot using an anti-V5 antibody. Red lines show BMPER protein, blue lines show BMPER-P370L protein and the three lanes represent technical repeats. A) Western blot of reduced samples of BMPER, BMPER-P370L and HEK293T cell conditioned media showing the cleavage products of BMPER and BMPER-P370L. B) Western blot of a non-reduced sample of BMPER, BMPER-P370L and HEK293T cell media showing the cleavage products of BMPER and BMPER-P370L. C) Western blot of reduced samples of BMPER, BMPER-P370L and HEK293T cell lysates showing the cleavage products of BMPER.

4.3.4   BMPER-P370L Inhibits BMP Signalling More than BMPER

With the knowledge that BMPER-P370L is secreted, we wished to determine how it affected BMP-2 mediated signalling by monitoring phosphorylation of SMAD proteins (pSMAD). Using cell lines produced in section 4.3.2, six well plates were seeded with BMP-2, BMPER & BMP-2 or BMPER-P370L & BMP-2 cells. While BMPER and BMPER-P370L were constitutively expressed, BMP-2 was only expressed in the presence of doxycycline. Cell media was exchanged to expression media containing

200 ng/mL of doxycycline to induce BMP-2 expression. Cells were incubated for 48 hours before the media was removed, and the cells were lysed. Cell lysates were Western blotted with the antibody for pSMAD-1/5/8, and compared to a control of SMAD-1 (section 2.7.3). Figure 4.9 shows the results of the densitometry analysis, where cells expressing only BMP-2 was normalised to 1. In Figure 4.9, cell-lines expressing BMPER and BMP-2 showed a decrease in pSMAD 1/5/8 by 30%, while cell lines expressing BMPER-P370L and BMP-2 showed a 45% decrease in pSMAD 1/5/8. The data here presents an interesting result, where the inhibition of BMP-2 by BMPER-P370L is increased, and could contribute to the pathological phenotype.
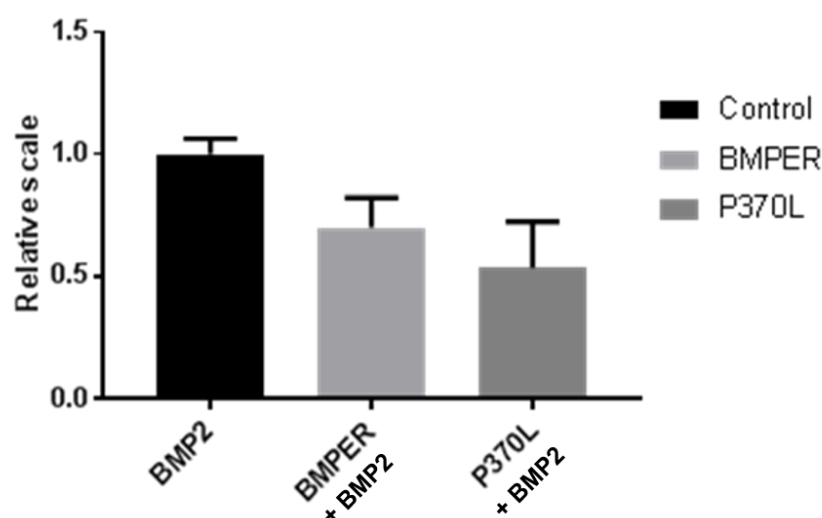
Figure 4.9 – BMPER-P370L Inhibits Signalling Greater than BMPER

Analysis of pSMAD assays, where BMP-2 is normalised to 1, calculated from the intensity of the bands whilst completing a Western blot analysis. BMPER showed an inhibition of BMP signalling by 30% where BMPER-P370L inhibits BMP signalling by 45% (n = 3). Error bars are shown from the standard deviation of the mean. Using an ordinary one-way ANOVA with multiple comparisons there were no statistically significant changes in signalling, although a trend can be seen.

4.4 Discussion

With the results described above it is possible to see how the interactions of BMPER and N-BMPER fit into the current model of BMPER signalling through biophysical and cell based assays. Additionally it has been possible to look at the effects of the BMPER-P370L mutation on the ability of BMPER to inhibit BMP signalling.

4.4.1 BMPER Binds to ECM Components

To build upon earlier work shown that BMPER binds to HSPGs (Rentzsch et al., 2006) where BMPER was bound to heparin beads, the binding was examined using recombinant human BMPER as opposed to zebrafish protein. The result was expected to be similar as the heparan sulphate binding region has been localised to residues '393-396' in zebrafish BMPER which are conserved in humans (Rentzsch et al., 2006). The heparin binding experiment also showed that N-BMPER was unable to bind to HS beads, localising the binding to the C-terminus. The experiment carried out using human BMPER constructs, shown in Figure 4.1, shows that human BMPER does indeed bind to heparin. When cells were dosed with increasing concentrations of unfractionated heparin, a dose response was observed for BMPER (Figure 4.1-A), but not for N-BMPER (Figure 4.1-B). This assay again confirms that human BMPER binds to HSPGs through its C-terminus.

With the conserved formation of a ternary complex comprising chordin, Tsg and BMPs in the BMP signalling pathway (De Robertis and Moriyama, 2016; Wharton and Serpe, 2013; Zakin and De Robertis, 2010), it has also been suggested that BMPER may play a role in the formation of a larger complex during signalling (Zakin and De Robertis, 2010). Although extensive binding studies of various homologues of BMPER and chordin have been completed, forming a consensus (Ambrosio et al., 2008; Rentzsch et al., 2006; Zhang et al., 2010a; Zhang et al., 2007b), and BMPER and BMPs (Ambrosio et al., 2008; Binnerts et al., 2004; Moser et al., 2003; Qiu et al., 2008a; Zhang et al., 2007a; Zhang et al., 2010a; Zhang et al., 2008), the literature is conflicted regarding the binding of BMPER to Tsg (Ambrosio et al., 2008; Zhang et al., 2007a). With this conflict in literature, binding studies between BMPER and Tsg were carried out using SPR analysis, which has proven useful for studying chordin and Tsg as well as chordin and BMPER in the past (Troilo et al., 2016a; Troilo et al., 2014; Zhang et al., 2010a). In section 4.1.2, it is shown that when Tsg is immobilised to the SPR sensor chip, BMPER is able to bind to Tsg with a K_D of 68.37 nM, contradicting a previous SPR studies (Zhang et al., 2007a). As chordin interacts with Tsg through its vWFC domains, the binding of Tsg and N-BMPER was investigated. Figure 4.2-B shows that N-BMPER binds to Tsg with a high affinity of K_D 0.55 nM.

This contrast in results to Zhang *et al.* may be due to a cross-species protein of the different expression system (recombinant mouse homologue of Tsg and a recombinant zebrafish homologue BMPER expressed in SF9 insect cells). Possible problems in previous experiments are twofold. Firstly, the use of SF9 cells to produce recombinant mammalian protein, rather than mammalian cells, could cause a problem as each of the proteins used in the experiment are glycosylated. It has been shown that insect cells produce smaller N-glycans than mammalian expression systems (Chang et al., 2007). The interaction between BMP and Tsg has been shown to be affected by changes in glycosylation (Billington et al., 2011). Tsg from mouse is predicted to have two glycosylation sites (Gupta and Brunak, 2002), where the human Tsg protein has three, and with the SF9 cells providing a reduced glycan, this may alter the binding further. Secondly, although human BMPER and zebrafish BMPER share 69% sequence identity, the differences may be enough to oblate the binding of the BMPER and Tsg. Figure 4.10 shows the sequence alignments of BMPER homologues from four commonly studies species, with Table 4-2 showing the percentage sequence identities.

```
Drosophila   MCCQSSGQWKFPAQQPRKSSAPRRRHTGFRPSTQLLILIAVLLALLQGRTVDAGAGDSLS 60
Zebrafish    -----------------------MLC-LSPPVT----YNLLFWFLILQTP-QSTAPLIT 30
Human        ------MLWFSGVGA------LAERYCRRSPGIT----CCVLLLLNCSGVPMSLASSFLT 44
Mouse        ------MLWFFSVRA------LAERPCRRSPGIT----CCVLLLLNCSGVPMSLASSFLT 44
                                    *      :*: :     :  .  :. .: ::

Drosophila   GVRQSCSNEGEEVQLKNQPQIFTCFKCECQNGFVNCRD-TCPP-VNDCYILDKSNGTCCR 118
Zebrafish    GTEASCENEGEVLHIPNITD-NPCISCVCLNQKAECKQEKCAPLAEDCALVVKQTGACCE 89
Human        GSVAKCENEGEVLQIPFITD-NPCIMCVCLNKEVTCKREKCPVLSRDCALAIKQRGACCE 103
Mouse        GSVAKCENEGEVLQIPFITD-NPCIMCVCLNKEVTCKREKCPVLSRDCALAIKQRGACCE 103
             *  .*.**** ::: .   .  *: * ** . *:  .  .** :  *. *:**.

Drosophila   RCKGCSFRGMSYESGSEWNDPEDPCKTYKCVATVVTETIQKCYSQCDNNQLQPPRPGECC 178
Zebrafish    KCKGCLLKGTSYNSSHHWISPVKPCVTYSCQEGVITEAEMRCVIHCKNPKI---HPKKCC 146
Human        QCKGCTYEGNTYNSSFKWQSPAEPCVLRQCQEGVVTESGVRCVVHCKNPLE---HLGMCC 160
Mouse        RCKGCTHEGRTYNSSFKWQTPAEPCVLRQCQEGVVTESEVRCVVHCKNPAE---HQGACC 160
             :****  .* :*:*. .*  * .**   .*  *:**:  :*  :*.*     :   **

Drosophila   PTCQGCKINGQTVAEGHEVDASIDDRCLVCQCRGTQLTCSKKTCPVLPCPMSKQIKRPDE 238
Zebrafish    PTCPGCIFEGNLYKEDEEFHP-EGNPCIKCVCTGGQSMCHKLVCPVLSCPSHLTHTPPGQ 205
Human        PTCPGCVFEGVQYQEGEEFQP-EGSKCTKCSCTGGRTQCVREVCPILSCPQHLSHIPPGQ 219
Mouse        PTCPGCVFEGVQYREGEEFQP-EGNKCITCSCVGGRTQCVREVCPILSCPQHLSHTPSGQ 219
             *** **  ::*       *..*..   ..  *  * * * * :   * : .**:* **      .:

Drosophila   CCPRCPQNH-SFLPVPGKCLFNKSVYPEKTQFMPDRCTNCTCLNGTSVCQRPTCPILECA 297
Zebrafish    CCPRCRGQRRVFDLSPGSCLFHSEVYENGSSISYDNCTTCTCVDSTVLCRKRCSPPGSCH 265
Human        CCPKCLGQRKVFDLPFGSCLFRSDVYDNGSSFLYDNCTACTCRDSTVVCKRKCSHPGGCD 279
Mouse        CCPKCLGQRKVFDLPFGSCLFRSDVYDNGASFVYDNCTVCTCKDSTMVCKKKCSHPGVCN 279
             ***:*  ::  *.***...**  :   *.** *** *:  * :  .**:* **        *

Drosophila   PEFQEPDGCCPRCAVA---EVRSECSLDGIVYQNNETWDMGPCRSCRCNGGTIRCAQMRC 354
Zebrafish    ---G--TACCEACQSHLKMEDVKYCRVKSKIYRDGDRWSSVNCSLCTCVKGNIQCQPKIC 320
Human        ---QGQEGCCEECLLRVPPEDIKVCKFGNKIFQDGEMWSSINCTICACVKGRTECRNKQC 336
Mouse        ---SDEDACCEDCLLRVPPEDIKVCKFGSKIFRDGEMWSSVNCSICACVKGKTECRKKQC 336
                      .**  *      *  .  *  . :::::.: *.   *  * * *   .*      *

Drosophila   PAVKCRANEELKQPPGECCQRCVETAGTCTVFGDPHFRTFDGKFFSFQGSCKYLLASDCM 414
Zebrafish    VPITSCPSNKIL-NRTGCCPVCTDKPGVCTVFGDPHYNTFDGRTFNFQGTCKYVLTKDCS 379
Human        IPISSCPQGKIL-NRKGCCPICTEKPGVCTVFGDPHYNTFDGRTFNFQGTCQYVLTKDCS 395
Mouse        VPVSSCPQGKIL-NRKGCCPICTEKPGVCTVFGDPHYNTFDGRTFNFQGTCQYVLTKDCS 395
             :..  . ::     **  *.:. *.********:.****: *.***:*:*:*:.**

Drosophila   G--KTFHIRLTNEGRGTRRSSWAKTVTLSLRNLKVNLGQRMRVKVNGTRVTLPYFVVAGG 472
Zebrafish    -PAASFTVLVKNDARRTRSFSWTKSVELHTSGLSISLHQHLTVRQNGTRIALPFHSTGV- 437
Human        SPASPFQVLVKNDARRTRSFSWTKSVELVLGESRVSLQQHLTVRWNGSRIALPCRAPHF- 454
Mouse        SPASPFQVLVKNDARRTRSFSWTKSVELMLGESTVSLQQHLTVRWNGSRIALPCHTPHF- 454
             * :  :.*:.* ** **:*:* *      :.* *:: *: **:*::**

Drosophila   QNVTIERLADGGAVMLRSEMGLTLEWNGAGFLQVSVPAKFKKRLCGLCGNFNGSSRDDLT 532
Zebrafish    -----HIDLDGYLLKLTTIAGLEITWDGDSLVEVVAAPHLKGQLCGLCGNYNGQRRDDSL 492
Human        -----HIDLDGYLLKVTTKAGLEISWDGDSFVEVMAAPHLKGKLCGLCGNYNGHKRDDLI 509
Mouse        -----HIDLDGYLLKVTTRAGLEISWDGDSFVEVMAAPHLKGKLCGLCGNYNGHKRDDLI 509
              . **  : : :  **  :  * :*:*  .  ::* :*******:** ***

Drosophila   GKDGRSHGDDEVWHFANSWKVGGPKSCSRKRE----FLAATPTRDKRKSNFYCHPLSVPA 588
Zebrafish    GGDGQFKFD--VDELAESWRVEDNELCEVQNRRPTSFLCAGTVKVKLRAHRTCQKLK-SW 549
Human        GGDGNFKFD--VDDFAESWRVESNEFCNRPQRKPVPELCQGTVKVKLRAHRECQKLK-SW 566
Mouse        GGDGNFKFD--VDDFAESWRVESNEFCNRPQRKPVPELCQGTVKVKLRAHRECQKLK-SW 566
             * **. : *  * .:*:**:* . : *.  ..     *.   .: * ::: *: *.

Drosophila   LFGECNERLNPENYKAACRMDVCECPS-GDCHCDSFAAYAHECRRLGVQLPDWRSATNCP 647
Zebrafish    EFQKCHSAVDFTSFYKSCVTDMCECPVHKNCFCESFIAYSRACEREGIHV-HWRPELTCM 608
Human        EFQTCHSTVDYATFYRSCVTDMCECPVHKNCYCESFLAYTRACQREGIKV-HWEPQQNCA 625
Mouse        EFQTCHSTVDYTTFYRSCVTDMCECPVHKNCYCESFLAYTRACQREGIKV-HWEPQQSCA 625
             *  *:. ::  .:  :*  *:**** :* *:** **:: *.* *::: .*.   .*

Drosophila   AGWRRNATLSSFKGNQFYGD-PSFSRMKGRRQKNHQLRLQLQQEQQQRSKQGQKGRHKPG 706
Zebrafish    STHCKHGAVYDTCGPGCTKSCHNWNEI-GPCQ----------------RPCIAGCHCPA 650
Human        ATQCKHGAVYDTCGPGCIKTCDNWNEI-GPCN----------------KPCVAGCHCPA 667
Mouse        ATQCKHGAVYDTCGPGCVKTCDNWNEI-GPCN----------------KPCIAGCHCPA 667
             :   ::.:: .  *      .:..: * :              :   * * *.

Drosophila   GHNQLDRQGHNGLDKDQLQKEFILKHVPSSFLYPRAPDRTPPPLH       751
Zebrafish    SLVM--FQGHC-----------IK----PTS---CPGR------        668
Human        NLVL--HKGRC-----------IK----PVL---CPQR------        685
Mouse        NLVL--HKGRC-----------IK----PVL---CPQR------        685
             .       :*:              :*          .* *
```

Figure 4.10 – Sequence alignment of BMPER homologues

Sequences of BMPER from Drosophila (UniProt ID Q9GTX3), Zebrafish (UniProt A6H8K2), Human (UniProt ID Q8N8U9) and Mouse (UniProt ID Q8CJ69) aligned using Clustal Omega (Sievers and Higgins, 2018); '*' indicates a fully conserved residue, ':' indicates residues with strongly similar properties and '.' Indicates residues with weakly similar properties.

	Drosophila	Zebrafish	Human	Mouse
Drosophila	100.00			
Zebrafish	34.10	100		
Human	33.83	67.37	100	
Mouse	33.53	68.41	92.70	100

Table 4-2 – Identity Matrix of BMPER homologues

Amino acid sequence identities of the BMPER homologues shown in Figure 4.10.

With the interactions of N-BMPER and Tsg being so markedly different from BMPER and Tsg the question of "why?" arose. In Figure 4.3, the dissociation rate of BMPER and N-BMPER from Tsg was examined and was similar at $2.03 \times 10^{-4}s^{-1}$ and $2.28 \times 10^{-4}s^{-1}$ respectively. Although the structure of BMPER has been shown to be elongated, the presence of the C-terminus of BMPER may be having some effects on the binding of Tsg due to conformational changes.

BMPER and chordin have been shown to bind tightly from a variety of SPR and pull down studies (Ambrosio et al., 2008), yet corroboration of the data by SPR was required using human homologues. Figure 4.4 shows the interaction of N-BMPER and chordin domains vWFC 2-3. These constructs were used as sub-domain 2 of vWFC1 to vWFC4 of BMPER, were required to bind to vWFC2 of chordin at a K_D of 140 nM (Zhang et al., 2010a). The proteins in the study by Zhang et al. were those previously described (zebrafish BMPER homologue and the mouse chordin homologue using SF9 insect cells). The resulting K_D of 0.2 nM, shown in Figure 4.4 was a much tighter interaction than that shown by Zhang et al. (Zhang et al., 2010a) but was similar to the K_D of 1.27 nM shown by Ambrosio et al. (Ambrosio et al., 2008) where recombinant mouse homologues of BMPER and chordin were used. The agreement of the K_D in the study shown and that of Ambrosio et al. (Ambrosio et al., 2008) are likely due to using recombinant protein from the same species and produced in mammalian expression systems.

With the validation of the interaction of BMPER to Tsg *in vitro*, the next step was to determine how this effects BMP signalling. Provisionally, Ambrosio *et al.* had shown that not only did BMPER and Tsg interact with one another; Tsg enabled BMPER to

bind tighter to BMP-4 and crosslinking showed a ternary complex (Ambrosio et al., 2008). It was also shown that the BMPER:Tsg:BMP-4 complex inhibited the binding of BMP-4 to BMPR-IB receptor. Around the same time as the *in vitro* studies, *in vivo* studies were carried out to elucidate how Tsg and BMPER may interact with one another on a genetic level (Ikeya et al., 2008; Zakin et al., 2008). By using double knock out studies in mice it was shown that pups with the double *BMPER*[-/-] and *Tsg*[-/-] mutations survived with little appearance of skeletal phenotypes (Ikeya et al., 2006). This showed that BMPER and Tsg were linked, with Tsg epistatic over BMPER. A further study showed that both Tsg and BMPER abolish any proangiogenic effect either protein had on their own, in terms of sprouting and branching (Heinke et al., 2013). However, the way in which BMPER and Tsg inhibit BMP signalling has not previously been described.

BMP signalling assays were completed to assess how the combination of BMPER and Tsg can affect BMP signalling. Previous BMP signalling assays, performed in collaboration with the Sengle Lab (Troilo et al., 2016a; Troilo et al., 2014), were repeated for BMPER and Tsg. As mentioned, ALP assays had been completed for Tsg and had found that at 28× molar excess inhibited BMP-4 signalling by 20% (Troilo et al., 2016a). It was discovered that when BMPER was titrated onto C2C12 cells, there was little change in signalling (Figure 4.5-A, red line). When coupled with Tsg there was a clear decrease in the percentage of BMP-4 signalling; BMP-4 signalling was decreased enough to determine an IC_{50} at 2.6× molar excess, of BMPER (Figure 4.5-A, blue line). Clearly the difference between the absence and presence of Tsg in Figure 4.5-A is greater than the 20% decrease in BMP-4 activity that had previously been shown (Troilo et al., 2016a) and suggests BMPER and Tsg act synergistically to bind, and inhibit, BMP-4. This clearly corroborated the information shown by Ambrosio *et al.* whereby BMP-4 is more tightly bound by Tsg and BMPER, but also this prevents BMP-4 binding to BMPR-IB (Ambrosio et al., 2008). When repeated with N-BMPER alone, there was a clear decrease in BMP-4 activity (Figure 4.5-B, red line). When N-BMPER was titrated onto the cells with Tsg, the activity decreased further, again allowing an IC_{50} at 1.4× molar excess to be determined (Figure 4.5-B, blue line). At 1.4× molar excess, the percentage inhibition of N-BMPER alone is approximately 20%. Again, when compared to the activity of Tsg, the combination cannot account for 50% inhibition of BMP-4 signalling and so suggests a synergistic effect. All values can be found in Table 4-1

Figure 4.11-A shows a brief schematic of the relative inhibition potential of BMPER, N-BMPER and Tsg. BMPER and N-BMPER have distinct differences in terms of binding to the cell surface through the C-terminal vWFD domain to HSPGs. When comparing the inhibition potential of BMPER and N-BMPER alone, N-BMPER is a more potent inhibitor. This is most likely due to the ability of N-BMPER to diffuse away from the cell surface and into the media, showing a higher inhibition potential, where BMPER, as shown by Serpe *et al.,* acts as a short range inhibitor, but also a facilitator of signalling by sequestering BMPs close to the cell surface (Serpe et al., 2008). These effects are amplified considerably by the addition of Tsg, enhancing the binding of BMPER to BMP-4 (Figure 4.11-B).
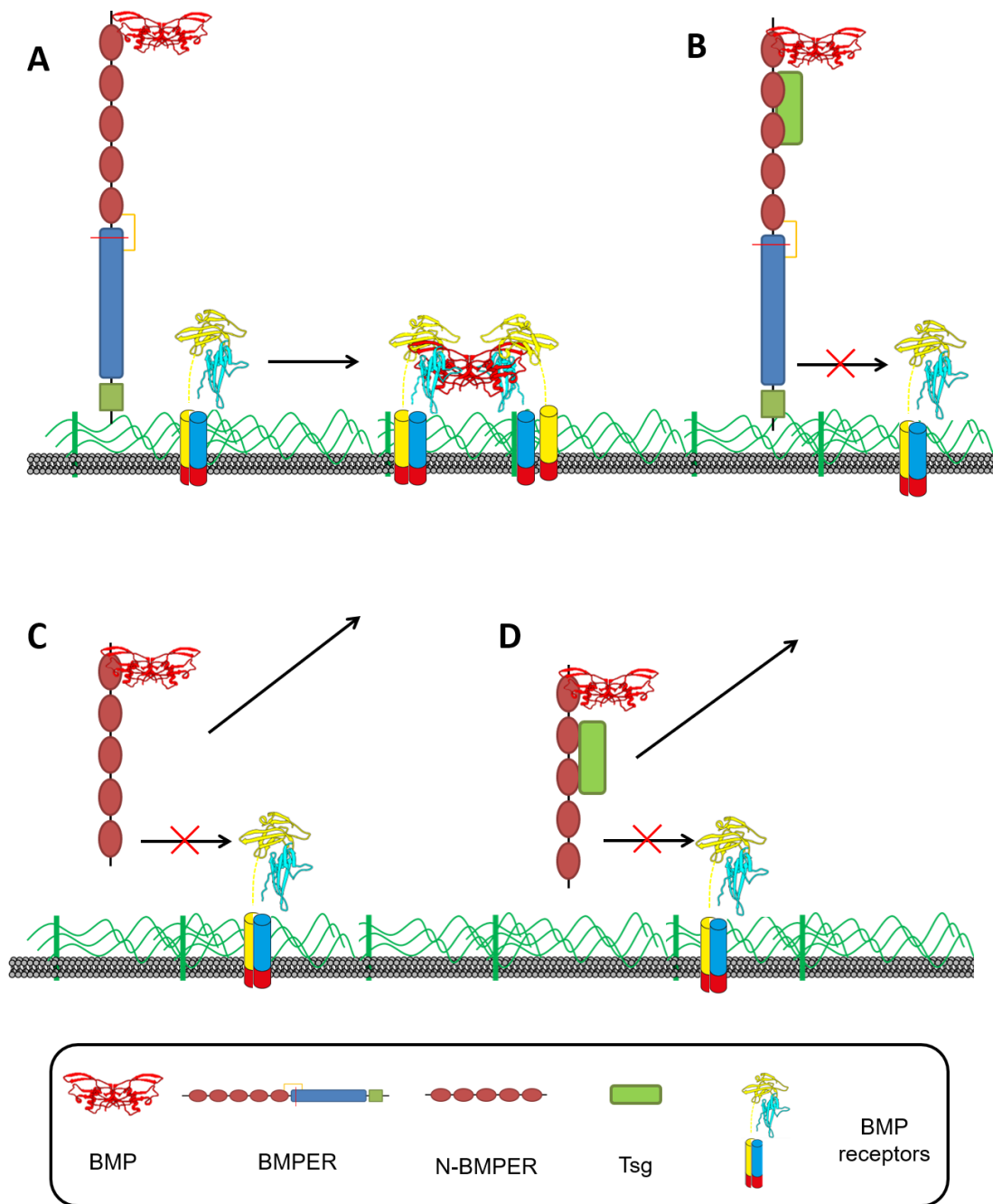
Figure 4.11 – Model of Action of BMPER and N-BMPER.

A simple schematic showing how BMPER, N-BMPER and Tsg can inhibit BMP signalling, but when BMPER or N-BMPER are bound to Tsg the signalling is inhibited further. Shown here are BMPER in the absence (A) and presence (B) of Tsg bound to HSPGs through the vWFD domain and N-BMPER in the absence (C) and presence (D) of Tsg, without the capacity to bind to the cell surface. This shows how BMPER, N-BMPER and Tsg have the ability to agonise and antagonise BMP signalling. When Tsg is bound to BMPER or N-BMPER the signalling is inhibited further. For N-BMPER there is no vWFD domain to mediate binding to the cell surface and so the N-BMPER:Tsg:BMP complex will diffuse into intercellular space, decreasing signalling more.

4.4.2 Mutations in BMPER cause DSD by increasing BMP inhibition

For this study the DSD mutation, BMPER-P370L, was generated in a lentiviral vector and transduced into HEK293-T cells, which were selected by sorting for GFP (Figure 4.6). It was hypothesised that BMPER-P370L would not be secreted as it had previously been shown that mutations in the acid-catalysed 'GDPH' motif of RGMs prevented secretion from HEK293-T cells (Bell et al., 2013). By interrupting the cleavage motif it is thought that the vWFD domain is unable to fold correctly and so is degraded. If BMPER-P370L were not secreted it would effectively represent a knockout of BMPER which would mimic phenotypes shown in mouse models of $BMPER^{-/-}$ mice which have very similar skeletal characteristics to patients suffering from DSD. Surprisingly, Western blots of the conditioned media of BMPER-P370L showed that BMPER-P370L was secreted (Figure 4.8-A & -B). Even more surprisingly BMPER-P370L appeared to be cleaved under both reducing and non-reducing conditions, but with a mass lower than that of the BMPER cleavage product, approximately 37 kDa. Western blots of the cell lysate were used to determine if this cleavage occurs intra- or extracellularly. Figure 4.8-C shows that BMPER is cleaved intracellularly, as would be expected due to the low pH of the secretory pathway (Ambrosio et al., 2008; Lidell and Hansson, 2006; Lidell et al., 2003), whereas BMPER-P370L does not appear to be cleaved within the cell lysate. This leads us to believe that BMPER-P370L is cleaved extracellularly by an ECM protease. As BMPER-P370L is secreted, cell based assays were used to determine how BMP signalling may be effected.

Further cell-lines were generated using the lentiviral vectors described above by co-transducing BMPER constructs with a doxycycline inducible BMP-2 construct (Figure 4.7). It had been shown by Dr Stuart Cain that when the BMP-2 vector was not induced, BMP-2 could not be detected above endogenous levels showing no 'leaky-ness' (data not shown). Cells were lysed and BMP-2 activity was probed by Western blotting against pSMAD-1/5/8. The values were normalised against SMAD-1, with the resulting graphs shown in Figure 4.9. Here we see that BMPER inhibits BMP-2 signalling, and BMPER-P370L is a more effective inhibitor of BMP-2 signalling. This increase in inhibition is reminiscent of the increased BMP inhibitory activity of N-BMPER when compared to BMPER. Disruption of the acid-catalysed 'GDPH' motif could cause the vWFD domain to fold incorrectly and expose a cryptic cleavage site. When BMPER-P370L is cleaved it liberates an N-terminal fragment, of approximately 63 kDa. This fragment would almost mimic N-BMPER, containing the vWFC domains

(Figure 4.12). This N-terminal cleavage product may account for the increased inhibition of BMP-2 signalling. It has been show that BMP signalling is extremely important in the early stages of embryogenesis, with homozygous null mice showing embryonic lethality and heterozygous BMP-2 null mice exhibiting cardiac malformation (Zhang and Bradley, 1996). As BMPER transcripts were shown to be high in early embryogenesis (Moser et al., 2003), the interference of BMP signalling in early embryogenesis may bring about the onset of the DSD phenotype.
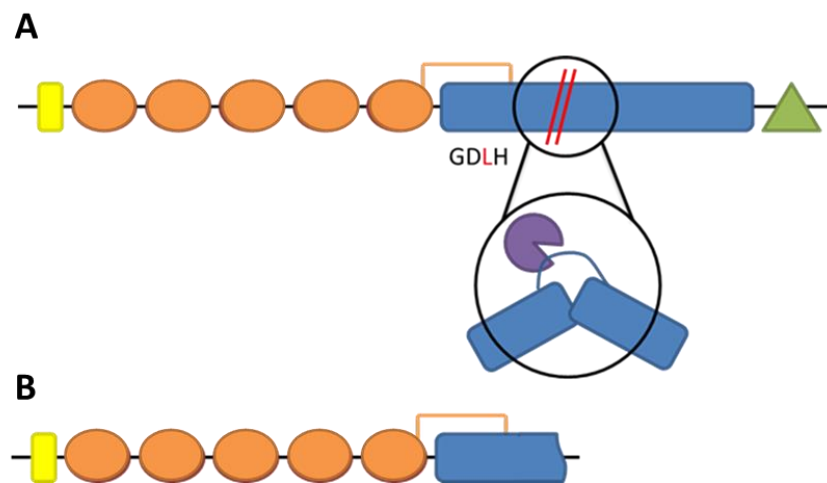
Figure 4.12 – Potential Mechanism of P370L Cleavage.

A) BMPER schematic shown with the P370L mutation. The inset shows now an exposed loop from the vWFD domain may be cleaved by a protease. B) The resulting N-terminal cleavage product of BMPER.

To conclude, BMPER functions by binding through its C-terminal vWFD domain to the ECM, mediated by HS, and bind BMPs and BMP regulators that are secreted into the ECM through the N-terminal vWFC domains. The results described within this chapter demonstrate the capacity of BMPER to bind through these domains, and how the binding may differ between species. It also presents the first biochemical study into the DSD causing mutation at the acid-catalysed cleavage site 'GDPH' and how this mutation impacts on BMP signalling.

5   Results Chapter 3: Beamline Upgrades and B21

This chapter will reflect upon the contributions to the work on several projects that were carried out at Diamond Light Source's beamline B21 throughout the time of my PhD.

5.1   Beamline Upgrades

Many experiments to study biological samples have been conducted using SAXS. It has traditionally been completed using batch mode, where protein samples are introduced into beamline X-rays in a capillary, usually quartz. Problems with this type of data collection arose with the formation of time, or radiation, dependant aggregates. As a result, beamlines have devised several ways to lessen effects of radiation damage induced by X-ray radiation. These methods include the flowing of samples slowly through the X-ray beam, utilised at beamlines BM29 at the ESRF (France) and P12 at DESY (Germany), or oscillating the sample back and forth, utilised at Station G1 at CHESS (USA). Further problems were seen with the measuring of complexes, where non-covalently complexed proteins, in a dynamic equilibrium, would form mixtures in the capillary. To overcome these problems many SAXS beamlines are employing SEC-SAXS where data is collected on the eluent of SEC purification runs. In addition to problems with protein complexes, the SEC-SAXS method can be used to remove any trace aggregates that would interfere with data collection. In order to improve the data collected using SEC-SAXS on beamline B21, custom samples cells with custom window materials have been designed and employed.

5.1.1   B21 Custom Sample Cells

With low flux beamlines (less than 10^{12} photons/s) there is a large challenge in collecting high quality data. Beamline B21 has a flux of 10^{11} photons/s and so to make use of this low flux to collect high quality data a custom sample environment was designed.

Beamline B21 was designed for use with the EMBL automated sample-changing robot using the custom quartz sample cell (Round et al., 2015), but upgrades to the beamline have begun to implement SEC-SAXS, using an Agilent HPLC. The HPLC

is able to use both the Superose and Superdex increase series of SEC columns (GE Healthcare) and the KW series of SEC columns (Shodex). The HPLC setup is fitted with two pumps to equilibrate a column with one pump whilst running the other with sample and vice versa.

It is possible to control the HPLC remotely from the control room while the HPLC runs in the experimental hutch of the beamline. The HPLC can complete runs using a sample injection mechanism, loading samples from a 96-well plate. The 96-well plate sits in a temperature controlled chamber so multiple samples can be run using a combination of the Agilent control software and GDA (http://www.opengda.org/). The standard EMBL capillary worked well with the HPLC setup. The eluent from the HPLC was used to flow through the sample cell and into waste, yet attempts were made to find a better solution for use with the stop flow system described below.

The main problem with the EMBL capillary is that it was designed to withstand the pressure of the vacuum used to draw samples through the beam in the Bio-SAXS robot (Round et al., 2015); the walls of the capillary are made of a relatively thick quartz to account for this pressure (Table 5-1), and would absorb a higher proportion of scattered X-rays.

	EMBL Sample Cell Capillary	Custom B21 Cell
External Diameter (mm)	1.8	N/A
Internal Diameter (mm)	1.7	1.5
Capillary Wall thickness (mm)	0.05	Varied
Volume at beam (µL)	11.35	12.75

Table 5-1 – Dimensions of the EMBL and B21 Custom Cells

The values presented are for the dimensions of the samples cells used at Diamond Light Source. The values allow for the comparison of dimensions of the quartz capillary to be compared to that the custom sample cell and the calculation of the volumes in the cells at the 5 mm beam.

Although the absorption of X-rays is not usually a problem in standard batch mode operation, it becomes a problem with the lower protein concentrations resulting from a SEC run. Samples are diluted by approximately one third (Figure 5.1) and the flow of the samples means that the signal per frame can only be recorded for the duration of the peak. A secondary problem was that the capillary has a large total volume (approximately 35 μL) (Zhang et al., 2007b). Preliminary studies found that this caused dilution of the sample by diffusion when the sample was static using the stop flow method, preventing effective buffer subtractions.
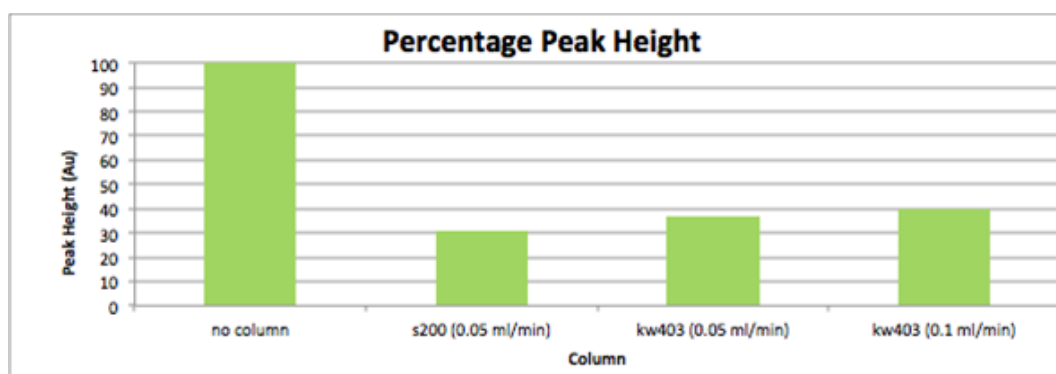
Figure 5.1 – Relative Dilution by Different SEC Columns

Comparison of A280 peak heights normalised to the injection of BSA through the HPLC system without a column attached and with an S200 increase column and KW403 column. The S200i column, run at 0.05 mL/min, provides a 3.2 times dilution where the KW403 column, run at 0.05 mL/min, provides a 2.7 times dilution. The dilution is lessened again for the KW403 column at a flow rate of 0.1 mL/min to 2.5 times allowing for a higher peak with a smaller base, and a higher concentration at the peak.

A new sample cell was designed in collaboration with the beamline B21 staff (Figure 5.2-A). The manufacturing of the sample cell was completed with stainless steel, taking approximately six weeks. Although the manufacturing time is not a limiting factor, a scheme was put into place for decreasing the time taken for sample cell optimisation. Using 3D printing the sample cell was optimised through various print jobs to produce a smaller cell (section 2.10). Optimisation proved successful as a sample cell, printed with a resolution of 30 μm, was competed in 3 hours and 20 minutes. The cells shown in Figure 5.2-C are the initial steel prototype next to a 3D printed cell with a shorter end. These optimised cells would decrease the cost and manufacturing time of making the stainless-steel cells due to a smaller material requirement. It could also allow the sample cells to be used for more hazardous substances like live virus and toxic proteins, by sending out the sample cells to labs

where they could be loaded under the appropriate safety conditions. The printed sample cells would then be sealed and returned to beamline B21 for data collection. This would be duplicated for measuring buffer too. Once the samples were measured the sample cells containing the hazardous materials could be disposed of appropriately without incurring large costs for acquiring new sample cells.

Figure 5.2-B shows how the B21 SEC-SAXS sample cell was printed, where a scaffold is printed around the cell in order to support the sample cell during construction process. The cells were printed on an angle so as to prevent pooling of the resin at areas critical for creating a vacuum in the flight-tube. The printed cells had the scaffold removed, the ends tapped, to allow valves to be screwed into them, and were sanded down at the points where the cell came into contact with O-rings to form a vacuum in the flight-tube. The cells were then baked at 50 °C in a vacuum oven overnight to harden before use. Window materials were then glued onto the cleaned surface of sample cells for measuring data.
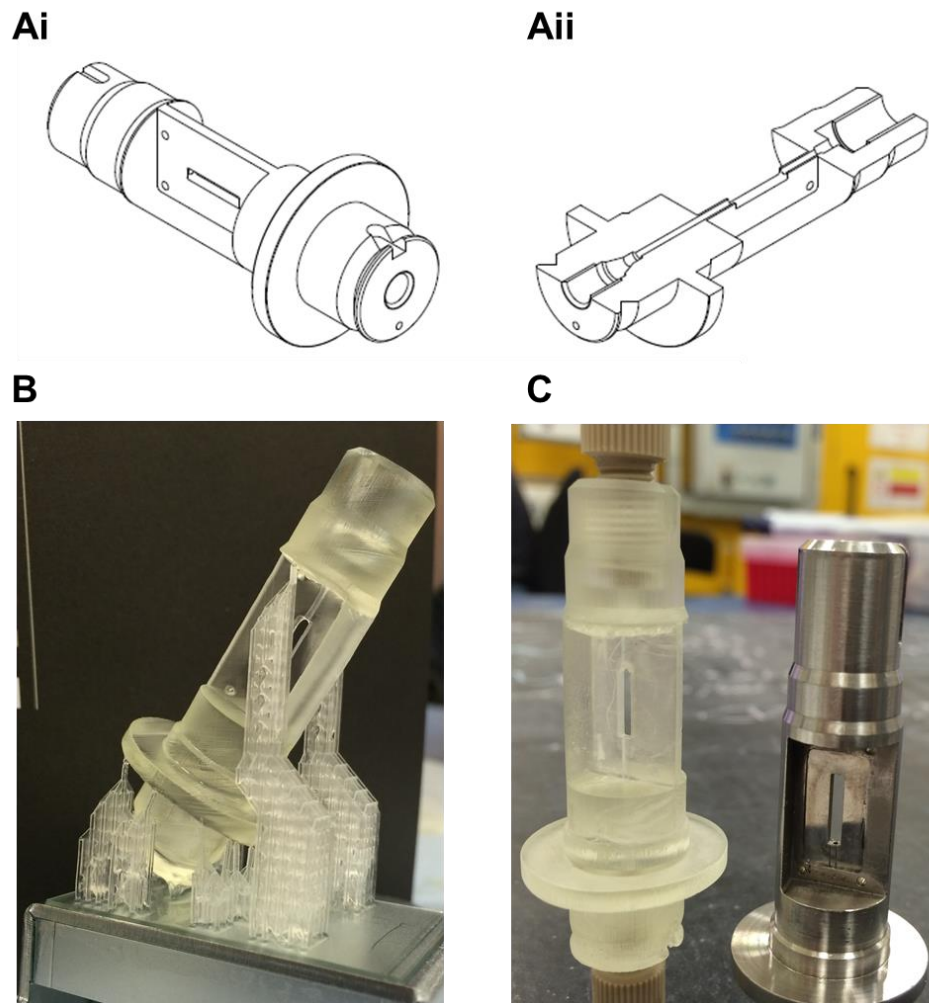
Figure 5.2 – Design of the B21 SEC-SAXS Sample Cell.

A) The engineering schematics of the B21 SEC-SAXS sample cell shown in full (Ai) and in cross section (Aii). B) A 3D printed sample cell shown with the scaffold required to support printing the sample cell at an angle. C) Comparison of the 3D printed sample cell after the scaffolding support was removed next to the machined stainless steel B21 SEC-SAXS sample cell. It can be seen here that the end length of the stainless steel sample cell is longer than that of the optimised 3D printed cell. The figure also represents how valves can be easily screwed into the sample cell using the tapped holes.

5.1.2   Testing Windows Samples

Collectively the beamline staff and I completed the measuring of the following data. Various window materials were glued onto the sample cell in order to identify the most appropriate material. The cell was placed into the vacuum and an exposure of 15 minutes was recorded (Figure 5.3-A). From these results it was determined that synthetic mica, with a width at approximately 25 µm had the lowest scatter; the background of the synthetic mica window is approximately 100 times lower than that

of a quartz window. Additionally, a silicon nitride (SiN) window, with a width of 2 μm, was measured (data not shown) with a further reduced background.

A SEC-SAXS study was carried out to determine how the quartz windows compared to that of the SiN windows by measuring a glucose isomerase (GI) standard with 14 one second exposures, subtracting buffer from the sample across the elution peak (Figure 5.3-B). From the scattering profiles of GI it is possible to see that the SiN windows (black curve) provided less absorbance than the quartz windows (green curve) and so a higher signal to noise ratio was achieved. With that, it was still possible to extract the GI scattering signal from the data by applying the noisy-coding channel theorem (Rambo and Tainer, 2013). Although SiN was a more efficient window material it was prone to breaking and so the sturdier synthetic mica was used, with only a small drop in performance.

A

B
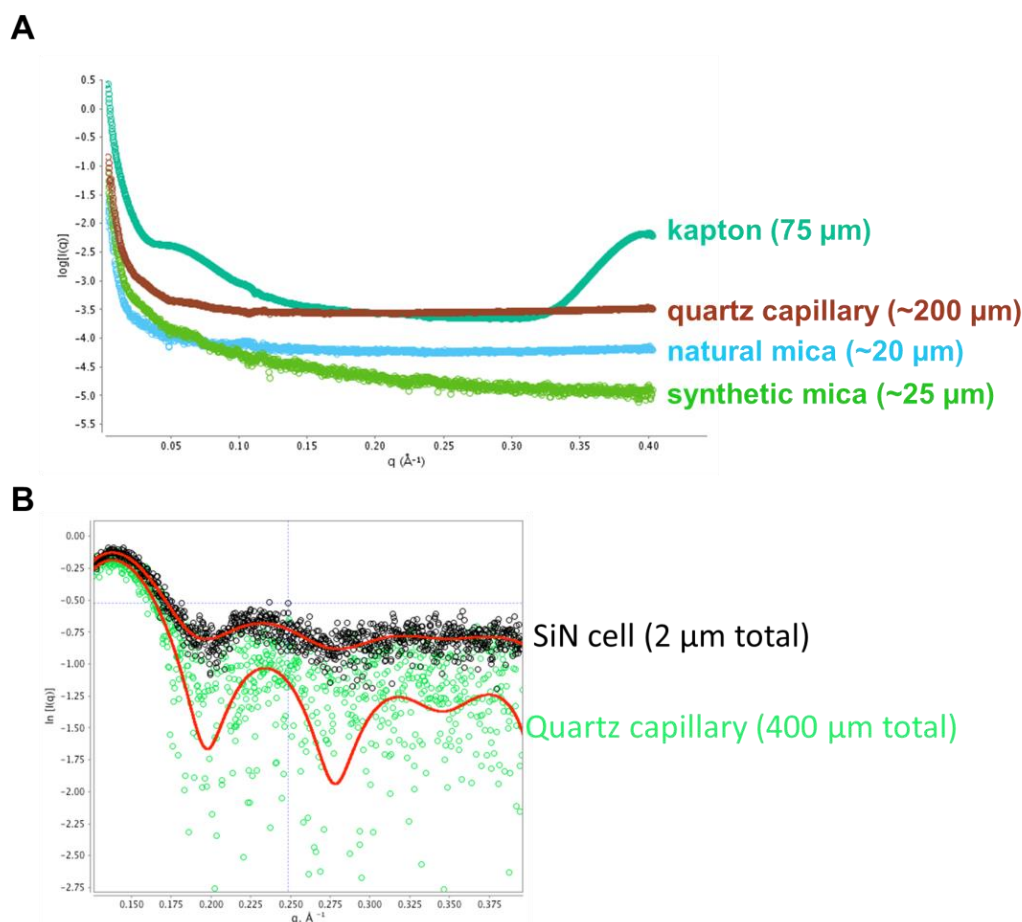
Figure 5.3 – Testing Different Window Materials for the Sample Cell

A) Measurements of different potential window materials were taken by gluing them onto the 3D printed sample cells. Each measurement consisted of a single 15 minute exposure on beamline B21 and the resulting scatter plotted without scaling. This data shows that synthetic mica, with a width of approximately ~25 μm, was the material with the lowest background scattering. B) SEC-SAXS experiment using GI with the processed scattering curves showing the improvements of the SiN windows (black) over the quartz windows (green) with an improved signal to noise ratio. The red line indicated the Fourier Transform of the PDDF for GI. To extract the GI scattering signal from the quartz capillary the noisy-coding channel theorem was applied (Rambo and Tainer, 2013).

5.1.3   HPLC with A Stop Flow System

Implementation of the newly customised sample cell with mica windows was tested with a stop-flow system (Figure 5.4-A) using three protein standards: xylanase, bovine serum albumin (BSA) and GI. The sample cell has a valve on the back end that opens under flow through route 1 and automatically closes when the flow is diverted through the valve into route 2 (Figure 5.4-A). This allows the peak of the sample to be held in the sample cell and, due to the shape of the cell, sample diffusion is minimised relative

136

to the quartz capillary (data not shown). Data were collected in the same way as batch mode, with 30 frames at 10-second exposures for each of the standards, and the corresponding buffers were collected by the same method after 1 full CV of buffer had passed through the column. The frames were reduced, subtracted (the buffer was treated in the same way as the sample and held in the sample cell) and averaged and the PDDF calculated in ScÅtter (Figure 5.4-B) (section 2.9.2.2). As these proteins are standards the R_g and the D_{max} were known. Whilst calculating the PDDF very little data needed truncating from the high q region due to the high signal to noise ratio achieved by the accumulation of frames (Figure 5.4-B). Similarly, very few points were removed from the low q region due to the use of the SEC column, although removal of parasitic scatter from the beamstop was required. The volume of correlation (V_c) plot gives an indication of the molecular mass of the protein (Rambo and Tainer, 2013), but can also be used as an estimate to show how well the buffer of the sample matches the subtracted buffer. This can be seen qualitatively in Figure 5.4-Biii where the integration of the $q.I(q)$ plot shows a flat line. This shows that the buffer was well matched to the buffer in the sample.

This technique takes advantage of the lower flux density of beamline B21 where the accumulation of frames, and therefore signal, can occur without the onset of radiation damage. Although other beamlines utilise SEC-SAXS, this method of stopping the flow of the HPLC by diverting the flow through a quick change valve appears to be the only type in the world.
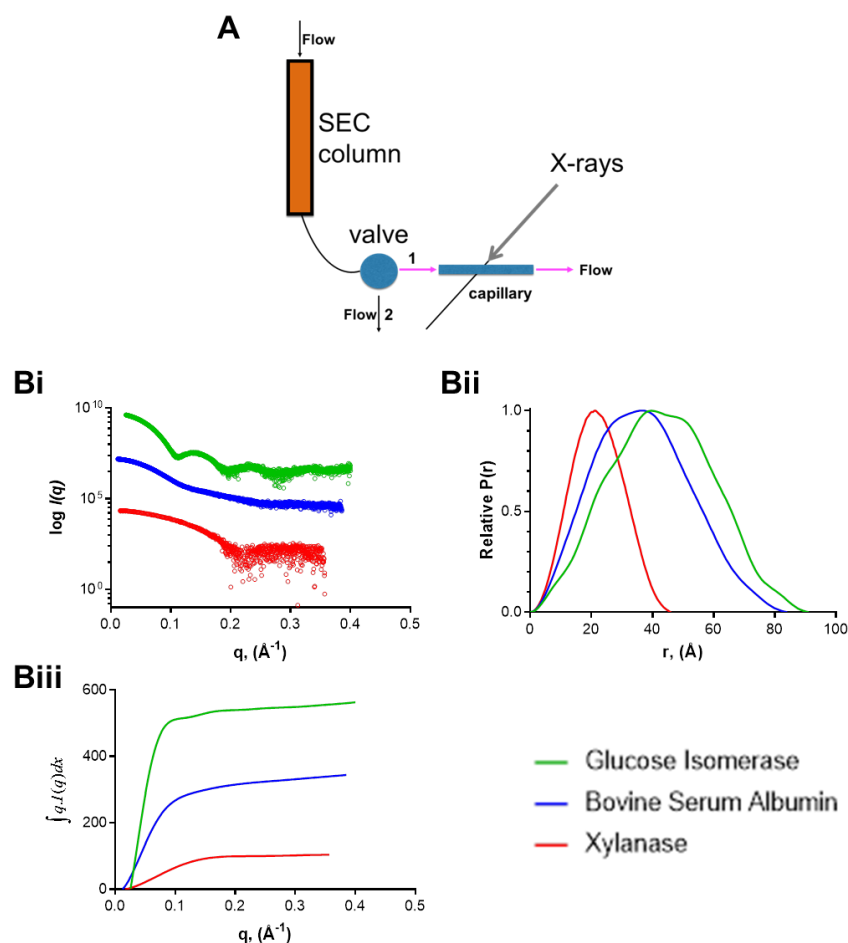
Figure 5.4 – Measurements of standard proteins using the stop-flow technique.

A) Schematic of the stop flow set-up where the eluent from the SEC column goes through a valve and onto the X-ray beam, following route 1. The valve can be changed remotely from the control room to capture the peak elution and allow the column to keep flowing to waste through route 2. B) 10 mg/mL samples Xylanase (red), bovine serum albumin (blue) and GI (green) standards were run down a KW-403 SEC column (Shodex). The peak was retained in the X-ray beam for 30 frames of 10-second exposures. Processed scattering data (Bi) and corresponding PDDF (Bii) are shown with volume of correlation (V_c) plot (Biii).

5.2 **Discussion**

Here the upgrades to the peripheral hardware of beamline B21 were described. These upgrades have had a beneficial effect on the ability to collect high quality SAXS data with long exposures without the onset of aggregation or the need for beam attenuators, like higher flux beamlines (Blanchet et al., 2015). While other beamlines with greater flux have become specialist in high throughput SAXS, B21 utilised the lower flux density in order to record high quality data from fewer samples. This is accompanied with software that allows users to interact with the SEC-SAXS setup effectively and without specialist knowledge.

In addition to SEC-SAXS, further improvements have been made to the sample environments in order to increase the signal to noise ratio without the need for beamline upgrades, although these are scheduled. By utilising 3D printing and testing window materials, beamline B21 has become the first user of the EMBL sample robot to utilise new sample cells (Blanchet et al., 2015; Round et al., 2015). This has allowed materials other than quartz to be used and with the addition of the stop flow-system, SEC-SAXS data can be collected at lower concentrations.

Amongst the developments that are ongoing at beamline B21, the new beamstop will provide a higher signal to noise ratio with a predicted background that is five times lower. Additionally the beamline will upgrade its optics and will see the inclusion of a double-multilayer monochromator (DMM) which is predicted to increase the flux by approximately 80 times. Increasing the flux means that data can both be collected faster and at a better quality on lower sample concentrations. This makes the technique more accessible and open to a larger audience in the biology community.

Overall these modifications to the beamline have allowed what might be considered disadvantages to be turned into beamline strengths. Beamline B21 now has the capacity to surpass many other beamlines (e.g. BM29 of the ESRF and P12 of PETRA III) with regard to data quality, although it lacks the throughput of higher flux beamlines; arguably the quality of information generated using B21 exceeds that of high throughput systems.

6 Results Chapter 4: Paper Contributions

This chapter contains contributions to projects that have used SAXS to address questions and have subsequently been published. As these projects are varied, from bacterial cell wall biosynthesis (Cleverley et al., 2016), to herpes viral transcription factors (Tunnicliffe et al., 2017), only the SAXS experiment will be described, and how the results have allowed the hypothesis in each situation to be tested will be discussed.

6.1 Twisted Gastrulation paper

This section contains a reprint of the article:

Troilo, H., Barrett, A.L., Zuk, A.V., <u>Lockhart-Cairns, M.P.</u>, Wohl, A.P., Bayley, C.P., Dajani, R., Tunnicliffe, R.B., Green, L., Jowitt, T.A*., et al.* (2016). Structural characterization of twisted gastrulation provides insights into opposing functions on the BMP signalling pathway. Matrix Biol.

My contribution to the paper is the analysis and interpretation of SAXS data shown in Figure 2. The full paper can be found in Appendix D, Section 12.1.

6.1.1 Twisted Gastrulation Paper Methods

Tsg is a secreted glycoprotein involved in the bone morphogenetic (BMP) signalling pathway. Currently no structural knowledge is available for Tsg. Due to the size of Tsg only NMR spectroscopy and X-ray crystallography are suitable methods for determining the high resolution structure experimentally but so far Tsg has not been resolved by these methods, presumably due to the low expression level and high degree of glycosylation. Here SAXS was employed as a technique to determine the low resolution structure of Tsg (Troilo et al., 2016a).

SAXS data were collected on both native recombinant Tsg and natively deglycosylated recombinant Tsg on beamline I22 at Diamond Light Source, and at PETRA III, on beamline P12. From the data, shown in Figure 2, it was possible to determine the R_g and the D_{max} for native Tsg as 31 Å and 90 Å, and for deglycosylated Tsg as 25 Å and 110 Å. The generation of *ab initio* models was completed for both of the samples using DAMMIN (Svergun, 1999), and average over the resulting 20 models using the DAMAVER suite (Volkov and Svergun, 2003). From the generated models it was possible to see that native Tsg had an additional feature (Figure 2-Ei), not shown by the deglycosylated Tsg (Figure 2-Eii). This was most likely due to the presence of the three N-linked glycans predicted for Tsg.

In order to map the location of the glycans, multi-phase *ab initio* models were generated with MONSA (Svergun, 1999) through the ATSAS online server. Contrasts of Tsg were set to the protein as 1, and the glycans as 1.6 to represent the differences in their scattering contribution. Phase-1 was specified as the volume of deglycosylated Tsg, and was assumed to be rigid; phase-2 was specified as the difference in volume between native Tsg and deglycosylated Tsg. It is phase-2 that has a contrast of 1.6. From the 20 models that were generated, three representative models were displayed showing the locations of the glycans (Figure 2-F).

By modelling the location of the glycans, it was possible to show that they cluster together on the 3D structure, although predictions show that they are spread throughout the primary sequence. The clustering of the glycans shows that this may contribute to their role in binding BMPs presented previously (Billington Jr et al., 2011). This paper showed that the truncation or removal of Tsg glycans perturbs the binding of BMPs by Tsg, and so the clustering of glycans may explain this change in binding affinity.

6.2 **GpsB paper**

This section contains a reprint of the article:

Cleverley, R.M., Rismondo, J., Lockhart-Cairns, M.P., Van Bentum, P.T., Egan, A.J., Vollmer, W., Halbedel, S., Baldock, C., Breukink, E., and Lewis, R.J. (2016). Subunit Arrangement in GpsB, a Regulator of Cell Wall Biosynthesis. Microb Drug Resist.

SAXS data forms the basis for figures 1-3. The full paper can be found in Appendix D, Section 12.2.

6.2.1 *GpsB Paper Methods*

GpsB is a key regulator in cell wall division in Gram-positive bacteria. GpsB is known to be involved in cell division by binding components of the cell division machinery. As the full length GpsB protein was resistant to crystallising, SAXS was used to determine the subunit arrangement of *Listeria monocytogenes* GpsB (*Lm*-GpsB) (Cleverley et al., 2016). By determining the full length structure, and coupling this with functional assays, it allows new insights into how GpsB is essential for efficient cell division.

The SAXS data of *Lm*GpsB and N-*Lm*GpsB were collected in either batch mode at beamline P12 at PETRA III (Germany), and also by SEC-SAXS on beamline B21 at Diamond Light Source (UK). The data were initially processed to attain the R_g and the D_{max} of *Lm*GpsB at 53.1 Å and 185 Å (Figure 2) and for N-*Lm*GpsB at 22.2 Å and 76 Å (Figure 1) respectively. Advice was provided for docking of crystal structures, with regards to the method of representing and validating the structures. Examples can be seen in Figure 1-A where CRYSOL (Svergun et al., 1995) was used to calculate the theoretical scatter of the crystal structure and showed a reasonable fit with a χ^2 of 1.8 to the experimental data. Additionally, the PDDF of N-*Lm*GpsB was determined using ScÅtter and directly compared to the experimental data, again resulting in a good fit (Figure 1-A). As the end result would be to dock models into the SAXS results, it was advised that the flexibility of the proteins were taken into account, if they did indeed exhibit flexibility. In Figure 2B, the Porod-Debye plots were plotted and show that N-*Lm*GpsB was rigid, as would be expected of a single, helical domain, and that *Lm*GpsB presented a higher degree of flexibility.

These data presented the first insight into the arrangement of *Lm*GpsB allowing for prediction of the importance of this structural arrangement for function, with scope for future experiments to test this.

6.3 **ICP4 paper**

This section contains a reprint of the article:

Tunnicliffe, R.B., Lockhart-Cairns, M.P., Levy, C., Mould, A.P., Jowitt, T.A., Sito, H., Baldock, C., Sandri-Goldin, R.M., and Golovanov, A.P. (2017). The herpes viral transcription factor ICP4 forms a novel DNA recognition complex. Nucleic Acids Res.

My contributions are shown in Figure 4A-D and Supplementary Figure 4 and 5. The full paper can be found in Appendix D, Section 12.312.1.

6.3.1 ICP4 paper methods

Infection by the Herpes simplex virus-1 (HSV-1) causes a sequential cascade of viral gene expression, including the 'immediate-early' (IE) genes. The IE gene in turn produces infected cell protein 4 (ICP4), which is able to act as a repressor by binding to the promoter of the IE genes. In the experiments the N-terminus of the ICP4 was used (ICP4N) in conjunction with a 19 base pair IE3 (IE3_19mer). The N-terminus contains an intrinsically disordered region (IDR) that was not resolved through structure determination by X-ray crystallography. A single phenylalanine from the IDR, intercalating into the DNA strand, was resolved suggesting the presence of the IDR region around the DNA strand. The IDR region proved important for DNA binding as removal decreased binding, calculated by SPR.

SAXS methods were employed for this project in order to determine how the ICP4N transcription factor dimer changed in conformation when bound to a DNA strand, analogous to the ICP4 promoter (IE3_19mer). Data for the ICP4N dimer, IE3_19mer and the ICP4N: IE3_19mer complex were collected using HPLC-SAXS at beamline B21 (Supplementary Figure 4A). Data were shown to be monomeric through the linearity of the Guinier analysis (Supplementary Figure 4E). Scattering data were processed to determine the PDDF resulting in a real space R_g and D_{max} for the ICP4N dimer, IE3_19mer and the ICP4N: IE3_19mer complex (Supplementary Figure 4B). Interestingly the R_g and D_{max} of the ICP4N dimer decreased from 31 Å to 25 Å and from 127 Å to 83 Å, respectively, during the formation of the ICP4N:IE3_19mer complex. This was confirmed by the dimensionless Kratky plot (Figure 4A), where the

ICP4N:IE3_19mer complex is shown to be almost spherical compared to the elongation of the ICP4N dimer.

Comparison of the scattering data to the computed scattering intensity calculated from the ICP4N:IE3_19mer complex crystal structure (Figure 4B) showed similarity, but the lack of the IDR in the resolved structure accounted for the poor fit, with a χ^2 of 2.4.

The *ab* initio models generated through DAMMIF (Franke and Svergun, 2009) fitted the data well (Supplementary Figure 4C). The docking of the crystal structure into the *ab initio* model (Supplementary Figure 5B) showed that there were regions of empty volume around the DNA and below the ICP4N:IE3_19mer complex. It was assumed that these were occupied by the IDR regions so these regions modelled for by protein multiphase *ab initio* modelling. The phases were accounted for by determining the volume of the DNA component from SAXS, and this was designated phase-1 as this was a more static structure than the protein dimer, as shown above. Phase-2 was calculated by subtracting the volume of the DNA from the volume of the ICP4N:IE3_19mer complex. Due to the inherent scattering differences of DNA and protein, phase-1 was assigned a contrast of 2, and phase-2 was assigned a contrast of 1. 20 MONSA (Svergun, 1999) runs were used to generate the multi-phase *ab initio* models through the ATSAS online server. The models fitted the data well (Supplementary Figure 4D) and methods described in Section 2.9 were used to generate a finalised model (Figure 4C). Figure 4D shows the crystal structure of the ICP4N:IE3_19mer docked within the multi-phase *ab initio* model with the DNA fitting within the phase-1 volume and the protein fitting within the phase-2 volume. Labels show the location of the IDR within the phase-2 volume.

Together with the hydrodynamic data and binding data shown within the rest of the article, a novel binding model was developed showing how the ICP4 dimer may 'search' for DNA. DNA can then be bound through the IDR regions working in a synergistic manner with the globular region of the dimer, enhancing the protein:DNA complex (Tunnicliffe et al., 2017).

6.4 **CAR paper**

This section contains a reprint of the article:

Gahloth, D., Dunstan, M.S., Quiaglia, D., Klumbys., E., Lockhart-Cairns, M.P., Hill, A.M., Derrington, S.R., Scutton, N.S., Turner, N.J., and Leys, D. (2017). Structure and mechanism of carboxylic acid reductase. Nature Chemical Biology.

SAXS data validated crystal structures and was presented in a number of supplementary figures. The full paper can be found in Appendix D, Section 12.4.

6.4.1 *CAR Paper Methods*

Carboxylic acid reductase (CAR) reduces carboxylic acids to aldehydes without proceeding to alcohols, a difficult task for small molecule catalysts. CAR presents a potential biotechnological tool that can easily be altered to increase specificities for different substrates. CAR is made up of an adenylation domain (A-domain), a peptidyl carrier protein (PCP) domain and a reductase termination domain (R-domain). Here a combination of SAXS and X-ray crystallography provided information resulting in a model of action (Gahloth et al., 2017).

All SAXS data were collected on beamline B21 at Diamond Light Source using SEC-SAXS. I mainly subtracted SAXS data, and processing of data was split between the lead author and me. Crystal structures were determined for several combinations of domain arrangements from two species: *Nocardia iowensis* (CAR*ni*) and *Segniliparus rugosus* (CAR*sr*). Firstly, the CAR A-PCP domain pair was crystallised in an open conformation and a closed conformation (Supplementary Figure 2B). To determine which of the conformations existed in solution, SAXS data were collected for A-PCP. Neither conformation fitted the SAXS data well, shown by the high χ^2 value (Supplementary Figure 4A and C), and so an ensemble method (Tria et al., 2015), was used to sample conformational space. The ensemble resulted in the open model present in 34% of occurrences and the closed model present in the remaining 66%. The ensemble fitted the SAXS data well, with a χ^2 value of 1.39 (Supplemental Figure 4D). A further crystal structure was determined of the PCP-R domains, and a similar process was carried out to determine the occurrence of multiple domain conformations (Supplementary Figure 7). In this case an ensemble of three crystal

structures was used to describe the SAXS curve (χ^2 1.21), where the determined crystal structure was present 10% of the time, and two open conformations were present between 20% and 70% of the time (Supplementary Figure 7D). Finally, ensemble modelling was carried out using SAXS data of the whole CAR protein (Supplementary Figure 8). This resulted in two distinct open conformations one at 60% occupancy and the other at 40% (χ^2 1.45).

Collectively, the data presented allowed a model showing how the domains of the CAR protein work to facilitate catalysis. The full CAR protein exists in two open conformations showing that there are degrees of flexibility of the protein that would allow for catalysis, for ligand exchange, and either the adenylation/reduction or thiolation conformations (Gahloth et al., 2017).

7   Final discussion

This thesis focusses on the BMP regulator BMPER which has essential roles in angiogenesis, embryonic development and homeostasis. The structure of BMPER is incompletely understood with only the structure of first vWFC domain bound to BMP-2 determined (Zhang et al., 2007a). The binding of BMP regulators BMPER, Tsg and chordin and the functional consequence of these interactions is unclear due to conflicts in the literature. For example, with regard to whether or not BMPER and Tsg interact directly (Ambrosio et al., 2008; Zhang et al., 2007a). This thesis provides data on the structure and binding partners of BMPER, which is supported by cell based assays and mutational studies. The latter part of this thesis documents my inputs into the upgrades and commissioning of the Diamond Light Source beamline B21.

7.1   Nanoscale Structure of BMPER

Previous structural analysis of BMPER had been limited to the first vWFC domain, which is responsible for the binding of BMPs (Fiebig et al., 2013; Zhang et al., 2008). Although this does not provide information on the structure of the full length BMPER molecule it has provided the first structure of a vWFC domain interacting with BMP-2 and shown how these domains interact with BMPs (Zhang et al., 2008). More recently, the crystal structure of another vWFC domain from CCN3 has been solved showing high levels of structural homology to the vWFC domains of BMPER and collagen 2A (O'Leary et al., 2004; Xu et al., 2017; Zhang et al., 2008). There had not been further advancements in this area of study for BMPER in some time, and knowledge of the global structure of BMPER would aid in further understanding its function. Here, expression, purification, structural and biophysical studies were employed to characterise the N-terminal and full length structure of BMPER.

It was possible to express and purify N- and FL-BMPER from a native mammalian expression system. Biochemical analysis of N-BMPER showed that there are four N-linked glycosylation sites using PNGase, in agreement with previous studies (Kamimura et al., 2004a). Biophysical analysis by MALS and AUC showed that N-BMPER and FL-BMPER were monomeric and elongated molecules (Figure 3.4). Structural measurements using SAXS enabled the generation of bead models to segment the volumes of N-BMPER and C-BMPER (Figure 3.10). Additionally homology models were used to generate a quasi-atomic model of N-BMPER (Figure

3.9) that were used to represent the bead model in the multiphase reconstruction. These SAXS data were corroborated by structural analysis of BMPER by negative-stain TEM (Figure 3.13). The biophysical and structural data was then compared using *in silico* techniques, where the hydrodynamic parameters were calculated from the structural models (Figure 3.15). Overall, the structure and biophysical characterisation of BMPER was completed with the segmentation of N- and C-terminal domains and the N-BMPER domains represented as quasi-atomic models.

7.2 Investigating BMPERs Binding Affinity to Tsg and How This Affects BMP Signalling

Many binding studies had been conducted to determine the binding affinity of BMPER to chordin (Ambrosio et al., 2008; Rentzsch et al., 2006; Zhang et al., 2007b; Zhang et al., 2010b; Zhang et al., 2008). We were able to reproduce these findings with similar binding affinities between BMPER and chordin (Figure 4.4). Interestingly there were conflicting data regarding whether BMPER binds to Tsg (Ambrosio et al., 2008; Zhang et al., 2007b). In these studies, recombinant zebrafish BMPER and mouse Tsg, both expressed in SF9 insect cells, did not interact in SPR assays (Zhang et al., 2007a). However, in our binding studies with mammalian expressed human BMPER, N-BMPER and Tsg they interacted with high affinity (K_D of 0.55 nM) (Figure 4.2). The differences in results observed between our experiments and the published work may be due to the truncated glycans added by sf9 cells for different proteins of different species i.e. zebrafish and mouse. Glycosylation has been shown to have large effects on the ability of ECM proteins to bind to each other (Billington et al., 2011; Troilo et al., 2016a).

Building upon the binding analysis, BMPER protein was provided for cell based assays conducted by the Sengle Lab (University of Cologne, Germany) with a similar approach to previous studies (Troilo et al., 2016a; Troilo et al., 2014). These assays showed that BMPER and Tsg worked in concert to decrease the activity of BMP-4 (Figure 4.5). Together with the results from Ambrosio et al. we see that it is possible to form a triparte complex with BMPER, Tsg and BMP-4 (Ambrosio et al., 2008). Not only does BMPER bind to both Tsg and BMP individually, and Tsg binds to BMPs, it appears that BMPER binds more strongly to BMP when Tsg is present (Ambrosio et

al., 2008). Together this shows the synergistic inhibitory effects of BMPER and Tsg to BMP signalling.

A model proposed by Zakin and De Robertis shows how BMPER may be a vital protein in the binding of the chordin:Tsg:BMP complex (Figure 7.1) through the N-terminal vWFC domains, but also binds to cell surface associated HSPG glypican and the BMPR-1B receptor (Zakin and De Robertis, 2010). The data presented can be used to dissect this model proposed and may eventually be used to improve it.
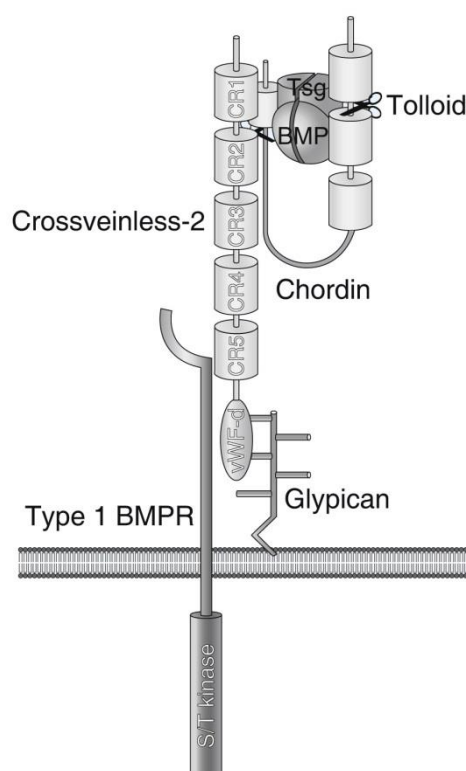
Figure 7.1 – Proposed Model of BMPER Binding

Model showing how the membrane tethered BMPER molecule binds to the Tsg:chordin:BMP complex diffusing through the ECM (Zakin and De Robertis, 2010). The chordin-BMP-Tsg ternary complex interacts with BMPER via chordin binding to the vWFC domains of BMPER. The vWFD domain of BMPER binds to the HSPG glypican and the BMP receptor type 1B.

In order for BMPER to simultaneously bind the cell surface and the Tsg:chordin:BMP complex it would have to be in an extended conformation. Both the SAXS and negative stain EM data supported by hydrodynamic measurements from AUC show

that BMPER is elongated. The extended conformation of BMPER provides a mode of action for how BMPER can bind through both the N- and C-terminus to the chordin-Tsg-BMP complex and the cell surface, respectively. To determine this experimentally, cells expressing BMPER were dosed with varying concentrations of heparin and the media analysed by Western blot. This resulted in a dose response of BMPER being competed from the cell surface by heparin. Conversely this did not occur with N-BMPER, as it lacks the vWFD domain. Interestingly chordin has also been shown to bind to HS, where Tsg cannot (Jasuja et al., 2004). This is interesting as it shows that when BMPER and Tsg work in concert to inhibit BMP signalling, where the vWFD domain is not present in SD mutations; there is no binding of the complex to the cell surface.

7.3 Biochemical Studies into the BMPER Mutation

DSD is a rare, recessive, embryonic lethal, skeletal disorder shown to be as a result of mutations in BMPER. DSD has a more mild form, ISD, which is known to be less frequently lethal. Through genomic sequencing it was first thought that mutations in *pax1* and *meox1* caused DSD (Vatanavicharn et al., 2007) but it was later shown that a mutation in *bmper* was also responsible for DSD (Funari et al., 2010). Subsequently it has been shown that multiple mutations in BMPER are involved in DSD (Scottoline et al., 2012), and are responsible for various levels of severity of this disorder (Kuchinskaya et al., 2016; Zong et al., 2015). Interestingly a mutation in the 'GDPH' cleavage site appeared to be closely linked to a possible mechanism of protein secretion as determined in mucins (Fahim et al., 1987; Lidell and Hansson, 2006; Lidell et al., 2003) and repulsive guidance molecule (RGM), where in the latter it was shown that mutations clustering around the cleavage site prevent cleavage and cause juvenile hemochromatosis (Bell et al., 2013). It was anticipated that the same mechanism would be required for the secretion of BMPER, and would thus explain the DSD phenotype shown in *bmper*$^{-/-}$ mice, we show that this is not the case.

We showed that BMPER-P370L was secreted when cloned into a vector with a mild promoter (UbiC) and the native signal peptide. While this contrasts with previous studies around the 'GDPH' motif mutation (Bell et al., 2013) it appears this may still be a cause for the disease. When secreted from HEK293-T cells, BMPER-P370L showed a C-terminal cleavage product that was smaller than the cleavage product of

native BMPER. This was also observed on non-reducing gels so the cleavage fragments are not stabilised by a disulphide bond. Furthermore the levels of BMP signalling in the presence of BMPER-P370L, by monitoring SMAD phosphorylation, were less than those for wildtype BMPER, but similar to that of N-BMPER. Previous studies on the 'GDPH' motif mutation (Bell et al., 2013) indicated that the 'GDPH' motif is required for the correct folding of the vWFD domain. From this we speculate that the cleavage of BMPER-P370L is through an extracellular protease, so far not identified. This could be due to incorrect folding of the vWFD domain due to the 'GDPH' mutation. As the vWFD domain is highly disulphide bonded, which occurs in the endoplasmic reticulum, providing a rigid structure, the lack of 'GDPH' cleavage in the latter stages of the secretory pathway still allows secretion. Furthermore the GPDH sequence is found at the N-terminus of the vWFD domain and the mutation may confer a conformational change or local misfolding, which could allow a proteolytic site that was previously buried within the domain to be exposed for cleavage. Once cleaved this N-terminal fragment would behave like N-BMPER. BMPs and Tsg bind to the N-terminal fragment strongly, and due to the lack of the vWFD domain there is no cell surface interaction allowing the BMPER complex to diffuse into the matrix and reducing BMP signalling. These findings could explain the reduced BMP signalling seen for the P370L mutant and it will be interesting to see if these data correlate with findings in DSD patients and models.

7.4 Aiding the Implementation and Optimisation of SEC-SAXS at Beamline B21

Improving the hardware and software for any set of techniques allows incremental performance boosts. Where X-ray crystallographers, NMR spectroscopers and more currently cryo-electron microscopists have generated a whole host of new methods cultivated around improved measurement techniques and software methodologies, small-angle X-ray scatterers are enjoying their renaissance, with an increase in usage through greater beamline access and better tools available to process data (Graewert and Svergun, 2013). The use of SEC through HPLC for SAXS is not a new method, but it is now becoming commonplace on beamlines and has room for improvements, with SEC-SANS also being implemented with some success (Jordan et al., 2016). The major improvements would be made with respect to the software. A current problem of SEC-SAXS is that although samples are separated from aggregates, there is always the possibilitiy of heterogeneity within the SEC peak. It is possible to

deconvolute the data within the peaks through rigorous statistical methods, and this would be extremely useful here.

Beamline B21 at Diamond Light Source had its first users at the beginning of 2014 using a BioSAXS robot based on the EMBL design (Round et al., 2015). Since then there have been marked changes throughout the commissioning period. While B21 has a modest flux, at 10^{11} photons per second, compared to other European SAXS beamlines, 10^{13} at BM29 at the ESRF (Pernot et al., 2013) and up to 10^{15} at P12 at DESY (Blanchet et al., 2015), the implementation of SEC-SAXS has been fairly unique. While BM29 and P12 have their own implementation of SEC-SAXS using HPLC, B21 has worked at improving the sample environment to increase the amount of scattering obtained from the sample (Figure 5.3). The sample cell was designed to utilise the maximum amount of beam available to scatter from the sample. This increase in the beam utilisation means that for each frame there is a higher exposure producing higher signal to noise ratios for even lower concentrations.

By increasing the signal to noise ratio, it means that proteins that cannot be highly concentrated or that cannot be produced in high quantities can be analysed with greater precision to higher q and thus providing less ambiguity when further modelling steps are applied. These results show that the current use of cylindrical quartz capillaries doesn't allow the full potential of the beamline to be utilised, as by changing the shape of the cell and the material of the capillary it would provide a lower background scatter with a better signal to noise ratio, illustrated by the window material experimentation.

7.5 Future Work

Following on from the work carried our throughout this thesis, future work can be categorised as either University of Manchester based, or Diamond Light Source based. Where work at University of Manchester would focus on the biological aspects of BMPER and work at Diamond Light Source would focus on the improvements of the beamline.

7.5.1 Complex formation of BMPER binding proteins

The work looking at BMPER in combination with additional regulators has a large scope for expansion. While we have shown that BMPER will bind to Tsg and chordin, with past works showing the binding of BMPER to BMPs (Ambrosio et al., 2008; Binnerts et al., 2004; Coles et al., 2004; Conley et al., 2000; De Robertis, 2009; Kamimura et al., 2004b; Moser et al., 2003; Qiu et al., 2008b; Rentzsch et al., 2006; Zhang et al., 2007b; Zhang et al., 2008), the formation of complexes for structural studies has not yet been attempted. Investigating the molecular details of how BMPER, Tsg and chordin interact and how these interactions influence BMP binding is integral to deconvolution of their regulation. Previously, the structural determination of a complex of around 220 kDa would have required high levels of protein for the formation of crystals for X-ray crystallography. However, the introduction of the high quantum efficiency direct detectors (DDD), responsible for the 'resolution revolution' in cryoEM means that high resolution structures of complexes in this mass range are becoming more routine, with an order of magnitude less protein (Khoshouei et al., 2017; Wu et al., 2016). This has been facilitated with the use of phase plates, and the improvement in cameras since the first generation of DDD cameras. This approach would not only provide information regarding the structure of the multi-part complexes, but it would also provide the structure of Tsg and the CHRD regions of chordin, both of which are novel folds, and have little homology to existing structures. Direct observation of the protein structure would allow a more complete visualisation of the interaction, and the residues involved. As vWFC domains are found in other ECM proteins, such as collagen 2A and CCN3, it may be useful to determine how other structures may bind to BMPs and other regulators if a specific motif can be found. Expression of the multipart complex could be completed using lentiviral vectors that have different fluorescent tags and affinity tags. It would then be possible to sort cells to select cells that express each single regulator, pairs of regulators and all three components. By cloning each construct with a different affinity tag, e.g. His_6, STREP-II and FLAG, it would allow for a more focused and rigorous purification method.

Furthermore, by expressing the regulators together it may provide a higher yield of the complexes as there is possibly a higher chance of complex formation through longer incubation within the media before expression. Furthermore, the expression of TGF-β, when co-expressed with LTBP-1, was shown to be much higher than that of just TGF-β alone; very little TGF-β being secreted and the majority of the TGF-β precursor being found within the cell (Miyazono et al., 1991).

Following the signalling assays conducted within this study of BMPER and Tsg, assays with BMPER and chordin would also be useful to observe their combined effect on BMP signalling. Chordin is a BMP antagonist but as BMPER has both pro- and anti-BMP activity modes, it would be interesting to see how these regulators work together, although it has been shown by SPR that BMPER and chordin are able to bind to BMPs even when the clip region of BMPER has been removed indicating that there may be a mechanism where chordin dominates competition for BMP interactions (Zhang et al., 2010a). Furthermore, all of BMPER, Tsg and chordin would be useful in determining how they would inhibit the signalling of BMPs and how different combinations of the regulators would provide insight into the overall mechanism. It would be possible to conduct these experiments using ALP assays to give a comparison to the data in Section 4.2. It may also be possible to determine the order in which BMPER, Tsg and chordin interact with one another and with BMPs using quartz crystal microbalance with dissipation monitoring (QCMD). This would allow the monitoring of complexation or displacement of proteins in real time and has been used to assess the increase of decrease both mass and structural properties of molecular layers (Dixon, 2008).

7.5.2 Further Investigating the Effects of BMPER-P370L

Initial studies of BMPER-P370L have yielded novel information regarding the proteolytic cleavage of this mutant and its inhibition of BMP inhibition, but much is still unknown about the function of BMPER cleavage and how pathology results from BMPER mutation. By utilising cell based experiments it would be possible to compare the published data on BMPER with those of the mutations. Although many of the pathologies affect the bones and cartilage, studies on BMPER function have primarily been angiogenic assays. By comparing BMPER-P370L with wildtype BMPER in angiogenic assays, it may be possible to provide more information on the effects of BMPER-P370L on signalling. The most efficient approach would be the use of

CRISPR/Cas 9 gene editing to insert the *bmper-P370L* gene into the *bmper* promoter so that cells would have native levels of BMPER transcription, and the same regulation. Cells could then be probed using angiogenic assays to determine efficiency of mutant BMPER in supporting vessel sprouting and network formation. Gene expression levels can also be monitored by looking at specific genes through q-PCR, or RNA-Seq to identify the effect of mutant BMPER.

Additionally, binding studies of BMPER-P370L with Tsg and BMPs would be important to show that there have not been changes in the vWFC domains in order to support the proposed mechanism of action for BMPER-P370L.

7.5.3   Improvements to Beamline B21

The upgrades to the beamline outlined in this thesis are downstream of the optics hutch and have been shown to increase the signal sensitivity of the sample. Further improvements to the experimental setup would be possible in a number of ways. Firstly, the detector could be upgraded from the Pilatus 2M detector (Dectris) to the equivalent Eiger detector. This provides a better signal to noise ratio, smaller pixel size and a more continuous readout. This would prevent the loss of time throughout the collection process; between frames the Pilatus detectors must pause for 10 ms in order to have non-overlapping frames, where the Eiger can continuously read out the data. Additionally, at the detector end of the flight tube the vacuum is maintained by a piece of Kapton. Due to the high vacuum the Kapton bows into the vacuum tube which means that the scattered photons must travel through the air before hitting the detector. Using a stronger, but similarly transparent material, to reduce this air gap would increase the number of photons hitting the detector, improving the signal. This could be further increased by putting the entire detector into a vacuum. Further up from the experimental hutch the optics could be upgraded by to a double-multilayer monochromator (DMM). The DMM uses a relatively wide bandpass for the energy range from the synchrotron source and can be used to increase the flux. Calculations carried out on beamline B21 show that it may be possible to increase the flux by 80 times, with the DMM alone. Combining all of the above, it would be possible to obtain high data signals for proteins at lower concentrations, therefore allowing the acceptable protein concentration to be decreased to below 1 mg/mL.

The upgrades that are mentioned here would not be trivial. The upgrades to the detector alone would cost more than £400,000, and to put this detector under vacuum would require a refitting of the vacuum tube and a specially designed Eiger detector that can withstand the vacuum, further increasing the cost to close to £1 million. Again the upgrade to the DMM would likely cost up to £1 million. That said, in the Diamond Light Source 10 year plan, developments have been scheduled to increase the flux of the beamline by up to 2 orders of magnitude.

8　References

Allendorph, G.P., Isaacs, M.J., Kawakami, Y., Izpisua Belmonte, J.C., and Choe, S. (2007). BMP-3 and BMP-6 structures illuminate the nature of binding specificity with receptors. Biochemistry *46*, 12238-12247.

Allendorph, G.P., Vale, W.W., and Choe, S. (2006). Structure of the ternary signaling complex of a TGF-β superfamily member. Proceedings of the National Academy of Sciences *103*, 7643-7648.

Almasri, M., Kishta, W., Abduljabbar, F.H., Arlet, V., Saran, N., and Oullet, J. (2017). Ischiospinal Dysostosis in a Child with Pierre-Robin Syndrome. Case reports in orthopedics *2017*, 8263536.

Ambrosio, A.L., Taelman, V.F., Lee, H.X., Metzinger, C.A., Coffinier, C., and De Robertis, E.M. (2008). Crossveinless-2 Is a BMP feedback inhibitor that binds Chordin/BMP to regulate Xenopus embryonic patterning. Developmental cell *15*, 248-260.

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta crystallographica Section D, Biological crystallography *62*, 1243-1250.

Arora, K., Levine, M.S., and O'Connor, M.B. (1994). The screw gene encodes a ubiquitously expressed member of the TGF-beta family required for specification of dorsal cell fates in the *Drosophila* embryo. Genes & development *8*, 2588-2601.

Attisano, L., Cárcamo, J., Ventura, F., Weis, F., Massagué, J., and Wrana, J.L. (1993). Identification of human activin and TGFβ type I receptors that form heteromeric kinase complexes with type II receptors. Cell *75*, 671-680.

Baldock, C., Oberhauser, A.F., Ma, L., Lammie, D., Siegler, V., Mithieux, S.M., Tu, Y., Chow, J.Y., Suleman, F., Malfois, M.*, et al.* (2011). Shape of tropoelastin, the highly extensible protein that controls human tissue elasticity. Proceedings of the National Academy of Sciences of the United States of America *108*, 4322-4327.

Basham, M., Filik, J., Wharmby, M.T., Chang, P.C., El Kassaby, B., Gerring, M., Aishima, J., Levik, K., Pulford, B.C., Sikharulidze, I.*, et al.* (2015). Data Analysis WorkbeNch (DAWN). Journal of synchrotron radiation *22*, 853-858.

Bell, C.H., Healey, E., van Erp, S., Bishop, B., Tang, C., Gilbert, R.J., Aricescu, A.R., Pasterkamp, R.J., and Siebold, C. (2013). Structure of the repulsive guidance molecule (RGM)-neogenin signaling hub. Science *341*, 77-80.

Bella, J., and Hulmes, D.J. (2017). Fibrillar Collagens. Sub-cellular biochemistry *82*, 457-490.

Ben-Neriah, Z., Michaelson-Cohen, R., Inbar-Feigenberg, M., Nadjari, M., Zeligson, S., Shaag, A., Zenvirt, S., Elpeleg, O., and Levy-Lahad, E. (2011). A deleterious founder mutation in the BMPER gene causes diaphanospondylodysostosis (DSD). American journal of medical genetics Part A *155a*, 2801-2806.

Bernado, P., and Svergun, D.I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Molecular bioSystems *8*, 151-167.

Berry, R., Jowitt, T.A., Ferrand, J., Roessle, M., Grossmann, J.G., Canty-Laird, E.G., Kammerer, R.A., Kadler, K.E., and Baldock, C. (2009). Role of dimerization and substrate exclusion in the regulation of bone morphogenetic protein-1 and mammalian tolloid. Proceedings of the National Academy of Sciences of the United States of America *106*, 8561-8566.

Berry, R., Jowitt, T.A., Garrigue-Antar, L., Kadler, K.E., and Baldock, C. (2010). Structural and functional evidence for a substrate exclusion mechanism in mammalian tolloid like-1 (TLL-1) proteinase. FEBS Lett *584*, 657-661.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L.*, et al.* (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res *42*, W252-258.

Billington, C.J., Jr., Fiebig, J.E., Forsman, C.L., Pham, L., Burbach, N., Sun, M., Jaskoll, T., Mansky, K., Gopalakrishnan, R., O'Connor, M.B.*, et al.* (2011). Glycosylation of Twisted Gastrulation is Required for BMP Binding and Activity during Craniofacial Development. Front Physiol *2*, 59.

Billington Jr, C.J., Fiebig, J.E., Forsman, C.L., Pham, L., Burbach, N., Sun, M., Jaskoll, T., Mansky, K., Gopalakrishnan, R., and O'Connor, M.B. (2011). Glycosylation of Twisted Gastrulation is Required for BMP Binding and Activity during Craniofacial Development. Frontiers in physiology *2*, 1-9.

Binnerts, M.E., Wen, X., Cante-Barrett, K., Bright, J., Chen, H.T., Asundi, V., Sattari, P., Tang, T., Boyle, B., Funk, W.*, et al.* (2004). Human Crossveinless-2 is a novel inhibitor of bone morphogenetic proteins. Biochemical and biophysical research communications *315*, 272-280.

Blader, P., Rastegar, S., Fischer, N., and Strahle, U. (1997). Cleavage of the BMP-4 antagonist chordin by zebrafish tolloid. Science *278*, 1937-1940.

Blanchet, C.E., Spilotros, A., Schwemmer, F., Graewert, M.A., Kikhney, A., Jeffries, C.M., Franke, D., Mark, D., Zengerle, R., Cipriani, F.*, et al.* (2015). Versatile sample environments

and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). J Appl Crystallogr *48*, 431-443.

Bragdon, B., Moseychuk, O., Saldanha, S., King, D., Julian, J., and Nohe, A. (2011). Bone morphogenetic proteins: a critical review. Cellular signalling *23*, 609-620.

Bragg, W.H., and Bragg, W.L. (1913). The Reflection of X-rays by Crystals. Proceedings of the Royal Society of London Series A *88*, 428-438.

Brookes, E., Demeler, B., and Rocco, M. (2010a). Developments in the US-SOMO bead modeling suite: new features in the direct residue-to-bead method, improved grid routines, and influence of accessible surface area screening. Macromolecular bioscience *10*, 746-753.

Brookes, E., Demeler, B., Rosano, C., and Rocco, M. (2010b). The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. European biophysics journal : EBJ *39*, 423-435.

Brown, M.A., Zhao, Q., Baker, K.A., Naik, C., Chen, C., Pukac, L., Singh, M., Tsareva, T., Parice, Y., Mahoney, A.*, et al.* (2005). Crystal structure of BMP-9 and functional interactions with pro-region and receptors. The Journal of biological chemistry *280*, 25111-25118.

Cai, J., Pardali, E., Sánchez-Duffhues, G., and ten Dijke, P. (2012). BMP signaling in vascular diseases. FEBS letters *586*, 1993-2002.

Caligur, V. (2008). Glycosaminoglycan Sulfation and Signaling. In BioFiles.

Chang, C., Holtzman, D.A., Chau, S., Chickering, T., Woolf, E.A., Holmgren, L.M., Bodorova, J., Gearing, D.P., Holmes, W.E., and Brivanlou, A.H. (2001). Twisted gastrulation can function as a BMP antagonist. Nature *410*, 483-487.

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I.*, et al.* (2007). Glycoprotein Structural Genomics: Solving the Glycosylation Problem. Structure *15*, 267-273.

Cleverley, R.M., Rismondo, J., Lockhart-Cairns, M.P., Van Bentum, P.T., Egan, A.J., Vollmer, W., Halbedel, S., Baldock, C., Breukink, E., and Lewis, R.J. (2016). Subunit Arrangement in GpsB, a Regulator of Cell Wall Biosynthesis. Microb Drug Resist.

Coffinier, C., Ketpura, N., Tran, U., Geissert, D., and De Robertis, E.M. (2002). Mouse Crossveinless-2 is the vertebrate homolog of a Drosophila extracellular regulator of BMP signaling. Mech Dev *119 Suppl 1*, S179-184.

Coles, E., Christiansen, J., Economou, A., Bronner-Fraser, M., and Wilkinson, D.G. (2004). A vertebrate crossveinless 2 homologue modulates BMP activity and neural crest cell migration. Development *131*, 5309-5317.

Conley, C.A., Silburn, R., Singer, M.A., Ralston, A., Rohwer-Nutter, D., Olson, D.J., Gelbart, W., and Blair, S.S. (2000). Crossveinless 2 contains cysteine-rich domains and is required for high levels of BMP-like activity during the formation of the cross veins in Drosophila. Development *127*, 3947-3959.

Crapo, J.D., Barry, B.E., Gehr, P., Bachofen, M., and Weibel, E.R. (1982). Cell number and cell characteristics of the normal human lung. The American review of respiratory disease *126*, 332-337.

Daly, N.L., and Craik, D.J. (2011). Bioactive cystine knot proteins. Current opinion in chemical biology *15*, 362-368.

De Robertis, E.M. (2009). Spemann's organizer and the self-regulation of embryonic fields. Mech Dev *126*, 925-941.

De Robertis, E.M., and Moriyama, Y. (2016). The Chordin Morphogenetic Pathway. Current topics in developmental biology *116*, 231-245.

Debye, P., Jr., H.R.A., and Brumberger, H. (1957). Scattering by an Inhomogeneous Solid. II. The Correlation Function and Its Application. Journal of Applied Physics *28*, 679-683.

Dectris (2017). EIGER X Detector Series for SAXS and WAXS (https://www.dectris.com/EIGER_X_Detectors.html: Dectris).

Dixon, M.C. (2008). Quartz Crystal Microbalance with Dissipation Monitoring: Enabling Real-Time Characterization of Biological Materials and Their Interactions. Journal of Biomolecular Techniques : JBT *19*, 151-158.

Donatelli, J.J., Sethian, J.A., and Zwart, P.H. (2017). Reconstruction from limited single-particle diffraction data via simultaneous determination of state, orientation, intensity, and phase. Proceedings of the National Academy of Sciences of the United States of America *114*, 7222-7227.

Donatelli, J.J., Zwart, P.H., and Sethian, J.A. (2015). Iterative phasing for fluctuation X-ray scattering. Proceedings of the National Academy of Sciences of the United States of America *112*, 10286-10291.

Durand, D., Vives, C., Cannella, D., Perez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010). NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. J Struct Biol *169*, 45-53.

Eldar, A., Dorfman, R., Weiss, D., Ashe, H., Shilo, B.-Z., and Barkai, N. (2002). Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning. Nature *419*, 304-308.

Esko, J.D., Kimata, K., and Lindahl, U. (2009). Proteoglycans and Sulfated Glycosaminoglycans. In Essentials of Glycobiology, A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, and M.E. Etzler, eds. (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press

The Consortium of Glycobiology Editors, La Jolla, California.).

Fahim, R.E., Specian, R.D., Forstner, G.G., and Forstner, J.F. (1987). Characterization and localization of the putative 'link' component in rat small-intestinal mucin. The Biochemical journal *243*, 631-640.

Feigin, L., and Svergun, D.I. (1987). Structure analysis by small-angle X-ray and neutron scattering (Springer).

Fiebig, J.E., Weidauer, S.E., Qiu, L.Y., Bauer, M., Schmieder, P., Beerbaum, M., Zhang, J.L., Oschkinat, H., Sebald, W., and Mueller, T.D. (2013). The Clip-Segment of the von Willebrand Domain 1 of the BMP Modulator Protein Crossveinless 2 Is Preformed. Molecules *18*, 11658-11682.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M.*, et al.* (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. Journal of Applied Crystallography *50*.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography *42*, 342-346.

Fratalocchi, A., and Ruocco, G. (2011). Single-molecule imaging with x-ray free-electron lasers: dream or reality? Physical review letters *106*, 105504.

Funari, V.A., Krakow, D., Nevarez, L., Chen, Z., Funari, T.L., Vatanavicharn, N., Wilcox, W.R., Rimoin, D.L., Nelson, S.F., and Cohn, D.H. (2010). BMPER mutation in diaphanospondylodysostosis identified by ancestral autozygosity mapping and targeted high-throughput sequencing. American journal of human genetics *87*, 532-537.

Gahloth, D., Dunstan, M.S., Quiaglia, D., Klumbys., E., Lockhart-Cairns, M.P.*,* Hill, A.M., Derrington, S.R., Scutton, N.S., Turner, N.J., and Leys, D. (2017). Structure and mechanism of carboxylic acid reductase. Nature Chemical Biology.

Garcia de la Torre, J., Llorca, O., Carrascosa, J.L., and Valpuesta, J.M. (2001). HYDROMIC: prediction of hydrodynamic properties of rigid macromolecular structures obtained from electron microscopy images. European biophysics journal : EBJ *30*, 457-462.

Glatter, O. (1977a). Data Evaluation In Small-Angle Scattering - Calculation Of Radial Electron-Density Distribution By Means Of Indirect Fourier Transformation. Acta Physica Austriaca *47*, 83-102.

Glatter, O. (1977b). A new method for the evaluation of small-angle scattering data. Journal of Applied Crystallography *10*, 415-421.

Glatter, O., and Kratky, O. (1982). Small angle X-ray scattering (Academic press).

Gonzales, M., Verloes, A., Saint Frison, M.H., Perrotez, C., Bourdet, O., Encha-Razavi, F., Joye, N., Taillemite, J.L., Walbaum, R., Pfeiffer, R.*, et al.* (2005). Diaphanospondylodysostosis (DSD): confirmation of a recessive disorder with abnormal vertebral ossification and nephroblastomatosis. American journal of medical genetics Part A *136a*, 373-376.

Graewert, M.A., and Svergun, D.I. (2013). Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). Current opinion in structural biology *23*, 748-754.

Groppe, J., Greenwald, J., Wiater, E., Rodriguez-Leon, J., Economides, A.N., Kwiatkowski, W., Affolter, M., Vale, W.W., Izpisua Belmonte, J.C., and Choe, S. (2002). Structural basis of BMP signalling inhibition by the cystine knot protein Noggin. Nature *420*, 636-642.

Groppe, J., Hinck, C.S., Samavarchi-Tehrani, P., Zubieta, C., Schuermann, J.P., Taylor, A.B., Schwarz, P.M., Wrana, J.L., and Hinck, A.P. (2008). Cooperative Assembly of TGF-β Superfamily Signaling Complexes Is Mediated by Two Disparate Mechanisms and Distinct Modes of Receptor Binding. Molecular cell *29*, 157-168.

Grudinin, S., Garkavenko, M., and Kazennov, A. (2017). Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. Acta Crystallogr D Struct Biol *73*, 449-464.

Guinier, A. (1939). La diffraction des rayons X aux tres petits angles: applications a l'etude de phenomenes ultramicroscopiques.

Gupta, R., and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 310-322.

Guttman, M., Weinkam, P., Sali, A., and Lee, K.K. (2013). All-atom ensemble modeling to analyze small-angle x-ray scattering of glycosylated proteins. Structure *21*, 321-331.

Healey, E.G., Bishop, B., Elegheert, J., Bell, C.H., Padilla-Parra, S., and Siebold, C. (2015). Repulsive guidance molecule is a structural bridge between neogenin and bone morphogenetic protein. Nat Struct Mol Biol *22*, 458-465.

Heinke, J., Juschkat, M., Charlet, A., Mnich, L., Helbing, T., Bode, C., Patterson, C., and Moser, M. (2013). Antagonism and synergy between extracellular BMP modulators Tsg and BMPER balance blood vessel formation. J Cell Sci *126*, 3082-3094.

Heng, S., Paule, S., Hardman, B., Li, Y., Singh, H., Rainczuk, A., Stephens, A.N., and Nie, G. (2010). Posttranslational activation of bone morphogenetic protein 2 is mediated by proprotein convertase 6 during decidualization for pregnancy establishment. Endocrinology *151*, 3909-3917.

Hynes, R.O., and Naba, A. (2012). Overview of the matrisome--an inventory of extracellular matrix constituents and functions. Cold Spring Harbor perspectives in biology *4*, a004903.

Ikeya, M., Kawada, M., Kiyonari, H., Sasai, N., Nakao, K., Furuta, Y., and Sasai, Y. (2006). Essential pro-Bmp roles of crossveinless 2 in mouse organogenesis. Development *133*, 4463-4473.

Ikeya, M., Nosaka, T., Fukushima, K., Kawada, M., Furuta, Y., Kitamura, T., and Sasai, Y. (2008). Twisted gastrulation mutation suppresses skeletal defect phenotypes in Crossveinless 2 mutant mice. Mech Dev *125*, 832-842.

Ikushima, H., and Miyazono, K. (2010). TGFβ signalling: a complex web in cancer progression. Nature Reviews Cancer *10*, 415-424.

Jasuja, R., Allen, B.L., Pappano, W.N., Rapraeger, A.C., and Greenspan, D.S. (2004). Cell-surface heparan sulfate proteoglycans potentiate chordin antagonism of bone morphogenetic protein signaling and are necessary for cellular uptake of chordin. The Journal of biological chemistry *279*, 51289-51297.

Jordan, A., Jacques, M., Merrick, C., Devos, J., Forsyth, V.T., Porcar, L., and Martel, A. (2016). SEC-SANS: size exclusion chromatography combined in situ with small-angle neutron scatteringThis article will form part of a virtual special issue of the journal, presenting some highlights of the 16th International Conference on Small-Angle Scattering (SAS2015). Journal of Applied Crystallography *49*, 2015-2020.

Kadler, K.E., Baldock, C., Bella, J., and Boot-Handford, R.P. (2007). Collagens at a glance. J Cell Sci *120*, 1955-1958.

Kamimura, M., Matsumoto, K., Koshiba-Takeuchi, K., and Ogura, T. (2004a). Vertebrate crossveinless 2 is secreted and acts as an extracellular modulator of the BMP signaling

cascade. Developmental dynamics : an official publication of the American Association of Anatomists *230*, 434-445.

Kamimura, M., Matsumoto, K., Koshiba-Takeuchi, K., and Ogura, T. (2004b). Vertebrate crossveinless 2 is secreted and acts as an extracellular modulator of the BMP signaling cascade. Developmental dynamics *230*, 434-445.

Kang, E.H., Mansfield, M.L., and Douglas, J.F. (2004). Numerical path integration technique for the calculation of transport properties of proteins. Phys Rev E *69*.

Kelley, R., Ren, R., Pi, X., Wu, Y., Moreno, I., Willis, M., Moser, M., Ross, M., Podkowa, M., and Attisano, L. (2009). A concentration-dependent endocytic trap and sink mechanism converts Bmper from an activator to an inhibitor of Bmp signaling. The Journal of cell biology *184*, 597-609.

Khoshouei, M., Radjainia, M., Baumeister, W., and Danev, R. (2017). Cryo-EM structure of haemoglobin at 3.2 A determined with the Volta phase plate. Nature communications *8*, 16099.

Kingsley, D.M. (1994). The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes & development *8*, 133-146.

Kišonaitė, M., Wang, X., and Hyvönen, M. (2016). Structure of Gremlin-1 and analysis of its interaction with BMP-2. Biochemical Journal *473*, 1593-1604.

Kohfeldt, E., Maurer, P., Vannahme, C., and Timpl, R. (1997). Properties of the extracellular calcium binding module of the proteoglycan testican. FEBS Lett *414*, 557-561.

Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. Journal of Applied Crystallography *39*, 277-286.

Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography *34*, 33-41.

Kratky, O., and Porod, G. (1949a). Diffuse small-angle scattering of X-rays in colloid systems. Journal of colloid science *4*, 35-70.

Kratky, O., and Porod, G. (1949b). Röntgenuntersuchung gelöster Fadenmoleküle. Recueil des Travaux Chimiques des Pays-Bas *68*, 1106-1122.

Kuchinskaya, E., Grigelioniene, G., Hammarsjo, A., Lee, H.R., Hogberg, L., Grigelionis, G., Kim, O.H., Nishimura, G., and Cho, T.J. (2016). Extending the phenotype of BMPER-related skeletal dysplasias to ischiospinal dysostosis. Orphanet journal of rare diseases *11*, 1.

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., and Lopez, R. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Larrain, J., Oelgeschlager, M., Ketpura, N.I., Reversade, B., Zakin, L., and De Robertis, E.M. (2001). Proteolytic cleavage of Chordin as a switch for the dual activities of Twisted gastrulation in BMP signaling. Development *128*, 4439-4447.

Larraín, J., Oelgeschläger, M., Ketpura, N.I., Reversade, B., Zakin, L., and De Robertis, E.M. (2001). Proteolytic cleavage of Chordin as a switch for the dual activities of Twisted gastrulation in BMP signaling. Development *128*, 4439-4447.

Legare, J.M., Seaborg, K., Laffin, J., and Giampietro, P.F. (2017). Diaphanospondylodysostosis and ischiospinal dysostosis, evidence for one disorder with variable expression in a patient who has survived to age 9 years. American journal of medical genetics Part A.

Lidell, M.E., and Hansson, G.C. (2006). Cleavage in the GDPH sequence of the C-terminal cysteine-rich part of the human MUC5AC mucin. The Biochemical journal *399*, 121-129.

Lidell, M.E., Johansson, M.E., and Hansson, G.C. (2003). An autocatalytic cleavage in the C terminus of the human MUC2 mucin occurs at the low pH of the late secretory pathway. The Journal of biological chemistry *278*, 13944-13951.

Lin, C.Y., Huang, Z., Wen, W., Wu, A., Wang, C., and Niu, L. (2015). Enhancing Protein Expression in HEK-293 Cells by Lowering Culture Temperature. PLoS One *10*.

Lippmann, E.S., Azarin, S.M., Kay, J.E., Nessler, R.A., Wilson, H.K., Al-Ahmad, A., Palecek, S.P., and Shusta, E.V. (2012). Human Blood-Brain Barrier Endothelial Cells Derived from Pluripotent Stem Cells. Nature biotechnology *30*, 783-791.

Ludtke, S.J. (2016). Single-Particle Refinement and Variability Analysis in EMAN2.1. Methods Enzymol *579*, 159-189.

Mertens, H.D., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. Journal of structural biology *172*, 128-141.

Mi, L.Z., Brown, C.T., Gao, Y., Tian, Y., Le, V.Q., Walz, T., and Springer, T.A. (2015). Structure of bone morphogenetic protein 9 procomplex. Proceedings of the National Academy of Sciences of the United States of America *112*, 3710-3715.

Miyazono, K., Olofsson, A., Colosetti, P., and Heldin, C.H. (1991). A role of the latent TGF-beta 1-binding protein in the assembly and secretion of TGF-beta 1. The EMBO journal *10*, 1091-1101.

Moser, M., Binder, O., Wu, Y., Aitsebaomo, J., Ren, R., Bode, C., Bautch, V.L., Conlon, F.L., and Patterson, C. (2003). BMPER, a novel endothelial cell precursor-derived protein, antagonizes bone morphogenetic protein signaling and endothelial cell differentiation. Mol Cell Biol *23*, 5664-5679.

Moser, M., and Patterson, C. (2003). Thrombin and vascular development: a sticky subject. Arteriosclerosis, thrombosis, and vascular biology *23*, 922-930.

Mulloy, B., and Rider, C.C. (2015). The Bone Morphogenetic Proteins and Their Antagonists. Vitamins and hormones *99*, 63-90.

Nisbet, D.L., Chitty, L.S., Rodeck, C.H., and Scott, R.J. (1999). A new syndrome comprising vertebral anomalies and multicystic kidneys. Clinical dysmorphology *8*, 173-178.

O'Leary, J.M., Hamilton, J.M., Deane, C.M., Valeyev, N.V., Sandell, L.J., and Downing, A.K. (2004). Solution structure and dynamics of a prototypical chordin-like cysteine-rich repeat (von Willebrand Factor type C module) from collagen IIA. The Journal of biological chemistry *279*, 53857-53866.

Pelikan, M., Hura, G.L., and Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. General physiology and biophysics *28*, 174-189.

Perkins, S.J., Wright, D.W., Zhang, H., Brookes, E.H., Chen, J., Irving, T.C., Krueger, S., Barlow, D.J., Edler, K.J., Scott, D.J., *et al.* (2016). Atomistic modelling of scattering data in the Collaborative Computational Project for Small Angle Scattering (CCP-SAS). J Appl Crystallogr *49*, 1861-1875.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., *et al.* (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. Journal of synchrotron radiation *20*, 660-664.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr *45*, 342-350.

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G., and Svergun, D.I. (2007). ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. Journal of Applied Crystallography *40*, S223-S228.

Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J *89*, 1237-1250.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF chimera - A visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Piccolo, S., Agius, E., Lu, B., Goodman, S., Dale, L., and De Robertis, E.M. (1997). Cleavage of Chordin by Xolloid metalloprotease suggests a role for proteolytic processing in the regulation of Spemann organizer activity. Cell *91*, 407-416.

Piccolo, S., Sasai, Y., Lu, B., and De Robertis, E.M. (1996). Dorsoventral patterning in *Xenopus*: inhibition of ventral signals by direct binding of chordin to BMP-4. Cell *86*, 589-598.

Piszkiewicz, D., Landon, M., and Smith, E.L. (1970). Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. Biochemical and biophysical research communications *40*, 1173-1178.

Prefumo, F., Homfray, T., Jeffrey, I., Moore, I., and Thilaganathan, B. (2003). A newly recognized autosomal recessive syndrome with abnormal vertebral ossification, rib abnormalities, and nephrogenic rests. American journal of medical genetics Part A *120a*, 386-388.

Putnam, C. (2016). Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data. Journal of Applied Crystallography *49*.

Putnam, D.K., Lowe, E.W., Jr., and Meiler, J. (2013). Reconstruction of SAXS Profiles from Protein Structures. Computational and structural biotechnology journal *8*, e201308006.

Qiu, L.-y., Zhang, J.-l., Kotzsch, A., Sebald, W., and Mueller, T.D. (2008a). Crystallization and preliminary X-ray analysis of the complex of the first von Willebrand type C domain bound to bone morphogenetic protein 2. Acta Crystallographica Section F: Structural Biology and Crystallization Communications *64*, 307-312.

Qiu, L.Y., Zhang, J.L., Kotzsch, A., Sebald, W., and Mueller, T.D. (2008b). Crystallization and preliminary X-ray analysis of the complex of the first von Willebrand type C domain bound to bone morphogenetic protein 2. Acta crystallographica Section F, Structural biology and crystallization communications *64*, 307-312.

Radcliffe, P.A., and Mitrophanous, K.A. (2004). Multiple gene products from a single vector: 'self-cleaving' 2A peptides. Gene Therapy *11*, 1673-1674.

Rambo, R.P. (2015). Resolving Individual Components in Protein-RNA Complexes Using Small-Angle X-ray Scattering Experiments. Methods Enzymol *558*, 363-390.

Rambo, R.P., and Tainer, J.A. (2011). Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers *95*, 559-571.

Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature *496*, 477-481.

Receveur-Brechot, V., and Durand, D. (2012). How random are intrinsically disordered proteins? A small angle scattering perspective. Current protein & peptide science *13*, 55-75.

Rentzsch, F., Zhang, J., Kramer, C., Sebald, W., and Hammerschmidt, M. (2006). Crossveinless 2 is an essential positive feedback regulator of Bmp signaling during zebrafish gastrulation. Development *133*, 801-811.

Rider, C.C., and Mulloy, B. (2010). Bone morphogenetic protein and growth differentiation factor cytokine families and their protein antagonists. The Biochemical journal *429*, 1-12.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320-324.

Ross, J.J., Shimmi, O., Vilmos, P., Petryk, A., Kim, H., Gaudenz, K., Hermanson, S., Ekker, S.C., O'Connor, M.B., and Marsh, J.L. (2001). Twisted gastrulation is a conserved extracellular BMP antagonist. Nature *410*, 479-483.

Round, A., Felisaz, F., Fodinger, L., Gobbo, A., Huet, J., Villard, C., Blanchet, C.E., Pernot, P., McSweeney, S., Roessle, M.*, et al.* (2015). BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. Acta crystallographica Section D, Biological crystallography *71*, 67-75.

Round, A., Franke, D., Moritz, S., Huchler, R., Fritsche, M., Malthan, D., Klaering, R., Svergun, D., and Roessle, M. (2008). Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. Journal of applied crystallography *41*, 913-917.

Salazar, V.S., Gamer, L.W., and Rosen, V. (2016). BMP signalling in skeletal development, disease and repair. Nat Rev Endocrinol *12*, 203-221.

Scheres, S.H. (2012a). A Bayesian view on cryo-EM structure determination. J Mol Biol *415*, 406-418.

Scheres, S.H. (2012b). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol *180*, 519-530.

Scheufler, C., Sebald, W., and Hülsmeyer, M. (1999). Crystal Structure of Human Bone Morphogenetic Protein-2 at 2.7 Å Resolution. J Mol Bio *287*, 103-115.

Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic acids research *38*, W540-W544.

Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2013). Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J *105*, 962-974.

Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2016). FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J *78*, 1606-1619.

Scott, I.C., Blitz, I.L., Pappano, W.N., Maas, S.A., Cho, K.W., and Greenspan, D.S. (2001). Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signalling. Nature *410*, 475-478.

Scottoline, B., Rosenthal, S., Keisari, R., Kirpekar, R., Angell, C., and Wallerstein, R. (2012). Long-term survival with diaphanospondylodysostosis (DSD): survival to 5 years and further phenotypic characteristics. American journal of medical genetics Part A *158a*, 1447-1451.

Sengle, G., Ono, R.N., Lyons, K.M., Bachinger, H.P., and Sakai, L.Y. (2008). A new model for growth factor activation: type II receptors compete with the prodomain for BMP-7. J Mol Biol *381*, 1025-1039.

Serpe, M., Umulis, D., Ralston, A., Chen, J., Olson, D.J., Avanesov, A., Othmer, H., O'Connor, M.B., and Blair, S.S. (2008). The BMP-binding protein Crossveinless 2 is a short-range, concentration-dependent, biphasic modulator of BMP signaling in Drosophila. Developmental cell *14*, 940-953.

Sharma, U., Carrique, L., Vadon-Le Goff, S., Mariano, N., Georges, R.N., Delolme, F., Koivunen, P., Myllyharju, J., Moali, C., Aghajari, N.*, et al.* (2017). Structural basis of homo- and heterotrimerization of collagen I. Nature communications *8*, 14671.

Shimmi, O., Umulis, D., Othmer, H., and O'Connor, M.B. (2005). Facilitated Transport of a Dpp/Scw Heterodimer by Sog/Tsg Leads to Robust Patterning of the *Drosophila* Blastoderm Embryo. Cell *120*, 873-886.

Sievers, F., and Higgins, D.G. (2014a). Clustal omega. Curr Protoc Bioinformatics *48*, 3.13.11-16.

Sievers, F., and Higgins, D.G. (2014b). Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol *1079*, 105-116.

Sievers, F., and Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many protein sequences. Protein Sci *27*, 135-145.

Sun, M., Forsman, C., Sergi, C., Gopalakrishnan, R., O'Connor, M.B., and Petryk, A. (2010). The expression of twisted gastrulation in postnatal mouse brain and functional implications. Neuroscience *169*, 920-931.

Svergun, D. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophysical journal *76*, 2879-2886.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography *28*, 768-773.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophysical Journal *80*, 2946-2953.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol *157*, 38-46.

Tasian, S.K., Kim, G.E., Miniati, D.N., and DuBois, S.G. (2012). Development of anaplastic Wilms tumor and subsequent relapse in a child with diaphanospondylodysostosis. Journal of pediatric hematology/oncology *34*, 548-551.

Tria, G., Mertens, H.D.T., Kachala, M., and Svergun, D.I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ *2*, 207-217.

Troilo, H. (2013). The Role of Chordin and Twisted Gastrulation in Regulatiing BMP Family Growth Factors. In Faculty of Life Science (University of Manchester), pp. 114-120.

Troilo, H., Barrett, A.L., Zuk, A.V., Lockhart-Cairns, M.P., Wohl, A.P., Bayley, C.P., Dajani, R., Tunnicliffe, R.B., Green, L., Jowitt, T.A.*, et al.* (2016a). Structural characterization of twisted gastrulation provides insights into opposing functions on the BMP signalling pathway. Matrix Biol.

Troilo, H., Bayley, C.P., Barrett, A.L., Lockhart-Cairns, M.P., Jowitt, T.A., and Baldock, C. (2016b). Mammalian tolloid proteinases: role in growth factor signalling. FEBS Lett *590*, 2398-2407.

Troilo, H., Steer, R., Collins, R.F., Kielty, C.M., and Baldock, C. (2016c). Independent multimerization of Latent TGFbeta Binding Protein-1 stabilized by cross-linking and enhanced by heparan sulfate. Scientific reports *6*, 34347.

Troilo, H., Zuk, A.V., Tunnicliffe, R.B., Wohl, A.P., Berry, R., Collins, R.F., Jowitt, T.A., Sengle, G., and Baldock, C. (2014). Nanoscale structure of the BMP antagonist chordin supports cooperative BMP binding. Proceedings of the National Academy of Sciences of the United States of America *111*, 13063-13068.

Tunnicliffe, R.B., Lockhart-Cairns, M.P., Levy, C., Mould, A.P., Jowitt, T.A., Sito, H., Baldock, C., Sandri-Goldin, R.M., and Golovanov, A.P. (2017). The herpes viral transcription factor ICP4 forms a novel DNA recognition complex. Nucleic Acids Res.

Tuukkanen, A.T., Kleywegt, G.J., and Svergun, D.I. (2016). Resolution of ab initio shapes determined from small-angle scattering. IUCrJ *3*, 440-447.

Urist, M.R. (1965). Bone: formation by autoinduction. Science *150*, 893-899.

Vatanavicharn, N., Graham, J.M., Jr., Curry, C.J., Pepkowitz, S., Lachman, R.S., Rimoin, D.L., and Wilcox, W.R. (2007). Diaphanospondylodysostosis: six new cases and exclusion of the candidate genes, PAX1 and MEOX1. American journal of medical genetics Part A *143a*, 2292-2302.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. Journal of Applied Crystallography *36*, 860-864.

Weber, D., Kotzsch, A., Nickel, J., Harth, S., Seher, A., Mueller, U., Sebald, W., and Mueller, T.D. (2007). A silent H-bond can be mutationally activated for high-affinity interaction of BMP-2 and activin type IIB receptor. BMC structural biology *7*, 6.

Wharton, K.A., and Serpe, M. (2013). Fine-tuned shuttles for bone morphogenetic proteins. Curr Opin Genet Dev *23*, 374-384.

Wohl, A.P., Troilo, H., Collins, R.F., Baldock, C., and Sengle, G. (2016). Extracellular Regulation of Bone Morphogenetic Protein Activity by the Microfibril Component Fibrillin-1. The Journal of biological chemistry *291*, 12732-12746.

Wu, S., Armache, J.-P., and Cheng, Y. (2016). Single-particle cryo-EM data acquisition by using direct electron detection camera. Microscopy *65*, 35-41.

Xu, E.R., Blythe, E.E., Fischer, G., and Hyvonen, M. (2017). Structural analyses of von Willebrand factor C domains of collagen 2A and CCN3 reveal an alternative mode of binding to bone morphogenetic protein-2. The Journal of biological chemistry.

Yadin, D., Knaus, P., and Mueller, T.D. (2016). Structural insights into BMP receptors: Specificity, activation and inhibition. Cytokine Growth Factor Rev *27*, 13-34.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics *21*, 3369-3376.

Zakin, L., Chang, E.Y., Plouhinec, J.L., and De Robertis, E.M. (2010). Crossveinless-2 is required for the relocalization of Chordin protein within the vertebral field in mouse embryos. Dev Biol *347*, 204-215.

Zakin, L., and De Robertis, E. (2010). Extracellular regulation of BMP signaling. Current Biology *20*, R89-R92.

Zakin, L., Metzinger, C.A., Chang, E.Y., Coffinier, C., and De Robertis, E.M. (2008). Development of the vertebral morphogenetic field in the mouse: interactions between Crossveinless-2 and Twisted Gastrulation. Dev Biol *323*, 6-18.

Zhang, H., and Bradley, A. (1996). Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. Development *122*, 2977-2986.

Zhang, J.-L., Huang, Y., Qiu, L.-Y., Nickel, J., and Sebald, W. (2007a). von Willebrand factor type C domain-containing proteins regulate bone morphogenetic protein signaling through different recognition mechanisms. Journal of Biological Chemistry *282*, 20002-20014.

Zhang, J.-L., Patterson, L.J., Qiu, L.-Y., Graziussi, D., Sebald, W., and Hammerschmidt, M. (2010a). Binding between Crossveinless-2 and Chordin Von Willebrand Factor Type C Domains Promotes BMP Signaling by Blocking Chordin Activity. PLoS ONE *5*, e12846.

Zhang, J.L., Huang, Y., Qiu, L.Y., Nickel, J., and Sebald, W. (2007b). von Willebrand factor type C domain-containing proteins regulate bone morphogenetic protein signaling through different recognition mechanisms. The Journal of biological chemistry *282*, 20002-20014.

Zhang, J.L., Patterson, L.J., Qiu, L.Y., Graziussi, D., Sebald, W., and Hammerschmidt, M. (2010b). Binding between Crossveinless-2 and Chordin von Willebrand factor type C domains promotes BMP signaling by blocking Chordin activity. PLoS One *5*, e12846.

Zhang, J.L., Qiu, L.Y., Kotzsch, A., Weidauer, S., Patterson, L., Hammerschmidt, M., Sebald, W., and Mueller, T.D. (2008). Crystal structure analysis reveals how the Chordin family member crossveinless 2 blocks BMP-2 receptor binding. Developmental cell *14*, 739-750.

Zhao, Y.G., Bishop, B., Clay, J.E., Lu, W.X., Jones, M., Daenke, S., Siebold, C., Stuart, D.I., Jones, E.Y., and Aricescu, A.R. (2011). Automation of large scale transient protein expression in mammalian cells. Journal of Structural Biology *175*, 209-215.

Zong, Z., Tees, S., Miyanji, F., Fauth, C., Reilly, C., Lopez, E., Tredwell, S., Paul Goldberg, Y., Delaney, A., Eydoux, P.*, et al.* (2015). BMPER variants associated with a novel, attenuated subtype of diaphanospondylodysostosis. Journal of human genetics *60*, 743-747.

9 Appendix A

9.1 Derivation of Normalised Kratky Plot

A derivation of the Normalised Kratky Plot (Durand et al., 2010; Receveur-Brechot and Durand, 2012) as derived from the GA.

$$I(q) = I(0).e^{-\frac{q^2.R_g^2}{3}} \tag{1}$$

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

$$(u)^2.\frac{I(q)}{I(0)} = (u)^2.e^{-\frac{(u)^2}{3}} \tag{3}$$

$$f(u)' = 2u.e^{-\frac{(u)^2}{3}} - u^2.\frac{2u}{3}.e^{-\frac{(u)^2}{3}} \tag{4}$$

$$2u.e^{-\frac{(u)^2}{3}} = \frac{2u^3}{3}.e^{-\frac{(u)^2}{3}} \tag{5}$$

$$3 = u^2 \tag{6}$$

$$u = \sqrt{3} \tag{7}$$

$$R_g.q = \sqrt{3} \approx 1.73 \tag{8}$$

$$\left(q.R_g\right)^2.\left(\frac{I(q)}{I(0)}\right) = (\sqrt{3})^2.e^{-\frac{(\sqrt{3})^2}{3}} \tag{9}$$

$$\left(q.R_g\right)^2.\left(\frac{I(q)}{I(0)}\right) = 3e^{-1} \tag{10}$$

$$3e^{-1} \approx 1.104 \tag{11}$$

Equation 9-1– Derivation of the Dimensionless Kratky Plot

Where $I(q)$ is the scattering intensity, q is the scattering angle, $I(0)$ is the scattering intensity where $q = 0$, R_g is the radius of gyration and $u = (q.R_g)$. Equation (1) is the GA, and is multiplied by $(q.R_g)$ to produce equation (2). From here $(q.R_g)$ is substituted for u for ease of following (3). The first derivative, with regards to u is taken in (4) to determine the gradient of a slope resulting in (5). When this is solved for 0, u has the value of $\sqrt{3}$ (6, 7, 8). This gives the value of the domain axis $(q.R_g)$. by substituting this value back into (2), and solving in (9, 10) we have the value of the range axis, $3e^{-1}$, (11).

10 Appendix B

10.1 Sequences and Primers

```
TCCTCCTTCTTGACAGGTTCTGTTGCAAAATGTGAAAATGAAGGTGAAGTCCTCCAGATTCCATTTATCACAGACAACCCTTGCATAATGTGTGTCTGCTTGAACAAGGAAGTGA
 S  S  F  L  T  G  S  V  A  K  C  E  N  E  G  E  V  L  Q  I  P  F  I  T  D  N  P  C  I  M  C  V  C  L  N  K  E  V
AAGAGAGAGAAGTGCCCCGTGCTGTCCCGAGACTGTGCCCTGGCCATCAAGCAGAGGGGAGCCTGTTGTGAACAGTGCAAAGGTTGCACCTATGAAGGAAATACCTATAACAGCT
 K  R  E  K  C  P  V  L  S  R  D  C  A  L  A  I  K  Q  R  G  A  C  C  E  Q  C  K  G  C  T  Y  E  G  N  T  Y  N  S
AAATGGCAGAGCCCGGCTGAGCCTTGTGTTCTACGCCAGTGCCAGGAGGGCGTTGTCACAGAGTCTGGGGTGCGCTGTGTTGTTCATTGTAAAAACCCTTTGGAGCATCTGGGAA
 K  W  Q  S  P  A  E  P  C  V  L  R  Q  C  Q  E  G  V  V  T  E  S  G  V  R  C  V  V  H  C  K  N  P  L  E  H  L  G
TGCCCCACATGTCCAGGCTGTGTGTTTGAGGGTGTGCAGTATCAAGAAGGGGAGGAATTTCAGCCAGAAGGAAGCAAATGTACCAAGTGTTCCTGCACTGGAGGCAGGACACAAT
 C  P  T  C  P  G  C  V  F  E  G  V  Q  Y  Q  E  G  E  E  F  Q  P  E  G  S  K  C  T  K  C  S  C  T  G  G  R  T  Q
AGAGAAGTCTGTCCCATTCTCTCCTGTCCCCAGCACCTTAGTCACATACCCCCAGGACAGTGCTGCCCCAAATGTTTGGGTCAGAGGAAAGTGTTTGACCTCCCTTTTGGGAGCT
 R  E  V  C  P  I  L  S  C  P  Q  H  L  S  H  I  P  P  G  Q  C  C  P  K  C  L  G  Q  R  K  V  F  D  L  P  F  G  S
TTTCGAAGTGATGTTTATGACAATGGATCCTCATTTCTGTACGATAACTGCACAGCTTGTACCTGCAGGGACTCTACTGTGGTTTGCAAGAGGAAGTGCTCCCACCCTGGTGGCT
 F  R  S  D  V  Y  D  N  G  S  S  F  L  Y  D  N  C  T  A  C  T  C  R  D  S  T  V  V  C  K  R  K  C  S  H  P  G  G
CAAGGCCAGGAGGGCTGTTGTGAAGAGTGCCTCCTACGAGTGCCCCCAGAAGACATCAAAGTATGCAAATTTGGCAACAAGATTTTCCAGGATGGAGAGATGTGGTCCTCTATCA
 Q  G  Q  E  G  C  C  E  E  C  L  L  R  V  P  P  E  D  I  K  V  C  K  F  G  N  K  I  F  Q  D  G  E  M  W  S  S  I
ACCATCTGTGCTTGTGTGAAAGGCAGGACGGAGTGTCGCAATAAGCAGTGCATTCCCATCAGTAGCTGCCCACAGGGCAAAATTCTCAACAGAAAAGGATGCTGTCCTATTTGCA
 T  I  C  A  V  K  G  R  T  E  C  R  N  K  Q  C  I  P  I  S  S  C  P  Q  G  K  I  L  N  R  K  G  C  C  P  I  C
AAGCCCGGCGTTTGCACGGTGTTTGGAGATCCCCACTACAACACTTTTGACGGTCGGACATTTAACTTTCAGGGGACGTGTCAGTACGTTTTGACAAAAGACTGCTCCTCCCCTG
 K  P  G  V  C  T  V  F  G  D  P  H  Y  N  T  F  D  G  R  T  F  N  F  Q  G  T  C  Q  Y  V  L  T  K  D  C  S  S  P
CCCTTCCAGGTGCTGGTGAAGAACGACGCCCGCCGGACACGCTCCTTCTCGTGGACCAAGTCGGTGGAGCTGGTGCTGGGCGAGAGCAGGGTCAGCCTGCAGCAGCACCTCACCG
 P  F  Q  V  L  V  K  N  D  A  R  R  T  R  S  F  S  W  T  K  S  V  E  L  V  L  G  E  S  R  V  S  L  Q  Q  H  L  T
TGGAACGGCTCGCGCATCGCGCTCCCCTGCCGCGCGCCACACTTCCACATCGACCTGGATGGCTACCTCTTGAAAGTGACCACCAAAGCAGGTTTGGAAATATCTTGGGATGGAG
 W  N  G  S  R  I  A  L  P  C  R  A  P  H  F  H  I  D  L  D  G  Y  L  L  K  V  T  T  K  A  G  L  E  I  S  W  D  G
TTTGTAGAAGTCATGGCTGCGCCGCATCTCAAGGGCAAGCTCTGTGGTCTTTGTGGCAACTACAATGGACATAAACGTGATGACTTAATTGGTGGAGATGGAAACTTCAAGTTTG
 F  V  E  V  M  A  A  P  H  L  K  G  K  L  C  G  L  C  G  N  Y  N  G  H  K  R  D  D  L  I  G  G  D  G  N  F  K  F
GATGACTTTGCTGAATCTTGGAGGGTGGAGTCCAATGAGTTCTGCAACAGACCTCAGAGAAAGCCAGTGCCTGAACTGTGTCAAGGGACAGTCAAGGTAAAGCTCCGGGCCCATC
 D  D  F  A  E  S  W  R  V  E  S  N  E  F  C  N  R  P  Q  R  K  P  V  P  E  L  C  Q  G  T  V  K  V  K  L  R  A  H
TGCCAAAAGCTCAAATCCTGGGAGTTTCAGACCTGCCACTCGACTGTGGACTACGCCACTTTCTACCGGTCCTGTGTGACAGACATGTGTGAATGTCCAGTCCATAAAAACTGTT
 C  Q  K  L  K  S  W  E  F  Q  T  C  H  S  T  V  D  Y  A  T  F  Y  R  S  C  V  T  D  M  C  E  C  P  V  H  K  N  C
GAGTCATTTTTGGCATATACCCGGGCCTGCCAGAGAGAGGGCATCAAAGTCCACTGGGAGCCTCAGCAGAATTGTGCAGCCACCCAGTGTAAGCATGGTGCTGTGTACGATACCT
 E  S  F  L  A  Y  T  R  A  C  Q  R  E  G  I  K  V  H  W  E  P  Q  Q  N  C  A  A  T  Q  C  K  H  G  A  V  Y  D  T
CCGGGATGTATCAAGACGTGTGACAACTGGAATGAAATTGGTCCATGCAACAAGCCGTGCGTTGCTGGGTGCCACTGTCCAGCAAACTTGGTCCTTCACAAGGGAAGGTGCATCA
 P  G  C  I  K  T  C  D  N  W  N  E  I  G  P  C  N  K  P  C  V  A  G  C  H  C  P  A  N  L  V  L  H  K  G  R  C  I
GTCCTTTGTCCCCAGCGGCTGGTGCCCCGCGGCTCCCATCATCACCACCACCATTAG
 V  L  C  P  Q  R  L  V  P  R  G  S  H  H  H  H  H  H  -
```

Supplemental Figure 10-1 – Sequence of BMPER

Translation of the BMPER DNA to protein sequence. Red sequence denotes the 5 vWFC domain, blue sequence denotes the vWFD domain and the purple sequence denotes the TIL domain. The sequence which is underlines denotes the thrombin cleavage site, the green sequence denotes the His6 tag and the bold text denotes the stop codon.

10.2 BMPER, Chordin and Tsg pHLsec Primers

BMPER

Number	Name	Forward	Reverse
1	S40 – G226		GCGGGTACCACCCAAACATTTGGGGCAGCACTG
2	S40 – G236		GCGGGTACCCCCAAAAGGGAGGTCAAACACTTTCCTC
3	S40 – L291		GCGGGTACCTAGGAGGCACTCTTCACAACAGCCCTC
4	S40 – V300	CGTAGCTGAAACCGGTTTCCTCCTT	GCGGGTACCTACTTTGATGTCTTCTGGGGGCACTCG
5	S40 – T358	CTTGACAGGTTCTGTTGC	GCGGGTACCCAGTGCAAATAGGACAGCATCCTTTTCTGTTG
6	S40 – L496		GCGGGTACCAAGACCACAGAGCTTGCCCTTGAG
7	S40 – H503		GCGGGTACCATGTCCATTGTAGTTGCCACAAAGACCAC
8	S40 – S530		GCGGGTACCGGACTCCACCCTCCAAGATTCAGC
9	S40 – E619		TGATGGTGGTGCTTGGTACCCTCCCAGTGGACTTTGATGCC
10			ACCTGGCTGGTCTCGGTACCCTCCCAGTGGACTTTGATGCC
11	S40 – R687		TGATGGTGGTGCTTGGTACCCCGCTGGGGACAAAGGACTG
12			ACCTGGCTGGTCTCGGTACCCCGCTGGGGACAAAGGACTG
13	V230 – L496	CGTAGCTGAAACCGGTGTGTTTGACCTCCCTTTTGGGAGC	GCGGGTACCAAGACCACAGAGCTTGCCCTTGAG
14	V230– H503		GCGGGTACCATGTCCATTGTAGTTGCCACAAAGACCAC
15	V230– S530		GCGGGTACCGGACTCCACCCTCCAAGATTCAGC
16	V230– E621		TGATGGTGGTGCTTGGTACCCTCCCAGTGGACTTTGATGCC
17			ACCTGGCTGGTCTCGGTACCCTCCCAGTGGACTTTGATGCC
18	V230– R687		TGATGGTGGTGCTTGGTACCCCGCTGGGGACAAAGGACTG
19			ACCTGGCTGGTCTCGGTACCCCGCTGGGGACAAAGGACTG
20	I298– L496	CGTAGCTGAAACCGGTATCAAAGTA	GCGGGTACCAAGACCACAGAGCTTGCCCTTGAG
21	I298– H503	TGCAAATTTGGCAACAAGATT	GCGGGTACCATGTCCATTGTAGTTGCCACAAAGACCAC
22	I298– S530		GCGGGTACCGGACTCCACCCTCCAAGATTCAGC
23	I298 – E621		TGATGGTGGTGCTTGGTACCCTCCCAGTGGACTTTGATGCC
24			ACCTGGCTGGTCTCGGTACCCTCCCAGTGGACTTTGATGCC
25	I298 – R687		TGATGGTGGTGCTTGGTACCCCGCTGGGGACAAAGGACTG
26			ACCTGGCTGGTCTCGGTACCCCGCTGGGGACAAAGGACTG
27	T358 – L496	CGTAGCTGAAACCGGTACTGAAAA	GCGGGTACCAAGACCACAGAGCTTGCCCTTGAG
28	T358– H503	GCCCGGCGTTTGC	GCGGGTACCATGTCCATTGTAGTTGCCACAAAGACCAC
29	T358– S530		GCGGGTACCGGACTCCACCCTCCAAGATTCAGC
30	T358– E619		TGATGGTGGTGCTTGGTACCCTCCCAGTGGACTTTGATGCC
31			ACCTGGCTGGTCTCGGTACCCTCCCAGTGGACTTTGATGCC
32	T358– R687		TGATGGTGGTGCTTGGTACCCCGCTGGGGACAAAGGACTG
33			ACCTGGCTGGTCTCGGTACCCCGCTGGGGACAAAGGACTG

Chordin

Number	Name	Forward	Reverse
34	A27 – A866	CGCACCGGTGCCGGCCCAGAGCCC	GCGGGTACCCAGCCTGCATGGGGTCCCC
35	A27 – S955	C	GCGGGTACCAGAGCCTTCGGCTTCTTTCTCCAGC
36	D154 – A866	CGCACCGGTGACCGCGGGGAGCCA	GCGGGTACCCAGCCTGCATGGGGTCCCC
37	D154 – S955	GG	GCGGGTACCAGAGCCTTCGGCTTCTTTCTCCAGC

Tsg

Number	Name	Forward	Reverse
38	C26 – F223	CGCACCGGTTGTAACAAAGCACTCTGTGCTAGTGATG	GCGGGTACCAAACATGCAGTTCATACATTTGACAG
39	N27 – F223	CGCACCGGTAACAAAGCACTCTGTGCTAGTGATGTGAG	

Supplemental Table 10-1 – Primers for pHLsec

Primers were used to generate 25 BMPER clones, 4 chordin clones and 2 Tsg clones for ligation into pHLsec vectors tagged with either His_6 or 1D4 tags resulting in 62 clones. Red sequence denotes In-Fusion tags and blue sequence denotes cloning sites.

10.3 **Sequence Alignment of BMPER and Chordin vWFC domains**


```
                  1          10               20                      30
sp|BMPER|238-289  ..CLFRS..DVYDNGSSFLYD........NCTACTCRD..........STVVCKR..KC.
sp|BMPER|299-358  KVCKFGN..KIFQDGEMWSSI........NCTICACVK..........GRTECRN..KQC
sp|Chordin|784-850 ..CYFDGDRSWRAAGTRWHPV.VPPFGLIKCAVCTCKG........GTGEVHCEK..VQC
sp|Chordin|872-932 ..CRFAG..QWFPESQSWHPS.VPPFGEMSCITCRCGA........G..VPHCER..DDC
sp|BMPER|108-163  ..CTYEG..NTYNSSFKWQSP.AEP.....CVLRQCQE..........GVVTESG..VRC
sp|Chordin|49-126  AGCTFGG..KVYALDETWHPDLGEPFGVMRCVLCACEAPQWGRRTRGPGRVSCKNIKPEC
sp|Chordin|703-763 NTCFFEG..QQRPHGARWAPNYDPL.....CSLCTCQR..........RTVICDP..VVC
sp|BMPER|50-105   ..CENEG..EVLQI...PFIT.DNP.....CIMCVCLN..........KEVTCKR..EKC
sp|BMPER|164-225  PGCVFEG..VQYQEGEEFQPE.GSK.....CTKCSCTG..........GRTQCVR..EVC
```

```
                  40          50
sp|BMPER|238-289  S....HPGGCD.QGQEGCCEEC.
sp|BMPER|299-358  IPISSCPQGKI.LNRKGCCPICT
sp|Chordin|784-850 PRL.ACAQPVR.VNPTDCCKQC.
sp|Chordin|872-932 SLPLSCGS....GKESRCCSRC.
sp|BMPER|108-163  .VV.HCKNPL..EHLGMCCPTC.
sp|Chordin|49-126  PTP.ACGQPR..QLPGHCCQTCP
sp|Chordin|703-763 PPP.SCPHPV..QAPDQCCPVCP
sp|BMPER|50-105   PVL.SRDCALAIKQRGACCEQC.
sp|BMPER|164-225  PIL.SCPQHLSHIPPGQCCPKCL
```

Supplemental Figure 10-2 – Sequence comparison of vWFC domain of BMPER and Chordin

vWFC domains of BMPER and chordin are aligned using Clustal (Larkin et al., 2007) and annotated with ESPript 3 (Robert and Gouet, 2014).

10.4 **Mutagenesis Conformation**
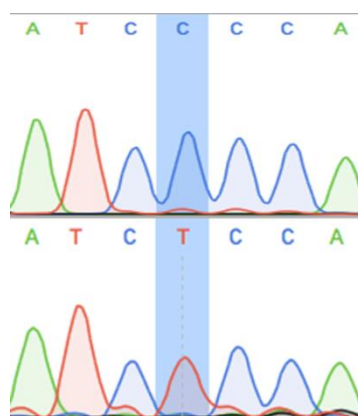
Supplemental Figure 10-3 – Mutagenesis Verification

Sequencing data providing evidence for the mutation of BMPER to BMPER-P370L.
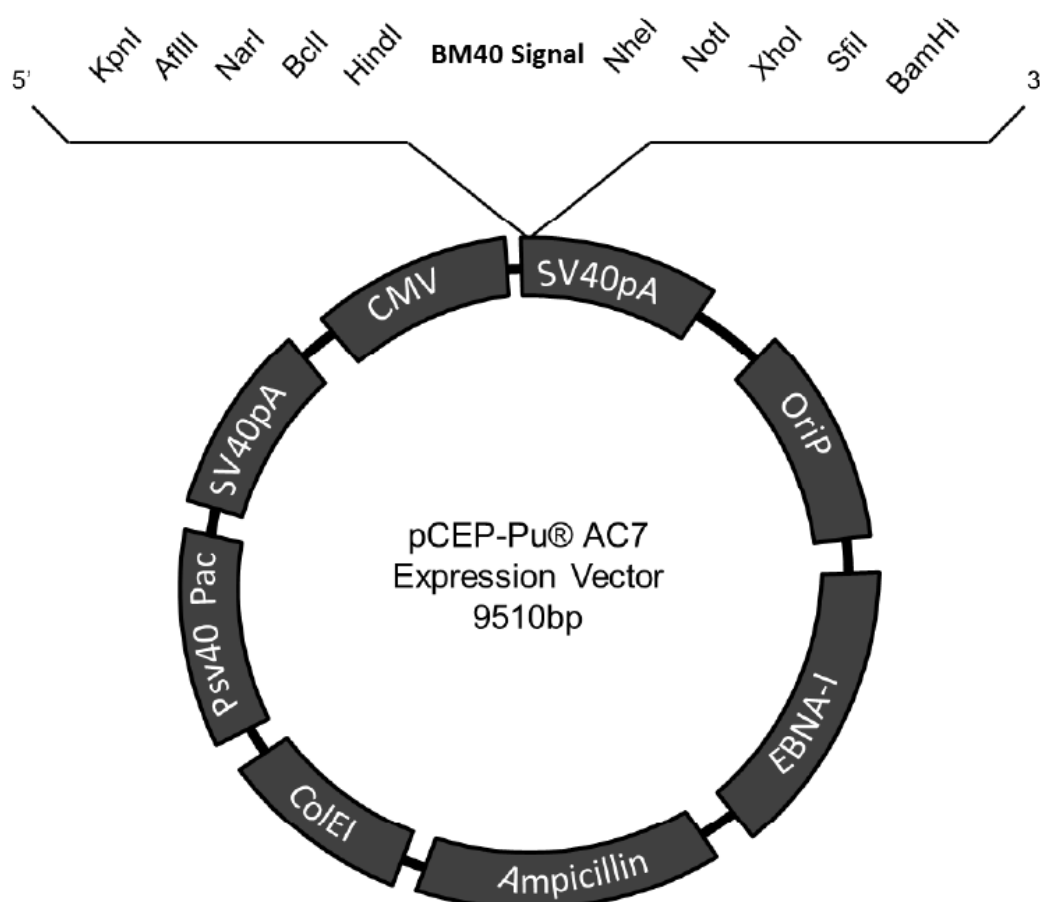
11 Appendix C

11.1 BMPER GeneArt Synthetic Gene

Supplemental Figure 11-1 – Synthetic BMPER Gene

A) BMPER construct was generated by GeneArt gene synthesis (Thermo Fischer) based on accession number NM_133468.4.

11.2 pCEP-Pu/AC7 Vector

- **OriP/EBNA-I:** Eukaryotic origin of replication induced by EBNA-I allowing for maintenance of this plasmid in stably-transfected cell lines
- **Ampicillin:** Ampicillin resistance open reading frame
- **ColE1:** Bacterial origin of replication
- **PSV40:** Promoter for Pac in eukaryotic cells
- **Pac:** Puromycin resistance open reading frame
- **SV40pA:** 3' polyadenylation signal
- **CMV:** Construct promoter sequence
- **BM40:** Signal peptide for construct secretion in eukaryotic cell lines

Supplemental Figure 11-2 – pCEP-Pu/AC7 Mammalian Expression Vector

The pCEP-Pu/Ac7 mammalian expression vector, allowing for expression and secretion of BMPER constructs in stably transfected HEK293-EBNA cells.
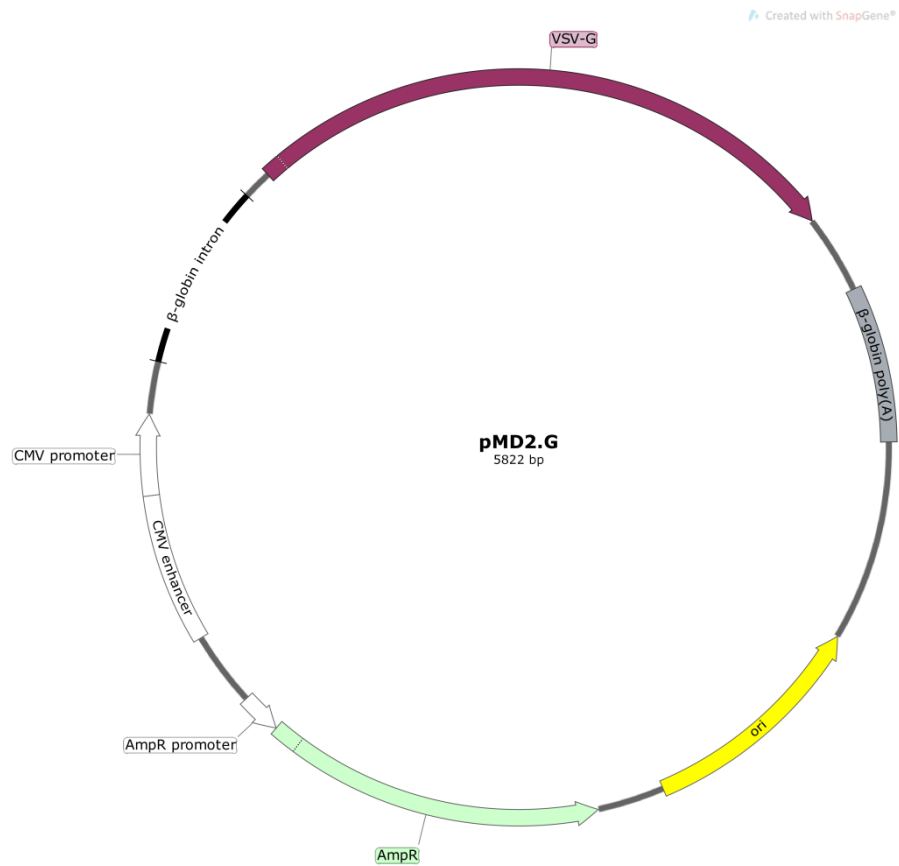
11.3 **Lentiviral Vectors**
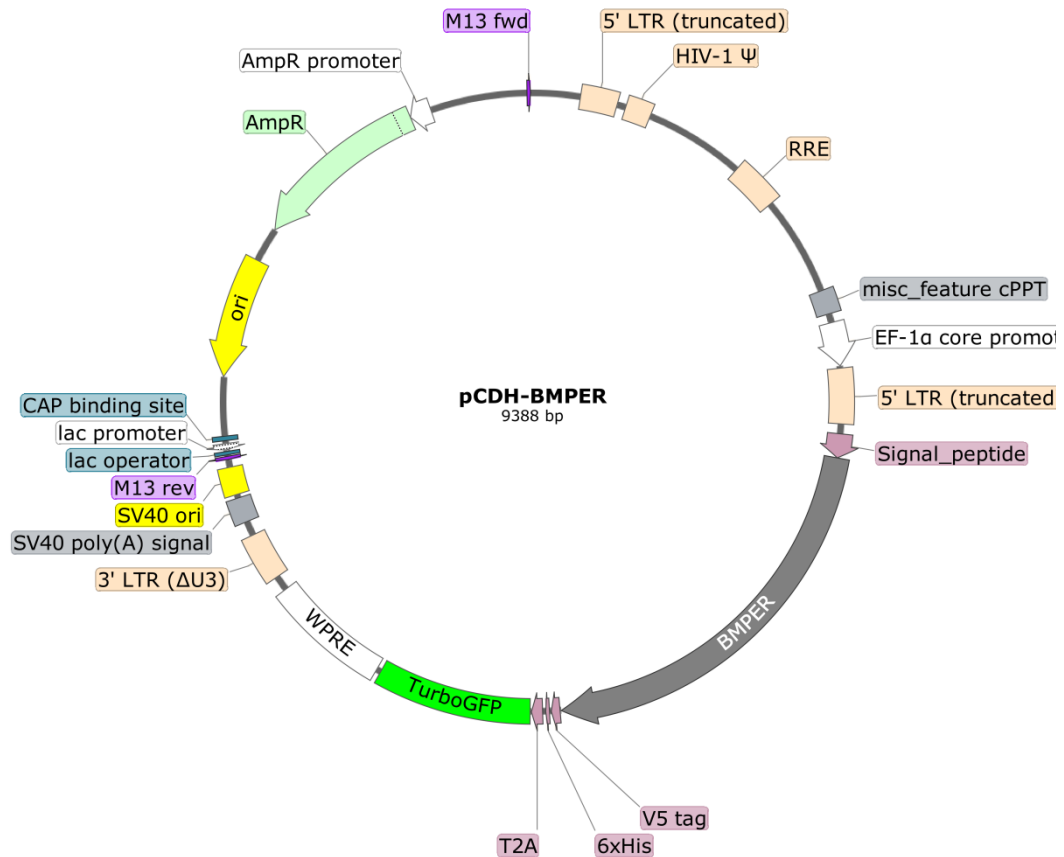
Supplemental Figure 11-3 - pCDH Lentiviral Expression Vector

The pCDH lentiviral mammalian expression vector allowing for expression and secretion of BMPER constructs when transduced into HEK293-T cells with pMD2.G and psPAX2 packaging vectors. The cells transduced with this vector can be selected for by sorting cells for GFP.

Supplemental Figure 11-4 – Doxycycline Inducible BMP-2 Expression Vector

The pCDH lentiviral mammalian expression vector allowing for expression and secretion of BMP-2, in the presence of doxycycline, when transduced into HEK293-T cells with pMD2.G and psPAX2 packaging vectors. The cells transduced with this vector can be selected for by sorting cells for blue fluorescent protein.

11.5 **Packaging Vectors**

Supplemental Figure 11-5 – pMD2.G Packaging Vector

Packaging vector used in conjunction with psPAX2 and pCDH vectors to form effecting lentiviral particles.
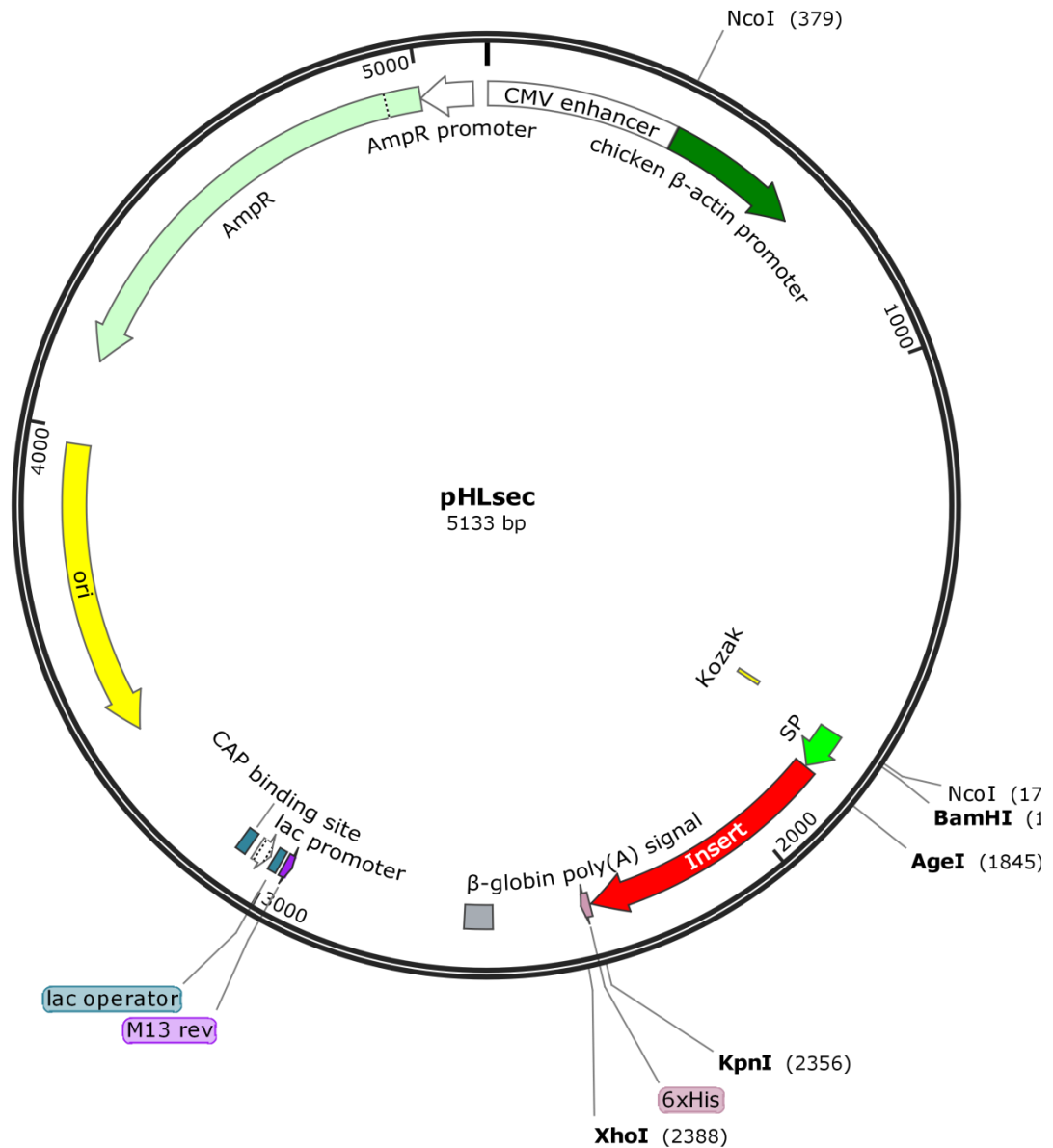
Supplemental Figure 11-6 – psPAX2 Packaging Vector

Packaging vector used in conjunction with pMD2.G and pCDH vectors to form effecting lentiviral particles.

11.6 **pHLsec Vector**

Supplemental Figure 11-7 – pHLsec Mammalian Expression Vector

The pHLsec mammalian expression vector, allowing for expression and secretion of BMPER constructs in transiently transfected HEK293-T cells.

12 Appendix D

12.1 Tsg Paper

Structural characterization of twisted gastrulation provides insights into opposing functions on the BMP signalling pathway

Helen Troilo[a], Anne L. Barrett[a], Alexandra V. Zuk[b], Michael P. Lockhart-Cairns[a,c], Alexander P. Wohl[b], Christopher P. Bayley[a], Rana Dajani[a], Richard B. Tunnicliffe[a], Lewis Green[a], Thomas A. Jowitt[a], Gerhard Sengle[b,d] and Clair Baldock[a]

a - *Wellcome Trust Centre for Cell-Matrix Research,* Faculty of Life Sciences, University of Manchester, M139PT, UK
b - *Center for Biochemistry,* Medical Faculty, University of Cologne, Cologne, Germany
c - *Beamline B21,* Diamond Light Source, Harwell Science & Innovation Campus, Didcot, Oxfordshire, UK
d - *Center for Molecular Medicine Cologne (CMMC),* University of Cologne, Cologne, Germany

Correspondence to Clair Baldock: Wellcome Trust Centre for Cell-Matrix Research, Faculty of Life Sciences, Michael Smith Building, University of Manchester, M139PT, UK. clair.baldock@manchester.ac.uk
http://dx.doi.org/10.1016/j.matbio.2016.01.019
Edited by R. Iozzo

Abstract

Twisted gastrulation (Tsg) and chordin are secreted glycoproteins that function together as BMP (bone morphogenetic protein) antagonists to regulate BMP growth factor signalling. Chordin binds to BMPs, preventing them from interacting with their receptors and Tsg is known to strengthen this inhibitory complex. Tsg also acts as a BMP agonist by promoting cleavage of chordin by tolloid-family proteinases. Here we explore the structural mechanism through which Tsg exerts this dual activity. We have characterized the nanoscale structure of human Tsg using in-solution biomolecular analysis and show that Tsg is a globular monomer with a flattened cross shape. Tsg has a high proportion of N-linked glycans, in relation to its molecular weight, which supports a role in solubilising BMPs. Tsg binds with high affinity to the C-terminal region of chordin and was also able to inhibit BMP-7 signalling directly but did not have an effect on BMP-4 signalling. Although both Tsg and mammalian tolloid are involved in chordin cleavage, no interaction could be detected between them using surface plasmon resonance. Together these data suggest that Tsg functions as a BMP-agonist by inducing conformational change in chordin making it more susceptible to tolloid cleavage and as a BMP-antagonist either independently or *via* a chordin-mediated mechanism. Following single cleavage of chordin by tolloids, Tsg continues to strengthen the inhibitory complex, supporting a role for partially cleaved chordin in BMP regulation.

Introduction

Bone morphogenetic proteins (BMPs) are secreted growth factors of the transforming growth factor-β family. They are essential for early embryonic patterning, most notably in forming the morphogen gradient controlling dorsal ventral patterning [1]. They also have roles in adult tissues including in muscle maintenance [2] and fracture healing [3]. During dorsoventral gradient formation, chordin is the primary BMP regulator in many organisms including *Xenopus* [4] and zebrafish [5]. Chordin antagonizes signalling by binding to BMPs and preventing them from interacting with their cell surface receptors. This regulation is modulated by the small secreted glycoprotein twisted gastrulation (Tsg) which can form a ternary complex with chordin and BMP [6–8]. Tsg increases the effectiveness of chordin as a BMP antagonist [6], either by facilitating the interaction between chordin and BMP or by blocking additional BMP receptor binding sites. It is also thought that Tsg stabilizes BMPs in the extracellular space and that N-linked glycosylation of Tsg has an important role in facilitating its interaction with BMPs [9].

BMP inhibition is relieved by cleavage of chordin by tolloid family metalloproteinases [10], removing a

BMP-binding von Willebrand Factor C (vWC) homology domain from each terminus [11]. The individual vWC domains have greatly reduced bioactivity *in vivo* and this cleavage has been shown to be an essential regulatory mechanism in patterning. Cleavage at each site is an independent event and fragments lacking both the N- and C-termini have been shown to exist *in vivo* at significant levels [12]. Partially cleaved chordin has been shown to be as effective as full length chordin at BMP inhibition with some gain of function when cleavage occurs at the C-terminal site [13]. Tsg increases the rate of chordin cleavage by tolloids [6] and competes with the residual fragments for BMP binding [14], as well as increasing their rate of degradation *in vivo* [15]. Since Tsg also strengthens the chordin-BMP inhibitory complex, it therefore has a dual effect on BMP signalling [15]. In most species the net effect is that Tsg functions as a BMP agonist but there are exceptions to this, which are thought to be based on endogenous tolloid concentration [16].

The effects of Tsg are well documented but less is known about its mechanism of interaction with tolloids and chordin. One possibility is that Tsg functions as a scaffolding molecule between tolloid and chordin (in a similar manner to Olfactomedin1 in *Xenopus* [17]) and that it increases the cleavage rate in this way. However data suggest that its mechanism of action may involve inducing conformational change in chordin. For example, mouse chordin acquires an additional tolloid cleavage site in the presence of Tsg [6], suggesting that a cryptic site may be exposed following Tsg binding. This would also explain why Tsg has not been shown to exert a broader influence on tolloid cleavage of other substrates in the matrix.

Tsg has previously been predicted based on sequence analysis to be a two-domain protein and is suggested to consist of an N-terminal cysteine-rich domain homologous to a vWC domain [14] and a C-terminal domain separated by a putative hinge region [18] (Fig. 1A). However, there is currently no structural evidence to support this prediction and the number and arrangement of cysteine residues in Tsg are not consistent with a vWC domain fold. This paper investigates the oligomeric state and shape of human Tsg using small-angle X-ray scattering (SAXS), multi-angle light scattering (MALS) and
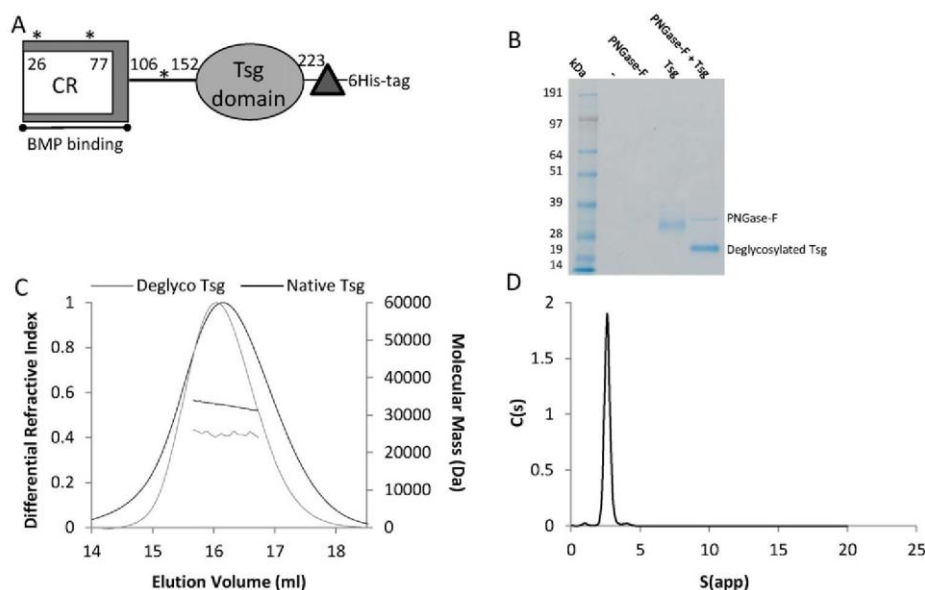

Fig. 1. Glycosylation and oligomeric state of Tsg. (A) Schematic diagram of Tsg. Tsg has previously been predicted to be composed of two domains, the first having homology to the cysteine-rich (CR) von Willebrand C-type domains in chordin and binds to BMPs [14] whereas the second (Tsg-specific domain of as yet unknown function) also contacts a high number of cysteine residues but has no known homologues. The domains are separated by a putative hinge region and both domains appear to be required for chordin binding [38]. The construct contains a 6-His-Tag for purification and a thrombin cleavage site for tag removal (triangle). Potential N-linked glycosylation sites indicated by asterisks. (B) Coomassie stained SDS-PAGE gel showing that purified Tsg can be deglycosylated under non-denaturing conditions. (C) MALS profile of native and deglycosylated Tsg. Refractive index and calculated molecular weight against elution volume from the size exclusion column (ml) showing the shift in elution volume and size following deglycosylation. Native Tsg (black line) has a molecular mass of 32.3 kDa (±0.13%), ~30% larger than deglycosylated Tsg (grey line) (25.0 kDa (±0.5%)) experimental errors from polydispersity. In both states the mass of Tsg is consistent with a monomer in solution (theoretical mass based on peptide sequence is 23.8 kDa). (D) Velocity AUC data for native Tsg showing a single species with mass ~ 36.7 kDa, $S_{20}W$ of 2.57, hydrodynamic radius of 3.44 nm and frictional ratio of 1.27.
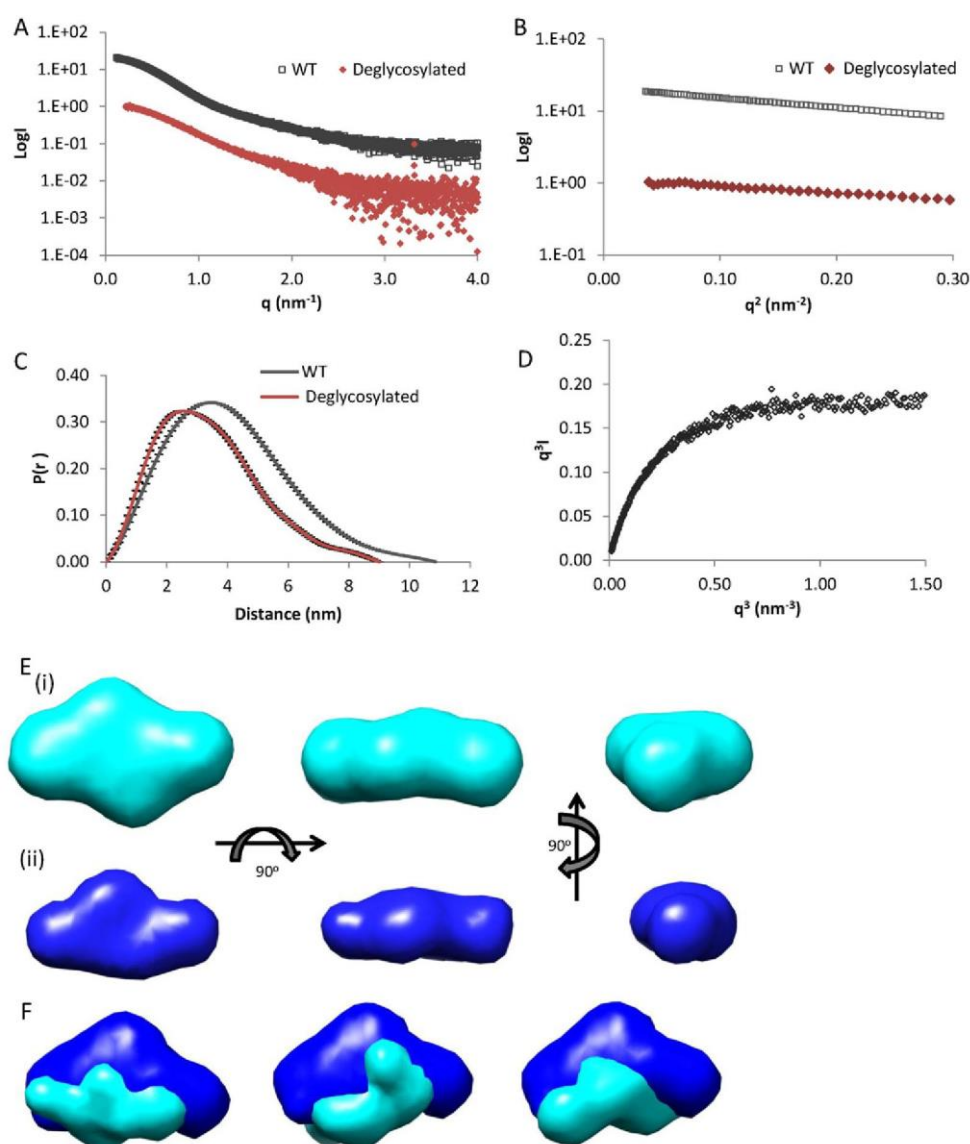
Fig. 2. SAXS data on native and deglycosylated Tsg. (A) The experimental SAXS data for the native (black) and deglycosylated (brown) Tsg are plotted as a function of q. (B) The low-angle region of the X-ray scattering data is show in the form of a Guinier plot, which is linear for values $q \leq 1/R_g$. From the Guinier plot, the Rg can be estimated as 3.1 nm for native and 2.5 nm for deglycosylated Tsg. (C) The distance distribution functions for native (black) and deglycosylated (brown) Tsg are shown. The maximum dimensions calculated with Gnom are 11 nm (native) and 9 nm (deglycosylated). (D) The q^3 plot has a linear plateau which indicates some flexibility in Tsg [22]. (E) Shapes were simulated *ab initio* by the programme DAMMIN [21]. Models for native (cyan (i)) and deglycosylated Tsg (blue (ii)) are each shown in 3 orthogonal views. (F) Modelling of the protein and glycan density was performed with MONSA [21]. The protein density was aligned and averaged and shown in blue, the additional density from the glycans is shown in cyan. Three representative models are shown.

analytical ultracentrifugation (AUC). In order to better understand the mechanism of Tsg in mediating interactions between chordin and BMPs, we used surface plasmon resonance (SPR) to determine binding locations. We also used SPR to test for direct interaction between Tsg and mammalian tolloid (mTld). To determine the extent of any direct anti-BMP-4 or -7 activity by Tsg, we used BMP activity assays including inhibition of alkaline phosphatase production, SMAD phosphorylation and
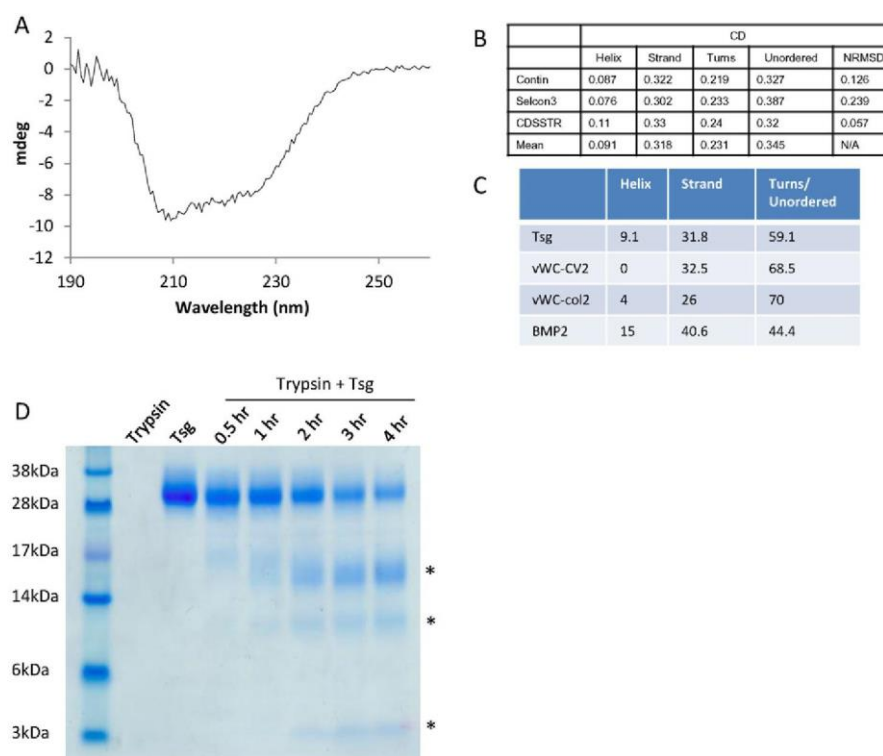
Fig. 3. Circular dichroism and secondary structure of Tsg. (A) Circular dichroism spectrum for 0.4 mg ml^{-1} native Tsg showing mdeg against wavelength. The graph shape is characteristic of a mix of α-helix and β-sheet secondary structure. (B) Table summarizing analysis of CD data for Tsg using the Contin, Selcon3, CDSSTR and K2d algorithms. Mean values exclude K2d owing to high normalized root-mean-square deviation (NRMSD) values. (C) Comparison of predicted secondary structure values for Tsg with secondary structure of known vWC domain structures and cysteine knot fold. (D) Limited proteolysis of Tsg with trypsin (1:50 *w*/w). After 30 min cleavage of Tsg was observed, leading to the appearance of 3 bands (marked by asterisks).

BMP-induced gene expression. Moreover we show that Tsg interacts with partially cleaved chordin to increase BMP inhibition and promote further cleavage by tolloids. These findings support the physiological relevance of partially cleaved chordin as an important BMP regulator [19]. Our results point to a monomeric Tsg interacting directly with chordin or BMP and suggest that Tsg has a type-specific regulation of BMPs enabling finely controlled regulation of BMP signalling in conjunction with other regulators.

Results

Expression and oligomeric state of twisted gastrulation

Human Tsg was expressed in a mammalian expression system and purified as a secreted protein (Fig. 1B). Tsg is frequently assumed to form a homo-dimer as immunoprecipitation studies have suggested that two molecules of Tsg bind to BMP-4 [14,15]. Therefore in order to determine the size and

oligomeric status of Tsg, multi-angle light scattering (MALS) was used in conjunction with size exclusion chromatography. The MALS profile (Fig. 1C) showed a single species that had a molecular mass of 32.3 kDa ±0.13%, which is larger than the predicted mass of a monomer (~23.8 kDa). However human Tsg has three predicted N-linked glycosylation sites and the glycosylation patterns of Tsg have previously been shown to have functional importance in BMP-binding [9]. Fig. 1B shows that Tsg can be deglycosylated by PNGase-F treatment in non-denaturing conditions. The difference in mobility on SDS-PAGE (*i.e.* molecular mass) following removal of the N-linked sugars appears to be considerable, especially in relation to the size of the protein. Therefore MALS was used to quantify the mass of the protein following deglycosylation. Fig. 1C shows the comparison of molecular mass for deglycosylated and native Tsg of 25.0 kDa and 32.3 kDa, respectively. The MALS data also show that in either the presence or absence of N-linked glycosylation Tsg is monomeric. We confirmed this finding by velocity sedimentation analytical ultracentrifugation (AUC) measurements showing a single species for native
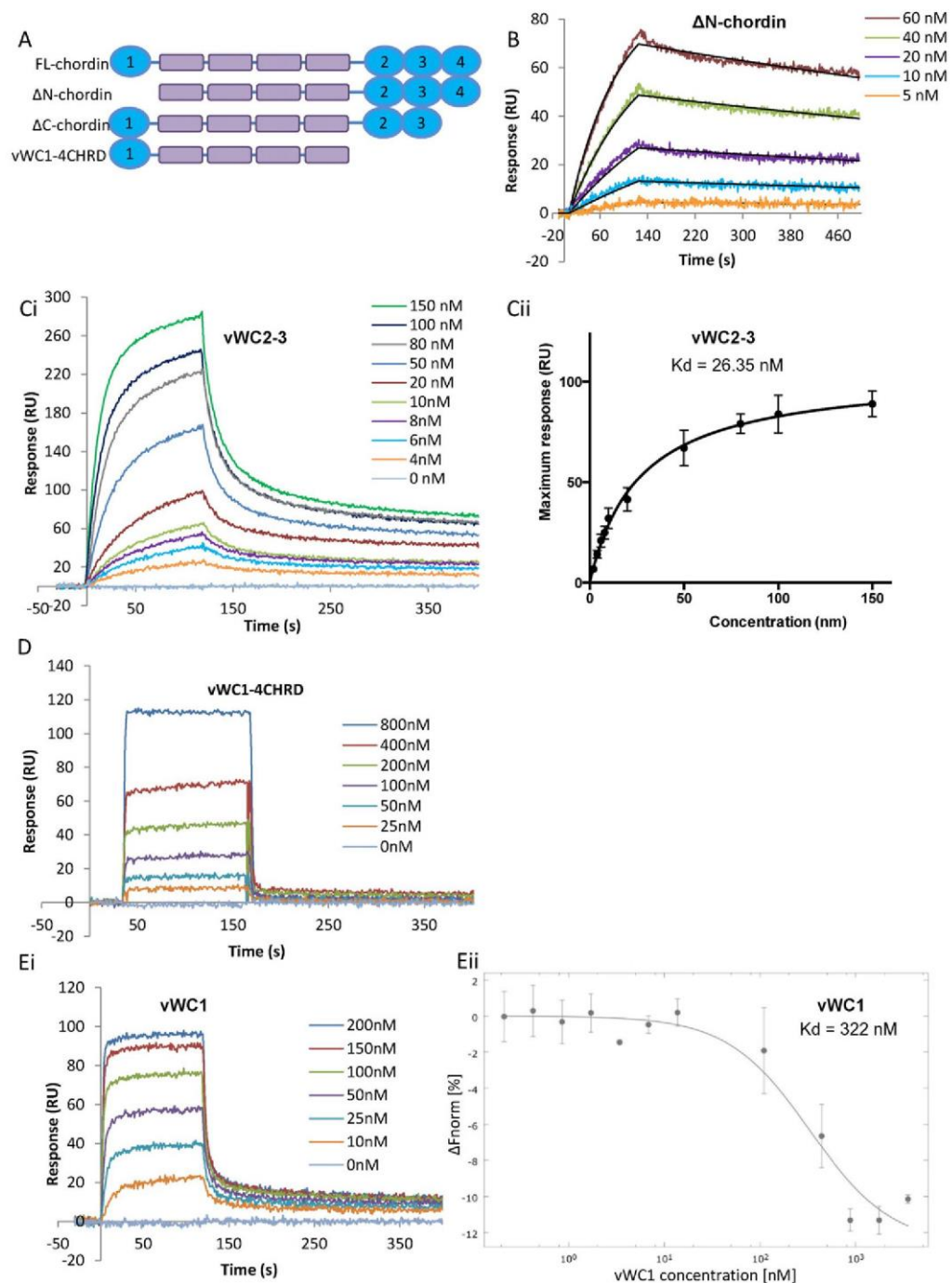
Fig. 4. SPR sensorgrams showing Tsg binding to chordin constructs (A) Schematic diagram of chordin constructs used in this analysis. Tsg was immobilised on a Proteon GLC chip *via* amine coupling. ΔN-chordin and vWC2–3were injected on to Tsg (300 RUs immobilised) at increasing concentrations. (B) ΔN-chordin fitted 1:1 Langmuir binding model shown in black. (C) The binding affinity for vWC2–3 was determined using equilibrium analysis. (D) vWC1-4CHRD did not bind to Tsg and vWC1 bound weakly in a highly biphasic manner (Ei) (for (D) and (Ei) 800 RUs of Tsg were immobilised). MST was performed to detect in solution binding between Tsg and vWC1. Tsg was labelled and added to increasing concentrations of vWC1, a K_D of 322 nM was determined (Eii). All experiments were performed in triplicate.

Tsg consistent with a glycosylated monomer of 36.7 kDa (Fig. 1D). The hydrodynamic radius of Tsg from MALS was 3.4 nm which is in agreement with the AUC data (3.44 nm) shown in Fig. 1D. The $S_{20}W$ was 2.57 S and the frictional ratio 1.27, indicative of a globular protein.

Nanostructure of Tsg

The 3D structures of both deglycosylated and native Tsg in solution were investigated using small-angle X-ray scattering (SAXS) (Fig. 2A). SAXS measurements were made at both Diamond Light Source (I22) and PETRAIII (P12) synchrotron radiation sources. The SAXS data quality was assessed using Guinier plots, to check for aggregation in the sample (Fig. 2B). The radii of gyration (Rg) obtained from the Guinier approximation were 2.5 nm and 3.1 nm for the deglycosylated and native Tsg, respectively. The radius of gyration is consistent with the hydrodynamic radius (3.4 nm from MALS and AUC for native Tsg) which supports the finding that Tsg is monomeric in solution even at high concentrations used for SAXS data collection (6 mg/ml). The maximum particle dimension for native Tsg was estimated as 11 nm using indirect Fourier transform with GNOM [20] (Fig. 2C) corresponding to the longest end-to-end distance in the protein. The maximum dimension for deglycosylated Tsg was lower (9 nm) than for native Tsg as a result of the larger hydration shell attracted by sugar groups. For native and deglycosylated Tsg, *ab initio* models were generated using DAMMIN [21]. The modelling allowed us to fit the experimental data with mean discrepancy factors (*x*) of 0.93 (native) and 1.09 (deglycosylated) and at least 20 simulations for each dataset were completed to determine the common structural features. The mean normalized spatial discrepancy factor between the solutions was 0.52 (native) and 0.60 (deglycosylated) indicating unique solutions. Fig. 2E shows a flattened cross shape for Tsg with dimensions 11 nm × 4.2 nm × 6.3 nm. Two distinct lobes of the protein can be distinguished in the model but the structure is relatively compact and appears to lack a distinct linker that was predicted to exist between domains [18]. However, the q^3 plot (Fig. 2D) suggests that there is flexibility in Tsg which may indicate that there is movement between the two lobes [22]. The superimposed *ab initio* models of deglycosylated and native Tsg show that the two models have a similar shape with additional density in the native structure close to the centre of the molecule where the sugars may be located (Fig. 2F).

Secondary structure and domain analysis

The structure of Tsg is currently unknown but it has been suggested that the cysteine-rich region has homology to a von Willebrand Factor-C domain [14].

To determine whether the secondary structure composition of Tsg is similar to a vWC domain, circular dichroism spectra were collected on native Tsg (Fig. 3A). The CD profile has a characteristic double dip with a deeper trough at 208 nm typical of α/β secondary structure. Deconvolution of the spectra predicted ~9% α-helix and ~32% β-sheet shown in Fig. 3B which is compared to the secondary structure content of vWC domains from CV-2 [23] and collagen IIa [24] as well as a cysteine knot protein (BMP-2) [23]. The CD data suggests that Tsg has a greater α-helical content than a single vWC domain and less coil/unordered regions which indicates Tsg may have more extensive secondary structural elements than a vWC domain (Fig. 3C). In order to determine whether Tsg has two domains separated by a flexible linker, limited proteolysis with trypsin was performed. After 30 min, some Tsg had been cleaved and could be separated into three prominent bands on SDS-PAGE. These species have molecular weights of approximately 16, 10 and 3 kDa (Fig. 3D). These sizes are consistent with cleavage at two potential trypsin cleavage sites in the putative linker. These data suggest that this linker sequence is accessible to trypsin and there are two main species following cleavage. However, only a proportion of Tsg is initially cleaved which indicates that there is some protection of this region in the folded protein.

A high-affinity Tsg binding site is localized towards the C-terminus of chordin

A previous study had shown that individual chordin domains bound Tsg with much lower affinity than the full length chordin molecule and suggested that vWC1, 3 and 4 all bound Tsg weakly [25]. We screened different fragments of chordin for Tsg binding using surface plasmon resonance (SPR) to determine which domains underpin higher affinity binding. Chordin constructs containing the C-terminal vWC domains (Fig. 4A) were able to bind to immobilised Tsg with high affinity (Fig. 4B and C; ΔN-chordin (K_D = 3.08 nM) and vWC2–3 (K_D = 26.35 nM)). The interaction between ΔN-chordin and Tsg fits a simple 1:1 Langmuir binding model, however the interaction between Tsg and vWC2–3 deviates from this fitting, especially at higher analyte concentrations therefore the dissociation constant was derived from equilibrium analysis (Fig. 3 Cii). The central region of chordin (vWC1-4CHRD region) does not bind to Tsg under these experimental conditions (Fig. 4D). The N-terminal BMP-binding domain (vWC1) bound in a highly biphasic manner, with large amounts binding but a very rapid dissociation (Fig. 4 Ei). It appears that there is an initial vWC1–Tsg interaction then a secondary interaction, which could be vWC1 self-association or binding of vWC1 to a second site on Tsg. This interaction was also analysed in solution with microscale
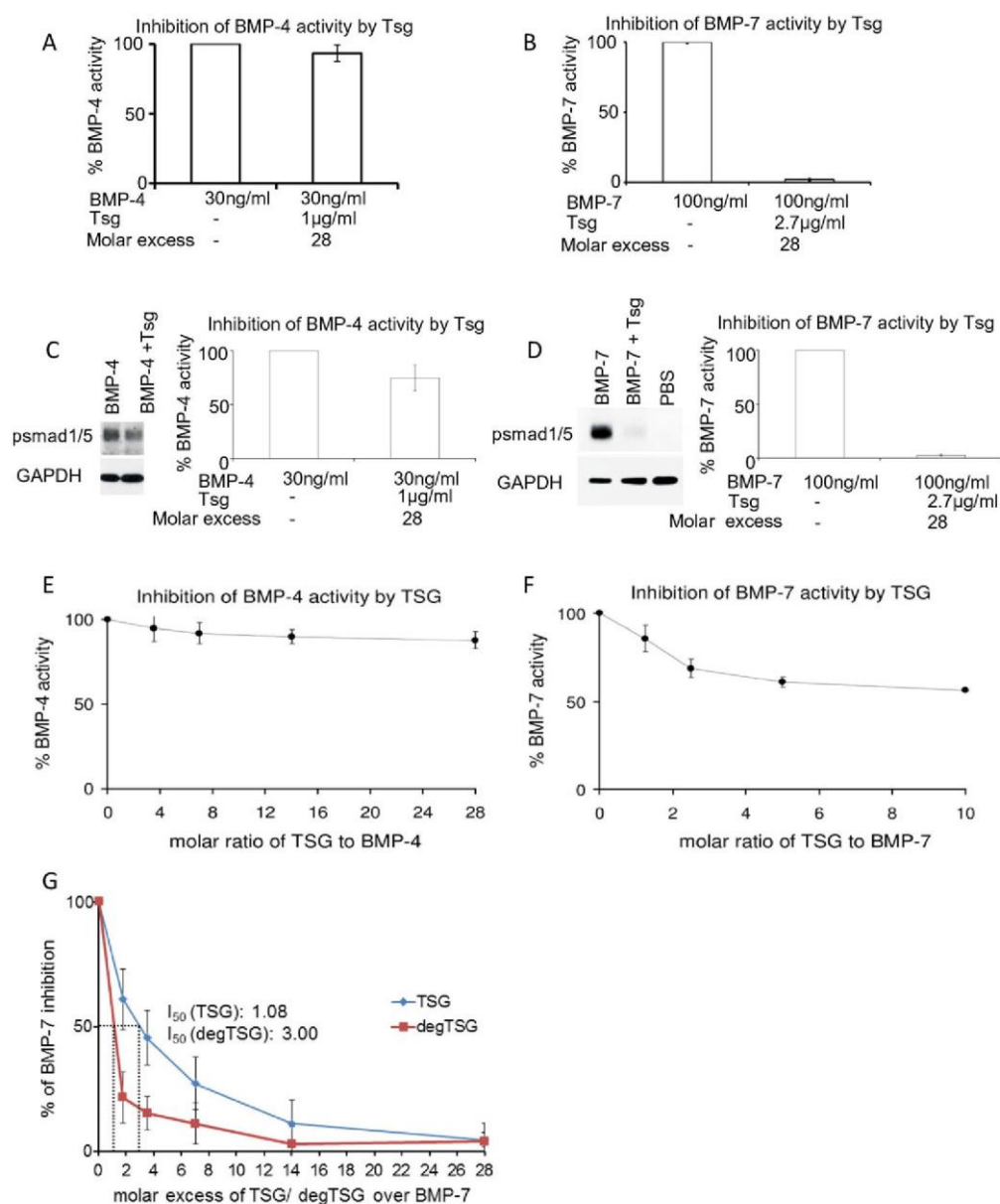
Fig. 5. Bioactivity of Tsg for BMP-4 and-7. Inhibition of BMP-4 and BMP-7 growth factor activity by Tsg was detected by inhibition of alkaline phosphatase production in C2C12 cells (A and B) and Smad-1/−5 phosphorylation in HEK293T cells (C and D) with a 28× molar excess of Tsg. A rapid qPCR assay based on the *Id3* expression in C2C12 cells after 5 h of incubation with growth factor was also used. A range of Tsg concentrations were tested with BMP-4 (E) and BMP-7 growth factor (F). (G) Natively deglycosylated Tsg was still able to inhibit BMP-7 as detected by inhibition of alkaline phosphatase production. Errors represent standard deviation from at least three independent experiments.

thermophoresis (MST) and a K_D of 322 nM determined (Fig. 4 Eii). Together these data suggest that high-affinity binding of Tsg to chordin occurs *via* the C-terminal vWC2–3 domains consistent with immunoprecipitation data from Larrain and co-workers [15], a second weaker interaction site is present on vWC1

consistent with previous binding studies [25,26] which appears more accessible following cleavage from the 4CHRD region. Natively deglycosylated Tsg still bound to full-length chordin with high affinity so the N-linked glycans do not appear to play a part in this interaction (Supplementary Fig. 1).
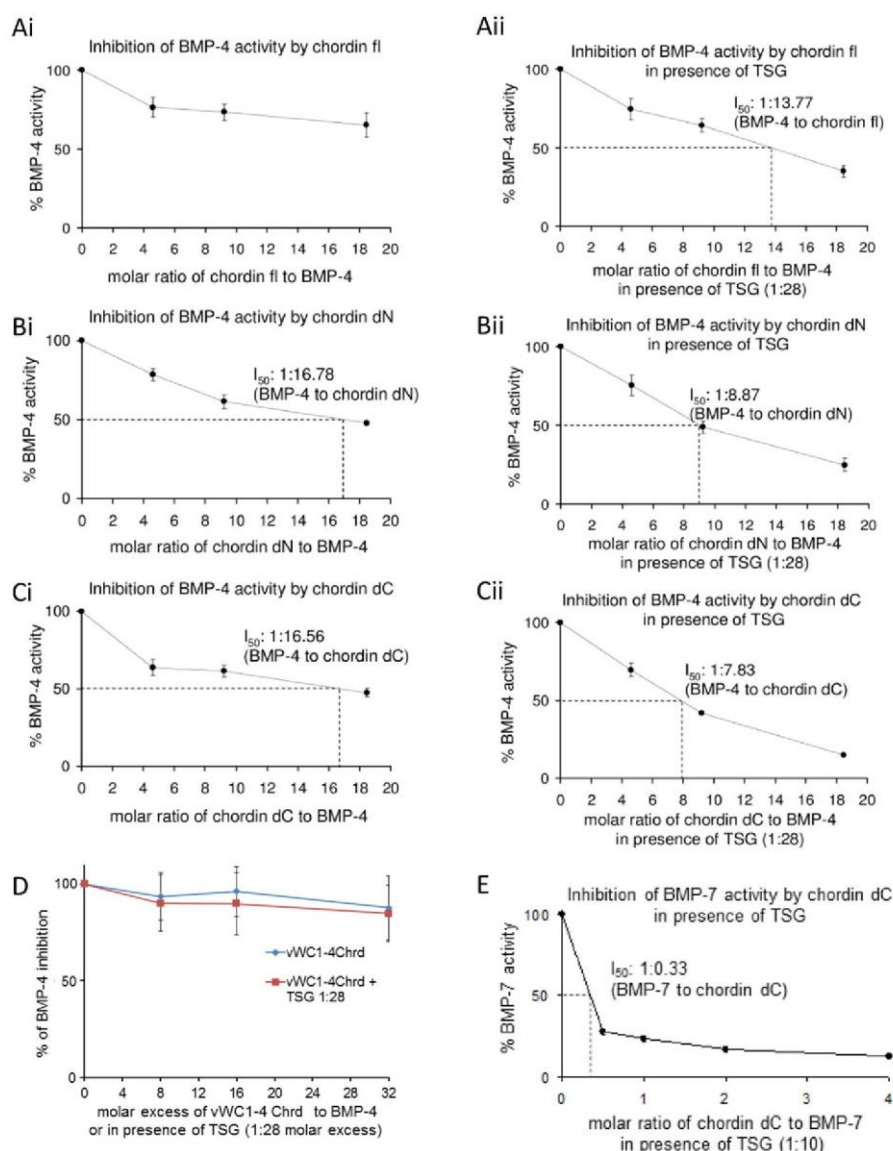
Fig. 6. Bioactivity of chordin and chordin fragments enhanced by Tsg (A) Inhibition of BMP-4 growth factor activity by chordin was detected by *Id3* reporter assay in the presence or absence of Tsg. Tsg increases inhibition of BMP-4 by both full length chordin and the chordin cleavage fragments (B) ΔN- and (C) ΔC-chordin. (D) vWC1-4CHRD only weakly inhibits BMP-4 and this is not enhanced by Tsg. (E) BMP-7 activity was inhibited by increasing amounts of ΔC-chordin in the presence of Tsg. Errors represent standard deviation from three independent experiments.

Tsg inhibition of BMP signalling

In order to investigate whether Tsg had an effect on BMP signalling in the absence of chordin, BMP-inhibition assays were performed. The inhibition of BMP-induced alkaline phosphatase production (Fig. 5A and B) and phosphorylation of Smad-1/–5 (Fig. 5C and D) by an excess of Tsg was analysed. Tsg was an ineffective BMP-4 inhibitor but showed complete inhibition of BMP-7 signalling. Using a qPCR based activity assay to determine the effect on BMP target gene expression, a range of Tsg:BMP molar ratios were screened and showed that little additional BMP inhibition was observed when increasing the molar ratio of Tsg:BMP-7 above 2 monomers:1 dimer (Fig. 5F). Native deglycosylation did not affect binding between BMP-7 and Tsg (Supplementary Fig. 1) and deglycosylated Tsg was a more effective inhibitor of BMP-7 (Fig. 5G).
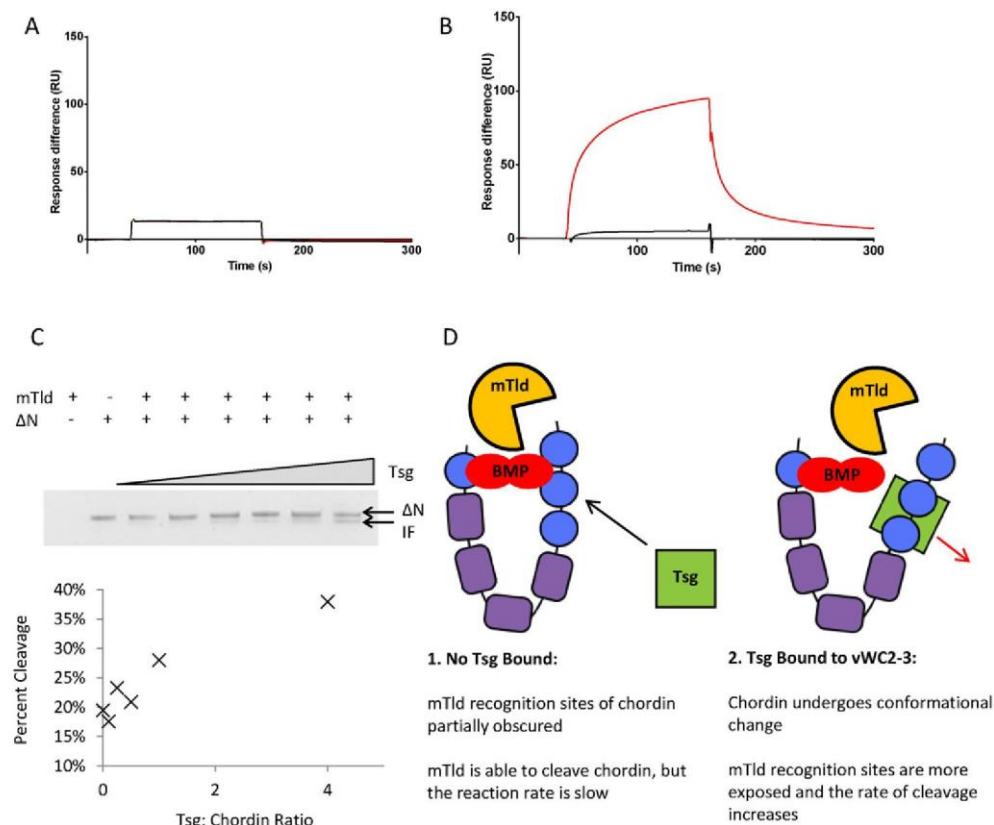
Fig. 7. SPR sensorgrams showing mTlD does not bind directly to Tsg. Immobilised proteins were cross-linked to a CM5 chip *via* amine coupling. 830 RU of Tsg was immobilised on the chip and two analytes were injected onto the immobilised Tsg: 500 nM mTlD (black response curve) and 500 nM mTlDC4C5 (red response curve). (B) Two analytes were injected onto the immobilised mTlD (2400 RU): 500 nM Tsg (black response curve) and 500 nM ΔN-chordin (red response curve). (C) ΔN-chordin cleavage assay by mTld in the presence of increasing amounts of Tsg (0:1, 0.1:1 0.25:1, 0.5:1, 1:1 and 4:1 molar ratio of Tsg:ΔN). The emergence of the chordin cleavage product, intermediate fragment (IF) is labelled. A plot of the percentage of cleaved chordin against Tsg:chordin ratio is shown. (D) Schematic diagram demonstrating proposed mechanism of conformational change in chordin following Tsg binding exposing tolloid recognition sites.

Tsg interaction with partially cleaved chordin

Tsg increases the inhibitory capacity of full length chordin, but competes with fully cleaved chordin for BMP binding. What is not clear is how Tsg interacts with partially cleaved chordin fragments; therefore as BMP-4 was not inhibited directly by Tsg, the effect of Tsg on BMP-4 inhibition by chordin was assessed. Fig. 6A shows that as expected Tsg increases the BMP-4 inhibitory effect of FL-chordin. Fig. 6B and C show that Tsg also enhances inhibition by ΔN- and ΔC-chordin, respectively. The observed effect must be mediated through chordin as Tsg did not inhibit BMP-4 directly (Fig. 5) and shows that Tsg increases the capacity for BMP inhibition of full length chordin and the larger cleavage fragments. In contrast, the shorter vWC1-4CHRD fragment only inhibited BMP-4 signalling weakly and Tsg was ineffective at

enhancing inhibition (Fig. 6D) which is consistent with the weaker binding seen between Tsg and vWC1. BMP-7 activity was inhibited effectively by increasing amounts ΔC-chordin with a 1:10 ratio of Tsg where only small amounts of ΔC-chordin were required to completely inhibit BMP-7 (Fig. 6E).

Tsg does not interact directly with mTld

Tsg increases the rate of chordin cleavage by tolloids but the mechanism for this is not clear. Given that this effect is also observed *in vitro* when Tsg, mTld and chordin are isolated from other pathway components, the two possibilities are that either Tsg binds to tolloid and chordin simultaneously acting as a scaffolding protein or Tsg induces a conformational change in chordin making it more accessible to tolloid cleavage. To investigate the former, we determined

whether there was an interaction between Tsg and mTld using SPR (Fig. 7A). The interaction of mTld with Tsg as both ligand and analyte was screened, however no binding was detected in either configuration (Fig. 7A and B). In addition, the C-terminal domains of mTld were also screened for binding as these domains are known to have the highest affinity interaction with tolloid substrates [27] but these domains showed no interaction with Tsg either (Fig. 7A) whereas on the same chip △N-chordin bound to mTld (Fig. 7B). This suggests that Tsg does not act as a scaffold between chordin and mTld and provides further support for a conformational change in chordin following Tsg binding. In order to demonstrate that Tsg is able to enhance mTld cleavage of chordin in the absence of other factors _e.g._ BMPs, △N-chordin was incubated with mTld in the presence of increasing concentrations of Tsg showing a dose dependent increase in rate of cleavage in the presence of Tsg (Fig. 7C). Fig. 7D shows a schematic diagram of the proposed interaction.

Discussion

Using small angle X-ray scattering we show the nanoscale structure of Tsg, an elongated globular-like molecule and is not, as previously postulated a dimer [14,15] but rather a stable monomer in solution. Although, it is possible that two monomers of Tsg may interact with a BMP dimer as cross-linking studies have detected a 2:2 complex of _Xenopus_ Tsg-BMP-4 [15]. A relatively high proportion of the molecular weight of Tsg is contributed by N-linked glycans. This property is likely to make Tsg highly soluble and stable in the matrix, supporting the hypothesis that it plays a role in stabilizing BMP and facilitates diffusion of Tsg containing complexes. In agreement with that is a study showing that the glycosylation patterns in human Tsg and its homologues play a key role in BMP binding specificity [9]. Our structure shows that the sugars are surface-accessible and located between the two lobes of the protein. The structure of Tsg has an elongated flat face where glycosylation adds a bulge to this surface and could provide multiple binding sites for BMPs and thereby supporting its role as a BMP solubilising factor [28].

Our SPR solid phase binding experiments showed Tsg bound with high affinity to the C-terminal vWC2–3 region of chordin and this binding was unaffected by natively deglycosylating Tsg. Tsg bound weakly to the vWC1 domain alone but not in the presence of the adjacent 4CHRD region which is consistent with previous studies where mouse Tsg has been shown by immunoprecipitation to bind to the vWC1 domain of mouse chordin [6] and weakly by SPR ($K_D = 0.2$–0.9 μM) [25,26]. We were unable to detect any direct binding using SPR between mammalian tolloid or

tolloid domains and Tsg, despite both binding to chordin and Tsg enhancing tolloid cleavage of chordin in the absence of other factors. This is consistent with other studies that did not detect a direct interaction by immunoprecipitation [6]. If Tsg were acting as a scaffold between chordin and tolloid, any interaction would be expected to be strong and relatively stable. This finding, combined with the ability of Tsg to introduce new cleavage sites in some species [6], is a strong indicator that the mechanism for enhancing chordin cleavage by tolloids is through inducing a conformational change in chordin. This could subsequently enhance tolloid cleavage of chordin perhaps by making cleavage sites more readily available or presenting a cryptic cleavage site. This mechanism seems more efficient because it excludes a scenario in which excess Tsg remains bound to tolloid while it processes other targets unrelated to the chordin-BMP pathway.

Tsg acts to strengthen the inhibitory complex between BMP and chordin but we also found that in the absence of chordin, Tsg has significant BMP-7 inhibitory capacity. However this was not the case for BMP-4. It has previously been demonstrated that Tsg has a weak inhibition of BMP-2, and could inhibit the BMP-2/BMPR-IA interaction but not BMP-2/ActR-IIB binding [25]. Together these findings raise the possibility of differential Tsg interaction and mechanism depending on the specific BMP type. It has previously been shown that Tsg and chordin mask different BMPR-binding sites on BMPs [25]. Taken together this suggests that, in addition to strengthening the ternary complex, Tsg also increases inhibition of BMP-7 by blocking binding sites which chordin cannot.

Our previous findings showed that partially cleaved chordin retains BMP inhibitory activity at similar or higher levels to full length chordin and may represent a mechanism to modulate BMP inhibition by chordin [13]. Our new data extend this mechanism, where the presence of Tsg allows differentiation between and selective inhibition of BMPs. We postulate that in the absence of Tsg, chordin cleavage resulting in △N-chordin is a way to abolish chordin's inhibitory capacity for BMP-7 whereas chordin cleavage resulting in △C-chordin may represent an alternative way to inhibit BMP activity more efficiently [13]. This effect is less pronounced for BMP-4 than for BMP-7 (comparing Fig. 6A-C to Fig. S7 from [13]). This mechanism is probably needed in the absence of Tsg as in the presence of Tsg, full-length, △C- and △N-chordin all show a similarly enhanced inhibition rate (50% inhibition at a molar excess of 1:8–13 over BMP-4). However in contrast to BMP-4, Tsg is inhibited directly by BMP-7 (Fig. 5) so chordin may not be required when Tsg is present. However, △C-chordin and Tsg show the best inhibition rate which could be additive. With this mechanism it is possible, for example, in the

presence of Tsg to block BMP-7 activity completely while BMP-4 is still active. For instance, at a 1:10 ratio of Tsg, only small amounts of ΔC-chordin are necessary to completely inhibit BMP-7 (Fig. 6E) while BMP-4 activity is not affected at these concentrations (Fig. 6 Cii). Native deglycosylation of Tsg even enhances BMP-7 inhibition which adds another layer of tissue-specific modulation. This in contrast to non-glycosylated Tsg expressed in bacteria that had reduced binding to BMP-2 and reduced activity in a mandibular explant culture system [9]. This may indicate further differences between BMP sub-types or that natively deglycosylated Tsg retains some glycans.

We show that Tsg increases inhibition of BMPs by chordin fragments, while promoting their further cleavage by mTld. This supports the requirement of dual cleavage of chordin for activation of BMP signalling. In addition chordin is expressed as a number of tissue specific splice variants [29] and our data suggest that their regulation by Tsg is likely to be similar to full length chordin, provided they retain the Tsg-binding vWC2–3 domains. In summary, we show a nanoscale structure for monomeric Tsg supported by MALS and AUC data, we propose that Tsg acts primarily as a BMP agonist by inducing conformational change in chordin to make it more accessible for cleavage by tolloids. Tsg increases BMP inhibition with both full length chordin and partially cleaved chordin, which suggests chordin continues to inhibit BMPs until complete cleavage has occurred.

Experimental procedures

Protein expression and purification

Human Tsg was generated by PCR from image clone Q9GZX9 (Bioscience Gene Service). The chordin vWC2–3 (residues 701–860), vWC1 (residues 50–128) and vWC1-4CHRD (residues 27–650) constructs were generated by PCR from full-length human chordin. His$_6$ tags were incorporated at the C-termini following a thrombin cleavage site. Constructs were ligated into a modified pCEP4 vector [30] and transfected into HEK 293-EBNA cells cultured as described previously [13]. Proteins were purified by nickel affinity chromatography and size-exclusion chromatography on an AKTA purifier FPLC using a Superdex 200 10/300GL column (GE Healthcare). Protein identity was confirmed by in-gel trypsin digestion and LC–MS/MS using a NanoAcquity LC (Waters) coupled to a LTQ Velos (Thermo Fisher Scientific). Additional human chordin constructs (ΔN- and ΔC-chordin) and mammalian tolloid were also expressed in HEK 293-EBNA cells as previously described [13,31]. Except where otherwise stated buffer conditions for all experiments were

10 mM Tris–HCl, 150 mM NaCl, pH 7.4. Protein concentration was estimated using absorbance at 280 nm corrected for the molar extinction coefficient.

Native deglycosylation and limited proteolysis

Tsg was digested in 0.04 U/ml of PNGase-F at 37 °C for 36 h. Removal of the sugars was verified using SDS-PAGE and the digested protein separated from the enzyme using size exclusion chromatography. Proteolysis with trypsin (1:50, w/w) was carried out at 37 °C. At the times indicated, samples were withdrawn, and the reaction was stopped by the addition of 2 × SDS sample buffer containing 100 mm dithiothreitol and heating at 95 °C for 4 min as described in [32].

Multi-angle light scattering

Tsg samples (0.5 ml at approximately 0.5 mg/ml) were loaded onto a Superdex 200 10/300GL column running at a flow rate of 0.5 ml/min. Proteins eluting from the column passed through a Wyatt DAWN Heleos II EOS 18 angle laser photometer with QELS detector (Wyatt Technologies) and an Optilab T-rEX refractive index detector. The molecular mass and concentrations of the resulting peaks were analysed using ASTRA 5.6.

Velocity sedimentation analytical ultracentrifugation

Velocity sedimentation AUC was performed on Tsg (~0.2 mg/ml) at 25,000 rpm using a Beckman XL-A analytical ultracentrifuge with an An60Ti 4-hole rotor at 20 °C. The sedimenting boundary was monitored at 230 nm every 90 s for 300 scans with a radial step size of 3 μm. Data were analysed using Sedfit [33] and the resulting apparent sedimentation coefficient was corrected for standard conditions using the programme SEDNTERP [34]. The hydrodynamic radius (R_h) and frictional ratio (f/f_o), which represents the deviation of the friction of the molecule from a theoretical sphere of the same molecular weight, were also calculated using SEDNTERP.

Small-angle X-ray scattering

SAXS data were collected at 1–6 mg/ml on native and deglycosylated Tsg at the Diamond Light Source (I22) and PETRA III (P12) beamlines. Samples were maintained at 10 °C during exposure using the standard sample holders at each beamline. Data was checked for radiation damage between subsequent frames and different concentrations merged using PRIMUS [35]. The forward scattering intensity, Rg and one-dimensional intraparticle distance distribution function P(r) in real

space were evaluated with the indirect Fourier transform programme GNOM [20] and particle shapes were restored *ab initio* using DAMMIN [21]. At least 20 *ab initio* models were calculated with each programme these were combined into a single model using DAMAVER suite [36]. Multiphase volumetric modelling was performed using MONSA [21] to analyse the difference in density between the native and deglycosylated Tsg data. Models were generated with MONSA online using a phase contrast of 1.6 for the glycans. 10 models were calculated then the volume corresponding to Tsg was aligned and averaged using the methods described in [37].

Circular dichroism

Circular dichroism measurements were performed on 0.4 mg/ml Tsg using a Jasco J715 spectropolar-imeter at 190–260 nm and 0.05 cm path length using the sample buffer as a reference. The mean of twenty measurements was taken at 100 mdeg sensitivity at a rate of 20 nm/min. The estimation of protein secondary-structure content was analysed using Contin-LL, Selcon 3, CDSSTR and K2d programmes (all available algorithms provided on the DICHROWEB server www.dichroweb.cryst.bbk. ac.uk). The results of each algorithm were averaged to obtain a quantitative analysis of the secondary structural elements.

Surface plasmon resonance

Binding analyses were performed using either a Biacore T200 (GE Healthcare, Little Chalfont, UK) with a Series S CM5 sensor chip (GE Healthcare) or a ProteOn XPR36 (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK) with a ProteOn GLC sensor chip (Bio-Rad). Ligands were immobilised on the sensor chips (concentrations ranging from 3 to 9 µg/ml in 50 mM sodium acetate pH 4.0) *via* amine-coupling using an EDC/NHS chemical cross-linker following the manufacturer's instructions and subsequently blocked with ethanolamine. The native Tsg interactions in Fig. 4 were examined using the ProteOn XPR36 at 25 °C in 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% Tween-20, pH 7.4 with a flow rate of 50 µl/min. All other interactions were examined using the Biacore T200 at 25 °C in 10 mM Hepes, 150 mM NaCl, 3 mM EDTA, 0.005% Tween-20, pH 7.4 with a flow rate of 30 µl/min. For kinetic analysis, analytes were injected at several concentrations onto immobilised ligands. Kinetic constants were calculated by nonlinear fitting of kinetic models (including 1:1 binding and heterogeneous analyte) to the experimental data (the recorded association and dissociation curves) using the ProteOn Manager. Calculated kinetic constants included the association rate constant (k_a), the dissociation rate constant (k_d)

and the equilibrium dissociation rate constant (K_D). K_D was derived from the ratio of k_d to k_a.

Microscale thermophoresis

Tsg was labelled using the Monolith NT Protein labelling Kit Red-NHS kit from NanoTemper Technologies. Unbound fluorophore was removed using a desalting column equilibrated with MST buffer (50 mM Tris, 150 mM NaCl, 10 mM $MgCl_2$ 0.05% Tween-20 pH 7.4). 1:1 serial dilutions of vWC1 at a starting concentration of 7 µM in MST buffer were mixed 1:1 with 15 nM NT647-labelled Tsg and loaded into standard treated capillaries and analysed using the Monolith NT.115Pico from Nano-Temper Technologies at 20% LED power and 20% MST power. The experiment was done in triplicate and the K_D determined.

Alkaline phosphatase assay

Dose response curves for BMP-4 and BMP-7 induced ALP production in C2C12 cells were generated to determine the ED50 for BMP-4 (27 ng/ml (1.0 nM)) and BMP-7 (100 ng/ml (3.2 nM)) (Supplementary Fig. 2). C2C12 cells were seeded into 96-well plates (Costar) at a density of 30,000 cells/well in DMEM (MediaTech) containing 2% *v*/v FCS. BMP stimulation was carried out in triplicates with 30 ng/ml BMP-4, or 100 ng/ml BMP-7 (both human from R&D Systems) in the presence or absence of 28-fold molar excess of Tsg for 72 h. Prior to stimulation proteins were dialyzed together in PBS, containing 0.01% BSA as carrier. After stimulation, cell layers were washed twice with PBS and lysed with 100 µl lysis buffer (0.1 M glycine pH 9.6, 1 mM $MgCl_2$, 1 mM $ZnCl_2$) containing 1% NP-40. Lysates were clarified by centrifugation at 18,800 *g*. 100 µl Alkaline Phosphatase Yellow (pNPP) Liquid Substrate (Sigma-Aldrich) were added to 50 µl of each lysate and incubated for 20 min at room temperature followed by measurement of activity in a TECAN infinite M1000 reader (Dynamic Instruments, Wilmington, DE) at 405 nm.

Phospho-smad Western blot assay

HEK 293T cells were seeded into 6-well plates at a density of 1000,000 cells/well in DMEM. The following day, cells were stimulated for 1 h with 30 ng/ml BMP-4 or 100 ng/ml BMP-7 in the presence or absence of 28-fold molar excess of Tsg. After stimulation cell layers were incubated with lysis buffer (10 mM NaCl, 1.5 mM $MgCl_2$, 20 mM HEPES pH 7.4, 20% glycerol, 0.1% triton X-100, and 1 mM DTT), centrifuged at 2000 rpm for 5 min at 4 °C, and washed with PBS. Aliquots of the supernatant were supplemented with 1× PhosSTOP (Roche) and western blotted. Phosphorylated Smad was

detected using pooled anti-phosphosmad1 (1:1250) and Smad5 (1:1000) antibodies (Abcam). Band intensities were quantitated with ImageJ and plotted against concentration.

BMP-induced expression of Id3

C2C12 cells were seeded into 96-well plates at a density of 30,000 cells/well in DMEM without serum. Proteins were dialyzed together in PBS, 0.01% BSA overnight. BMP stimulation was carried out in the presence of 30 ng/ml BMP-4, or 100 ng/ml BMP-7 for 5 h. After incubation, total RNA from three wells was harvested using TRIzol® reagent (Invitrogen). Total RNA preparations were quantified by photospectrometry. 0.1 μg of RNA per sample was reverse-transcribed using the Bio-Rad iScript™ cDNA synthesis kit. Reverse-transcribed samples were amplified in triplicates with primers for *Id3*, a BMP-responsive element, using the iTaq™ SYBR Green Supermix (Bio-Rad). Analysis of data was performed using the $2^{-\Delta\Delta Ct}$ method and quantitated relative to the *Arbp PO* gene. Gene expression was normalized to samples where cells were incubated with corresponding amounts of BSA, which provided an arbitrary constant for comparative fold expression.

mTld cleavage assay

50 $\mu g\,ml^{-1}$ ΔN-chordin was incubated with 3 $\mu g\,ml^{-1}$ mTld at 37 °C for 0.5 h. Tsg was added in a 0.1:1, 0.25:1, 0.5:1, 1:1 and 4:1 molar ratio of Tsg:ΔN-chordin. At 1:1 or higher rations of Tsg:ΔN-chordin an increase in tolloid cleavage was observed. Gels were digitised and bands quantitated using ImageJ.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.matbio.2016.01.019.

References

[1] L. Dale, F.C. Wardle, A gradient of BMP activity specifies dorsal–ventral fates in early *Xenopus* embryos, Semin. Cell Dev. Biol. 10 (1999) 319–326.

[2] R. Sartori, et al., BMP signaling controls muscle mass, Nat. Genet. 45 (2013) 1309–1318.

[3] C.G. Wilson, F.M. Martin-Saavedra, N. Vilaboa, R.T. Franceschi, Advanced BMP gene therapies for temporal and spatial control of bone regeneration, J. Dent. Res. 92 (2013) 409–417.

[4] Y. Sasai, et al., *Xenopus* chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes, Cell 79 (1994) 779–790.

[5] S. Schulte-Merker, K.J. Lee, A.P. McMahon, M. Hammerschmidt, The zebrafish organizer requires chordino, Nature 387 (1997) 862–863.

[6] I.C. Scott, et al., Homologues of twisted gastrulation are extracellular cofactors in antagonism of BMP signalling, Nature 410 (2001) 475–478.

[7] C. Chang, et al., Twisted gastrulation can function as a BMP antagonist, Nature 410 (2001) 483–487.

[8] J.J. Ross, et al., Twisted gastrulation is a conserved extracellular BMP antagonist, Nature 410 (2001) 479–483.

[9] C.J. Billington Jr., et al., Glycosylation of twisted gastrulation is required for BMP Binding and activity during craniofacial development, Front. Physiol. 2 (2011) 59.

[10] S. Vadon-Le Goff, D.J. Hulmes, C. Moali, BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling, Matrix Biol. 44–46 (2015) 14–23.

[11] S. Piccolo, et al., Cleavage of Chordin by Xolloid metalloprotease suggests a role for proteolytic processing in the regulation of Spemann organizer activity, Cell 91 (1997) 407–416.

[12] J. Xie, S. Fisher, Twisted gastrulation enhances BMP signaling through chordin dependent and independent mechanisms, Development 132 (2005) 383–391.

[13] H. Troilo, et al., Nanoscale structure of the BMP antagonist chordin supports cooperative BMP binding, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 13063–13068.

[14] M. Oelgeschlager, J. Larrain, D. Geissert, E.M. De Robertis, The evolutionarily conserved BMP-binding protein twisted gastrulation promotes BMP signalling, Nature 405 (2000) 757–763.

[15] J. Larrain, et al., Proteolytic cleavage of chordin as a switch for the dual activities of twisted gastrulation in BMP signaling, Development 128 (2001) 4439–4447.

[16] I.L. Blitz, K.W. Cho, C. Chang, Twisted gastrulation loss-of-function analyses support its role as a BMP inhibitor during early *Xenopus* embryogenesis, Development 130 (2003) 4975–4988.

[17] H. Inomata, T. Haraguchi, Y. Sasai, Robust stability of the embryonic axial pattern requires a secreted scaffold for chordin degradation, Cell 134 (2008) 854–865.

[18] P. Vilmos, K. Gaudenz, Z. Hegedus, J.L. Marsh, The twisted gastrulation family of proteins, together with the IGFBP and CCN families, comprise the TIC superfamily of cysteine rich secreted factors, Mol. Pathol. 54 (2001) 317–323.

[19] H. Troilo, et al., The role of chordin fragments generated by partial tolloid cleavage in regulating BMP activity, Biochem. Soc. Trans. 43 (2015) 795–800.

[20] D.I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495–503.

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

[22] R.P. Rambo, J.A. Tainer, Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod–Debye law, Biopolymers 95 (2011) 559–571.

[23] J.L. Zhang, et al., Crystal structure analysis reveals how the chordin family member crossveinless 2 blocks BMP-2 receptor binding, Dev. Cell 14 (2008) 739–750.

[24] J.M. O'Leary, et al., Solution structure and dynamics of a prototypical chordin-like cysteine-rich repeat (von Willebrand Factor type C module) from collagen IIA, J. Biol. Chem. 279 (2004) 53857–53866.

[25] J.L. Zhang, Y. Huang, L.Y. Qiu, J. Nickel, W. von Sebald, Willebrand factor type C domain-containing proteins regulate bone morphogenetic protein signaling through different recognition mechanisms, J. Biol. Chem. 282 (2007) 20002–20014.

[26] T. Fujisawa, Y. Huang, W. Sebald, J.L. Zhang, The binding of von Willebrand factor type C domains of Chordin family proteins to BMP-2 and Tsg is mediated by their SD1 subdomain, Biochem. Biophys. Res. Commun. 385 (2009) 215–219.

[27] V. Hintze, et al., The interaction of recombinant subdomains of the procollagen C-proteinase with procollagen I provides a quantitative explanation for functional differences between the two splice variants, mammalian tolloid and bone morphogenetic protein 1, Biochemistry 45 (2006) 6741–6748.

[28] L. Zakin, E.M. De Robertis, Extracellular regulation of BMP signaling, Curr. Biol. 20 (2010) R89–R92.

[29] C. Millet, P. Lemaire, B. Orsetti, P. Guglielmi, V. Francois, The human chordin gene encodes several differentially expressed spliced variants with distinct BMP opposing activities, Mech. Dev. 106 (2001) 85–96.

[30] E. Poschl, J.W. Fox, D. Block, U. Mayer, R. Timpl, Two non-contiguous regions contribute to nidogen binding to a single EGF-like motif of the laminin gamma 1 chain, EMBO J. 13 (1994) 3741–3747.

[31] R. Berry, et al., Role of dimerization and substrate exclusion in the regulation of bone morphogenetic protein-1 and mammalian tolloid, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 8561–8566.

[32] J.Y. Suk, et al., Structural consequences of cysteine substitutions C1977Y and C1977R in calcium-binding epidermal growth factor-like domain 30 of human fibrillin-1, J. Biol. Chem. 279 (2004) 51258–51265.

[33] P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling, Biophys. J. 78 (2000) 1606–1619.

[34] J.S. Philo, An improved function for fitting sedimentation velocity data for low-molecular-weight solutes, Biophys. J. 72 (1997) 435–444.

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

[36] V.V. Volkov, D.I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering, J. Appl. Crystallogr. 36 (2003) 860–864.

[37] R.P. Rambo, Resolving individual components in protein-RNA complexes using small-angle X-ray scattering experiments, Methods Enzymol. 558 (2015) 363–390.

[38] M. Oelgeschlager, et al., The pro-BMP activity of Twisted gastrulation is independent of BMP binding, Development 130 (2003) 4047–4056.
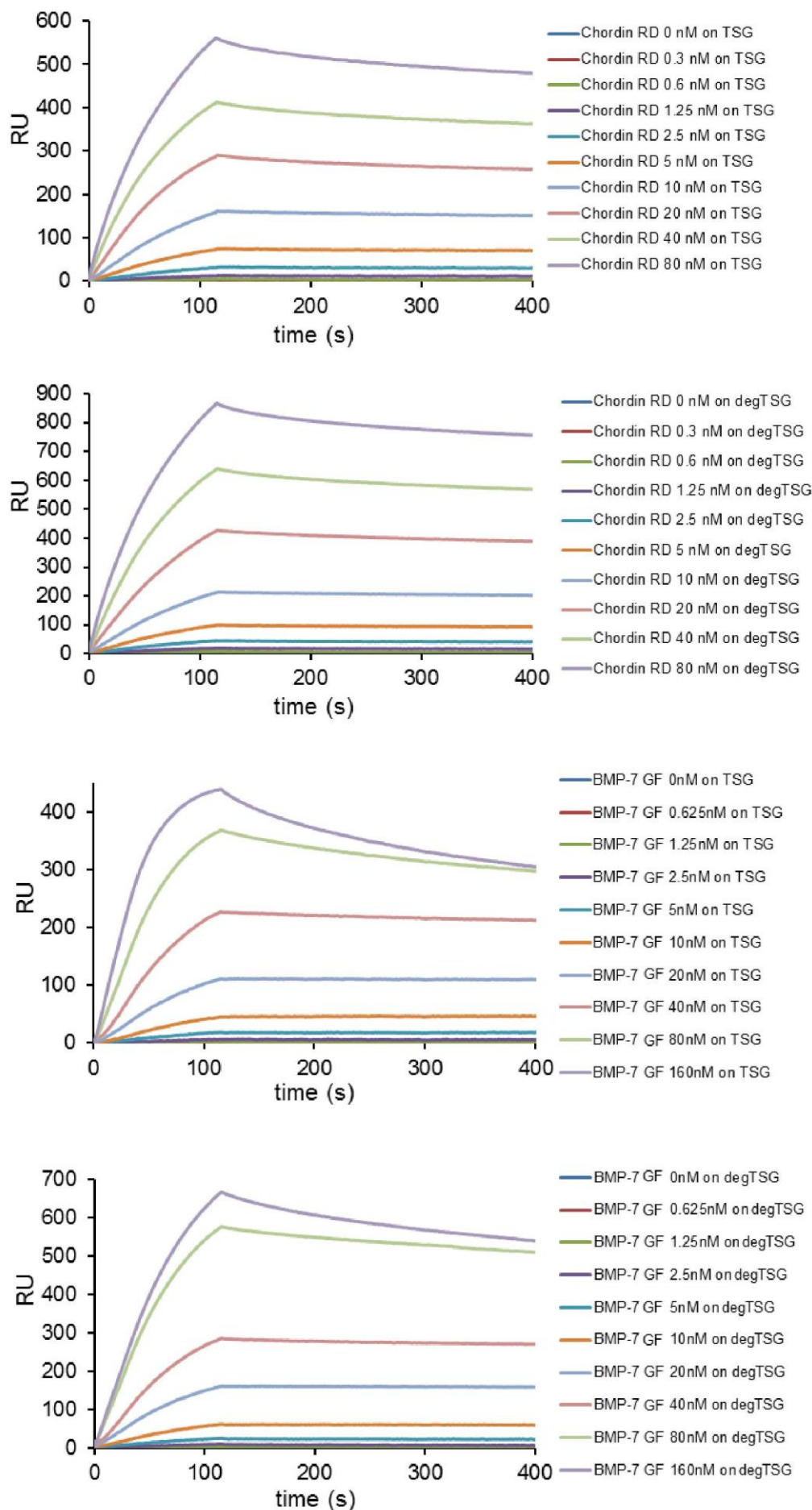
Supplementary Figure 1: SPR sensorgrams showing native and deglycosylated Tsg binding to chordin and BMP-7.
Tsg was immobilised on Biacore CM5 chip via amine coupling. Full-length chordin and BMP-7 (both R&D systems) were injected on to either native or deglycosylated Tsg (1200 RUs immobilised) at increasing concentrations.
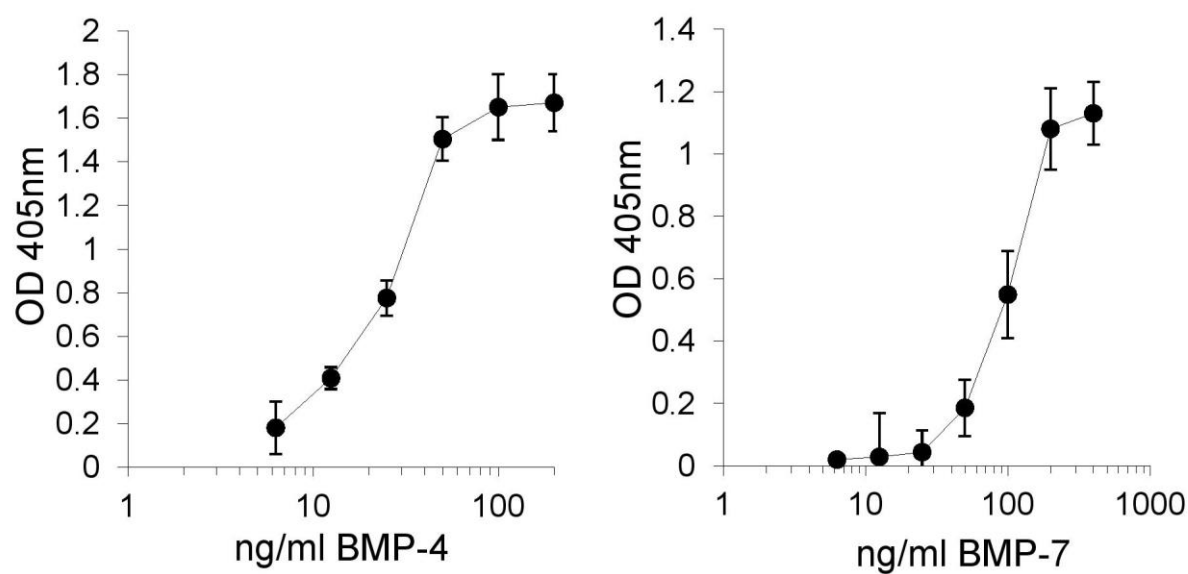
Supplementary Figure 2: Dose response curves of BMP-4 and -7 in ALP assays.
Titration experiments were performed to generate dose response curves for BMP-4 and BMP-7 induced ALP production in C2C12 cells. The ED50 were determined for BMP-4 (27 ng/ml (1.0 nM)) and BMP-7 (100 ng/ml (3.2 nM)).

12.2 GspB Paper

Subunit Arrangement in GpsB, a Regulator of Cell Wall Biosynthesis

Robert M. Cleverley,[1] Jeanine Rismondo,[2] Michael P. Lockhart-Cairns,[3,4] Paulien T. Van Bentum,[5] Alexander J.F. Egan,[6] Waldemar Vollmer,[6] Sven Halbedel,[2] Clair Baldock,[4] Eefjan Breukink,[5] and Richard J. Lewis[1]

GpsB, a key regulator of cell division in Gram-positive bacteria, interacts with a key peptidoglycan synthase at the cell division septum, the penicillin binding protein PBP1 (*a.k.a.* PonA). *Bacillus subtilis* GpsB has been reported to interact with other components of the cell division machinery, including EzrA, MreC, and PrkC. In this study, we report an analysis of the arrangement of subunits in *Listeria monocytogenes* GpsB by small-angle X-ray scattering. The resulting model has an elongated shape with residues critical for interaction with PBP1 and the cell membrane clustered at one end of the molecule. Mutations that destabilize the hexameric assembly of the wild-type protein have a *gpsB* null phenotype, indicating that oligomerization is critical for the correct function of GpsB. We suggest a model in which a single GpsB hexamer can interact with multiple PBP1 molecules and can therefore influence the arrangement of PBP1 molecules within the cell division machinery, a dynamic multiprotein complex called the divisome, consistent with a role for GpsB in modulating the synthesis of the cell wall.

Introduction

BACTERIAL CELL DIVISION has attracted increasing interest in recent years as a potential target of action for novel antibiotics in the battle against the rise of antimicrobial resistance to current therapies.[1] Cell division is undertaken by the coordinated action of a dynamic assembly of proteins termed the divisome.[2] A central component of the divisome is FtsZ, the bacterial homologue of tubulin, which drives division at midcell by forming the polymeric dynamic Z-ring that exerts a contractile force on the cytosolic face of the membrane during septal closure.[3] Z-ring contraction is accompanied, on the opposite side of the membrane, by the remodeling of the essential peptidoglycan sacculus by penicillin binding proteins (PBPs) and peptidoglycan hydrolases, respectively.[4] Division thus requires the coordination of processes on opposite faces of the cytoplasmic membrane. The membrane-associated proteins EzrA and GpsB appear to mediate this coordination in Gram-positive bacteria.[5] FtsZ binds directly to the cytoplasmic domain of EzrA,[6,7] while bacterial two hybrid studies have revealed the interaction of both EzrA and GpsB with PBPs[5,8,9]; the interaction of GpsB with PBP1 (PonA), the major peptidoglycan synthase at the cell division septum in *Bacillus subtilis*,[10] has also been observed *in vitro* by surface plasmon resonance and fluorescence polarization.[11] The deletion of *ezrA* is lethal in *Listeria monocytogenes*[12] and *Streptococcus pneumoniae*[13–15] and is accompanied by the mislocalization of FtsZ and PBPs in both *B. subtilis*[5] and *Staphylococcus aureus*.[9] It is not clear whether *ezrA* is essential in *S. aureus*, as opposing results have been obtained in different laboratories.[8,9] Similarly, *gpsB* has been found to be essential in the pathogenic *S. pneumoniae* strain D39, where its depletion causes severe cell division defects,[16] but not in other closely-related strains.[13,14,17,18] GpsB is important for virulence in *L. monocytogenes*, and here its knockout causes growth defects and increased susceptibility to lysis.[11] Severe growth defects and perturbed localization of PBPs occur in *B. subtilis* upon the concurrent deletion of *ezrA* and *gpsB*.[5]

[1]Institute for Cell and Molecular Biosciences, University of Newcastle, Newcastle upon Tyne, United Kingdom.
[2]FG11 Division of Enteropathogenic Bacteria and Legionella, Robert Koch Institute, Wernigerode, Germany.
[3]Diamond Light Source Ltd, Harwell Science & Innovation Campus, Didcot, United Kingdom.
[4]Wellcome Trust Centre for Cell Matrix Research, Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom.
[5]Membrane Biochemistry and Biophysics, Bijvoet Centre for Biomolecular Research, University of Utrecht, Utrecht, the Netherlands.
[6]Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, University of Newcastle, Newcastle upon Tyne, United Kingdom.

The importance of GpsB in mediating the coordination of PBP activity with other divisome components is underlined by the interaction network in which GpsB participates. First, GpsB has been reported to interact with EzrA in two hybrid assays, although only unidirectionally,[5] potentially providing an indirect link between FtsZ and PBPs, given that numerous *in vitro* and *in vivo* studies have shown that EzrA interacts with FtsZ (*e.g.*, 6–9). Second, it has been reported that GpsB is phosphorylated in *B. subtilis* by PrkC,[19] a eukaryotic-like membrane-embedded Ser/Thr kinase with an extracellular PASTA domain[20] that binds peptidoglycan fragments.[21] The relationship between GpsB and PrkC suggests that GpsB could be involved in transducing alterations in cell wall structure on the outside of the cell to the interior. Finally, a major portion of the GpsB sequence is homologous to the lipid binding domain of the DivIVA protein; the lipid binding domain promotes the localization of DivIVA to negatively curved lipid membranes,[22] which is pivotal to DivIVA's role in promoting the assembly of the divisome at midcell.[23] Whether GpsB regulates cell wall synthesis by similarly localizing to curved membranes remains to be clarified.

To help elucidate further the functional role of GpsB, we recently solved crystal structures of its component domains from *L. monocytogenes* (*Lm*GpsB) and *B. subtilis* (*Bs*GpsB).[11] GpsB proteins are typically approximately 100 amino acids in length and contain two domains, with the 65–70 amino acid N-terminal domain (N-GpsB) resembling the lipid binding domain of DivIVA, a parallel coiled-coil dimer.[23] The 20–25 amino acid C-terminal domain (C-GpsB) forms a parallel triple helical coiled-coil.[11] While the N- and C-terminal domains in isolation assemble into dimers and trimers, respectively, the full-length GpsB protein is a hexamer,[11] which to date has been recalcitrant to successful crystallization. In this study, we present a study of the arrangement of subunits in *Lm*GpsB using small-angle X-ray scattering (SAXS) and a mutational analysis of candidate oligomer stabilizing residues. To understand the role of GpsB in cell division, we have also analyzed the enzymatic properties of PBP1 in the presence and absence of *Bs*GpsB. In combination, these data are consistent with a model where GpsB serves to localize multiple copies of PBP1, a process that appears to be critical for efficient cell division.

Materials and Methods

Bacterial strains and growth conditions

All bacterial strains used in this study are listed in Table 1. *L. monocytogenes* strains were grown in brain–heart infusion (BHI) broth or on BHI agar plates at 37°C. When necessary, antibiotics and supplements were added to the growth media at concentrations of 50 µg/ml for kanamycin and 1 mM for IPTG. The *Escherichia coli* strain TOP10 was used for all cloning procedures.[24]

Construction of plasmids and strains

The T88A, T88D, F91A, L94A, and F105A amino acid exchanges were introduced into *L. monocytogenes* *gpsB*, cloned previously into pSH254, by Quikchange mutagenesis[25] using the primer pairs JR290/JR291, JR291/JR293, JR311/JR312, JR313/JR314, and JR315/JR316, respectively (Table 2). The presence of the desired mutations was verified by DNA sequencing. The resulting plasmids are listed in

TABLE 1. STRAINS AND PLASMIDS USED IN THIS STUDY

Name	Relevant characteristics	Source[a]/ reference
Plasmids		
pSH254	P*help*-lacO-gpsB lacI neo	11
pJR133	P*help*-lacO-gpsB T88A lacI neo	This work
pJR134	P*help*-lacO-gpsB T88D lacI neo	This work
pSH433	P*help*-lacO-gpsB F91A lacI neo	This work
pSH434	P*help*-lacO-gpsB L94A lacI neo	This work
pSH435	P*help*-lacO-gpsB F105A lacI neo	This work
Listeria monocytogenes strains		
EGD-e	wild-type, serovar 1/2a strain	54
LMJR19	Δ*gpsB* (lmo1888)	11
LMS56	Δ*gpsB* attB::P*help*-lacO-gpsB lacI neo	11
LMJR161	Δ*gpsB* attB::P*help*-lacO-gpsB T88A lacI neo	pJR133 → LMJR19
LMJR162	Δ*gpsB* attB::P*help*-lacO-gpsB T88D lacI neo	pJR134 → LMJR19
LMS185	Δ*gpsB* attB::P*help*-lacO-gpsB F91A lacI neo	pSH433 → LMJR19
LMS186	Δ*gpsB* attB::P*help*-lacO-gpsB L94A lacI neo	pSH434 → LMJR19
LMS187	Δ*gpsB* attB::P*help*-lacO-gpsB F105A lacI neo	pSH435 → LMJR19

[a]The arrow (→) stands for a transformation event.

Table 1 and were introduced into *L. monocytogenes* EGD-e by electroporation.[26] Kanamycin-resistant clones were selected, and plasmid insertion at the *attB* site of the tRNA^Arg locus was verified by PCR.

To generate the plasmid for expressing C-*Lm*GpsB, the open reading frame encoding *L. monocytogenes* GpsB residues 87–113 was PCR amplified from genomic DNA from strain EGD-e with primers RC87f and RC113f, which contain flanking *Nco*I and *Xho*I restriction sites, and cloned between the *Nco*I and *Xho*I sites of pMAT11, a modified form of pHAT4.[27] The plasmids for expressing the full-length *Lm*GpsB_F91A, *Lm*GpsB_F105A mutants and the double mutant *Lm*GpsB_F91AF105A were prepared by Quikchange mutagenesis using primers RC91f, RC91r, RC105f, and RC105r; for the double mutant, two rounds of mutagenesis were used—the first round with primer pairs RC91f and RC91r and the second round with primer pairs RC105f and RC105r (Table 2).

Similarly, the full-length *Bs*GpsB_T75E and *Bs*GpsB_T75D mutants were prepared by Quikchange mutagenesis with primers RC75Ef and RC75Er, and RC75Df and RC75Dr, respectively; the plasmid for expressing full-length wild-type *Bs*GpsB was used as a template for the Quikchange reaction to generate the T75D mutation. By contrast, for the T75E Quikchange reaction, the plasmid for expressing *Bs*GpsB_T75D was used as a template. The resultant plasmids were sequenced to verify the successful introduction of the desired mutations and were subsequently used for the overproduction of the full-length GpsB variants by the same procedures as wild-type GpsB proteins.[11]

Isolation of cellular proteins and Western blotting

Cells were harvested by brief centrifugation in a table-top microfuge and washed with ZAP buffer (10 mM Tris-HCl

TABLE 2. OLIGONUCLEOTIDES USED IN THIS STUDY

Name	Sequence (5′ → 3′)
JR290	GTTGGTTGCTCCAGCAGGTTGTG
JR291	CTGCTGGAGCAACCAACTTTGATATTC
JR292	GTTGGTGTCTCCAGCAGGTTGTG
JR293	CTGCTGGAGACACCAACTTTGATATTC
JR311	GGAACAACCAACGCTGATATTCTAAAG
JR312	CTTTAGAATATCAGCGTTGGTTGTTCC
JR313	GATATTGCAAAGCGTCTTTCTAATTTAG
JR314	CTAAATTAGAAAGACGCTTTGCAATATC
JR315	GAAAAACATGTTGCTGGAAATAAGCTG
JR316	CAGCTTATTTCCAGCAACATGTTTTTC
RC91f	CACAACCTGCTGGAACAACCAACGCTGATATTCTAAAGCGTCTTTC
RC91r	GAAAGACGCTTTAGAATATCAGCGTTGGTTGTTCCAGCAGGTTGTG
RC105f	CGTCTTTCTAATTTAGAAAAACATGTTGCTGGAAATAAGCTGGACGATAACG
RC105r	CGTTATCGTCCAGCTTATTTCCAGCAACATGTTTTTCTAAATTAGAAAGACG
RC75Ef	CAGCCTGTGCAATCTAACGAAACAAACTTTGATATTTTAAAACG
RC75Er	CGTTTTAAAATATCAAAGTTTGTTTCGTTAGATTGCACAGGCTG
RC75Df	CAGCCTGTGCAATCTAACGATACAAACTTTGATATTTTAAAACGGCTGTCTAACTTAG
RC75Dr	CTAAGTTAGACAGCCGTTTTAAAATATCAAAGTTTGTATCGTTAGATTGCACAGGCTG
RC87f	GCACAACCCATGGGAACAACCAACTTTGATATTCTAAAGCG
RC113r	CATAAGCTCGAGTTTCTCGCTCGTGGGGTTTAC

pH 7.5, 200 mM NaCl). Cells were resuspended in 1 ml ZAP buffer containing 1 mM PMSF and disrupted by sonication. The cell debris was removed by centrifugation, and the supernatant was used as the soluble protein extract. Protein samples were separated by SDS polyacrylamide gel electrophoresis and then transferred onto positively charged polyvinylidene fluoride (PVDF) membranes using a semidry transfer unit. GpsB and DivIVA were detected using polyclonal rabbit antisera raised against *L. monocytogenes* GpsB[11] and *B. subtilis* DivIVA,[28] respectively, with anti-rabbit immunoglobulin G conjugated to horseradish peroxidase as the secondary antibody. The ECL chemiluminescence detection system (Thermo Scientific) was then used for detection of the peroxidase conjugates on the PVDF membranes in a chemiluminescence imager (Vilber Lourmat).

Protein purification

PBP1 and GpsB proteins were purified as described previously.[11] C-*Lm*GpsB proteins were expressed as fusion proteins with N-terminal His$_6$- and MBP-tags; the protocols for expression and purification of the fusion protein by Ni-NTA chromatography were as described previously for C-*Bs*GpsB.[11] Following Ni-NTA chromatography, the fusion protein was cleaved with TEV protease (ratio of fusion protein:TEV 50:1 by mass) overnight at 4°C while simultaneously dialyzing into a buffer of 50 mM Tris-HCl pH 8.0, 300 mM NaCl using a 2 kDa molecular weight cutoff dialysis membrane. The dialysate was passed back over a 5 ml Ni-NTA cartridge (Qiagen) to separate TEV protease, C-*Lm*GpsB released by cleavage from the fusion protein and residual uncleaved His$_6$-MBP-C-*Lm*GpsB. The flow through from the Ni-NTA column was concentrated in a centrifugal concentrator to 1 ml and loaded onto a Superdex75 XK16/60 column (GE Healthcare) equilibrated in a buffer of 10 mM Tris-HCl pH 8.0, 250 mM NaCl. Elution fractions containing C-*Lm*GpsB were pooled, concentrated to *c*. 5 mg/ml in a centrifugal concentrator, and then dialyzed further into 20 mM sodium phosphate pH

7.8, 250 mM NaCl before finally flash freezing the protein in liquid nitrogen and storing at −80°C.

Size exclusion chromatography

The full-length wild-type *Lm*GpsB and *Lm*GpsB$_{F91A}$, *Lm*GpsB$_{F105A}$ and *Lm*GpsB$_{F91AF105A}$ proteins were analyzed on a Superdex200 Increase 10/300 GL column (GE Healthcare) equilibrated in 10 mM Tris-HCl pH 8.0, 250 mM NaCl, at a flow rate of 0.5 ml/min. Protein samples at 3 mg/ml concentration were injected onto the column through a 100 µl sample loop. Wild-type *Bs*GpsB and the *Bs*GpsB$_{T75D}$ and *Bs*GpsB$_{T75E}$ mutants were analyzed under the same conditions, except the protein concentration was 1 mg/ml.

Size exclusion chromatography-MALS

GpsB samples (500 µl) at concentrations of 8 mg/ml or 0.5 mg/ml were loaded onto a Superdex200 Increase 10/300 GL column (GE Healthcare) equipped with a Jasco UV-2077 detector, Wyatt DAWN Heleos II EOS 18-angle laser photometer (with the 13th detector replaced with the QELS in-line dynamic light scattering detector) coupled to a Wyatt Optilab rEX refractive index detector. The flow rate was 0.75 ml/min. Molecular mass and concentrations of the peaks eluting from the column in a running buffer of 10 mM Tris-HCl pH 8.0, 250 mM NaCl were analyzed using Astra 6.2 (www.wyatt.com/products/software/astra.html).

Circular dichroism

CD thermal melts were measured in a 1-mm path length quartz cuvette in a 20 mM sodium phosphate pH 7.8, 200 mM NaCl buffer for wild-type *Lm*GpsB, N-*Lm*GpsB, and C-*Lm*GpsB and in 20 mM sodium phosphate pH 7.8, 250 mM NaCl for wild-type *Bs*GpsB, *Bs*GpsB$_{T75D}$, and *Bs*GpsB$_{T75E}$. The temperature was increased at a rate of 1°C/min, and ellipticity was monitored at a wavelength of 222 nm with a response time of 8 seconds and bandwidth of 2 nm. The

reversibility of the melts was verified by recording spectra in the wavelength range 200–240 nm (scan speed 20 nm/min, response 2 seconds) before and after the melt, to check that the spectra were superimposable. Protein concentrations were 23 μM for wild-type LmGpsB, N-LmGpsB, and C-LmGpsB, 29 μM for LmGpsB$_{F91AF105}$, and 5 μM for wild-type BsGpsB, BsGpsB$_{T75D}$, and BsGpsB$_{T75E}$.

The CD spectra of wild-type and C-LmGpsB$_{F105A}$ proteins were recorded at protein concentrations of 23 and 33 μM, respectively, in a buffer of 20 mM sodium phosphate pH 7.8 at 20°C. The scan speed was 10 nm/min, and the response time was 4 seconds. Protein concentrations were determined based on calculated extinction coefficients at 280 nm[29] for LmGpsB, N-LmGpsB, and BsGpsB proteins; for the C-LmGpsB proteins, which lack tryptophan and tyrosine residues, amino acid analysis was used instead to quantify concentration. Secondary structure composition was evaluated from the spectra using the program CDSSTR[30] within the DICHROWEB[31] server.

SAXS data collection and processing

The LmGpsB and N-LmGpsB proteins were analyzed in a buffer of 10 mM Tris-HCl pH 8.0, 150 mM NaCl at a protein concentration of 40 mg/ml. SAXS intensity data, $I(q)$ versus q, ($q = 4\pi sin2\theta/\lambda$, where 2θ is the scattering angle) were collected using batch-mode SAXS on beamline P12 at PETRAIII (DESY; Hamburg), or batch mode and inline size exclusion chromatography (SEC)-SAXS on beamline B21, Diamond Light Source. Batch-mode SAXS was completed using 30 μl of LmGpsB or N-LmGpsB in a dilution series, which was loaded onto the beamline using the EMBL automated sample changer.[32] SAXS data were collected at PETRAIII with 0.5 second exposures onto a 2 M Pilatus detector (Dectris) at a distance of 3.0 m and an X-ray wavelength of 1.24 Å; these data were reduced and subtracted using in-house software.[33]

For SEC-SAXS data collection at Diamond, 50 μl of LmGpsB was loaded onto a Superdex 200 Increase 3.2/300 column and the eluent flowed through the SAXS beam at a flow rate of 0.1 ml/min; the buffer used as the blank in these measurements was the eluent after one SEC column volume. SAXS data were constantly collected at 1 second intervals using a 2 M Pilatus detector (Dectris) at a distance of 3.9 m and an X-ray wavelength of 1 Å. Subtraction of the SEC-SAXS data was completed for each frame across the elution peak, and the radius of gyration (R_g) and the integral of ratio to background were plotted. The data were scaled, merged, and averaged for each frame with a consistently similar R_g. All further processing and analysis of data were carried out using ScÅtter (www.bioisis.net/scatter).

Pair distance distribution functions were calculated with GNOM,[34] with D$_{max}$ values for LmGpsB and N-LmGpsB of 185 and 76 Å; the D$_{max}$ values were estimated initially using DATGNOM and the atomic coordinates of N-LmGpsB,[11] from which the pair distance distribution function was calculated with ScÅtter; all waters and heteroatoms were removed before the calculation.

SAXS model generation

Dummy atom models were constructed in DAMMIN in slow mode using scattering data up to a maximum q value of 8/R$_g$ (0.15Å^{-1} for LmGpsB, 0.36Å^{-1} for N-LmGpsB). For N-LmGpsB, the chi-squared values calculated by DAMMIN, corresponding to the agreement between experimental scattering curves and scattering curves calculated from the dummy atom models, ranged between 1.60 and 1.64. For LmGpsB, the chi-squared values from individual DAMMIN run were in the range 1.11–1.14. For N-LmGpsB, the average normalized structural discrepancy (NSD) between models from 10 independent DAMMIN runs, after averaging them together in the DAMAVER suite,[35] was 0.539 ± 0.09. For LmGpsB, the models from 15 DAMMIN runs were superimposed and averaged in the DAMAVER suite and the averaged model was then input into DAMSTART to generate a starting model for a further 15 DAMMIN runs. The final models were then again averaged together using the DAMAVER suite; the NSD between models was 0.552 ± 0.09.

Dummy atom models were visualized in CHIMERA[36] by representing the atoms as beads, with the bead scale adjusted to a value equivalent to the dummy atom radius divided by 0.7. The surface mesh representation of dummy atom models was generated by using the CHIMERA "molmap" command to convert the model to the equivalent electron density map at 25 Å resolution—the contouring of this map was then adjusted to enclose the surface of the dummy atoms. The crystal structures of the N- and C-terminal domains of GpsB[11] PDBids 4ug1 and 5an5, respectively, were also represented as 25 Å resolution electron density maps using the "molmap" command in CHIMERA. For the LmGpsB sample, the molecular weight of the scattering particles was calculated from the volume of correlation[37] using ScÅtter.

Production of lipid II-meso-diaminopimelic acid

Lipid II-meso-diaminopimelic acid (Lipid II-m-DAP) was synthesized using UDP-MurNAc-pentapeptide isolated from Bacillus cereus, essentially as described previously[38,39] with the following modifications. Purification was performed over a DEAE-cellulose column using a linear gradient of chloroform/methanol/water (2:3:1 v/v/v) to chloroform/methanol/ 1 M ammonium bicarbonate (2:3:1 v/v/v). The fractions containing Lipid II-m-DAP were collected and dried under vacuum. The resulting lipid II-m-DAP was then dissolved in 1:1 chloroform:methanol and stored at −20°C until use.

Production of [^{14}C]-amidated Lipid II-m-DAP

The ability of MurG to exchange the GlcNAc group between UDP-GlcNAc and Lipid II was used to radioactively label Lipid II. Since the E. coli MurG used here displayed much lower affinity for amidated Lipid II-m-DAP than the nonamidated form, we first synthesized [^{14}C]-Lipid II-m-DAP and then amidated this using AsnB from B. subtilis. First, 1.22 μmol of purified lipid II-m-DAP was incubated with 12.5 μCi of UDP-N-acetyl-D-[1-^{14}C]glucosamine (specific activity 55 mCi/mmol) (Hartmann Analytic GmbH). The labeling reaction was performed in 1 ml of 100 mM Tris-HCl, pH 8.0, 1 mM MgCl$_2$, and 2% (w/v) Triton X-100. The reaction was started by the addition of 1 μL purified recombinant E. coli MurG, which is able to exchange the N-acetyl-D-glucosamine of lipid II with that of UDP-N-acetyl-D-[1-^{14}C]glucosamine. After incubation at room temperature for 2.5 hours, the reaction was complete as determined by liquid scintillation counting.

To amidate the carboxylic acid group of the *m*-DAP, lipid II-DAP was incubated with *B. subtilis* AsnB in the presence of ATP and glutamine (to be published elsewhere), followed by extraction using butanol/pyridine acetate, pH 4.2, and another purification step over a DEAE cellulose column using a gradient of chloroform/methanol/water (2:3:1 v/v/v) to chloroform/methanol/0.5 M ammonium bicarbonate (2:3:1 v/v/v).

In vitro *peptidoglycan synthesis assays*

The continuous fluorescence glycosyltransferase (GTase) assay using dansyl-labeled lipid II as the substrate was performed essentially as described previously,[40] with only minor modifications to buffer conditions and temperature. We also used an endpoint assay to measure the GTase and transpeptidase (TPase) activity under the same experimental conditions. The activity of *Bs*PBP1 (0.4 μM) was measured at 37°C in the absence and presence of 10 μM *Bs*GpsB proteins in 50 mM HEPES.NaOH pH 7.5, 20 mM NaCl, 10 mM CaCl$_2$, and 5% (v/v) glycerol. The endpoint assay used [^{14}C]-labeled native amidated lipid II-*m*-DAP substrate followed by the quantification of both GTase and TPase products by high-pressure liquid chromatography (HPLC) as described previously.[41]

Results

Low resolution model of subunit arrangement in solution by SAXS

In the absence of diffracting crystals of full-length *Lm*GpsB, SAXS was used to determine its low resolution structure since the correct assembly of subunits in *Lm*GpsB is highly pertinent to the protein's function. To validate this approach, we first analyzed N-*Lm*GpsB for which the crystal structure has already been solved.[11] The measured scattering

FIG. 1. SAXS analysis of N-*Lm*GpsB. **(A)** SAXS scattering curves (*top*), Guinier plots (*middle*), and pair distribution functions (*bottom*) for N-*Lm*GpsB. The experimentally observed scattering curve is shown as a *black solid line* and the scattering curve, calculated from the coordinates with CRYSOL (42), as a *red dashed line*. The chi-squared value for the agreement between the two curves in the *q* range 0–0.25 Å^{-1} is 1.8. For the pair distribution function of the N-*Lm*GpsB, the *red dashed plot* represents the profile calculated from the coordinates with the ScÅtter software. **(B)** SAXS-derived *ab-initio* dummy atom models of N-*Lm*GpsB calculated with DAMMIN from small-angle X-ray scattering profiles. Representative models are shown from three independent DAMMIN calculations. The dummy atom model is shown in surface representation rendered in mesh. The coordinates of the N-*Lm*GpsB dimer (*red*) have been manually docked inside the mesh, represented either in ribbon form or as the equivalent electron density maps would appear at 25 Å. SAXS, small-angle X-ray scattering.

profile for N-*Lm*GpsB matched favorably with the theoretical profile calculated from the corresponding atomic coordinates (Fig. 1A). From a Guinier analysis of the scattering profile, an R_g of 21.4 Å was obtained, while the pair distribution function calculated by inverse Fourier transformation with GNOM[34] gave an R_g of 22.2 Å. These values compare extremely favorably with an R_g value of 21.4 Å that was calculated from the atomic coordinates with CRYSOL.[42] Finally, *ab initio* dummy atom molecular models built from the scattering data with DAMMIN[43] reproduced well the shape and dimensions of the dimeric N-*Lm*GpsB (Fig. 1B).

Analysis of *Lm*GpsB yielded the best quality data on SAXS beamlines with an in-line SEC facility. The monodispersity of *Lm*GpsB samples after in-line SEC was indicated by the linearity of Guinier plots at *q* values less than $1.3/R_g$ (Fig. 2A). The molecular mass of the scattering particles after in-line SEC, calculated from the volume of correlation,[37] was 75,900 Da, which is within 5% of the actual mass of the expressed *Lm*GpsB protein when assembled as a hexamer.

A Porod–Debye analysis of the comparative flexibility of *Lm*GpsB and N-*Lm*GpsB suggested that the full-length protein is more flexible than the isolated N-terminal domain. The Porod–Debye plot for N-*Lm*GpsB rose to a plateau (Fig. 2B), which is characteristic of a folded compact protein.[44] By contrast, the Porod–Debye plot for *Lm*GpsB had a less pronounced plateau and instead conformed to a hyperbole. The relatively shallow gradient of the asymptote, however, indicated that *Lm*GpsB is markedly less flexible than a completely unfolded protein. Given the successful crystal structure determinations of both N- and C-terminal domains of GpsB, the flexibility within *Lm*GpsB is most likely attributable to the linker region between the domains.

The pair distribution function of *Lm*GpsB had, like N-*Lm*GpsB, the characteristics of an elongated molecule,[45] with a steep initial increase in P(r) to a maxima followed by a more gradual decrease to zero at D_{max}. From the Guinier plot and pair distribution functions, R_g values of 51.2 and 53.1 Å were obtained for *Lm*GpsB. *Ab initio* dummy atom molecular models were calculated in DAMMIN without symmetry constraints, however, no obvious symmetric relationships were apparent from the resultant models. A notable feature of the models is two distinct lobes of different sizes connected by a narrower central region (Fig. 3).

The most logical arrangement of the individual domains in the GpsB hexamer resembles a tripod with the three dimeric N-GpsB domains aligned as in the legs in a tripod with the two C-GpsB trimeric domains encompassing its base (Fig. 4A). Such an arrangement is logical on the basis that it presents a relatively straight path and near equivalent distance between connected C- and N-terminal domains of
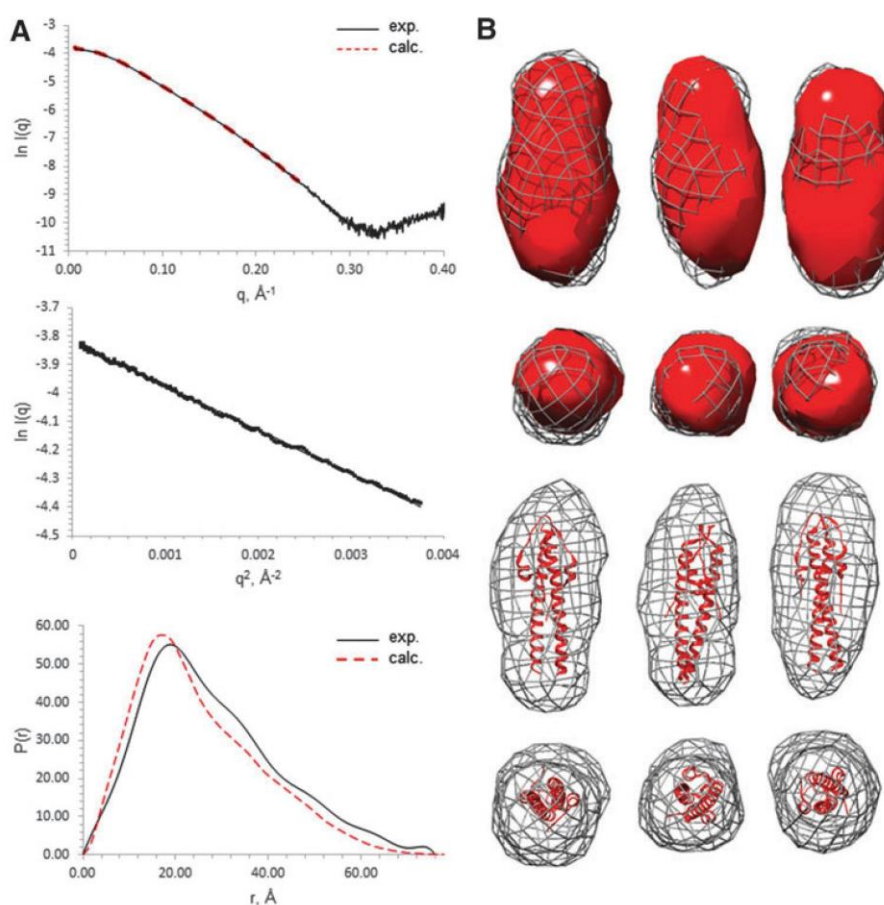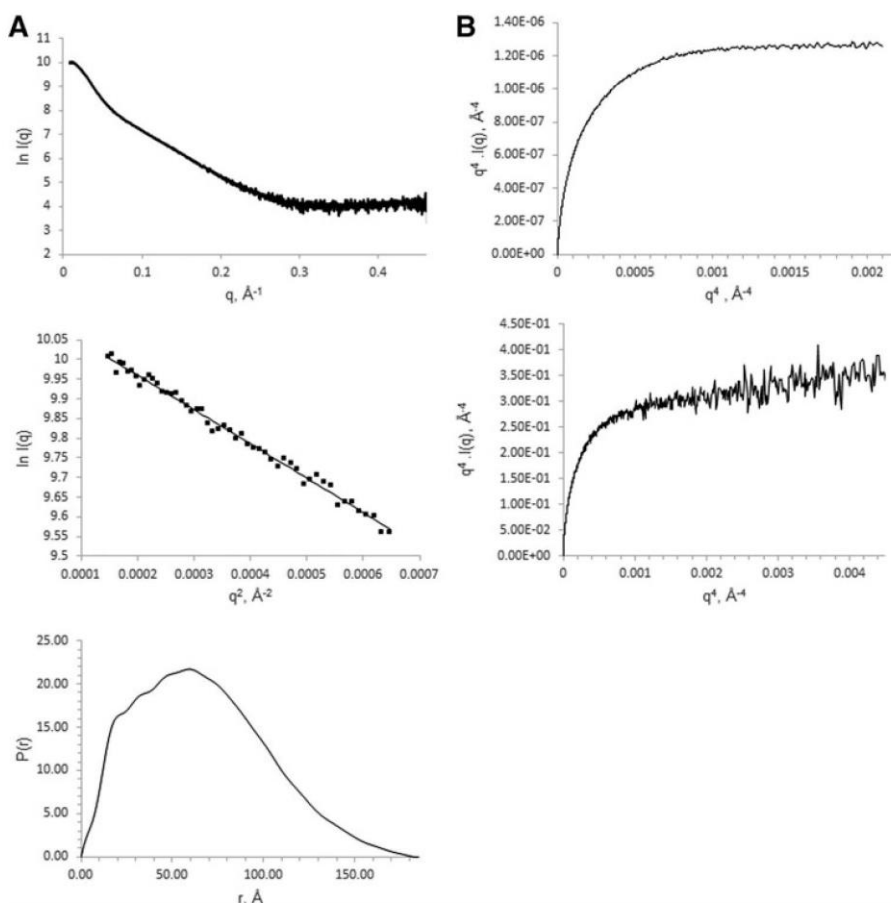

FIG. 2. SAXS analysis of *Lm*GpsB. (A) SAXS scattering curves (*top*), Guinier plots (*middle*), and pair distribution functions (*bottom*) for *Lm*GpsB. (B) Analysis of the flexibility of the N-*Lm*GpsB (*top*) and the *Lm*GpsB protein (*bottom*) by Porod–Debye plots. The plateau in the plot for N-*Lm*GpsB is consistent with a compact folded protein.

FIG. 3. SAXS molecular envelopes for *Lm*GpsB. **(A)** The average surface of 15 independently generated *ab initio Lm*GpsB dummy atom models is represented as a mesh. Three of these models are represented in **(B)** to illustrate the level of variation. The coordinates of N-*Lm*GpsB dimers (*red and cyan*) and the C-*Lm*GpsB trimers (*blue*) were docked manually and are shown in ribbon form (*right column*) or as equivalent electron density maps would appear at 25 Å resolution (*left, middle columns*). The coordinates shown correspond to residues 5–67 (N-terminal domain) and 89–106 (C-terminal domain) of *Lm*GpsB; *Lm*GpsB has a total of 113 residues. The models do not include the linker between domains. In the arrangements shown in **(B)**, with the simplest of the possible connectivities between domains, the chain termini are separated by distances of up to 55 Å. By comparison, a fully extended 21 polypeptide residue chain would be ~80 Å in length, and therefore, the models are compatible with the domains being connected by an appropriate length linker in a semi-compact conformation.
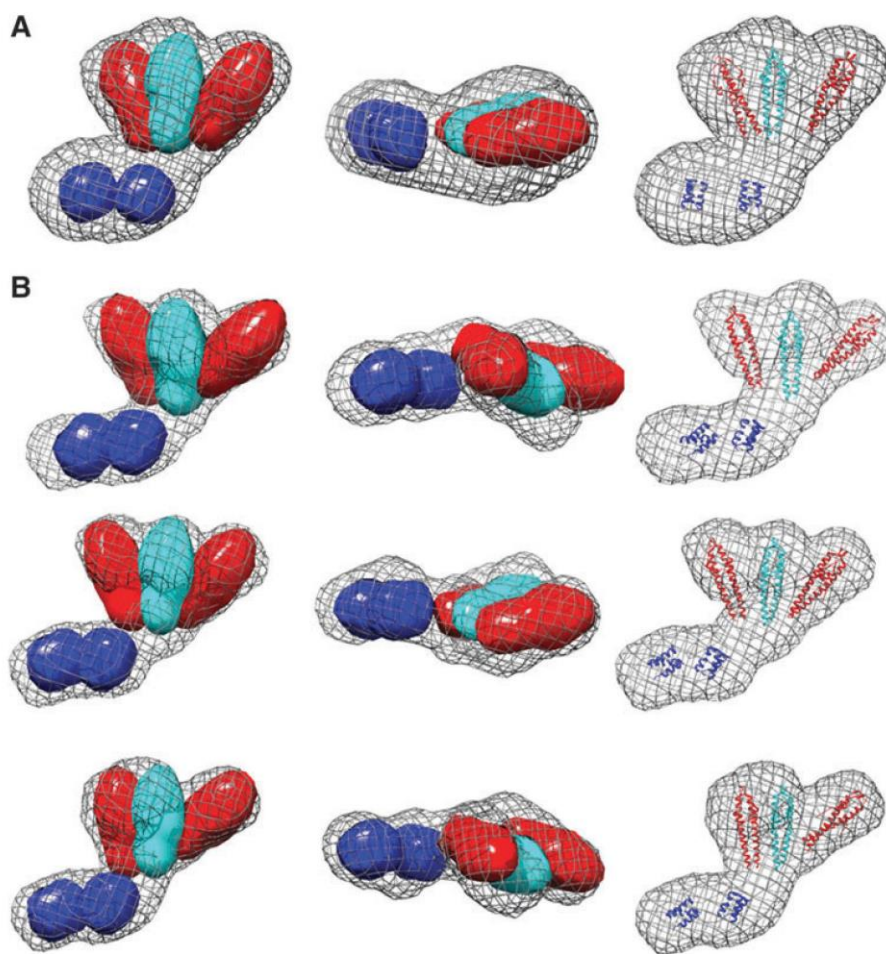
GpsB. Manually positioning the atomic coordinates of the N- and C-terminal domains of GpsB into the *ab initio* dummy atom model supports such a tripod arrangement (Fig. 3); the three N-GpsB dimers encompass the larger of the two lobes, and the two C-GpsB trimers correspond to the smaller one.

The importance of hexamer formation for GpsB activity in vivo

The tripod model positions the two C-*Lm*GpsB trimers in close proximity. The lattice interactions in the crystal structure of C-*Bs*GpsB (encompassing residues 76–98 of *Bs*GpsB) are thus of interest in considering how the subunits assemble to form the GpsB hexamer. Two highly conserved phenylalanines (F78, F92) stand out for being highly solvent exposed on the surface of the C-*Bs*GpsB trimer; within the crystal lattice both F78 and F92 are buried at the interface between adjacent trimers (Fig. 4B). To probe whether burial of these residues drives the assembly of subunits within the full-length hexameric GpsB, the equivalent residues in *Lm*GpsB, F91, and F105 were mutated to alanine, both singly and in combination. The oligomeric state of the wild-type and the mutant proteins was analyzed by SEC. Mutation of F105 to alanine both alone and in combination with

the F91A mutation substantially increased the retention volume of *Lm*GpsB mutants on SEC analysis relative to the wild-type *Lm*GpsB (Fig. 5A). Subsequent SEC-MALS analysis of the *Lm*GpsB$_{F91AF105A}$ double mutant confirmed this is due to a change in the oligomeric state; the double mutant formed a trimer at high protein concentrations and a dimer at lower concentrations (Fig. 5B). Mutation of F91A alone had an intermediate effect on assembly, with SEC analysis revealing two species; the predominant peak has a similar retention volume as the hexameric wild-type protein, while the other peak has a retention volume more similar to the F91AF105A double mutant (Fig. 5B).

The structural importance of F91 and F105 is further supported by thermal stability analysis using circular dichroism spectroscopy. Wild-type *Lm*GpsB has a T_m (the temperature of the midpoint of thermal unfolding) of 50°C, whereas the *Lm*GpsB$_{F91AF105A}$ double mutant has a T_m of 45°C (Fig. 6A). At the protein concentrations used for the CD analysis, the *Lm*GpsB$_{F91AF105A}$ double mutant was observed to be a dimer on SEC-MALS analysis (Fig. 6B), which may explain why *Lm*GpsB$_{F91AF105A}$ has a similar T_m as the dimeric, isolated N-*Lm*GpsB domain (Fig. 6A).

To further investigate the structural role of F105, the alanine substitution was made within the context of the
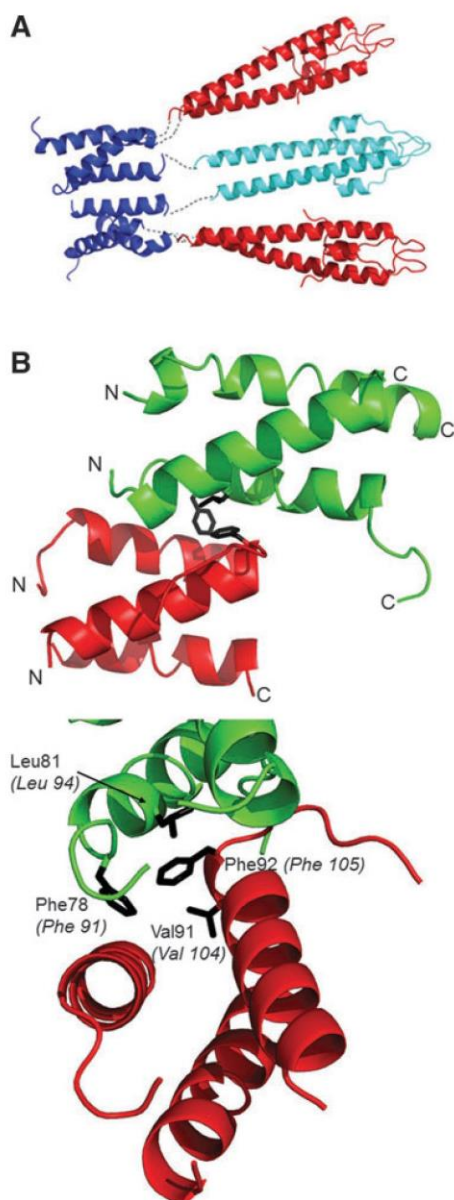
FIG. 4. Interaction between GpsB domains. **(A)** *A priori* hypothetical arrangement of the N- and C-terminal domains of GpsB, as described in the text. N-*Lm*GpsB dimers are colored *cyan and red*; C-*Lm*GpsB trimers *blue*. The putative path taken by interdomain linkers is represented as a *black dotted line*. Note the N-*Lm*GpsB dimer colored *cyan* is covalently linked to two separate C-*Lm*GpsB trimers and thus forms a bridge between two trimers that would stabilize association into a hexamer. **(B)** The interface between adjacent (*red and green*) C-*Bs*GpsB molecules as observed in its crystal lattice, with the key hydrophobic interfacial residues F78, L81, and F92 drawn as *sticks* and colored *black*. The *lower panel* also shows V91, a residue in the hydrophobic core of the C-*Bs*GpsB trimer, which packs against the interfacial residue F92. In the *lower panel*, the numbering of the *Listeria monocytogenes* GpsB equivalents of *Bs*GpsB F78, L81, V91, and F92 is shown in *italics* in parentheses.

isolated C-terminal domain. The isolated C-terminal domain forms trimers, rather than hexamers, in solution,[11] and F105 is exposed on the surface of this trimer, rather than in the hydrophobic core. F105 is unlikely to have an important structural role within the isolated C-terminal domain, but will stabilize the hexameric full-length protein. However, the C-*Lm*GpsB$_{F105A}$ fold was markedly destabilized as the alpha helical content was reduced from 70% in the wild-type C-*Lm*GpsB protein to 10% in the F105A mutant. This effect can perhaps be reconciled on closer analysis of the structure of C-*Bs*GpsB; F92 (equivalent to F105 in *L. monocytogenes* GpsB) is in close proximity to the hydrophobic core residue V91, and while one face of the F92 aromatic ring is surface exposed, the other face packs against V91 (Fig. 4B). The structural importance of F105 may therefore have two components; first, F105 stabilizes indirectly the folding of the isolated C-terminal domain because of its interaction with Val104, and second, F105 is likely to stabilize further interactions between C-GpsB trimers within the GpsB hexamer, by packing against Phe91 and Leu94 in another protomer in the hexamer (Fig. 4B). The dimers and trimers observed with the *Lm*GpsB$_{F91AF105A}$ double mutant likely arose from misfolding of at least the C-terminal domain, given the loss of secondary structure associated with the F105A mutation (Fig. 6A).

To validate the effect of mutations in the interface between C-GpsB trimers, we used a complementation assay to measure the effect of *gpsB* mutations on GpsB activity in *L. monocytogenes*. Briefly, the *L. monocytogenes* Δ*gpsB* mutant grew equally well as the wild-type strain at 30°C and was still viable at 37°C; however, the *L. monocytogenes* Δ*gpsB* mutant was not viable at 42°C, neither were *gpsB* mutants affected in self-interactions or interactions with the *L. monocytogenes* orthologue of *B. subtilis* PBP1, PBP A1 (for simplicity referred to herein as PBP1).[11] Mutations that individually exchanged the pair of phenylalanines (F91, F105, the equivalent *B. subtilis* residues are highlighted on the structure in Figure 4) and the adjacent L94 into alanines were introduced to *L. monocytogenes* *gpsB*, and the resulting alleles were tested in the complementation assay. As shown in Fig. 7A, the wild-type *L. monocytogenes* strain EGD-e proliferated readily at 42°C, whereas the Δ*gpsB* mutant (strain LMJR19) was unable to grow at all at this temperature. Reintroduction of wild-type *gpsB* into the Δ*gpsB* mutant (strain LMS56) repaired this defect, demonstrating successful complementation. The point mutations introduced fell into two classes: strain LMS185, expressing *gpsB*$_{F91A}$ was able to grow at 42°C, although with increased autolysis in the stationary phase as observed by phase contrast microscopy (not shown) and by growth curve analysis (Fig. 7A), suggesting that the *in vivo* activity of *Lm*GpsB$_{F91A}$ was partially impaired. By contrast, mutations L94A and F105A completely prevented growth of strains LMS186 and LMS187, respectively (Fig. 7A), indicating that these amino acid exchanges generated biologically inactive GpsB proteins. Western blotting showed that all the GpsB mutant proteins were expressed both during growth at 37°C, where all strains are viable (Fig. 7C), and also 2 hours after a temperature shift to 42°C (Fig. 7D). The loss of growth at 42°C is therefore more likely to result from a loss of a functional property of GpsB rather than a loss of protein production or its degradation.

Role of the phosphorylation of GpsB

It has been reported recently that GpsB was phosphorylated in *B. subtilis* cells at T75 (T88 in *Lm*GpsB) by the kinase PrkC.[19] The phosphomimetic mutations T75D and T75E appeared to have the same salt-sensitive phenotype as

a *gpsB* deletion.[19] Likewise, we found that the introduction of the analogous T88D exchange into *Lm*GpsB almost completely inactivated the protein. However, the phosphoablative T88A mutation was without effect on complementation activity of *Lm*GpsB (Fig. 7B, D). T75 in *Bs*GpsB and T88 in *Lm*GpsB are located in the linker between the two domains. To investigate whether the T75E and T75D mutations had any impact on the hexamerization of *Bs*GpsB, full-length, wild-type, T75E and T75D mutant *Bs*GpsB proteins (*Bs*GpsB$_{T75E}$ and *Bs*GpsB$_{T75D}$) were analyzed by SEC in conditions under which wild-type *Bs*GspB was shown by SEC-MALS to form a stable hexamer.[11] The T75E and the T75D mutations did not have a major effect on the size exclusion chromatograms of *Bs*GpsB; the major species in wild-type and both mutant proteins had an identical retention volume (Fig. 5C). Similarly, thermal denaturation analysis by circular dichroism did not reveal any marked difference in the structural stability of wild-type *Bs*GpsB and the T75D and T75E mutants (Fig. 6C).

Effect of BsGpsB on the glycosyltransferase and transpeptidase activity of BsPBP1 in vitro

To explore the functional role of GpsB and the relationship to its structure, the effect of *Bs*GpsB on the GTase and TPase activities of *Bs*PBP1 was investigated. *Bs*PBP1 was used as both the TGase and TPase activities of this enzyme have successfully been recapitulated *in vitro*,[46] which is not yet the case for the listeria PBPs. Both wild-type and phosphomimetic T75D/T75E mutants of *Bs*GpsB were used in these experiments in case any modulation of PBP1 activity by GpsB requires prior phosphorylation of GpsB. The GTase activity of *Bs*PBP1 was monitored using a fluorescence assay, in which the polymerization of fluorescently labeled lipid II followed by digestion of the resulting glycan chains by a muramidase results in a reduction of fluorescence intensity with time, correlated with the rate of the reaction. The addition of a twenty five-fold excess of either wild-type or the phosphomimetic T75D/T75E mutants of
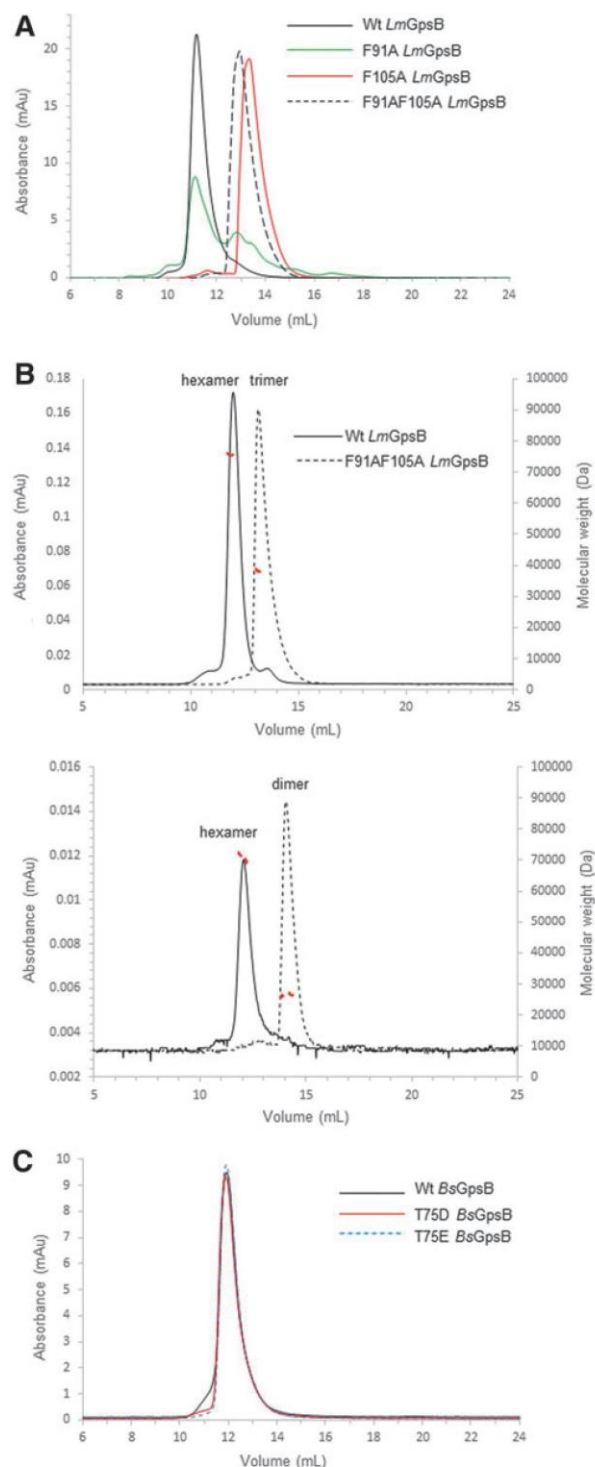
FIG. 5. SEC and SEC-MALS analysis of GpsB proteins. **(A)** SEC analysis of wild-type *Lm*GpsB (*solid black line*), *Lm*GpsB$_{F91A}$ (*solid green line*), *Lm*GpsB$_{F105A}$ (*solid red line*), and *Lm*GpsB$_{F91AF105A}$ (*dashed black line*) proteins at 3 mg/ml concentration. **(B)** SEC-MALS analysis of wild-type *Lm*GpsB (*solid black line*) and *Lm*GpsB$_{F91AF105A}$ (*dashed black line*) proteins. The chromatograms represent analysis at two injected protein concentrations of 8 mg/ml (*top*) and 0.5 mg/ml (*bottom*). The deconvoluted molecular masses of the eluting species (*red dashed lines*) are plotted on the *right hand* axis. The average of 76 kDa and 38.9 kDa across the major peak for wild-type *Lm*GpsB and *Lm*GpsB$_{F91AF105A}$ at 8 mg/ml is consistent with the theoretical mass of a hexamer (79 kDa) and trimer (38.6 kDa). At lower protein concentrations, the average mass of *Lm*GpsB$_{F91AF105A}$ is 26 kDa, consistent with the theoretical mass of a dimer (25.8 kDa), whereas the wild-type *Lm*GpsB remains hexameric. **(C)** SEC of wild-type *Bs*GpsB (*solid black line*), *Bs*GpsB$_{T75E}$ (*dashed blue line*), and *Bs*GpsB$_{T75D}$ (*solid red line*) mutants, all at 1 mg/ml. These phosphomimetic mutations do not introduce any significant change to the oligomeric state of *Bs*GpsB. SEC, size exclusion chromatography.
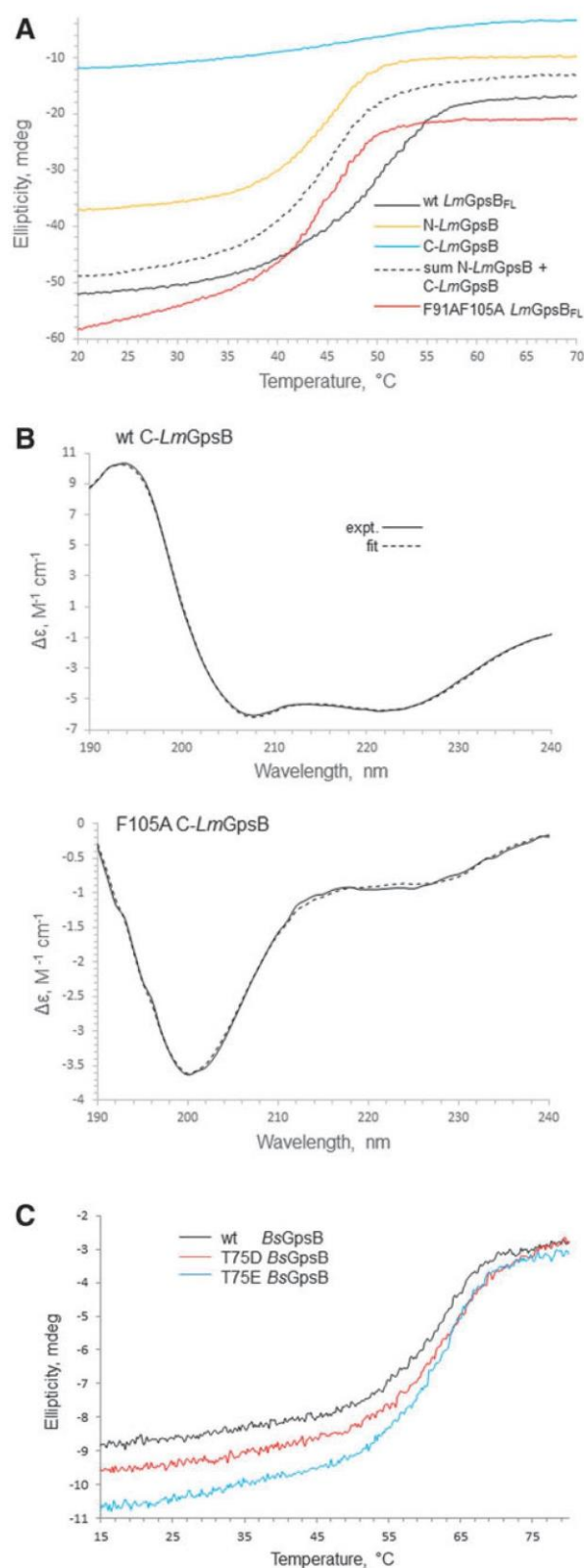
*Bs*GpsB had no apparent effect on the rate of the GTase reaction catalyzed by *Bs*PBP1 (Fig. 8A).

The effect of *Bs*GpsB on *Bs*PBP1 was explored further using a HPLC-based *in vitro* PG synthesis assay[41] to monitor the composition of peptidoglycan synthesized by *Bs*PBP1 in the presence and absence of *Bs*GpsB. Consistent with others' previous observations,[46] *Bs*PBP1 was active as a TPase when assayed with a native lipid II substrate that had been amidated at the ε-carboxylate of *meso*-diaminopimelic acid (lipid II-*m*-DAP). In this endpoint assay, PBP1 produced peptidoglycan with approximately 30% of the peptides present in cross-links, irrespective of the absence or the presence of a twenty five-fold molar excess of either wild-type or the phosphomimetic T75D/T75E mutants of *Bs*GpsB (Fig. 8B). As observed with other PBPs,[39] *Bs*PBP1 also exhibited a weak DD-carboxypeptidase activity that was unaffected by the presence of any *Bs*GpsB protein tested. Since *Bs*GpsB interacts with *Bs*PBP1 with a K_d of 0.7 μM,[11] at the protein concentrations used in both assays *Bs*PBP1 should be bound to *Bs*GpsB, suggesting that the interaction has no discernible effect on either the GTase or the TPase activities of *Bs*PBP1 *in vitro*.

Discussion

The quaternary structure and the connectivity between domains in GpsB are an intriguing geometric puzzle; GpsB is a hexamer, while the isolated N- and C-terminal domains are dimeric and trimeric, respectively. The arrangement of the domains from the SAXS analysis resembles a logical tripod arrangement, but with a more asymmetric shape than perhaps predicted *a priori*. The asymmetry may be driven, at least in part, by the interactions between C-GpsB trimers; the SAXS envelope suggests a "staggered" alignment of the two C-GpsB trimers, in which one of the trimers is closer to the center of mass of the GpsB hexamer than the other (Fig. 3). A similar staggered alignment is observed in the crystal packing of C-*Bs*GpsB. In this case, each trimeric C-*Bs*GpsB coiled-coil packs in a parallel manner against an adjacent trimer, but offset such that the N-terminal end of one coiled-coil packs against the C-terminal end of the other (Fig. 4B). This arrangement involves hydrophobic interactions between the highly conserved amino acids F78, L81, and F92 in *Bs*GpsB, equivalent to F91, L94, and F105 in

FIG. 6. Circular dichroism spectra. **(A)** Thermal denaturation of *Lm*GpsB (*black line*), N-*Lm*GpsB (*orange line*), C-*Lm*GpsB (*cyan line*), and *Lm*GpsB$_{F91AF105A}$ (*red line*), monitored by circular dichroism; the *dashed line* represents the summation of the individual thermal melts measured for N-*Lm*GpsB and C-*Lm*GpsB. Unfolding of the secondary structure is observed by monitoring ellipticity at 222 nm. The F91AF105A mutation reduces the stability of *Lm*GpsB. **(B)** Circular Dichroism spectra of wild-type C-*Lm*GpsB (*top*) and C-*Lm*GpsB$_{F105A}$ (*bottom panel*). The *dashed* spectrum represents the reconstructed spectra after fitting the secondary structural content with the program CDSStr (30). **(C)** Thermal denaturation of wt *Bs*GpsB (*black line*), *Bs*GpsB$_{T75D}$ (*red line*), and *Bs*GpsB$_{T75E}$ (*cyan line*) monitored by circular dichroism ellipticity at 222 nm. The midpoint of the unfolding transition is similar for the three proteins (*c.* 62°C).
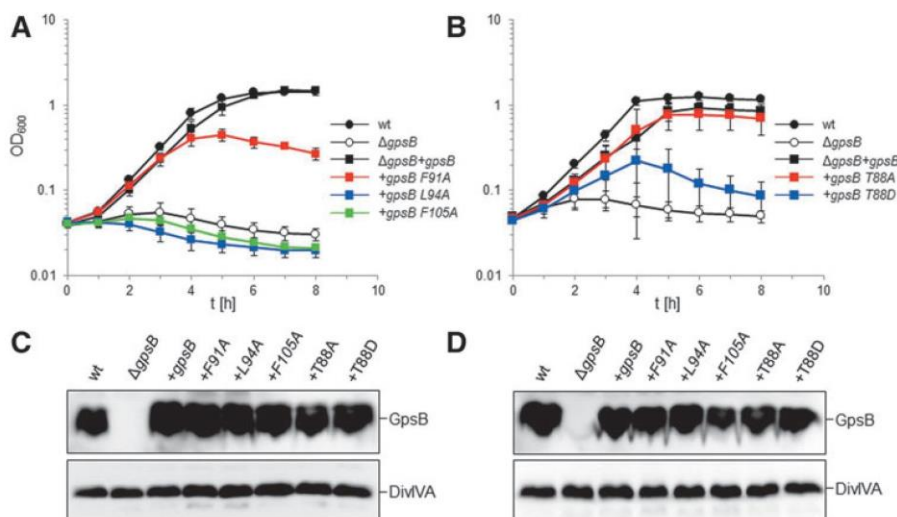
FIG. 7. Effect of mutations in the GpsB C-terminus on growth of *L. monocytogenes* at 42°C. **(A)** *L. monocytogenes* strains EGD-e (wt), LMJR19 (Δ*gpsB*), LMS56 (Δ*gpsB*+*gpsB*), LMS185 (Δ*gpsB*+*gpsBF91A*), LMS186 (Δ*gpsB*+*gpsBL94A*), and LMS187 (Δ*gpsB*+*gpsBF105A*) were grown in BHI broth containing 1 mM IPTG at 42°C, and growth was recorded in hourly intervals. **(B)** Growth of *L. monocytogenes* strains EGD-e (wt), LMJR19 (Δ*gpsB*), LMS56 (Δ*gpsB*+*gpsB*), LMJR161 (Δ*gpsB*+*gpsBT88A*), and LMJR162 (Δ*gpsB*+*gpsBT88D*) was measured as described in **(A)**. All average values and standard deviations were calculated from experiments performed in triplicate. **(C)** Western blots showing expression of *Lm*GpsB proteins in the same set of strains as in **(A, B)** during growth at 37°C (*top panel*). Cells were harvested for protein isolation at an optical density of 1.0 ($\lambda = 600$ nm). A parallel Western blot shows DivIVA expression for control (*bottom panel*). **(D)** Stability of *Lm*GpsB proteins after temperature upshift to 42°C. Western blot showing amounts of *Lm*GpsB proteins in cells that were grown at 37°C to an optical density of 1.0 and then shifted to 42°C for 2 further hours before protein isolation (*top panel*). Again, a parallel Western blot shows DivIVA protein levels for control (*bottom panel*).

*Lm*GpsB. From the SEC analysis, F105 is clearly more important for the stability of the hexamer than F91 (Fig. 5A) and this is fully consistent with the complementation assay, in which mutation of F105, as well as L94, has a larger impact than mutation of F91. The milder effect of the F91A mutation may be attributed to the F105A mutation having a greater impact on the concentration of GpsB hexamers *in vivo*. Overall, these observations support the conclusion that the interactions between C-GpsB domains in the GpsB hexamer are similar to those observed in the C-*Bs*GpsB crystal lattice, which is not predicted to be stable on analysis of the crystal packing with the PDB-PISA web server. However, weak interactions between isolated C-GpsB domains will be stabilized within GpsB by the covalently attached dimeric N-GpsB domains that will bridge two trimers in a hexamer. Such an effect could explain why the full-length GpsB protein has a higher T_m than the combined CD thermal melts of the isolated domains (Fig. 6A). The stabilizing effect is illustrated schematically in Fig. 4A, in which cyan is used to denote the N-*Lm*GpsB dimer that forms a bridge between two C-*Lm*GpsB trimers within the *Lm*GpsB hexamer. Previous mutagenesis results have demonstrated that the C-terminal domain is essential for formation of GpsB hexamers and that mutations in residues critical for C-GpsB trimer formation render GpsB inactive.[11] It would therefore appear from the loss of GpsB function in strains harboring *gpsB* alleles that affect GpsB hexamer formation that the correct assembly of GpsB is absolutely critical for its function *in vivo*.

With the arrangement of domains in the SAXS envelope, the distances between covalently linked domains are not equivalent. This asymmetric arrangement requires the linker between domains to adopt different conformations, and indeed secondary structure prediction analyses predict that the linker is disordered. The flexibility and bendability of the linker can be further reconciled with the high frequency of the disorder promoting residues[47] proline, lysine, methionine serine, and threonine in this region; the 23 amino acid *L. monocytogenes* GpsB linker has twelve such residues, which are also abundant in the linkers of orthologous GpsB proteins.[11]

The position of F91 and L94 close to or at the interface between subunits is particularly interesting in view of the recent identification that, in *B. subtilis*, the nearby T75 (equivalent to T88 in *L. monocytogenes*) is phosphorylated by the kinase PrkC[19] and that phosphomimetic mutations at this position produce a *gpsB* null-like phenotype both in *B. subtilis* and in *L. monocytogenes*. Similarly, phosphoablative mutations have no demonstrable effect in either species. The straddling of T88 by F91 and L94 suggests that the interaction between GpsB and PrkC will overlap with the interface between C-GpsB subunits, and hence, the interaction with PrkC could adjust the arrangement of subunits in the hexamer. The interaction with PrkC could therefore, in turn, influence the interaction with PBP1 by altering the GpsB quaternary structure and the arrangement of PBP1 binding sites. However, the phosphorylation of GpsB, mediated by PrkC, is not in itself sufficient to affect the interaction with PBP1 as both wild-type and phosphomimetic variants of GpsB behaved identically when combined with *Bs*PBP1 in activity assays (Fig. 8). We have shown here and elsewhere[11] that correct hexamer formation is essential for
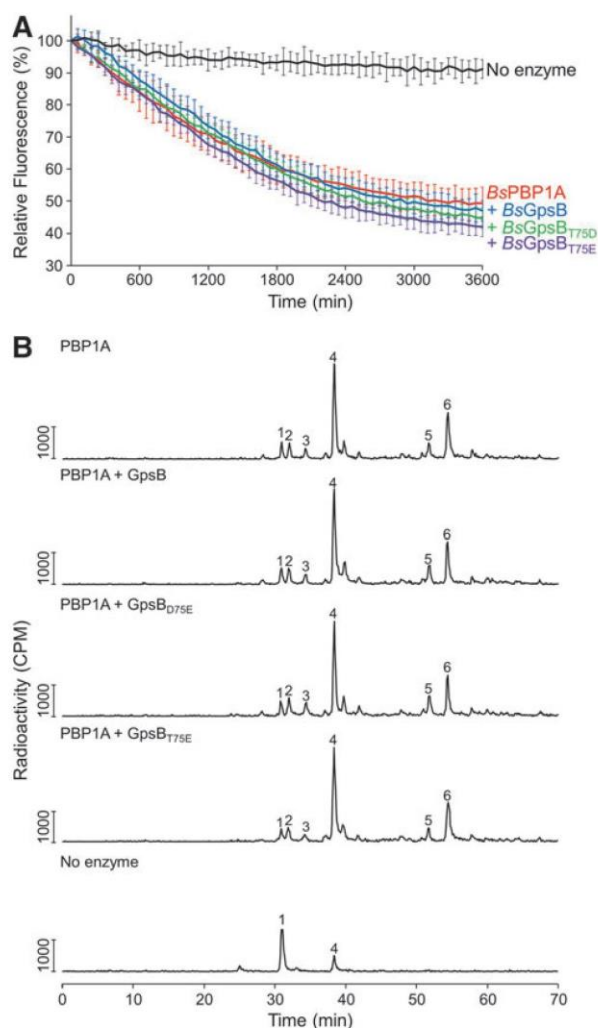
FIG. 8. GpsB does not affect the enzyme activities of PBP1 *in vitro*. (**A**) The GTase activity of *Bs*PBP1 in the presence (*blue line*) and absence of wild-type *Bs*GpsB (*red line*), *Bs*GpsB$_{T75D}$ (green line), and *Bs*GpsB$_{T75E}$ (*magenta line*) was measured using fluorescently labeled lipid II as described previously.[40] In the absence of *Bs*PBP1 (*black line*), no lipid II consumption is observed. Each measurement is shown as the mean $\pm$ SD ($n = 3$). (**B**) HPLC chromatograms from *in vitro* PG synthesis assays. *Bs*PBP1 in the presence and absence of *Bs*GpsB was incubated with radiolabeled amidated lipid II-*m*-DAP. The resultant PG was digested, boiled, and reduced before the resulting muropeptides were separated by HPLC. Peak 1—disaccharide pentapeptide(NH$_2$) monophosphate resulting from unused substrate and glycan chain ends; peak 2—disaccharide tetrapeptide(NH$_2$) derived from GTase and carboxypeptidase activity; peak 3—disaccharide pentapeptide (nonamidated) from the GTase activity on contaminating nonamidated substrate; peak 4—disaccharide pentapeptide(NH$_2$) resulting from the GTase activity; peak 5—bis-disaccharide tetrapeptide(2NH$_2$) resulting from GTase, TPase, and carboxypeptidase activities; peak 6—bis-disaccharide tetrapentapeptide(2NH$_2$) resulting from GTase and TPase activities. HPLC, high-pressure liquid chromatography.
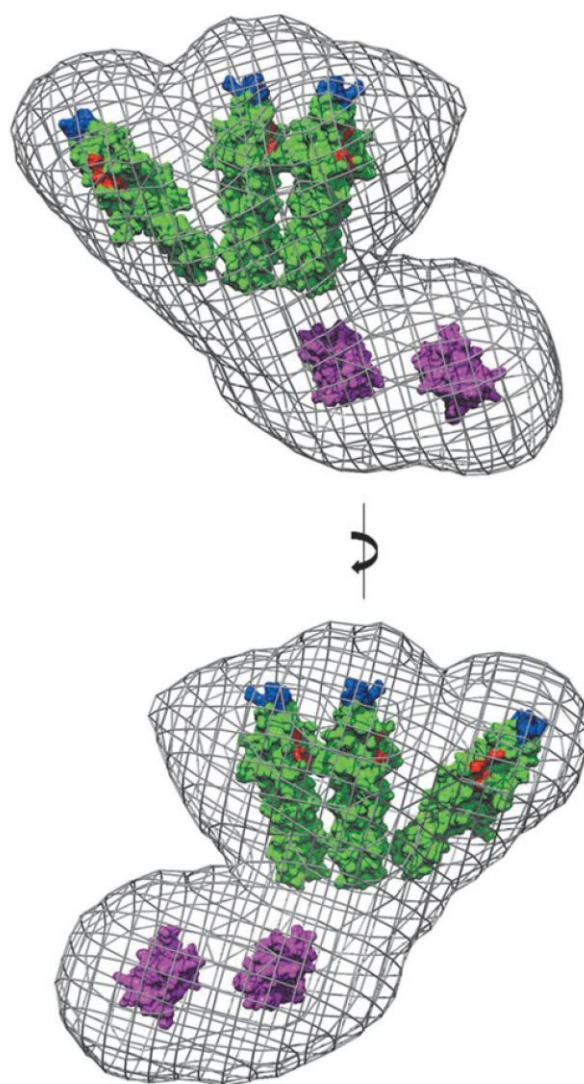
FIG. 9. PBP1 and membrane binding sites in the SAXS model of GpsB. The atomic coordinates of N-GpsB dimers and C-GpsB trimers, represented as molecular surfaces (*green and magenta*, respectively), have been manually positioned in the averaged *ab initio* molecular envelope of the *Lm*GpsB hexamer, shown as a *gray mesh*. Surface exposed residues critical for the interaction of *Lm*GpsB with PBP1[11] are colored *red* (D33, D37, I40). Residues critical for association of *Lm*GpsB with the cell membrane[11] are colored *blue* (L24, R25). The molecular envelope represented here is the average of 15 independently generated *ab initio* dummy atom models.

GpsB function in *L. monocytogenes*, which invites the suggestion that the phosphomimetic mutation instead affects GpsB function by altering its quaternary structure. However, a major effect on GpsB quaternary structure was not evident here as the two phosphomimetic mutants of *Bs*GpsB, T75D and T75E, behaved indistinguishably from wild-type *Bs*GpsB on SEC and had an identical thermal stability on circular dichroism analysis. Perhaps the *gpsB* null phenotype

of phosphomimetic mutations reflects the introduction of a net negative charge in this region affecting the ability of GpsB to interact with PrkC or another divisome component. A molecular rationale for the *gpsB* null phenotype associated with the phosphomimetic mutation reported earlier[19] and herein remains to be determined.

An important functional consequence of a closed tripod-like arrangement, as opposed to a more open "splayed out" arrangement of subunits, is the potential for the GpsB hexamer to interact simultaneously with multiple membrane-embedded PBP1 molecules. In the SAXS model of GpsB, the binding sites for PBP1 (a surface cleft in the N-terminal domain[11]) are clustered at one end of the elongated hexamer (Fig. 9). Similarly, the residues critical for the association of GpsB with the cell membrane (L24 and R25 in *Lm*GpsB[11]) are also clustered at the same end of the hexamer (Fig. 9). GpsB, therefore, has the implicit capacity to interact with multiple PBP1 molecules and thus to impact upon the arrangement of PBP1 molecules in the divisome, and this capacity may explain phenotypic and genetic evidence that GpsB regulates cell wall synthesis.[5,11] In *E. coli*, the essential cell division protein FtsN, which is unrelated to GpsB, may play a similar role as it stabilizes dimeric or multimeric forms of PBP1B *in vitro*.[48]

At 37°C, the Δ*gpsB* null mutant makes the overexpression of PBP1 lethal in *L. monocytogenes*,[11] which suggests that GpsB acts as a negative regulator of at least one of the enzymatic activities of PBP1. A simple regulatory mechanism could entail GpsB promoting the oligomerization of PBP1, a feasible scenario given that the GpsB hexamer has the potential to interact with multiple PBP1 molecules. PBP1B from *E. coli* forms dimers, and dimerization enhances both GTase and TPase activities *in vitro*.[41] The activities of *E. coli* PBP1B are stimulated or modulated by interactions with LpoB, FtsN, TolA, and CpoB, and these interactions are essential for the function of this peptidoglycan synthase in the cell.[39,48–53] However, there was no effect of *Bs*GpsB on either the GTase or the TPase activity of PBP1 *in vitro* (Fig. 8), which would argue against a regulatory mechanism. Any modulation of PBP1 activity by GpsB *in vivo* must therefore require other divisome components and/or membrane-bound elements that are not recapitulated in these experiments *in vitro*. In particular, simply by controlling the spatial arrangement of PBP1 molecules in the membrane *in vivo*, GpsB may influence both the arrangement of peptidoglycan strands and the pattern of intrastrand cross-links in the cell wall.

In this regard, the hexameric GpsB, and its flexible nature, could potentiate PBP1 activity or cellular localization by changes to the quaternary structure of GpsB: movements of the GpsB domains relative to one another could translate to large changes in the spatial arrangement of PBP1 binding sites. GpsB might thus act as an allosteric sensor, with its interaction with PBP1 modulated by changes in quaternary structure induced by interactions with other purported binding partners such as PrkC, EzrA, MreC, or DivIVA. These ideas will form the focus of our future experiments.

Acknowledgments

Disclosure Statement

No competing financial interests exist.

References

1. den Blaauwen, T., J.M. Anreu, and O. Monasterio. 2014. Bacterial cell division proteins as antibiotic targets. Bioorg. Chem. 55:27–38.
2. Gamba, P., J. Veening, N. Saunders, L. Hamoen, and R. Daniel. 2009. Two-step assembly dynamics of the *Bacillus subtilis* divisome. J. Bacteriol. 191:4186–4194.
3. Osawa, M., D.E. Anderson, and H.P. Erickson. 2008. Reconstitution of contractile FtsZ rings in liposomes. Science 320:792–794.
4. Typas, A., M. Banzhaf, C.A. Gross, and W. Vollmer. 2012. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. Nat. Rev. Microbiol. 10:123–136.
5. Claessen, D., R. Emmins, L. Hamoen, R. Daniel, J. Errington, and D. Edwards. 2008. Control of the cell elongation-division cycle by shuttling of PBP1 protein in *Bacillus subtilis*. Mol. Microbiol. 68:1029–1046.
6. Cleverley, R.M., J.R. Barrett, A. Baslé, N.K. Bui, L. Hewitt, A. Solovyova, Z.Q. Xu, R.A. Daniel, N.E. Dixon, E.J. Harry, A.J. Oakley, W Vollmer, and RJ Lewis. 2014. Structure and function of a spectrin-like regulator of bacterial cytokinesis. Nat. Commun. 5:5421.
7. Singh, J.K., R.D. Makde, V. Kumar, and D. Panda. 2007. A membrane protein, EzrA, regulates assembly dynamics of FtsZ by interacting with the C-terminal tail of FtsZ. Biochemistry 46:11013–11022.
8. Steele, V.R., A.L. Bottomley, J. Garcia-Lara, J. Kasturiarachchi, and S.J. Foster. 2011. Multiple essential roles for EzrA in cell division of *Staphylococcus aureus*. Mol. Microbiol. 80:542–555.
9. Jorge, A.M., E. Hoiczyk, J.P. Gomes, and M.G. Pinho. 2011. EzrA contributes to the regulation of cell size in *Staphylococcus aureus*. PLoS One 6:e27542.
10. Pedersen, L.B., E.R. Angert, and P. Setlow. 1999. Septal localization of penicillin-binding protein 1 in *Bacillus subtilis*. J. Bacteriol. 181:3201.
11. Rismondo, J., R.M. Cleverley, H.V. Lane, S. Großhennig, A. Steglich, L. Möller, G.K. Mannala, T. Hain, R.J. Lewis, and S. Halbedel. 2016. Structure of the bacterial cell division determinant GpsB and its interaction with penicillin binding proteins. Mol. Microbiol. 99:978–998.
12. Considine, K.M., R.D. Sleator, A.L. Kelly, G.F. Fitzgerald, and C. Hill. 2011. Identification and characterization of an

essential gene in *Listeria monocytogenes* using an inducible gene expression system. Bioeng. Bugs 2:150–159.

13. Song, J.H., K.S. Ko, J.Y. Lee, J.Y. Baek, W.S. Oh, H.S. Yoon, J.Y. Jeong, and J. Chun. 2005. Identification of essential genes in *Streptococcus pneumoniae* by allelic replacement mutagenesis. Mol. Cells 19:365–374.

14. Massidda, O., L. Nováková, and W. Vollmer. 2013. From models to pathogens: how much have we learned about *Streptococcus pneumoniae* cell division? Environ. Microbiol. 15:3133–3157.

15. Pinho, M.G., M. Kjos, and J.W. Veening. 2013. How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. Nat. Rev. Microbiol. 11:601–614.

16. Land, A.D., H.C. Tsui, O. Kocaoglu, S.A. Vella, S.L. Shaw, S.K. Keen, L.T. Sham, E.E. Carlson, and M.E. Winkler. 2013. Requirement of essential Pbp2x and GpsB for septal ring closure in *Streptococcus pneumoniae* D39. Mol. Microbiol. 90:939–955.

17. Thanassi, J.A., S.L. Hartman-Neumann, T.J. Dougherty, B.A. Dougherty, and M.J. Pucci. 2002. Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae.* Nucl. Acids Res. 30:3152–3162.

18. Fleurie, A., S. Manuse, C. Zhao, N. Campo, C. Cluzel, J.P. Lavergne, C. Freton, C. Combet, S. Guiral, B. Soufi, B. Macek, E. Kuru, M.S. VanNieuwenhze, Y.V. Brun, A.M. Di Guilmi, J.P. Claverys, A. Galinier, and C. Grangeasse. (2014). Interplay of the serine/threonine-kinase StkP and the paralogs DivIVA and GpsB in pneumococcal cell elongation and division. PLoS Genet. 10:e1004275.

19. Pompeo, F., E. Foulquier, B. Serrano, C. Grangeasse, and A. Galinier. 2015. Phosphorylation of the cell division protein GpsB regulates PrkC kinase activity through a negative feedback loop in *Bacillus subtilis*. Mol. Microbiol. 97:139–150.

20. Yeats, C., R.D. Finn, and A. Bateman. 2002. The PASTA domain: a beta-lactam-binding domain. Trends Biochem. Sci. 27:438.

21. Shah, I.M., M.H. Laaberki, D.L. Popham, and J. Dworkin. 2008. A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. Cell 135:486–496.

22. Lenarcic, R., S. Halbedel, L. Visser, M. Shaw, L.J. Wu, J. Errington, D. Marenduzzo, and L.W. Hamoen. 2009. Localisation of DivIVA by targeting to negatively curved membranes. EMBO J. 28:2272–2282.

23. Yun, R.H., A. Anderson, and J. Hermans. 1991. Proline in alpha-helix: stability and conformation studied by dynamics simulation. Proteins 10:219–228.

24. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

25. Zheng, L., U. Baumann, and J.L. Reymond. 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucl. Acids Res. 32:e115.

26. Monk, I.R., C.G. Gahan, and C. Hill. 2008. Tools for functional postgenomic analysis of *Listeria monocytogenes*. Appl. Environ. Microbiol. 74:3921–3934.

27. Thomaides, H.B., M. Freeman, M. El Karoui, and J. Errington. 2001. Division site selection protein DivIVA of *Bacillus subtilis* has a second distinct function in chromosome segregation during sporulation. Genes Dev. 15:1662–1673.

28. Peränen, J., M. Rikkonen, M. Hyvönen, and L. Kääriäinen. 1996. T7 vectors with modified T7lac promoter for expression of proteins in *Escherichia coli*. Anal. Biochem. 236:371–373.

29. Gill, S.C., and P.H. von Hippel. 1989. Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 182:319–326.

30. Sreerama, N., and R.W. Woody. 2000. Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. Anal. Biochem. 287:252–260.

31. Whitmore, L., and B.A. Wallace. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89:392–400.

32. Round, A.R., D. Franke, S. Moritz, R. Huchler, M. Fritsche, D. Malthan, R. Klaering, D.I. Svergun, and M. Roessle. 2008. Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. J. Appl. Cryst. 41:913–917.

33. Franke, D., A.G. Kikhney, and D.I. Svergun. 2012. Automated acquisition and analysis of small angle X-ray scattering data. Nucl. Instrum. Methods Phys. Res. A689:52–59.

34. Svergun, D.I. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Cryst. 25:495–503.

35. Volkov, V.V., and D.I. Svergun. 2003. Uniqueness of ab initio shape determination in small-angle scattering J. Appl. Cryst. 36:860–864.

36. Pettersen, E., T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comp. Chem. 25:1605–1612.

37. Rambo, R.P., and J.A. Tainer. 2013. Accurate assessment of mass, models and resolution by small-angle scattering. Nature 496:477–481.

38. Breukink, E., H.E. van Heusden, P.J. Vollmerhaus, E. Swiezewska, L. Brunner, S. Walker, A.J. Heck, and B. de Kruijff. 2003. Lipid II is an intrinsic component of the pore induced by nisin in bacterial membranes. J. Biol. Chem. 278:19898–19903.

39. Egan, A.J., J. Biboy, I. van't Veer, E. Breukink, and W. Vollmer. 2015. Activities and regulation of peptidoglycan synthases. Phil. trans. R. Soc. Lond. B Biol. Sci. 370:20150031.

40. Banzhaf, M., B. van den Berg van Saparoea, M. Terrak, C. Fraipont, A. Egan, J. Philippe, A. Zapun, E. Breukink, M. Nguyen-Distèche, T. den Blaauwen, and W. Vollmer. 2012. Cooperativity of peptidoglycan synthases active in bacterial cell elongation. Mol. Microbiol. 85:179–194.

41. Bertsche, U., E. Breukink, T. Kast, and W. Vollmer. 2005. In vitro murein (peptidoglycan) synthesis by dimers of the bifunctional transglycosylase-transpeptidase PBP1B from *Escherichia coli*. J. Biol. Chem. 280:38096–38101.

42. Svergun, D.I., C. Barberato, and M.H.J. Koch. 1995. CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Cryst. 28:768–773.

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

44. Rambo, R.P., and J.A. Tainer. 2011. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers 95:559–571.

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

46. Lebar, M.D., J.M. May, A.J. Meeske, S.A. Leiman, T.J. Lupoli, H. Tsukamoto, R. Losick, D.Z. Rudner, S. Walker, and D. Kahne. 2014. Reconstitution of peptidoglycan cross-linking leads to improved fluorescent probes of cell wall synthesis. J. Am. Chem. Soc. 136:10874–10877.

47. Linding, R., L. Jensen, F. Diella, P. Bork, T. Gibson, and R. Russell. 2003. Protein disorder prediction: implications for structural proteomics. Structure 11:1453–1459.

48. Müller, P., C. Ewers, U. Bertsche, M. Anstett, T. Kallis, E. Breukink, C. Fraipont, M. Terrak, M. Nguyen-Distèche, and W. Vollmer. 2007. The essential cell division protein FtsN interacts with the murein (peptidoglycan) synthase PBP1B in *Escherichia coli*. J. Biol. Chem. 282:36394–36402.

49. Typas, A., M. Banzhaf, B. van den Berg van Saparoea, J. Verheul, J. Biboy, R.J. Nichols, M. Zietek, K. Beilharz, K. Kannenberg, M. von Rechenberg, E. Breukink, T. den Blaauwen, C.A. Gross, and W. Vollmer. 2010. Regulation of peptidoglycan synthesis by outer-membrane proteins. Cell 143:1097–1109.

50. Egan, A.J., N.L. Jean, A. Koumoutsi, C.M. Bougault, J. Biboy, J. Sassine, A.S. Solovyova, E. Breukink, A. Typas, W. Vollmer, and J.P. Simorre. 2014. Outer-membrane lipoprotein LpoB spans the periplasm to stimulate the peptidoglycan synthase PBP1B. Proc. Natl. Acad. Sci. USA. 111:8197–8202.

51. Gray, A.N., A.J. Egan, I.L. Van't Veer, J. Verheul, A. Colavin, A. Koumoutsi, J. Biboy, A.F. Altelaar, M.J. Damen, K.C. Huang, J.P. Simorre, E. Breukink, T. den Blaauwen, A. Typas, C.A. Gross, and W. Vollmer. 2015. Coordination of peptidoglycan synthesis and outer membrane constriction during *Escherichia coli* cell division. Elife 4:e07118.

52. Paradis-Bleau, C., M. Markovski, T. Uehara, T.J. Lupoli, S. Walker, D.E. Kahne, and T.G. Bernhardt. 2010. Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases. Cell 143:1110–1120.

53. Lupoli, T.J., M.D. Lebar, M. Markovski, T. Bernhardt, D. Kahne, and S. Walker. 2014. Lipoprotein activators stimulate *Escherichia coli* penicillin-binding proteins by different mechanisms. J. Am. Chem. Soc. 136:52–55.

54. Glaser, P., L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloecker, P. Brandt, T. Chakraborty, A. Charbit, F. Chetouani, E. Couve, A. de Daruvar, P. Dehoux, E. Domann, G. Dominguez-Bernal, E. Duchaud, L. Durant, O. Dussurget, K.D. Entian, H. Fsihi, F. Garcia-del Portillo, P. Garrido, L. Gautier, W. Goebel, N. Gomez-Lopez, T. Hain, J. Hauf, D. Jackson, L.M. Jones, U. Kaerst, J. Kreft, M. Kuhn, F. Kunst, E. Kurapkat, E. Madueno, A. Maitournam, J.M. Vicente, E. Ng, H. Nedjari, G. Nordsiek, S. Novella, B. de Pablos, J.C. Perez-Diaz, R. Purcell, B. Remmel, M. Rose, T. Schlueter, N. Simoes, A. Tierrez, J.A.Vazquez-Boland, H. Voss, J. Wehland, and P. Cossart. 2001. Comparative genomics of Listeria species. Science 294:849–852.

Address correspondence to:
Richard J. Lewis, MA, DPhil
Institute for Cell and Molecular Biosciences
University of Newcastle
Newcastle upon Tyne NE2 4HH
United Kingdom

E-mail: r.lewis@ncl.ac.uk

12.3 **ICP4 Paper**

The herpes viral transcription factor ICP4 forms a novel DNA recognition complex

Richard B. Tunnicliffe[1], Michael P. Lockhart-Cairns[2,3], Colin Levy[1], A. Paul Mould[4], Thomas A. Jowitt[4], Hilary Sito[1], Clair Baldock[2], Rozanne M. Sandri-Goldin[5] and Alexander P. Golovanov[1,*]

[1]Manchester Institute of Biotechnology, School of Chemistry, Faculty of Science and Engineering, The University of Manchester, Manchester M1 7DN, UK, [2]Wellcome Trust Centre for Cell-Matrix Research, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, M13 9PT, UK, [3]Diamond Light Source, Harwell Science and Innovation Campus, Fermi Ave, Didcot OX11 0QX, UK, [4]Biomolecular Analysis Core Facility, Faculty of Biology, Medicine and Health, University of Manchester, M13 9PT, UK and [5]Department of Microbiology and Molecular Genetics, School of Medicine, University of California, Irvine, CA 92697-4025, USA

ABSTRACT

The transcription factor ICP4 from herpes simplex virus has a central role in regulating the gene expression cascade which controls viral infection. Here we present the crystal structure of the functionally essential ICP4 DNA binding domain in complex with a segment from its own promoter, revealing a novel homo-dimeric fold. We also studied the complex in solution by small angle X-Ray scattering, nuclear magnetic resonance and surface-plasmon resonance which indicated that, in addition to the globular domain, a flanking intrinsically disordered region also recognizes DNA. Together the data provides a rationale for the bi-partite nature of the ICP4 DNA recognition consensus sequence as the globular and disordered regions bind synergistically to adjacent DNA motifs. Therefore in common with its eukaryotic host, the viral transcription factor ICP4 utilizes disordered regions to enhance the affinity and tune the specificity of DNA interactions in tandem with a globular domain.

INTRODUCTION

Herpes simplex virus-1 (HSV-1) causes lifelong infections, typified by the sporadic appearance of acute localized symptoms such as cold sores, inter-dispersed by prolonged asymptomatic periods where the virus remains in a latent state. HSV-1 and other alphaherpesviruses can also cause more severe diseases, such as keratitis and encephalitis, and have been linked with the development of Alzheimer's disease (1). Due to HSV's persistence in certain types of cells and life-long infection, it has also been modified for the de-velopment of gene delivery systems for the treatment of genetic diseases and cancer, and therefore a detailed understanding of gene regulation within this virus is invaluable (2,3). During herpes infection a sequential cascade of viral gene expression is triggered. Initially five 'immediate-early' (IE) genes (4) followed by more numerous early (E) and then late (L) genes are transcribed (5). The IE gene product, infected cell protein 4 (ICP4) is a transcriptional regulator with a prominent role within this cascade (6,7). ICP4 can induce the expression of E and L genes (8,9), while conversely it can act as a repressor notably of itself and also other IE genes (10–12). It carries out these functions by interacting with DNA and modulating the activity of the cellular RNA polymerase II on viral genes (13–16).

HSV-1 ICP4 is a 1298 amino acid nuclear phosphoprotein that has been the subject of extensive biochemical studies, which have established that it homo-dimerizes and adopts an elongated conformation (17–19). ICP4 is composed of four major domains: N-terminal activation, DNA binding (DBD), linker region and C-terminal activation (CTA) (Figure 1A). Sequence homology to helix-turn-helix and uracil-DNA glycosylase domains was observed for the DBD and CTA domains respectively (16,20), other domains are predicted to be predominately disordered. ICP4 homo-dimerization is mediated by the DBD, this region interacts preferentially with a bi-partite and asymmetric DNA consensus sequence RTCGTCNNYNYSG (where R is a purine, Y is a pyrimidine, S is a C or G and N is any base) (17,20–22). Extensive studies using ICP4 point mutants have probed the functional significance of residues within the protein (Figure 1B and Supplementary Table S1) (13,14,16). These studies highlighted the functional importance of the DBD as mutations that disrupted DNA interactions and also negatively affected both the transactiva-
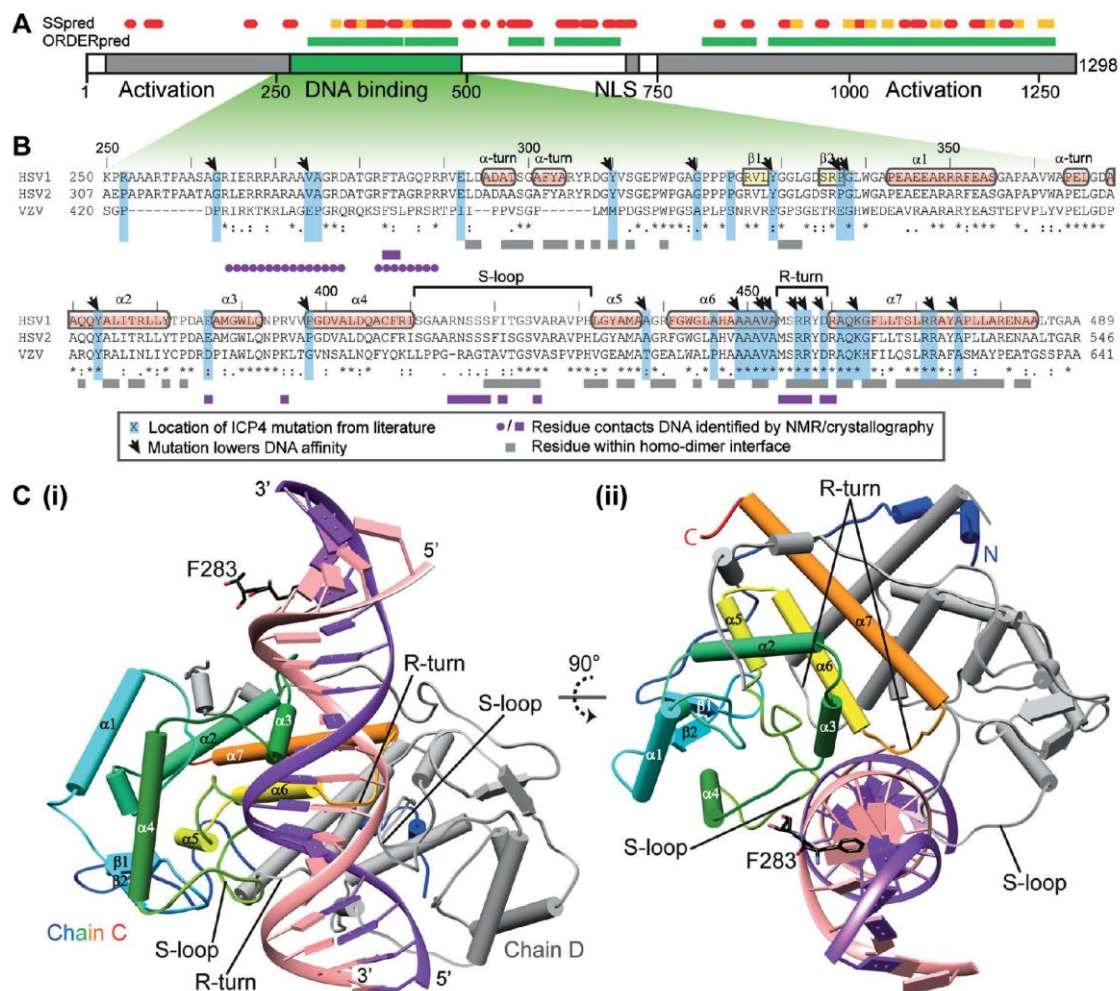
Figure 1. Summary of ICP4 domains and sites of DNA interaction and homo-dimerization. (**A**) The domain composition of HSV1 ICP4 with predicted secondary structure (SSpred) and ordered regions (ORDERpred) from PSIPRED (37). Predicted α-helices and β-sheets are shown as red and yellow respectively, with ordered regions colored green. (**B**) Sequence alignment of ICP4 DBDs from HSV1, HSV2 and VZV (Uniprot codes: P08392, P90493, Q8AZM1 respectively) from Clustal omega (36). Secondary structure elements determined here are labeled on the HSV-1 sequence. Residues previously probed by mutagenesis are highlighted blue with black arrows pointing to those with lower affinity for DNA (13,14,16). Below the sequences, blocks colored gray or purple indicate residues which form homo-dimer or DNA contacts respectively as observed in the crystallography data. Purple circles indicate protein–DNA contacts derived from NMR chemical shift perturbations in intrinsically disordered regions (IDRs). (**C**) Cartoon representation of the crystal structure with helices, sheets and loops shown as cylinders, arrows and coil respectively: (i) ICP4N·IE3_19mer with DNA colored purple and pink for sense and anti-sense strands respectively, one protein chain is colored gray and other blue through red from N-to-C termini. (ii) As with panel i with view rotated 90°.

tion and transrepression functions of ICP4, along with viral replication (13). Interpretation of these effects in the context of the structure of ICP4 was however not possible thus far.

The transrepression function occurs via the interaction of ICP4 with a consensus DNA sequence in viral promoter regions, for example ICP4's own gene (IE3) contains a consensus site located in proximity to the transcription start site (23,24), to which it can bind with nano-molar affinity (25,26). At these repression sites, ICP4 forms a tripartite complex (TPC) with the cellular proteins TFIIB plus TATA-binding protein (TBP) or TFIID (27). Formation of this complex is functionally important as mutants of ICP4 that cannot form a TPC but retain an ability to bind DNA

are unable to repress transcription (27,28). The genes of the remaining four HSV-1 IE proteins also contain sites matching the ICP4 consensus sequence. Similarly, ICP4 mediated transactivation also involves the DBD (13), likely via interaction with consensus DNA sites present in E and L genes, however intriguingly these sites are not essential for ICP4 function possibly due to the protein's ability to bind non-consensus DNA (29–31). ICP4 constructs containing both the DBD and the CTA domains can oligomerize on DNA, a property that may increase DNA affinity or specificity for the E or L genes (26). In addition, activation is mediated by interactions with cellular transcription factors via the N-terminal activation domain, specifically TFIID and

mediator, and further enhanced by additional interactions with the CTA domain (32,33). A further layer of complexity to ICP4 function is provided by the herpes simplex IE protein ICP0 and the ORF O protein, both of which have been identified as antagonists to the ICP4–DNA interaction (34,35). Therefore the DNA binding domain of ICP4 has a role in transcriptional regulation of viral genes throughout viral replication during lytic infection, but its sequence-specificity is most crucial for interactions with IE promoters.

Despite the general importance of ICP4 for HSV infection and the prominent role of the DBD in gene regulatory functions, no structural data were available for ICP4, and the mode of viral DNA recognition was unclear. To understand the details of ICP4–DNA recognition, we have solved crystal structures of the DBD in complex with DNA fragments from its own promoter. Additionally, we used a combination of solution techniques (small angle X-Ray scattering, nuclear magnetic resonance (NMR), multi-angle light scattering and analytical ultracentrifugation) plus surface-plasmon resonance experiments to determine the contribution to DNA binding of flanking intrinsically disordered regions (IDRs) not observed in the crystallography data. Together, the data revealed the details of both specific DNA recognition and dimerization of ICP4, and finally clarifies the results of previous mutational studies. The reported results should inform future functional studies in HSV-1 and provide an example of the synergistic action of globular and disordered regions for tuning DNA binding specificity.

MATERIALS AND METHODS

Cloning and expression

Sequence conservation (36), predictions of secondary structure and disorder (37) (Figure 1A and B) along with data from the literature (13,14,16,22,38) suggested the complete DNA binding domain (DBD) of ICP4 is comprised within residues 258–487. Therefore DNA encoding an HRV3C protease cleavable N-terminal Strep-tag with ICP4 residues 258–487 (ICP4N) was obtained by gene synthesis (Invitrogen), codon optimized for expression in *Escherichia coli*, a shorter construct of residues 288–487 (ICP4NΔIDR) was similarly obtained. The DNA fragments were individually cloned into the NdeI and XhoI restriction sites of pET-21a(+) (Merek). Both proteins were expressed in the same conditions using *E. coli* strain T7 Express LysY (New England Biolabs). Terrific Broth (Sigma) supplemented with 50 μg/mL ampicillin was inoculated with 1% v/v overnight pre-culture. Culture density was monitored at 600 nm until OD 0.6, at which point protein expression was induced with 1 mM IPTG and incubation continued for 5 h at 37 °C. Cells were pelleted by centrifugation (5000 *g*, 20 min). Selenomethionine (SeMet) labeled ICP4N was obtained by growing cells in M9 minimal media in place of Terrific Broth, using the protocol described by Van Duyne *et al.* (39). Uniformly ^{15}N-labeled proteins were obtained by growth in M9 minimal supplemented with ^{15}N-ammonium chloride.

Protein purification

Cell pellets were resuspended in ice cold running buffer (50 mM HEPES, 500 mM NaCl, 50 mM L-Arg, 50 mM L-Glu (40), 0.5 mM TCEP, pH 7.9) supplemented with 0.5% v/v Triton X-100, DNase, RNase and ethylenediaminetetraacetic acid free protease inhibitor (Roche). The cell suspension was lysed by sonication and clarified by centrifugation (38000 *g*, 30 min, 4°C) then the supernatant was passed through a 0.2 μm filter. The supernatant was purified using Strep-Tactin Superflow high capacity resin (IBA life sciences) in a gravity flow column, and bound material was eluted with 5 mM d-desthiobiotin in running buffer. The N-terminal Strep-tag was cleaved by incubating eluted protein with HRV3C protease (Sigma) for 16 h at 4°C. For surface plasmon resonance (SPR) experiments, to ensure complete Strep-tag removal, cleavage was carried out on column for 16 h at 4°C, and then the cleaved protein eluted from the column with running buffer and passed through a clean Strep-Tactin column. Finally the protein was purified on a Superdex 75 26/600 column (GE healthcare) pre-equilibrated in gel filtration buffer (20 mM HEPES, 150 mM NaCl, 50 mM L-Arg, 50 mM L-Glu, 0.5 mM TCEP, pH 7.9).

In order to study the DNA interactions of ICP4N constructs, synthetic IE3 DNA oligos were purchased (Invitrogen), namely IE3_19mer forward: CCGATCGTCCAC ACGGAGC and reverse-complement: GCTCCGTGTG GACGATCGG, IE3_19merMUT forward: CCGATCGT CCAAGATTAGC and reverse complement: GCTAATCT TGGACGATCGG, plus IE3_12mer forward: CCGATC GTCCAC and reverse-complement: GTGGACGATCGG. DNA oligos were solubilized in water and mixed in a 1:1 molar ratio, then annealed by heating to 90°C for 10 min then cooled to 20°C at 1°C/min. For the formation of protein–DNA complexes, 1 mg/ml ICP4N or ICP4NΔIDR were incubated with annealed DNA for 16 h at 4°C, at a molar ratio of 1:1.3 (protein dimer: DNA duplex). The protein–DNA solution was concentrated 10-fold in a Vivaspin 500 centrifugal device with a 5 kDa MWCO (Sartorius Stedim Biotech GmbH) prior to crystallization screens.

Crystallization

All ICP4N·IE3_19mer and ICP4NΔIDR·IE3_12mer crystals were obtained by the same method: Protein–DNA mixtures (at 1:1.3 molar ratio) concentrated to 10 mg/ml were used to set up 5 × 96 crystal trials and screened by the sitting drop vapor diffusion method. A 200 nl drop of protein–DNA concentrate was mixed with 200 nl of the screen condition using a TTP Mosquito Crystal nanolitre pipetting robot. Following 72 h incubation at 4 °C the plates were manually inspected and single crystals suitable for X-ray diffraction analysis were observed in a range of conditions. SeMet derivatized and native crystals of ICP4N·IE3_19mer grew from reservoir solutions consisting of 0.2 M ammonium sulphate, 0.1 M Bis/Tris pH 5.5, 25% w/v PEG 3350 (SG1 HT96 B8 Molecular Dimensions), crystals were cryoprotected with 20% PEG 200. ICP4NΔIDR·IE3_12mer grew from 0.2 M ammonium acetate trihydrate, 0.1 M Sodium HEPES pH 7.5, 25% w/v PEG 3350 (SG1 HT96 F2 Molecular Dimensions) and cryoprotected with Perfluoropolyether Cryo Oil. All crystals were flash frozen by

plunge freezing in liquid nitrogen prior to data collection at Diamond Light Source Ltd.

Data collection, structure determination, model building and refinement

Data were collected from single cryo frozen crystals of ICP4N·IE3_19mer and ICP4NΔIDR·IE3_12mer at beamlines i04 and i02 respectively (Diamond Light Source). A high redundancy dataset was collected for a selenomethionine derivatized ICP4N·IE3_19mer crystal to a resolution of 2.45 Å. In addition native data were collected for both ICP4N·IE3_19mer (2.28 Å) and ICP4NΔIDR·IE3_12mer (2.12 Å). All data were indexed, scaled and integrated with Xia2 (41).

Phases for the SeMet derivative of ICP4N·IE3_19mer were determined by the single-wavelength anomalous diffraction (SAD) method using Fast EP as implemented at Diamond Light Source (42,43). Three selenium sites per monomer were located, 12 in total with a CC_{all}/CC_{weak} of 36.65/25.42 in SHELXE (44). An automated build against the phased map in Phenix AutoBuild produced a partial model which was used as the basis for iterative cycles of rebuilding and refinement in COOT and Phenix.refine against the two high resolution native datasets (45,46). Complete data collection and refinement statistics are available in Table 1. Validation with both MolProbity and PDB_REDO were integrated into the iterative rebuild process (47,48).

Solution small angle X-ray scattering

Samples of free ICP4N, free IE3_19mer and ICP4N·IE3_19mer complex were prepared as previously described; to remove any un-bound DNA, the complex was then further purified by an additional size exclusion chromatography (SEC) step using a Superdex 75 26/600 column equilibrated in gel filtration buffer. Samples were then exhaustively dialyzed into 20 mM HEPES pH 7.4, 150 mM NaCl, 0.1 mM TCEP and concentrated in a Vivaspin 500 centrifugal device with a 5 kDa MWCO to 10 mg/ml. SAXS intensity data, $I(q)$ versus q ($q = 4\pi.sin2\theta\lambda$), of ICP4N IE3_19mer complex and ICP4N were collected using SEC-SAXS and the BioSAXS robot, respectively, on beamline B21 at Diamond Light Source (Didcot, UK) and the IE3_19mer on beamline BM29 at the ERSF (Grenoble, France). At B21, the ICP4N IE3_19mer complex was further purified using a Shodex KW-403 SEC column and Agilent HPLC before exposure to X-rays to isolate the ICP4N IE3_19mer complex from any dissociated monomer. A total of 50 μl of ICP4N·IE3_19mer complex was loaded onto the Shodex column and the eluent was flowed through the SAXS beam at 0.15 ml/min; the buffer used as the background was collected after one SEC column volume. SAXS data were collected at 1 s intervals using a Pilatus 2M detector (Dectris, Switzerland) at a distance of 3.9 m and an X-ray wavelength of 1 Å. A total of 30 μl of ICP4N was loaded into a 96 well plate and loaded into the BioSAXS robot. The sample was exposed to X-rays for eighteen 10-s frames, with buffer being exposed pre- and post-sample to ensure the sample cell is free of contamination. At BM29 samples

for SAXS were purified using a Superdex 200 increase 3.2/300 SEC column and Shimadzu HPLC before exposure to X-rays. A total of 50 μl of IE3_19mer was loaded onto the Superdex column and the eluent was flowed through the SAXS beam at 0.075 ml/min; the buffer used as the background was collected after one SEC column volume. The SAXS data were collected at 1-s intervals using a Pilatus 1M detector (Dectris, Switzerland) at a distance of 2.9 m and an X-ray wavelength of 0.992 Å. For each beamline data were reduced using in-house software. Subtractions of the SEC-SAXS data were completed for each frame across the elution peak and the radius of gyration (R_g) and the integral of intensity ratio to background were plotted. The data were scaled, merged and averaged for each frame with a consistently similar R_g. All further processing and analysis of data was carried out using ScÅtter (http://www.bioisis.net/scatter). Comparison of the crystal structure to the SAXS data was completed using the FoXS online server for computation and fitting (49,50).

Ab initio model generation

Dummy atom models (DAMs) were generated for the ICP4N·IE3_19mer complex using DAMMIF (51) in slow mode through ScÅtter. The calculated curves from DAMs were compared to the experimental data and the agreement was shown by chi squared (χ^2) values ranging from 1.34 to 1.36. The models from 17 independent DAMMIF runs were averaged using the DAMAVER suite with a mean normalized spatial discrepancy (NSD) of 0.66 ± 0.05 (standard deviation). Thirty five biphasic MONSA models were generated using the ATSAS online server. Phase one was defined by the DNA and phase two defined by the volume difference of the complex and DNA, equating to the protein contribution. The runs were split into phases before averaging the models using adapted scripts from (52).

DAMs of DAMMIF and MONSA were visualized by generating an electron density map at a resolution of 15 Å via the 'molmap' command in UCSF Chimera (53). The crystal structure was docked into the electron density using Chimeras 'Fit in map' command. For analysis of the conformation of free ICP4N, protein chains C and D were extracted from the ICP4N·IE3_19mer coordinates and missing residues in the N-terminal protein chains generated as a random coil using the Modeler function in UCSF Chimera. This seed structure was used by MultiFoXS to generate 10000 conformations and then select an ensemble that best represented the free ICP4N SAXS profile, residues 258–289 were defined as mobile and the dimer formed by residues 290–487 as a single rigid body (50).

Biophysical characterization of ICP4N dimerization

Samples of ICP4N·IE3_19mer complexes were prepared as previously described, prior to size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) analysis. Samples of ICP4N, ICP4NΔIDR and ICP4N·IE3_19mer (0.5 ml at 1 mg/ml) were loaded onto either a Superdex 75 10/300GL or a Superdex 200 10/300GL column (GE life-sciences, 0.75 ml/min in gel filtration buffer) and passed through a Wyatt DAWN Heleos II EOS

Table 1. Data collection and refinement statistics for ICP4NΔIDR·IE3_12mer and ICP4N·IE3_19mer complex structure

	ICP4NΔIDR·IE3_12mer	ICP4N·IE3_19mer
Data collection		
Space group	P 2₁2₁2₁	P 2₁2₁2₁
Cell dimensions		
a, b, c (Å)	127.3, 39.1, 90.4	61.5, 100.7, 201.9
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	45.22–2.12 (2.19–2.12) *	71.30–2.28 (2.36–2.28) *
R_{merge}	0.12 (0.77)	0.16 (1.33)
$I/\sigma I$	9.49 (2.20)	8.69 (1.39)
Completeness (%)	95 (100)	100 (100)
Redundancy	6.4 (6.6)	6.5 (6.5)
Total reflections	160 119 (17 313)	380 094 (37 510)
Unique reflections	25 175 (2606)	58 110 (5730)
Wavelength (Å)	0.920	0.979
Refinement		
Resolution (Å)	45.22–2.12	71.30–2.28
No. reflections	25 167 (2606)	58 104 (5729)
R_{work}	0.197 (0.278)	0.211 (0.291)
R_{free}	0.235 (0.323)	0.243 (0.325)
CC1/2	0.98 (0.96)	0.99 (0.88)
No. atoms		
Protein	3271	7240
Ligand	9	37
Water	196	236
B-factors		
Protein	46.5	53.0
Ligand/ion	40.3	69.9
Water	39.5	44.9
R.m.s. deviations		
Bond lengths (Å)	0.002	0.003
Bond angles (°)	0.47	0.56
Ramachandran		
Favored (%)	99.5	98.7
Allowed (%)	0.5	1.0
Outliers (%)	0.0	0.3

*Values in parentheses are for highest-resolution shell.

18-angle laser photometer coupled to a Wyatt Optilab rEX refractive index detector. Data were analyzed using Astra 6 software (Wyatt Technology Corp.). For sedimentation analytical ultracentrifugation, samples (20 μM protein or 20 μM protein dimer: IE3 1:1 co-purified) were buffer exchanged into 20 mM HEPES, 150 mM NaCl, pH 7.4 by exhaustive dialysis. The sedimentation coefficients for ICP4N in a DNA-free and DNA-bound state were determined from velocity experiments using the Optima XL-I ultracentrifuge (Beckman Instruments) and interference optics. The experiments were performed using double sector cells and sapphire windows and a rotor speed of 48000 rpm, taking 500 scans at 1 min intervals at a temperature of 20°C. The sedimenting boundaries were analyzed using the program Sedfit v8.7.

Surface plasmon resonance

Purified ICP4N and ICP4NΔIDR were exhaustively dialyzed into buffer B (20 mM HEPES, 150 mM NaCl, 2 mM MgCl₂, pH 7.4) and the concentration determined by UV absorption (280 nm) using an extinction coefficient of 40910 M⁻¹cm⁻¹ for each monomer. Synthetic DNA oligos were purchased (Invitrogen), with a biotin tag attached to the 5' end of the forward strand; these were solubilized and annealed into duplexes as described for co-crystallization experiments.

Experiments were performed using the ProteOn XPR36 SPR instrument (Bio-Rad Laboratories). The ProteOn XPR36 is a multiplex system that can be used to provide simultaneous flow of up to six different analyte concentrations (channels A1–A6) over up to six different ligand channels (L1–L6). Running buffer (RB) was 200 mM NaCl, 20 mM HEPES, 2 mM MgCl₂, 0.05% (w/v) Tween-20, pH 7.4. All experiments were performed at 25°C. Immobilization of NeutrAvidin was performed on a GLC chip (Bio-Rad Laboratories) in the vertical orientation. Three channels (L1–L3) were activated with 150 μl of a 1:1 mixture of 20 mM N-ethyl-N′-(3-dimethylaminopropyl) carbodiimide (EDC) and 6.5 mM sulfo-N-hydroxysuccinimide (sulfo-NHS) in water at a flow rate of 30 μl/min. NeutrAvidin was diluted in 10 mM sodium acetate buffer pH 5 to a final concentration of 50 μg/ml, and 150 μl was injected, followed by an injection of 150 μl of 1 M ethylenediamine-HCl, pH 8.5, at a flow rate of 30 μl/min. The immobilization level of NeutrAvidin was ∼3000 resonance units (RU). Next, 200 μl of 1:5000 dilution of biotinylated wild-type (WT) (L2) or mutant DNA (L3) in RB were injected at 200 μl/min for 60 s to allow their capture by the immobilized NeutrAvidin. Immobilization levels were ∼43 RU for WT DNA and 40 RU for mutant DNA. The L1 channel (NeutrAvidin only) was used as a reference. Measurements of equilibrium binding were made using five different concentrations of recombinant proteins (ICP4N and ICP4NΔIDR) in channels A2–

A6, channel A1 was used as a buffer only control. A short pulse of 2 M NaCl (50 μl/min for 60 s) was used for regeneration between measurements. Each measurement was repeated at least three times. Non-specific binding of recombinant proteins to the reference channel precluded the use of analyte concentrations above 800 nM.

All binding sensorgrams were collected, processed and analyzed using the integrated ProteOn Manager software (Bio-Rad Laboratories). Plots of maximum binding versus analyte concentration were used to calculate K_D values. Where required, additional data processing was carried out using SigmaPlot version 8 (Systat Software Inc). Short black segments on some sensorgrams represent artifact (spike) removal from the data.

NMR

Purified uniformly ^{15}N labeled proteins (ICP4N and ICP4ΔIDR) were dialyzed into NMR buffer (20 mM MES, 50 mM NaCl, 50 mM L-Arg, 50 mM L-Glu, 1 mM TCEP, 2 mM MgCl$_2$, pH 6.6) at 4°C. Proteins were concentrated in Vivaspin centrifugal devices to 0.05 mM and 500 μl samples were supplemented with 5% v/v D$_2$O. NMR spectra were recorded on a Bruker Advance 800 MHz spectrometer equipped with a TCI cryoprobe, data were acquired at 25°C. To assess signal perturbations observed on the sharp signals from flexible regions of ICP4N, the protein dimer was mixed 1:1 with IE3_19mer DNA duplex and to ensure no changes in pH occurred, dialyzed against NMR buffer. A control sample of protein lacking DNA was dialyzed in parallel in the same setup, then comparative ^{1}H-^{15}N TROSY were recorded. To facilitate the possible assignment of sharp backbone amide signals within the free ICP4N protein, a sample was concentrated further to 0.3 mM and 3D TOCSY-HSQC and NOESY-HSQC spectra were acquired, with mixing times of 45 and 120 ms respectively.

RESULTS

Structure of the ICP4N·IE3 self-regulation complex

The combination of the ICP4 DBD, residues 258–487 (ICP4N) with a 19mer DNA duplex matching the consensus site from the ICP4 promoter (IE3_19mer), resulted in the formation of a stoichiometric complex, which can be readily separated by gel filtration. It was also possible to form a smaller ICP4:DNA complex using a combination of truncated ICP4 DBD, residues 288–487 (ICP4NΔIDR), which eliminated a predicted IDR (Figure 1A), mixed with a shorter 12mer duplex (IE3_12mer). The structures of the ICP4N·IE3_19mer and ICP4NΔIDR·IE3_12mer complexes were solved to 2.28 Å and 2.12 Å resolution respectively (Table 1), and coordinates were submitted to the Protein Data Bank (5MHK for ICP4N·IE3_19mer and 5MHJ for ICP4NΔIDR·IE3_12mer).

The regions of DNA and polypeptide in the ICP4N and the ICP4NΔIDR structures with clearly interpretable electron density are listed in Supplementary Table S2 and the composition of the ICP4N·IE3_19mer asymmetric unit is illustrated in Supplementary Figure S1A. The most complete molecular assembly was present in the ICP4N·IE3_19mer data and comprised protein chains C, D and J with DNA

chains G and H which were chosen as the reference ICP4 DBD structure unless stated otherwise. An isolated region of electron density tentatively assigned to F283 and T284 was observed associated with a kink in DNA chains G+H. It is feasible that F283-T284 are part of chain D as the N-termini were within 11 Å, but due to a lack of clear continuous polypeptide density, the dipeptide was assigned a separate chain J. The DNA bases were numbered 1–19 in the sense stands (chain H) while the complementary base-pairs in the anti-sense strands are numbered 1′–19′ (chain G), and for convenience, numbers are shown as subscripts next to the nucleotide name.

The crystal structures indicated that the ICP4N·IE3 interaction is formed by a protein homo-dimer that adopts a closely complementary structure to the shape of the DNA duplex (Figure 1). Superposition of chains C and D, which together form a homo-dimer from the same molecular assembly indicated that the individual ICP4N polypeptide chains adopt the same overall fold (Supplementary Figure S1B) with a backbone RMSD of 1.15 Å for aa301–484 (Supplementary Figure S1C). Viewed individually, each polypeptide chain contains an N-terminal tail with two helical turns (aa295–298 and 301–305) that precede a globular region formed by residues 310–485. Structure homology searches indicated that the protein fold lacked homology to previously determined structures beyond trivial similarities with shorter helical hairpin fragments (Supplementary Table S3). The secondary structure of this globular region comprises a short poly-proline region (aa321–324), a small β-sheet (aa325–340) and then a more substantial 7 α-helical bundle (Figure 1). There is a notable 21 residue loop comprising aa411–431, connecting helices α4 and α5, which due to a triple-serine sequence we have named the 'S-loop'. Each monomer structure contains a prominent planar hydrophobic face, formed in the majority by α-helices 5, 6 and 7, two of these faces contact each other to form the homo-dimer. The face-to-face dimer-interface contains numerous symmetrically reciprocated intermolecular interactions that form a solvent excluded hydrophobic core (Figure 2A and B; Supplementary Table S4). Outside of this dimer-mediated hydrophobic core, further intermolecular contacts are present between the loops connecting helix α4 to α5 and α6 to α7 (and also the adjacent DNA duplex). Additionally, the N-terminal tail residues aa293–317 and 330–332 wrap around the globular region and contact helices α2 and α7 (Figure 2A and B), the sidechains D308 and R472 form a pair of intermolecular salt bridges. Characterization of ICP4 in solution by analytical ultracentrifugation (AUC) and SEC-MALS exclusively detected an ICP4N homodimer in the presence and absence of DNA (Table 2 and Figure 2C and D), and the truncated construct aa288–485 (ICP4NΔIDR) was also a homodimer (Supplementary Figure S2). The AUC data also indicated that ICP4N in complex with IE3_19mer is more compact than the free protein (Figure 2C and D).

The globular domain recognizes an upstream segment of the IE3 consensus DNA site

The crystal structure data indicated that the globular region of the ICP4N homo-dimer contacts the upstream base
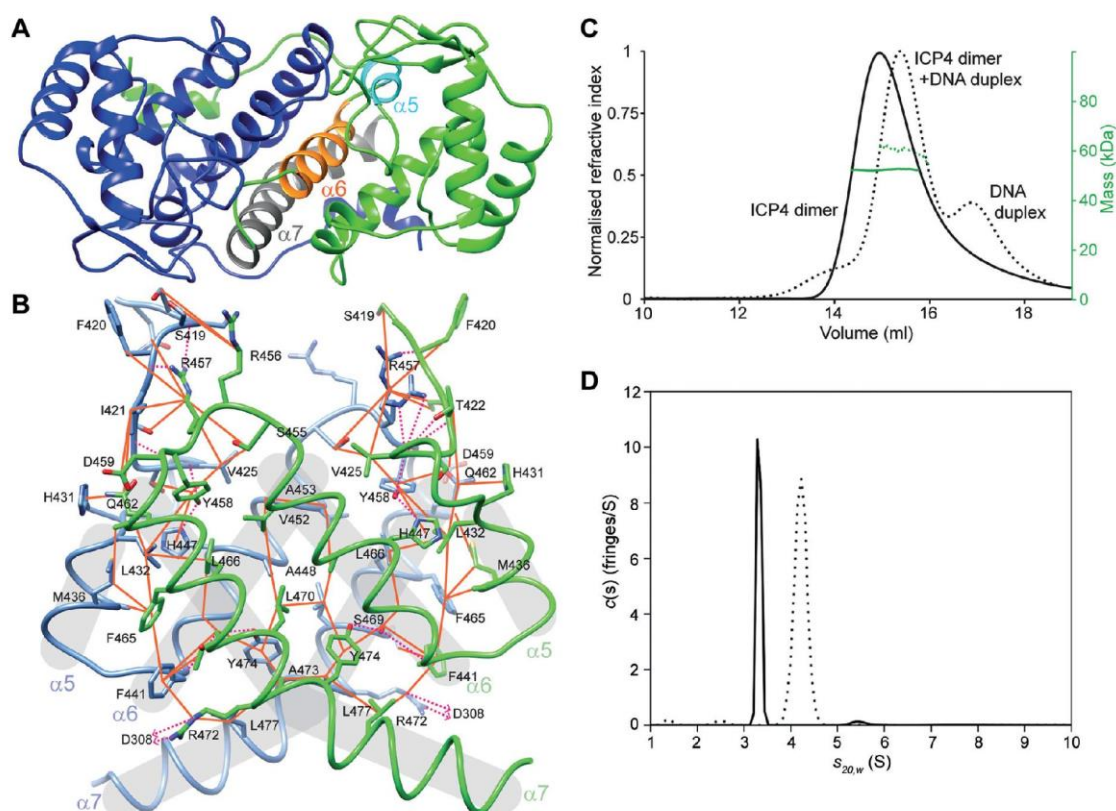
Figure 2. Structural details of and biophysical characterization of ICP4N dimerization. (**A**) Cartoon of protein chains C and D (colored blue and green respectively) from the ICP4N·IE3_19mer structure with α-helices 5, 6 and 7 highlighted which form the major hydrophobic homo-dimer interface. (**B**) Details of the residues within the major homo-dimer interface, hydrophobic contacts are indicated by orange lines and hydrogen bonds by pink dashes. (**C**) SEC-MALS profile of ICP4N with and without IE3_19mer DNA, shown as dashed or solid lines respectively and refractive index (black lines) and molecular mass (green lines), plotted against elution volume. (**D**) Velocity AUC analysis of ICP4N with and without IE3_19mer DNA, shown as dashed or solid lines respectively. For each sample a major peak was observed corresponding to a dimeric protein, free or in complex with DNA.

Table 2. Biophysical characterization of free ICP4N, free ICP4NΔIDR and the ICP4N·IE3_19mer complex

Construct	Predicted MW (kDa)	MALS		AUC		
		MW (kDa)	R_h (nm)	MW (kDa)	f/f_0	$S_{20,W}$ (S)
ICP4N	24.4	50.7	ND	48.8	1.45	3.32
ICP4N + DNA	24.4 + 11.6	60.1	4.83	60.6	1.22	4.27
ICP4NdIDR	21.3	44.0	ND	ND	ND	ND

ND indicates data were not determined

pairs 1–13 of the IE3_19mer DNA duplex (Figure 3A). The IE3_19mer DNA is bound across one edge of the ICP4N dimer interface and the upstream region is partially enveloped by complementary structural features of the protein (Figure 3). The ICP4 structure contains a number of apparent sequence-specific hydrogen bonds mediating binding to the upstream segment of the DNA. The protein residues which contact the DNA are mainly clustered within two regions in both protein chains: first the S-loop (aa411–431) and second the short loop connecting helix α6 to α7, which contains an arginine pair (R456, R457) and therefore we have called aa454–459 the 'R-turn' (Figures 1B and C and 3B). The S-loop forms intermolecular contacts with the R-turn of their dimeric partner; the same loops of different

monomers adopt a different conformation depending on their location within the major or minor DNA groove (Figure 3B).

The specificity of ICP4 to the RTCGTCNNYNYSG consensus sequence (where R is a purine, Y is a pyrimidine, S is a C or G and N is any base) appears to be determined by base-pair readout via the following structural features: (i) the first base 'R' of the recognition site, which is $A_{(4)}$ in the structure, base-pairs with a $T_{(4')}$ which forms a hydrogen bond with R456 (chain D) in the minor groove, this arrangement of the thymine could equally be accommodated by a cytosine but not a pyrimidine. (ii) The same R456 (chain D) sidechain forms a hydrogen bond with the carbonyl of the second consensus base $T_{(5)}$, and mutation
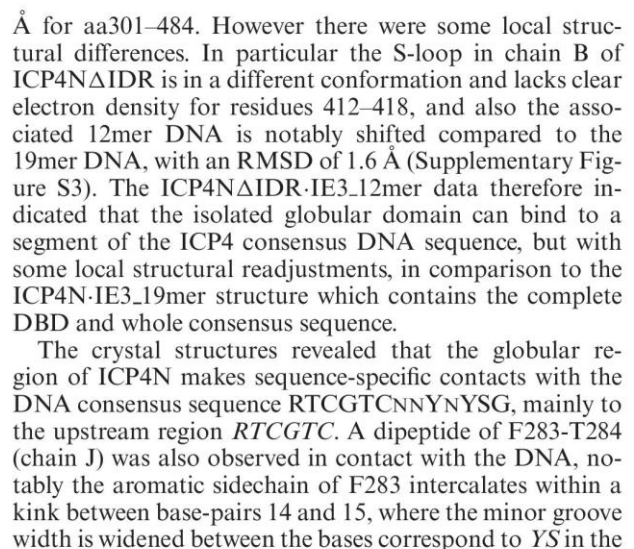
Figure 3. Details of the protein–DNA interface in the ICP4N·IE3_19mer structure. (**A**) Schematic of ICP4N-IE3 DNA interaction model. Protein–DNA hydrogen bonds identified from the crystal structure are indicated by dashed lines with locations of other contacts indicated with dark gray lines. Protein residues are colored blue or green when corresponding to chain C or D respectively. The vertical dashed arrow marks the DNA region that NMR, SAXS and SPR data suggested is bound by an IDR of the protein (residues 258–289). ICP4 consensus is sequence shown to the left side. (**B**) Overall ICP4N·IE3_19mer structure, with protein chains C and colored blue and green respectively and DNA space fill surface is shown colored dark and light orange for the sense (chain H) and antisense strands (chain G) respectively. Hydrogen bonds are indicated by dashes. Details of base-pair interactions, shown for (i) major groove bound by the ICP4N globular homo-dimer residues 416–419 and 456–457, (ii) DNA kink intercalated by F283 and (iii) minor groove bound by residues 416–419 and 456–457. (**C**) Surface of the ICP4N dimer colored by electrostatic potential (red through blue for acidic to basic charge) calculated by Adaptive Poisson-Boltzmann Solver module in Chimera (53). DNA is shown in cartoon form with the ICP4 consensus sequence labeled on sense strand. (**D**) Plot of DNA major- and minor-groove widths measured in both models in the ICP4N·IE3_19mer asymmetric unit (54).

of this thymine for a cytosine would eliminate this contact as it would change the carbonyl group, the H-bond acceptor, for an amide. (iii) The third base-pair $C_{(6)}$ along with $G_{(6')}$ form hydrogen bonds with residues in both S-loops, chain C into the minor groove and chain D into the major groove. (iv) The fourth, $G_{(7)}$ forms a pair of hydrogen bonds with R456 (chain C) in the major groove, an arrangement not compatible with an adenine base. Further protein contacts occur, but are limited to the phosphodiester backbone of IE3_19mer bases 1–13 and therefore are not expected to carry out any base-pair readout. A clear feature that mediates this protein–DNA interaction is the prominently positive protein surface charge at the DNA binding site (Figure 3C). In addition, the overall conformation of ICP4N DNA binding site complements the structure of the IE3 DNA duplex which deviates from an ideal B-form with a widened minor groove around $T_{(8)}$ (Figure 3D).

Superposition of the truncated ICP4NΔIDR polypeptides with ICP4N (chains C and D) indicated no global conformational changes and a backbone RMSD of 0.8

Å for aa301–484. However there were some local structural differences. In particular the S-loop in chain B of ICP4NΔIDR is in a different conformation and lacks clear electron density for residues 412–418, and also the associated 12mer DNA is notably shifted compared to the 19mer DNA, with an RMSD of 1.6 Å (Supplementary Figure S3). The ICP4NΔIDR·IE3_12mer data therefore indicated that the isolated globular domain can bind to a segment of the ICP4 consensus DNA sequence, but with some local structural readjustments, in comparison to the ICP4N·IE3_19mer structure which contains the complete DBD and whole consensus sequence.

The crystal structures revealed that the globular region of ICP4N makes sequence-specific contacts with the DNA consensus sequence RTCGTCNNYNYSG, mainly to the upstream region *RTCGTC*. A dipeptide of F283-T284 (chain J) was also observed in contact with the DNA, notably the aromatic sidechain of F283 intercalates within a kink between base-pairs 14 and 15, where the minor groove width is widened between the bases correspond to *YS* in the

consensus sequence (Figure 3D) (54). Therefore these data were indicative of a link between DNA within the $Y_N YSG$ downstream segment and aa258–289 of ICP4, a predicted IDR. In order to further investigate the involvement of the N-terminal IDR in sequence specific recognition of the downstream base-pairs, we utilized a combination of solution techniques.

Solution SAXS indicates additional protein density near downstream DNA base pairs

SAXS data were obtained by SEC-SAXS, removing any trace aggregates that may artificially elongate the molecules. Analysis of the dimensionless Kratky plot (55,56) showed that the unbound ICP4N dimer was more elongated than that of the complex due to the upward right shift from the Guinier-Kratky point relative to the protein–DNA complex (Figure 4A). Also the protein–DNA complex rested on the Guinier-Kratky point showing it to be a more compact species, which is consistent with the AUC data. Additionally, it was observed that when ICP4N binds to DNA IE3_19mer the R_g and D_{max} decrease from 31 Å to 25 Å and 127 Å to 83 Å, respectively (Supplementary Figure S4). This, coupled with the information shown on the dimensionless Kratky plot, suggests that ICP4N wraps around the DNA to make a more compact conformation upon binding. With the assumption that the changes in bound and free form within ICP4N are mostly within the predicted intrinsically disordered N-termini, an ensemble of structures was generated fitting the free ICP4N SAXS profile (20 structures, $\chi = 1.76$). This free ICP4N ensemble was clearly more expanded relative to the DAMMIF *ab initio* model of ICP4N·IE3_19mer (Supplementary Figure S5A) and was illustrative of a conformational change and compaction which accompanies DNA binding.

Comparison of the ICP4N·IE3_19mer SAXS data and the theoretical scattering curve, computed from the crystal structure, showed that the predicted scattering was similar to the measured data, although there is some discrepancy in the fitting indicated by the $\chi^2 = 2.4$ (Figure 4A and B). This disparity is likely caused by the regions of the protein unresolved in the crystal structure, mainly the N-terminal residues aa258–289. Docking of the crystal structure into the generated *ab initio* SAXS model showed that the model has two areas of unoccupied volume, one above the ICP4N globular dimer and the other below the dimer and in contact with the DNA (Supplementary Figure S5B). In order to confirm the volume is attributed to the ICP4N dimer, biphasic *ab initio* models were generated with MONSA using the scattering contrast between the DNA and protein (Figure 4C). Fitting the ICP4N·IE3_19mer crystal structure within the biphasic model clearly shows that the volume corresponding to that of the DNA maps to the location of the DNA seen in the crystal structure, and that the unoccupied volumes do indeed belong to the protein. SAXS data were submitted to the SASBDB with accession codes SASDB68, SASDB58 and SASDB48 for IE3_19mer, ICP4N and the complex respectively.

ICP4N residues 258–289 are intrinsically disordered and bind to DNA

Within the X-ray crystallography derived maps, electron density was not observed for ICP4 residues 258–286, a region previously predicted to be disordered, with the possible exception of F283 and T284 in chain J. Therefore, in order to investigate if the N-termini indeed contain IDRs in solution and if they contribute to DNA binding, we analyzed ICP4 samples by NMR, and compared spectra of the constructs with (ICP4N) and without (ICP4NΔIDR) the suspected disordered N-terminal region. ^{1}H-^{15}N correlation NMR spectra acquired on ICP4NΔIDR contained broad and dispersed amide signals characteristic for a large globular protein, whereas in the longer ICP4N construct, we observed additional prominent sharp, poorly dispersed amide signals characteristic of the presence of an IDR (Supplementary Figure S6), while signals from the globular domain were also observed with much lower signal-to-noise ratio. Analysis of 3D TOCSY-HSQC and NOESY-HSQC spectra allowed the amino acid type of 34 sharp backbone amide signals to be determined based on comparison with typical random coil chemical shifts. It was also possible to assign backbone amides to 21 residues within the N-terminal region 259–289, plus the C-terminal residues 485–487 (Figure 4E). While the remaining 10 peaks could not be attributed to specific sequence positions, their sidechain chemical shifts in the TOCSY spectra were characteristic of five arginines and five alanines, which matched the numbers of remaining unassigned residues scattered within regions 258–289 and 482–487. Assignment data were submitted to the BMRB with accession code 26957.

The addition of a stoichiometric amount of unlabeled IE3_19mer DNA duplex to [^{15}N]-ICP4N caused chemical shift perturbations within the sharp signals in ^{1}H-^{15}N correlation spectra (Figure 4E). Changes were observed for signals throughout the IDR, and notably to those assigned to residues 265–278 and 283–289 (including F284 and T285). All unassigned sharp signals were also perturbed by the presence of DNA to some extent. Together the NMR data confirmed that residues 259–289 of ICP4 are intrinsically disordered and interact with DNA.

The high affinity ICP4·DNA interaction requires both globular and disordered regions

In order to quantify the effect of mutations and truncations on the affinity of the ICP4N·IE3_19mer interaction, SPR binding studies were performed. Biotinylated DNA duplexes were immobilized on the sensor chip allowing screening with ICP4 constructs. Two DNA sequences were used, the WT IE3_19mer duplex as used in the crystal structure and a mutant 19mer duplex (IE3_19merMUT) where the *RTCGTC* part of the consensus sequence bound by the globular domain was left unchanged, while the downstream base-pair part $Y_N YSG$ was altered (Figure 5). Notably for ICP4N we measured nanomolar dissociation constants for the WT DNA interaction, whereas the affinity for IE3_19merMUT was weakened by two orders of magnitude; the data therefore supports that the ICP4N construct interacts with the complete ICP4 DNA consensus motif.
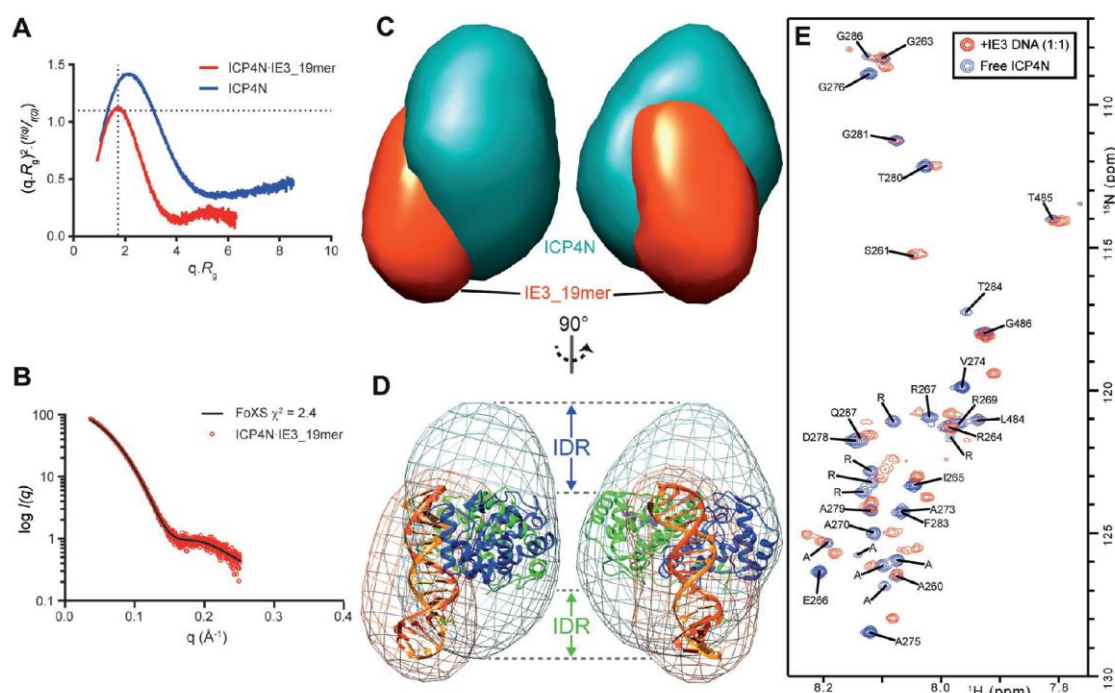
Figure 4. SAXS and NMR analysis of the ICP4N·IE3_19mer complex. (**A**) Dimensionless Kratky plot of the ICP4N bound (red) and unbound (blue) to DNA. Cross-hairs denote the Guinier–Kratky point (1.732, 1.104), the peak position for an ideal, globular particle. As indicated by the upward-right shift of the peaks in the dimensionless Kratky plot, ICP4N is more globular in the presence of DNA. (**B**) The calculated solution-state SAXS profile for the crystal structures of ICP4N·IE3_19mer complex (black line) compared to the measured scatter data (red circles). (**C**) Multi-phase *ab initio* model generated from SAXS data using MONSA show the presence of DNA (orange) and protein (teal) and their arrangement. (**D**) The crystal structure of the complex docked into the *ab initio* model revealing unoccupied volume around the DNA as well as above and below the protein dimer, assigned to the N-terminal IDRs. (**E**) NMR characterization of IDRs of the ICP4N dimer upon addition of equimolar amount of IE3_19mer duplex. ^{1}H-^{15}N TROSY spectrum of ICP4N showing sharp peaks assigned to residues within the unstructured N- and C-terminal regions in free and IE3 DNA bound forms, colored blue and red respectively. Peaks are labeled with assignments; when an unambiguous assignment was not possible the peaks are labeled with their amino acid type.

In comparison, the affinity of the shorter ICP4ΔIDR construct (aa288–487) for both WT and mutant DNA was three orders of magnitude weaker relative to that measured for ICP4N·WT. These results suggest that an absolute requirement for tight specific binding is the presence of the intrinsically disordered N-terminal region, whereas if this region is deleted, the affinity for DNA consensus sequence is not only reduced, but approximately half of this consensus sequence is no longer recognized.

DISCUSSION

ICP4 is a multi-domain protein that has been extensively studied due to its central role in HSV gene regulation, but structural data were lacking. ICP4 interacts with numerous sites within the HSV genome, with affinity highest for viral DNA fitting the bi-partite consensus sequence RTCGTCNNYNYSG (where R is a purine, Y is a pyrimidine, S is a C or G and N is any base), although it can also interact with non-consensus sites, a property which may contribute to its transactivation function (29–31). Due to the essential functional role of the DNA binding domain of ICP4, we have characterized the structure of this region revealing the details for sequence-specific DNA recognition and additionally the data allowing us to consider how ICP4 can bind to alterative sequences.

The combination of a crystal structure (see the overview stereo image on Supplemental Figure S7) with solution studies identified three regions of ICP4N (aa258–487), which form the majority of sequence-specific DNA contacts: (i) an N-terminal IDR comprised of residues 258–289, and within the globular homo-dimer, residues (ii) 415–427 and (iii) 455–457, which constitute the S-loop and the R-turn, respectively. ICP4 does not adopt a classic helix-turn-helix fold as previously predicted (16,20) but has a novel, more complex DNA binding fold. The crystal structure indicated that the R-turn and S-loops from different protein chains contact each other across the dimer interface, and adopt different conformations depending on their location in the DNA major or minor groove, which allows S418, S419 and R456 in particular to make alternative, yet specific contacts with the first 4 bp within the ICP4 consensus sequence RTCGTCNNYNYSG (Figure 3). To our knowledge this specific mode of DNA binding has not been observed before, however structural asymmetry in transcription factor homo-dimers is an inherent feature observed in examples that interact with an asymmetric DNA sequence (57). Away from the globular domain, with the exception of a DNA intercalating dipeptide F283 and T284, residues 258–286 were absent from the electron density; this region was shown by solution NMR data to be intrinsically disor-
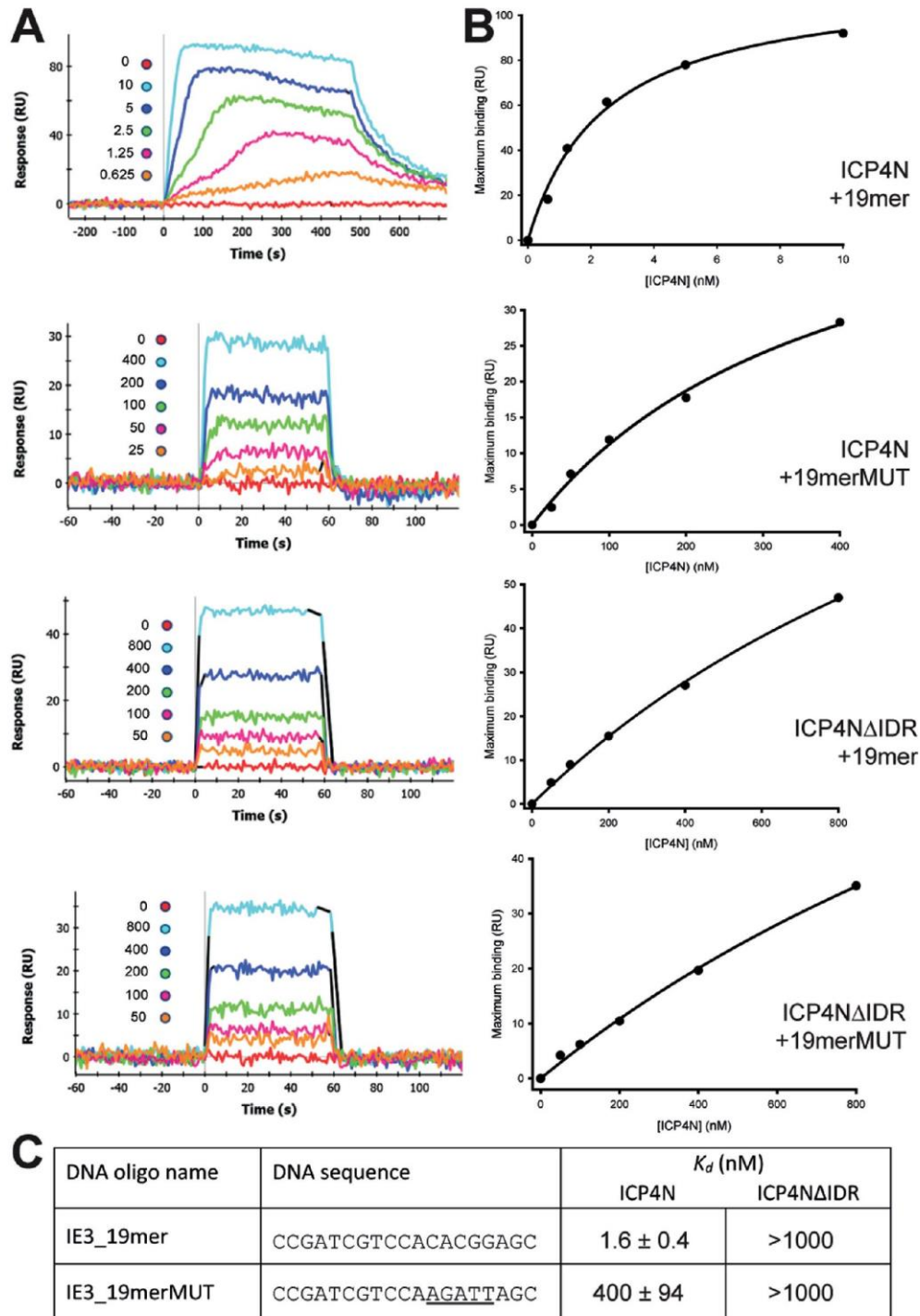
Figure 5. Binding of ICP4N and ICP4ΔIDR to biotinylated DNA duplexes measured by SPR. (**A**) Sensorgrams of different concentrations (nano-molar concentrations indicated on each plot) of ICP4 proteins binding to IE3 DNA duplexes. (**B**) Equilibrium analysis of SPR. (**C**) Mean dissociation constants (±SD) measured for each interaction, non-WT bases are underlined.

dered, but involved in DNA binding. Also SPR data were supportive of the hypothesis that the IDR interacts with the five downstream bases of the consensus sequence (*Y*N*YSG*), as the loss of the IDR prevented high affinity DNA interactions. In comparison to *RTCGTC*, the relatively less stringent nature of the *Y*N*YSG* region correlates with previous studies of DNA binding IDRs, in that they have an ability to form 'fuzzy' complexes with somewhat less specificity than the binding of globular domains (58–60). Fitting the ICP4N·IE3_19mer crystal structure coordinates within the biphasic *ab initio* model derived from SAXS revealed extra density assignable to protein but not observed the crystal structure (Figure 4D); the unoccupied volumes above and below the folded domain therefore likely correspond to the location of the two N-terminal IDRs. As only one of these volumes is located in the vicinity of the downstream part (*Y*N*YSG*) of the ICP4 consensus sequence, the data suggest that only a single IDR is required for high affinity DNA binding of the DNA oligo used here. It is currently unclear whether the 'spare' IDR will contribute to non-specific DNA binding upstream of the consensus sequence in longer DNA constructs. Together the data indicate that the ICP4N globular dimer and disordered regions act in synergy to bind to the *RTCGTC* and *Y*N*YSG* regions respectively, providing an explanation of the bi-partite nature of the DNA consensus sequence.

In common with ICP4, disordered regions in transcription factors have been shown to be functionally important for the search for specific DNA sites. For example; Hox proteins contain a mobile N-terminal arm preceding the helix-turn-helix fold of the DBD, the arm is utilized when searching for and enhancing the affinity to specific DNA bind sites (61). The IDRs in ICP4 DBD likely have a similar dual role in enhancing affinity for DNA (Figure 5) and also locating consensus sites, in the latter the two IDRs likely act independently allowing a broad search area (Figure 6). Interestingly, while the IDR is represented by sharp NMR signals consistent with increased flexibility of this region, primary sequence predictions (Figure 1A) and NMR secondary chemical shifts suggest that part of the IDR (residues 261–275) have a weak propensity to adopt a transient α-helical conformation; this region also shares sequence homology to the major groove binding α-helix of the cellular transcription factor Aryl hydrocarbon receptor nuclear translocator (HIF-1β), which interacts with DNA as a dimer (62). Homologs of HIF-1β undergo a disorder to order transition, becoming α-helical, upon binding to DNA (63). One can speculate that the ICP4 IDR may become more ordered and adopt an α-helical conformation upon binding with some DNA motifs, although not stable enough to be entirely rigid in a crystal form in a case studied here.

The functional importance of conserved residues within the DNA binding domain of ICP4 has previously been probed using a number of mutants of both the protein and the DNA sequence; in light of the data presented here, the structural significance of these mutations can now be examined (Figure 1B and Supplementary Table S1). Notably the R-turn point mutants R456L and R457L caused a loss of binding or lower affinity for DNA, respectively (13). Studies on the effect of alterations within the DNA sequence on the interaction of isolated ICP4 DNA bind-
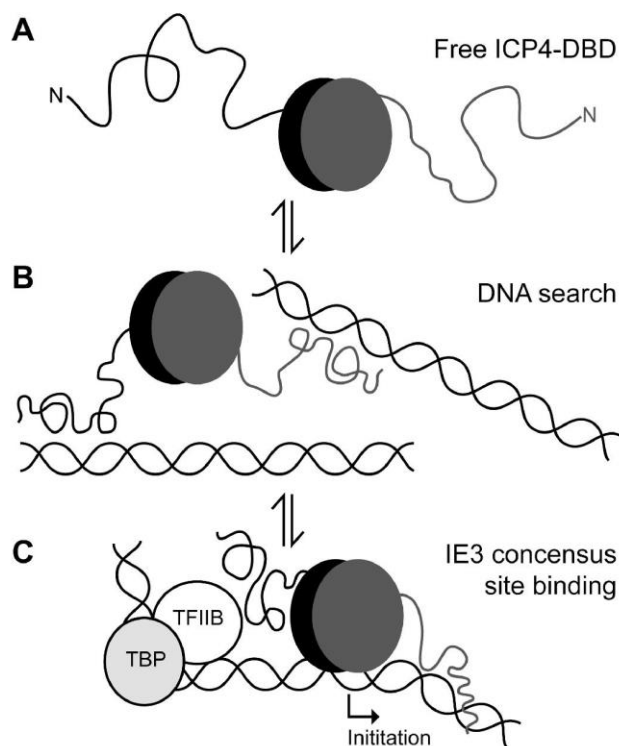

Figure 6. Model of action of the ICP4 DNA binding domain (DBD). The globular homo-dimer is represented by gray and black ovals and N-terminal IDRs by lines. (**A**) Free protein adopts an expanded conformation with the IDRs extended. (**B**) When not in contact with a DNA consensus site, ICP4-DBD and particularly the IDRs search DNA strands for sequence motifs. (**C**) Binding to the IE3 consensus site which overlaps with the transcription initiation site, ICP4 forms an asymmetric complex by the synergistic action of the globular region to the RTCGTC motif and an IDR with the downstream YNYSG motif. One IDR is not involved in specific DNA recognition and points upstream toward the TATA box, which is compatible with tripartite complex formation by the TATA binding protein, TFIIB and ICP4.

ing domains suggested that the first four bases within RTCGTC_{NN}Y_NYSG are the most important for binding efficiency (38,64). This correlates with our structural data as only the first 4 bp form clear sequence-specific contacts with ICP4N. The globular domain therefore likely binds to sites containing this minimal 4 bp motif in preference to those lacking it, for example, the motif underlined in the sequence GCTAGC<u>ATCG</u>ATCCATGGA bound by ICP4 when it multimerizes on the late gene encoding glycoprotein C; notably the 4 bp are part of a palindrome (ATC-GAT) and therefore the motif is duplicated on the sense and anti-sense strands (26). The *Y*N*YSG* region of the ICP4 consensus sequence was also shown in DNA mutation studies to contribute to specificity (64). A possible general role of the N-terminal IDR was previously suggested by truncating this region (22,38) and constructing point mutations in this region (14,16), which showed reduced DNA affinity. However the link was not previously identified between the IDR (aa258–289) and the *Y*N*YSG* region; our data identifies them as binding partners.

The homo-dimerization of ICP4 is a property resulting from the DNA-binding domain contained within the ICP4N construct, which was studied in detail here. The AUC and SEC-MALS data indicate that the domain forms a stable homo-dimer in solution regardless whether it is interacting with DNA or not (Figure 2A and B). The crystal structure revealed that each monomer contains a relatively flat hydrophobic dimer interface, composed of helices α5, α6 and α7 (aa398–452) (Figure 2C). Previously a short motif of residues 343–376 was implicated as responsible for the dimer interface, and a truncated construct of residues 343–490 could hetero-dimerize with a complete ICP4 DNA binding domain (21). The ICP4N·IE3_19mer structure indicates that residues 343–376 in isolation are not a dimerization motif, as they cannot form the major hydrophobic dimerization interface and so cannot promote a native-like ICP4 homo-dimer; these residues however do contribute to the dimer indirectly and they are therefore likely essential for correct protein folding. The importance of a dimerization interface observed in the ICP4N structure for stable DNA binding is supported by the previous studies of the temperature sensitive mutant A475V (tsK). At non-permissive temperatures, the A475V mutant protein poorly recognizes the IE3 consensus DNA site and similarly the tsK virus cannot repress IE gene expression and activate early or late gene expression (7,13). Our structural data indicates that A475 is located away from the DNA binding site, but is at the periphery of the homo-dimer interface, making intermolecular contacts in particular with Y306. Thus the additional steric bulking upon mutating the alanine to a valine would likely disrupt the sidechain packing and destabilize the protein, causing it at higher temperatures to lose a stable dimeric structure required for DNA recognition.

The sequence-specific interaction of transcription factors with DNA is an essential cellular process, and in eukaryotes often involves multiple binding domains and IDRs (65,66). Our data indicate that while herpes simplex virus transcription factor ICP4 utilizes a protein fold which is unknown in eukaryotes, the principle of the synergy between folded and flanking unfolded regions is the same. The combination of the ICP4N globular homo-dimer and IDRs tune the specificity and maximize affinity for a DNA motif. Conversely, the combination of the IDRs with the apparent conformational plasticity of the S-loop and R-turn likely allows non-consensus DNA interactions with ICP4. The flexibility of these regions may facilitate the linear movement of ICP4 along a DNA duplex in search of high affinity consensus sites, or hopping along longer distances mediated by the IDRs. The structure of ICP4N·IE3_19mer complex along with the corroborative solution and quantification data therefore bring a new level of insight into the function of this essential HSV transcription factor. This information should prove valuable for improving our understanding of the function of this prevalent virus and aid its utilization in disease therapies (2,3). In addition, due to their viral origins the high affinity ICP4–DNA complexes along with antagonists such as ICP0 could be useful components of synthetic biology circuits (34,67,68) within mammalian or bacterial systems.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Itzhaki,R.F. (2014) Herpes simplex virus type 1 and Alzheimer's disease: increasing evidence for a major role of the virus. *Front. Aging Neurosci.*, **6**, 202.
2. Manservigi,R., Argnani,R. and Marconi,P. (2010) HSV recombinant vectors for gene therapy. *Open Virol. J.*, **4**, 123–156.
3. Russell,S.J., Peng,K.W. and Bell,J.C. (2012) Oncolytic virotherapy. *Nat. Biotechnol.*, **30**, 658–670.
4. Honess,R.W. and Roizman,B. (1975) Regulation of herpesvirus macromolecular synthesis: sequential transition of polypeptide synthesis requires functional viral polypeptides. *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 1276–1280.
5. Gruffat,H., Marchione,R. and Manet,E. (2016) Herpesvirus late gene expression: a viral-specific pre-initiation complex is key. *Front. Microbiol.*, **7**, 869.
6. DeLuca,N.A., McCarthy,A.M. and Schaffer,P.A. (1985) Isolation and characterization of deletion mutants of herpes simplex virus type 1 in the gene encoding immediate-early regulatory protein ICP4. *J. Virol.*, **56**, 558–570.
7. Preston,C.M. (1979) Control of herpes simplex virus type 1 mRNA synthesis in cells infected with wild-type virus or the temperature-sensitive mutant tsK. *J. Virol.*, **29**, 275–284.
8. Godowski,P.J. and Knipe,D.M. (1986) Transcriptional control of herpesvirus gene expression: gene functions required for positive and negative regulation. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 256–260.
9. Watson,R.J. and Clements,J.B. (1980) A herpes simplex virus type 1 function continuously required for early and late virus RNA synthesis. *Nature*, **285**, 329–330.
10. DeLuca,N.A. and Schaffer,P.A. (1985) Activation of immediate-early, early, and late promoters by temperature-sensitive and wild-type forms of herpes simplex virus type 1 protein ICP4. *Mol. Cell. Biol.*, **5**, 1997–2008.
11. Gu,B., Rivera-Gonzalez,R., Smith,C.A. and DeLuca,N.A. (1993) Herpes simplex virus infected cell polypeptide 4 preferentially represses Sp1-activated over basal transcription from its own promoter. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 9528–9532.
12. O'Hare,P. and Hayward,G.S. (1985) Evidence for a direct role for both the 175,000- and 110,000-molecular-weight immediate-early proteins of herpes simplex virus in the transactivation of delayed-early promoters. *J. Virol.*, **53**, 751–760.
13. Allen,K.E. and Everett,R.D. (1997) Mutations which alter the DNA binding properties of the herpes simplex virus type 1 transactivating protein Vmw175 also affect its ability to support virus replication. *J. Gen. Virol.*, **78**, 2913–2922.
14. Paterson,T. and Everett,R.D. (1988) The regions of the herpes simplex virus type 1 immediate early protein Vmw175 required for site specific DNA binding closely correspond to those involved in transcriptional regulation. *Nucleic Acids Res.*, **16**, 11005–11025.

15. Paterson,T. and Everett,R.D. (1988) Mutational dissection of the HSV-1 immediate-early protein Vmw175 involved in transcriptional transactivation and repression. *Virology*, **166**, 186–196.

16. Shepard,A.A., Imbalzano,A.N. and DeLuca,N.A. (1989) Separation of primary structural components conferring autoregulation, transactivation, and DNA-binding properties to the herpes simplex virus transcriptional regulatory protein ICP4. *J. Virol.*, **63**, 3714–3728.

17. Metzler,D.W. and Wilcox,K.W. (1985) Isolation of herpes simplex virus regulatory protein ICP4 as a homodimeric complex. *J. Virol.*, **55**, 329–337.

18. Shepard,A.A., Tolentino,P. and DeLuca,N.A. (1990) Trans-dominant inhibition of herpes simplex virus transcriptional regulatory protein ICP4 by heterodimer formation. *J. Virol.*, **64**, 3916–3926.

19. Xia,K., DeLuca,N.A. and Knipe,D.M. (1996) Analysis of phosphorylation sites of herpes simplex virus type 1 ICP4. *J. Virol.*, **70**, 1061–1071.

20. Wyrwicz,L.S. and Rychlewski,L. (2007) Fold recognition insights into function of herpes ICP4 protein. *Acta Biochim. Pol.*, **54**, 551–559.

21. Gallinari,P., Wiebauer,K., Nardi,M.C. and Jiricny,J. (1994) Localization of a 34-amino-acid segment implicated in dimerization of the herpes simplex virus type 1 ICP4 polypeptide by a dimerization trap. *J. Virol.*, **68**, 3809–3820.

22. Wu,C.L. and Wilcox,K.W. (1990) Codons 262 to 490 from the herpes simplex virus ICP4 gene are sufficient to encode a sequence-specific DNA binding protein. *Nucleic Acids Res.*, **18**, 531–538.

23. Faber,S.W. and Wilcox,K.W. (1988) Association of herpes simplex virus regulatory protein ICP4 with sequences spanning the ICP4 gene transcription initiation site. *Nucleic Acids Res.*, **16**, 555–570.

24. Muller,M.T. (1987) Binding of the herpes simplex virus immediate-early gene product ICP4 to its own transcription start site. *J. Virol.*, **61**, 858–865.

25. Kattar-Cooley,P. and Wilcox,K.W. (1989) Characterization of the DNA-binding properties of herpes simplex virus regulatory protein ICP4. *J. Virol.*, **63**, 696–704.

26. Kuddus,R.H. and DeLuca,N.A. (2007) DNA-dependent oligomerization of herpes simplex virus type 1 regulatory protein ICP4. *J. Virol.*, **81**, 9230–9237.

27. Smith,C.A., Bates,P., Rivera-Gonzalez,R., Gu,B. and DeLuca,N.A. (1993) ICP4, the major transcriptional regulatory protein of herpes simplex virus type 1, forms a tripartite complex with TATA-binding protein and TFIIB. *J. Virol.*, **67**, 4676–4687.

28. Kuddus,R., Gu,B. and DeLuca,N.A. (1995) Relationship between TATA-binding protein and herpes simplex virus type 1 ICP4 DNA-binding sites in complex formation and repression of transcription. *J. Virol.*, **69**, 5568–5575.

29. Gu,B. and DeLuca,N. (1994) Requirements for activation of the herpes simplex virus glycoprotein C promoter in vitro by the viral regulatory protein ICP4. *J. Virol.*, **68**, 7953–7965.

30. Sampath,P. and Deluca,N.A. (2008) Binding of ICP4, TATA-binding protein, and RNA polymerase II to herpes simplex virus type 1 immediate-early, early, and late promoters in virus-infected cells. *J. Virol.*, **82**, 2339–2349.

31. Smiley,J.R., Johnson,D.C., Pizer,L.I. and Everett,R.D. (1992) The ICP4 binding sites in the herpes simplex virus type 1 glycoprotein D (gD) promoter are not essential for efficient gD transcription during virus infection. *J. Virol.*, **66**, 623–631.

32. Bruce,J.W. and Wilcox,K.W. (2002) Identification of a motif in the C terminus of herpes simplex virus regulatory protein ICP4 that contributes to activation of transcription. *J. Virol.*, **76**, 195–207.

33. Wagner,L.M., Lester,J.T., Sivrich,F.L. and DeLuca,N.A. (2012) The N terminus and C terminus of herpes simplex virus 1 ICP4 cooperate to activate viral gene expression. *J. Virol.*, **86**, 6862–6874.

34. Liu,M., Rakowski,B., Gershburg,E., Weisend,C.M., Lucas,O., Schmidt,E.E. and Halford,W.P. (2010) ICP0 antagonizes ICP4-dependent silencing of the herpes simplex virus ICP0 gene. *PLoS One*, **5**, e8837.

35. Randall,G., Lagunoff,M. and Roizman,B. (1997) The product of ORF O located within the domain of herpes simplex virus 1 genome transcribed during latent infection binds to and inhibits in vitro binding of infected cell protein 4 to its cognate DNA site. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 10379–10384.

36. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Soding,J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.

37. Buchan,D.W., Minneci,F., Nugent,T.C., Bryson,K. and Jones,D.T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.*, **41**, W349–W357.

38. Pizer,L.I., Everett,R.D., Tedder,D.G., Elliott,M. and Litman,B. (1991) Nucleotides within both proximal and distal parts of the consensus sequence are important for specific DNA recognition by the herpes simplex virus regulatory protein ICP4. *Nucleic Acids Res.*, **19**, 477–483.

39. Van Duyne,G.D., Standaert,R.F., Karplus,P.A., Schreiber,S.L. and Clardy,J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.*, **229**, 105–124.

40. Golovanov,A.P., Hautbergue,G.M., Wilson,S.A. and Lian,L.Y. (2004) A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.*, **126**, 8933–8939.

41. Winter,G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.*, **43**, 186–190.

42. Skubák,P. and Pannu,N.S. (2013) Automatic protein structure solution from weak X-ray data. *Nat. Commun.* **4**, 2777.

43. Terwilliger,T.C., Adams,P.D., Read,R.J., Mccoy,A.J., Moriarty,N.W., Grosse-Kunstleve,R.W., Afonine,P.V., Zwart,P.H. and Hung,L.W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D*, **65**, 582–601.

44. Sheldrick,G.M. (2002) Macromolecular phasing with SHELXE. *Zeitschrift für Kristallographie-Crystalline Materials*, **217**, 644–650.

45. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D*, **66**, 213–221.

46. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*, **60**, 2126–2132.

47. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D*, **66**, 12–21.

48. Joosten,R.P., Long,F., Murshudov,G.N. and Perrakis,A. (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*, **1**, 213–220.

49. Schneidman-Duhovny,D., Hammel,M., Tainer,J.A. and Sali,A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.*, **105**, 962–974.

50. Schneidman-Duhovny,D., Hammel,M., Tainer,J.A. and Sali,A. (2016) FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.*, **44**, W424–W429.

51. Petoukhov,M.V., Franke,D., Shkumatov,A.V., Tria,G., Kikhney,A.G., Gajda,M., Gorba,C., Mertens,H.D., Konarev,P.V. and Svergun,D.I. (2012) New developments in the program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **45**, 342–350.

52. Rambo,R.P. (2015) Resolving individual components in protein-RNA complexes using small-angle X-ray scattering experiments. *Methods Enzymol.*, **558**, 363–390.

53. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

54. El Hassan,M.A. and Calladine,C.R. (1998) Two distinct modes of protein-induced bending in DNA. *J. Mol. Biol.*, **282**, 331–343.

55. Rambo,R.P. and Tainer,J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, **95**, 559–571.

56. Receveur-Brechot,V. and Durand,D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.*, **13**, 55–75.

57. Swapna,L.S., Srikeerthana,K. and Srinivasan,N. (2012) Extent of structural asymmetry in homodimeric proteins: prevalence and relevance. *PLoS One*, **7**, e36688.

58. van der Lee,R., Buljan,M., Lang,B., Weatheritt,R.J., Daughdrill,G.W., Dunker,A.K., Fuxreiter,M., Gough,J., Gsponer,J.,

Jones,D.T. *et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, **114**, 6589–6631.

59. Crane-Robinson,C., Dragan,A.I. and Privalov,P.L. (2006) The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem. Sci.*, **31**, 547–552.

60. Vuzman,D. and Levy,Y. (2012) Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol. Biosyst.*, **8**, 47–57.

61. Bondos,S.E., Swint-Kruse,L. and Matthews,K.S. (2015) Flexibility and disorder in gene regulation: LacI/GalR and Hox proteins. *J. Biol. Chem.*, **290**, 24669–24677.

62. Wu,D., Potluri,N., Lu,J., Kim,Y. and Rastinejad,F. (2015) Structural integration in hypoxia-inducible factors. *Nature*, **524**, 303–308.

63. Fieber,W., Schneider,M.L., Matt,T., Krautler,B., Konrat,R. and Bister,K. (2001) Structure, function, and dynamics of the dimerization and DNA-binding domain of oncogenic transcription factor v-Myc. *J. Mol. Biol.*, **307**, 1395–1410.

64. Everett,R.D., Elliott,M., Hope,G. and Orr,A. (1991) Purification of the DNA binding domain of herpes simplex virus type 1 immediate-early protein Vmw175 as a homodimer and extensive mutagenesis of its DNA recognition site. *Nucleic Acids Res.*, **19**, 4901–4908.

65. Smith,N.C. and Matthews,J.M. (2016) Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Curr. Opin. Struct. Biol.*, **38**, 68–74.

66. Sharma,R., Raduly,Z., Miskei,M. and Fuxreiter,M. (2015) Fuzzy complexes: specific binding without complete folding. *FEBS Lett.*, **589**, 2533–2542.

67. Leavitt,J.M. and Alper,H.S. (2015) Advances and current limitations in transcript-level control of gene expression. *Curr. Opin. Biotechnol.*, **34**, 98–104.

68. Rao,C.V. (2012) Expanding the synthetic biology toolbox: engineering orthogonal regulators of gene expression. *Curr. Opin. Biotechnol.*, **23**, 689–694.

The herpes viral transcription factor ICP4 forms a novel DNA recognition complex

Richard B Tunnicliffe, Michael P Lockhart-Cairns, Colin Levy, A Paul Mould, Thomas A Jowitt, Hilary Sito, Clair Baldock, Rozanne M Sandri-Goldin & Alexander P Golovanov

Supplementary Table 1. Summary of effects of ICP4 mutations within the ICP4 DNA binding domain measured in a selection of previous studies. Abbreviations: ins=insertion mutant, NT=not tested, '+'=native-like property, '±'=property impaired, '-'=loss of property, TS= temperature sensitive. The references to the relevant studies are provided in the last column. The activity columns state the protein ability to trans-activate or -repress genes compared to the wild type.

Mutant	Mutant name used in the original study	Construct used (ICP4 residues)	DNA binding	Repression activity	Activation activity	Viable virus	Study
ins263	pi10.5	1-774	-	+	+	NT	(1)
ins274	pi11	1-774	±	+	-	NT	(1)
ins275	pI10	1-1298	+	-	26%	+	(2-4)
ins292	pI11	1-1298	+	+	31%	+	(2-4)
ins310	pI12	1-1298	±	±	20%	-	(2-4)
ins320	pi12	1-774	-	±	-	NT	(1)
ins324	pI13	1-1298	+	-	10%	-	(2-4)
ins329	pI14	1-1298	-	-	5%	-	(2-4)
ins337	pI15	1-1298	-	±	5%	-	(2-4)
ins338	pi13	1-774	-	±	-	NT	(1)
ins373	pI16	1-1298	-	-	17%	-	(2-4)
ins386	pi14	1-774	+	+	+	NT	(1)
ins398	pI17	1-1298	±	±	12%	-	(2-4)
ins438	pI18	1-1298	±	±	16%	-	(2-4)
A446S	446A/S	276-523	+	NT	NT	NT	(4)
A449S	449A/S	276-523	± TS	NT	NT	NT	(4)
ins449	pi16	1-774	-	±	-	NT	(1)
A450S	450A/S	276-523	+	NT	NT	NT	(4)
A451S	451A/S	276-523	+	NT	NT	NT	(4)
A451S+A453S	451/3AA/SS	276-523	± TS	NT	NT	NT	(4)
ins452	pi17	1-774	-	-	-	NT	(1)
A453S	453A/S	252-523 & 1-1298	± TS	NT	±	+	(4)
R456L	456R/L	252-523 & 1-1298	-	NT	-	-	(4)
R457L	457R/L	276-523 & 1-1298	± TS	NT	-	-	(4)
D459Y	459D/Y	252-523 & 1-1298	-	NT	NT	-	(4)
A461S	461A/S	276-523	+	NT	NT	NT	(4)
Q462H	462Q/H	276-523	+	NT	NT	NT	(4)
K463N+G464C	463/4KG/NC	276-523 & 1-1298	-	NT	-	-	(4)
G464C	464G/C	276-523	+	NT	NT	NT	(4)
R471L	471R/L	276-523	+	NT	NT	NT	(4)
R472L	472R/L	276-523	± TS	NT	NT	NT	(4)
A475V	475A/V tsK	276-523 & 1-1298	± TS	-	-	NT	(4,5)

Supplementary Table 2. Composition of the asymmetric units of the ICP4N·IE3_19mer and ICP4NΔIDR·IE3_12mer crystal structures. Regions of protein and DNA constructs observable in electron density (ED) are indicated.

Structure	Construct	Chain ID	Biological assembly	Regions observable in ED
ICP4N·IE3_19mer	ICP4N	A	A+B+E+F	300-485
	ICP4N	B	A+B+E+F	287-416, 419-485
	ICP4N	C	C+D+G+H+J	292-486
	ICP4N	D	C+D+G+H+J	289-485
	IE3_19mer	E	A+B+E+F	1'-15'
	IE3_19mer	F	A+B+E+F	1-16
	IE3_19mer	G	C+D+G+H+J	1'-19'
	IE3_19mer	H	C+D+G+H+J	1-18
	ICP4N	J	C+D+G+H+J	283-284
ICP4NΔIDR·IE3_12mer	ICP4NΔIDR	A	A+B+E+F	293-486
	ICP4NΔIDR	B	A+B+E+F	294-411, 419-487
	IE3_12mer	E	A+B+E+F	2-10
	IE3_12mer	F	A+B+E+F	1'-9'

Supplementary Table 3. Top 10 results from structural homology searches indicate weak similarly and low percentage identity of ICP4-DBD with known protein folds. Searches were performed using ICP4N·IE3_19mer coordinates and (A) DALI on chain C and (B) PDBeFOLD on chains C and D (6,7).

A

No:	Chain	Z	RMSD	lali	Nres	%id	Description
1:	1oxk-G	3.7	3.6	89	156	8	YPD1P;
2:	1oxk-K	3.7	3.6	89	156	8	YPD1P;
3:	1oxk-C	3.7	3.6	89	156	8	YPD1P;
4:	1oxk-E	3.7	3.6	89	156	8	YPD1P;
5:	1or3-A	3.7	3.4	86	136	13	PROTEIN(APOLIPOPROTEINE);
6:	1oxk-A	3.6	3.6	89	156	8	YPD1P;
7:	1oxk-I	3.6	3.6	89	156	8	YPD1P;
8:	1or2-A	3.6	3.5	86	130	12	APOLIPOPROTEINE;
9:	1c02-A	3.5	4.7	93	166	8	PHOSPHOTRANSFERASEYPD1P;
10	:1oxb-A	3.5	4.6	93	166	6	YPD1P;

B

No	Query Scoring			Query RMSD	N_{align}	N_g	$\%_{seq}$	Target $\%_{sse}$	Target Match	$\%_{sse}$	N_{res}	Target Title
	Q	P	Z									
1	0.021	-0.0	0.5	4.10	87	16	13	21	1xos:A	29	322	Catalytic domain of human phosphodiesterase 4b in complex with sildenafil
2	0.021	-0.0	0.1	5.64	119	13	4	21	3mzk:B	23	385	Sec13/sec16 complex, s.cerevisiae
3	0.020	-0.0	0.3	5.70	99	11	12	21	1mpg:A	26	282	3-methyladenine dna glycosylase ii from escherichia coli
4	0.019	-0.0	1.0	4.19	79	10	5	21	2grl:C	31	287	Crystal structure of dct/icf10 complex
5	0.019	-0.0	0.8	4.29	80	8	13	21	2grl:B	31	286	Crystal structure of dct/icf10 complex
6	0.019	-0.0	1.4	3.53	78	8	6	21	5dbk:B	28	352	Apo form of the quorum sensor nprr from b. Thuringiensis
7	0.019	-0.0	0.3	5.38	92	10	11	21	1mpg:B	26	282	3-methyladenine dna glycosylase ii from escherichia coli
8	0.019	-0.0	1.0	4.40	84	8	2	21	1x81:A	31	315	Farnesyl transferase structure of jansen compound
9	0.018	-0.0	1.1	4.58	87	8	6	21	3mzl:F	24	329	Sec13/sec31 edge element, loop deletion mutant
10	0.018	-0.0	1.0	4.61	87	8	5	21	3mzl:H	24	329	Sec13/sec31 edge element, loop deletion mutant

Supplementary Table 4. Episa analysis (8) of the extensive homo-dimer interface between ICP4N protein chains A and B, listing the intermolecular hydrogen binds (H-bond) and salt bridges (SB) detected within the crystal structure.

| Residue	[atom]	Type
Chain A	Chain B	
TYR 306 [OH]	ARG 471 [NH1]	H-bond
ASP 308 [OD1]	ARG 472 [NE]	SB
ASP 308 [OD1]	ARG 472 [NH2]	SB
ASP 308 [OD2]	ARG 472 [NE]	SB
ASP 308 [OD2]	ARG 472 [NH2]	SB
TYR 310 [O]	ARG 472 [NH2]	H-bond
GLY 330 [O]	GLN 462 [NE2]	H-bond
ALA 385 [O]	ARG 290 [NH1]	H-bond
SER 417 [O]	ARG 457 [NH1]	H-bond
PHE 420 [O]	ARG 457 [NH1]	H-bond
PHE 441 [O]	SER 469 [OG]	H-bond
ARG 457 [O]	THR 422 [OG1]	H-bond
ARG 457 [O]	THR 422 [N]	H-bond
ARG 457 [O]	GLY 423 [N]	H-bond
TYR 458 [OH]	VAL 425 [N]	H-bond
TYR 458 [OH]	HIS 447 [ND1]	H-bond
ASP 459 [OD2]	LEU 332 [N]	H-bond
SER 469 [OG]	TYR 474 [OH]	H-bond
TYR 474 [OH]	SER 469 [OG]	H-bond
ASN 482 [ND2]	ALA 297 [O]	H-bond
ARG 472 [NE]	ASP 308 [OD1]	SB
ARG 472 [NH2]	ASP 308 [OD1]	SB
ARG 472 [NE]	ASP 308 [OD2]	SB
ARG 472 [NH2]	ASP 308 [OD2]	SB
ARG 472 [NH2]	TYR 310 [O]	H-bond
GLN 462 [NE2]	GLY 330 [O]	H-bond
ARG 457 [NH1]	PHE 420 [O]	H-bond
SER 469 [OG]	PHE 441 [O]	H-bond
GLY 423 [N]	ARG 457 [O]	H-bond
THR 422 [N]	ARG 457 [O]	H-bond
HIS 447 [ND1]	TYR 458 [OH]	H-bond
VAL 425 [N]	TYR 458 [OH]	H-bond
LEU 332 [N]	ASP 459 [OD2]	H-bond
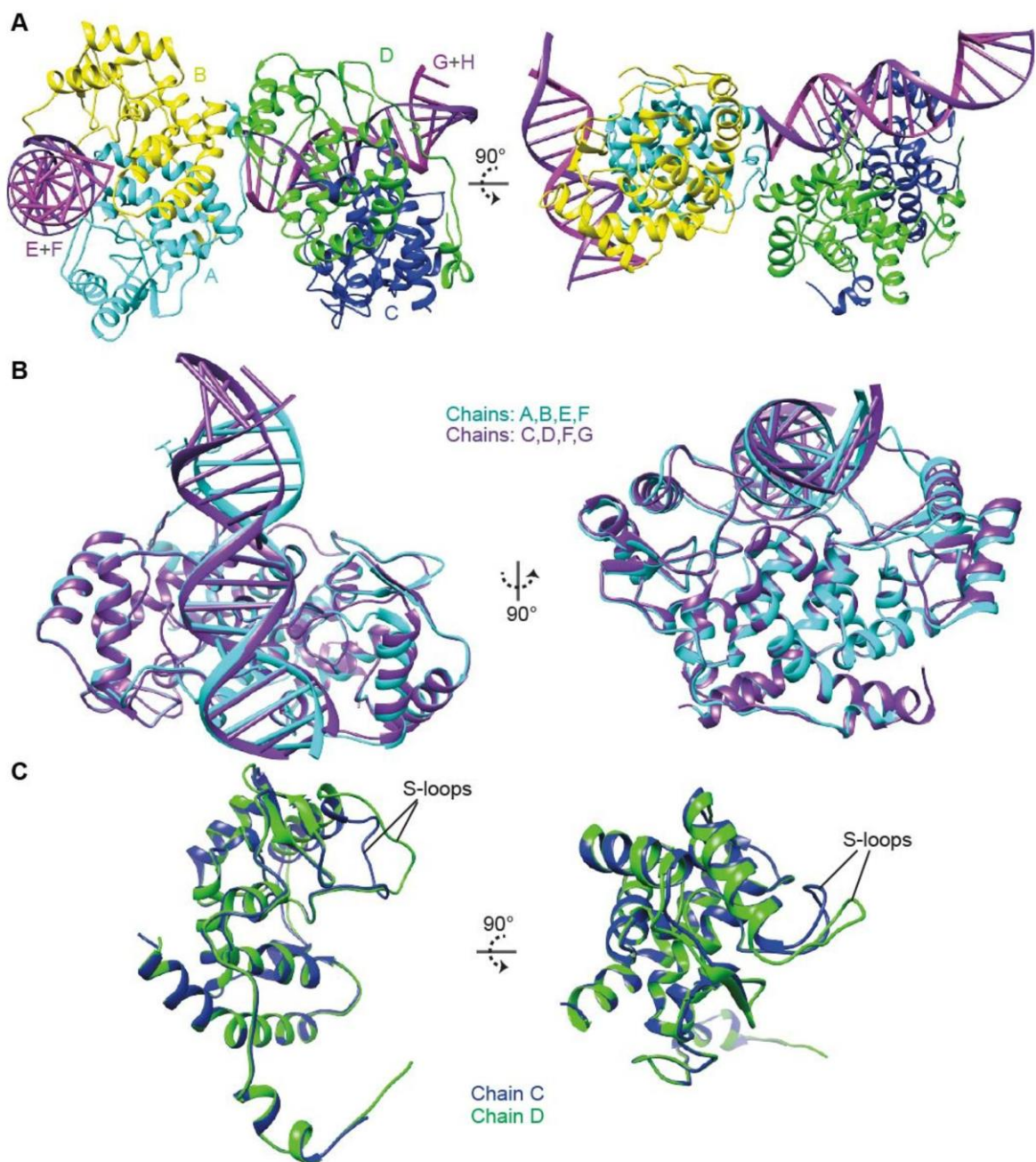
Supplementary Figure 1. Views of the ICP4N DNA-binding domain fold (A) Front and side views of a cartoon representation of the asymmetric unit (B) Superposition of assemblies comprising chains A,B,E,F with C,D,G,H. Superposition indicated a backbone RMSD of 0.8 Å for aa301-484 and 1.9 Å for the DNA base-pairs 1-13. (C) Superposition of chains C and D, which together form a homo-dimer.
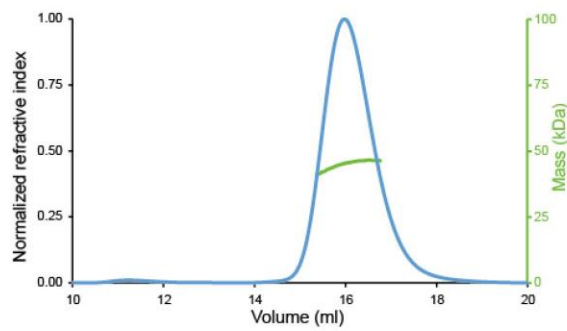
Supplementary Figure 2. MALS trace of ICP4ΔIDR, indicating a homo-dimer in solution.

ICP4N·IE3_19mer
ICP4NΔIDR·IE3_12mer
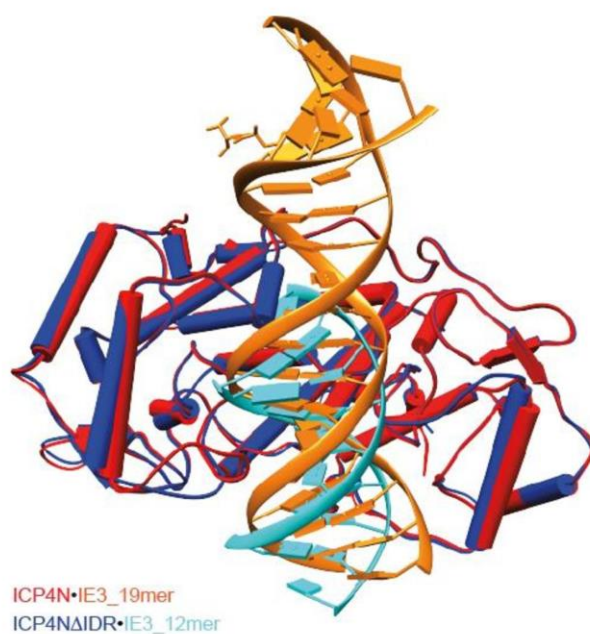
Supplementary Figure 3. Cartoon representation of the crystal structures with helices, sheets and loops shown as cylinders, arrows and coil respectively: (i) Superposition of ICP4N·IE3_19mer (protein red, DNA orange) and ICP4NΔIDR·IE3_12mer (protein blue, DNA cyan).
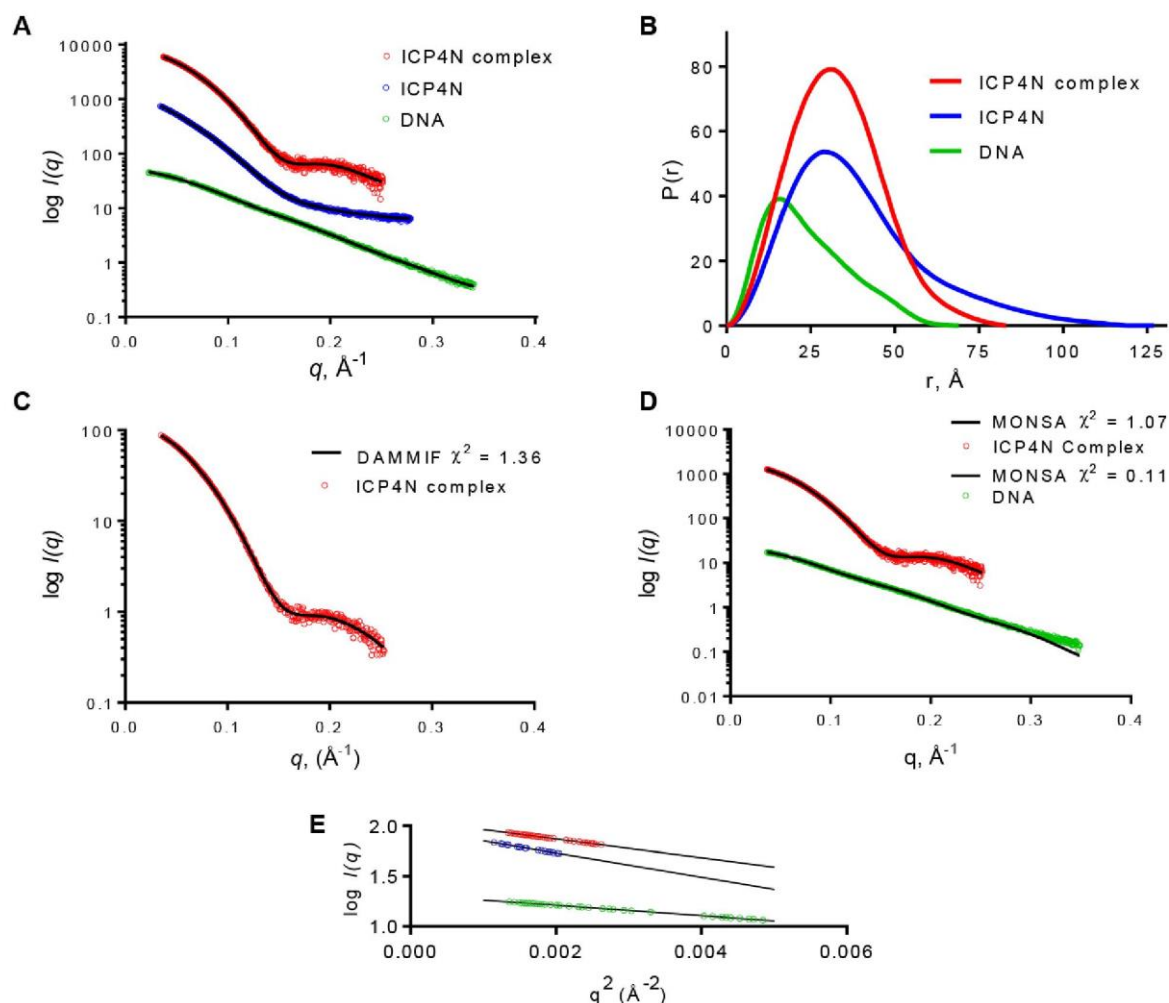
Supplementary Figure 4. Solution small angle X-ray scattering (SAXS) data. Presented for ICP4N·IE3_19mer complex, free ICP4N and free IE3_19mer DNA, colored red, blue and green respectively in all panels. (A) Processed experimental scatter curves of with their GNOM fit in black showing good agreement for each experiment. (B) Pair distance distribution functions P(r) obtained from each of the samples. (C) Scattering data showing the fit of one of the DAMMIF models to the data. (D) Scattering data showing the fit of one of the MONSA models to the data. (E) The low-angle region of the X-ray scattering data is show in the form of a Guinier plot, black lines indicate linear fit to each data set.
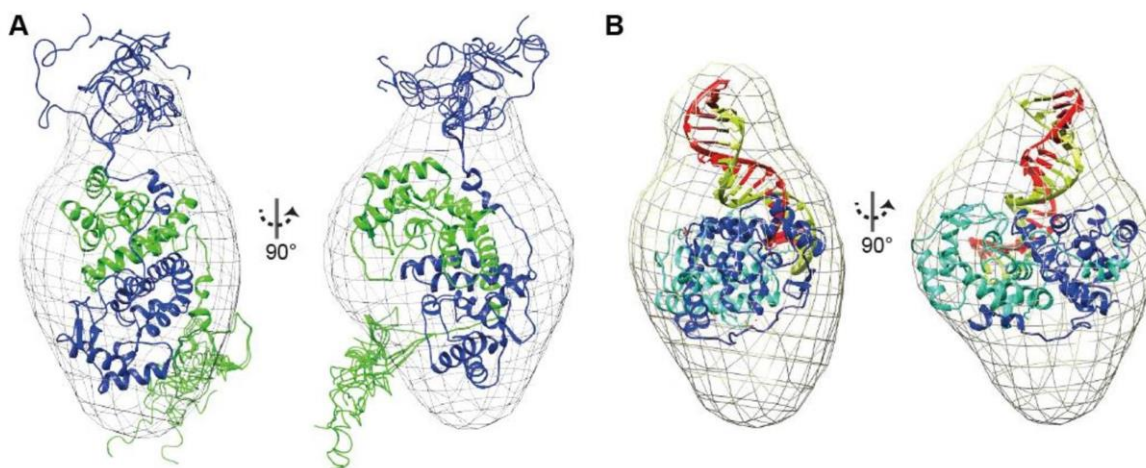
Supplementary Figure 5. DAMMIN *ab initio* model of the ICP4N·IE3_19mer complex. (A) Comparison of the ICP4N·IE3_19mer complex *ab initio* envelope with a SAXS derived structural ensemble representing free ICP4N (cartoon chains) illustrating the expanded structure of the free protein does not fit the more compact structure of the complex. (B) The ICP4N·IE3_19mer crystal structure of docked into the *ab initio* model of the complex generated revealing unoccupied volume around the DNA and above and below the protein dimer.
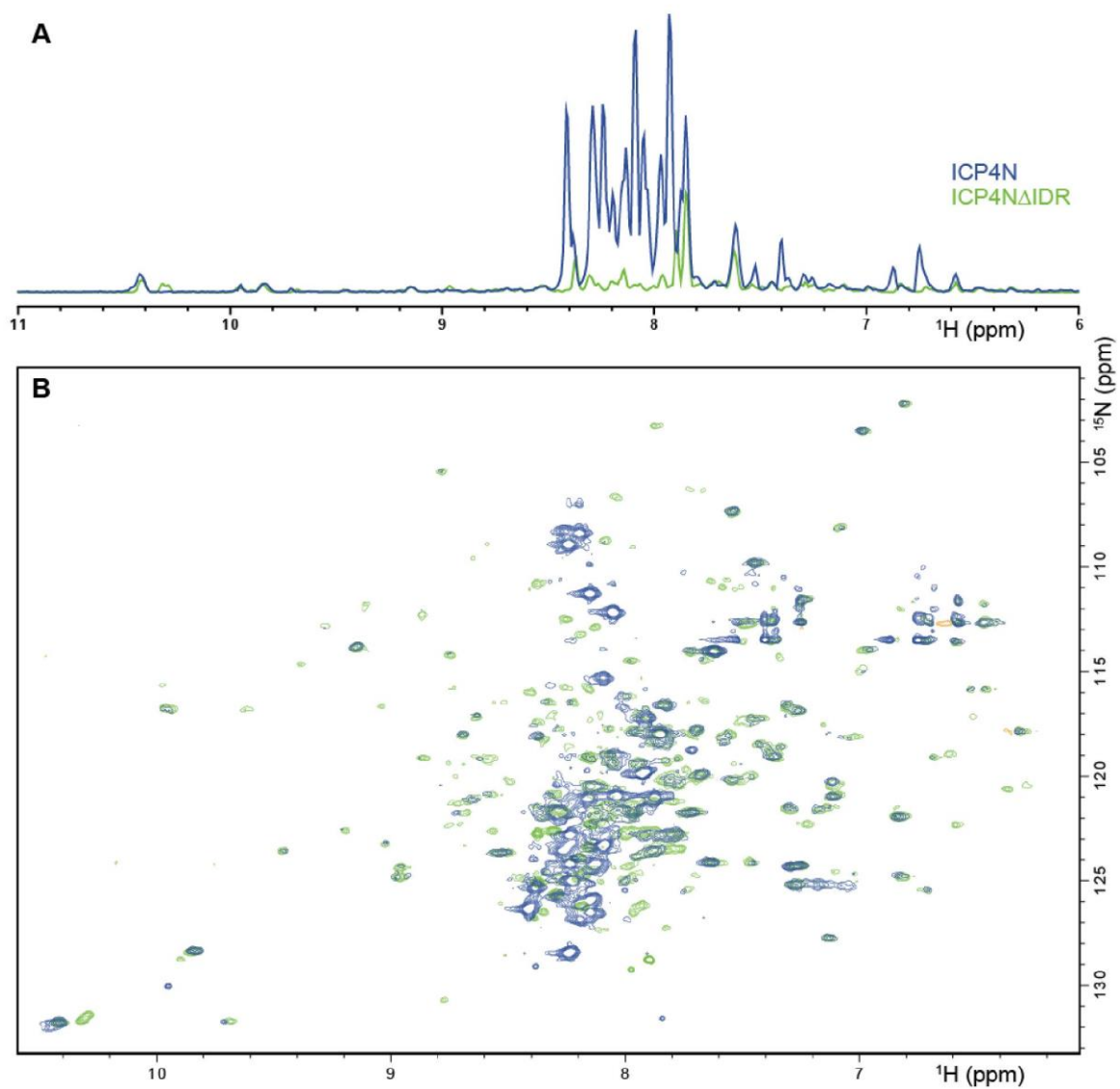
Supplementary Figure 6. ^{1}H-^{15}N correlation TROSY spectra of ICP4N and ICP4NΔIDR, colored blue and green respectively. (A) 1D ^{1}H projections superimposed, emphasizing the sharp, poorly dispersed backbone amide peaks of the intrinsically disordered region present in ICP4N but not ICP4NΔIDR. (B) 2D ^{1}H-^{15}N spectra superimposed.

A

B
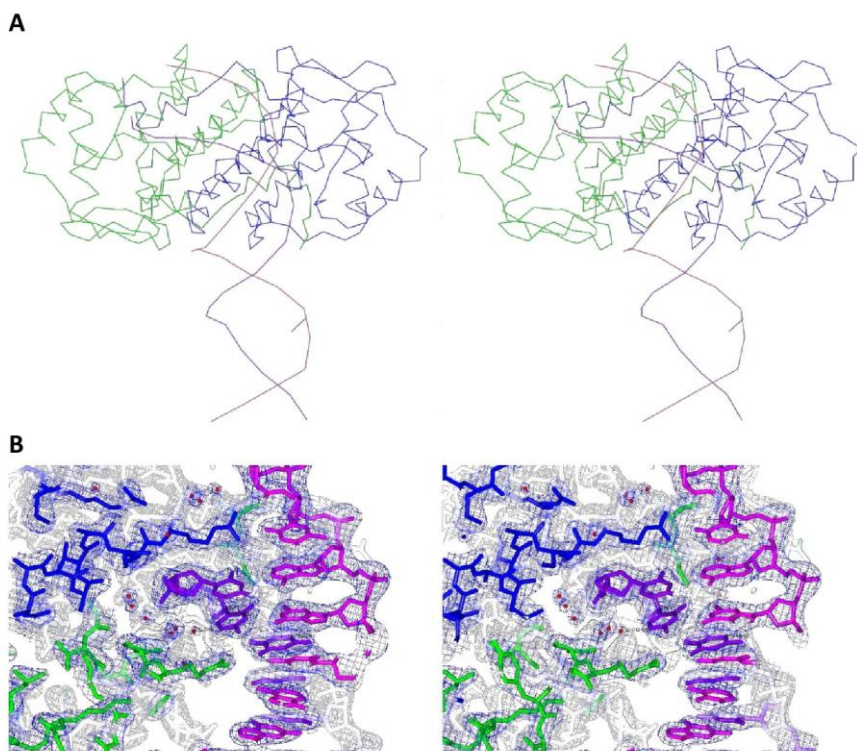
Supplementary Figure 7. Stereo images of the ICP4N·IE3_19mer complex showing the Cα chain trace (A) and a sample of the electron density map (B); the latter was created in CCP4 MG with the $2F_o$-F_c map at 1.5 σ. Chains C,D,G,H,J are colored blue, green, pink, purple and black respectively, waters shown as red spheres.

Supplemental References

1. Shepard, A.A., Imbalzano, A.N. and DeLuca, N.A. (1989) Separation of primary structural components conferring autoregulation, transactivation, and DNA-binding properties to the herpes simplex virus transcriptional regulatory protein ICP4. *J Virol*, **63**, 3714-3728.
2. Paterson, T. and Everett, R.D. (1988) The regions of the herpes simplex virus type 1 immediate early protein Vmw175 required for site specific DNA binding closely correspond to those involved in transcriptional regulation. *Nucleic Acids Res*, **16**, 11005-11025.
3. Paterson, T. and Everett, R.D. (1988) Mutational dissection of the HSV-1 immediate-early protein Vmw175 involved in transcriptional transactivation and repression. *Virology*, **166**, 186-196.
4. Allen, K.E. and Everett, R.D. (1997) Mutations which alter the DNA binding properties of the herpes simplex virus type 1 transactivating protein Vmw175 also affect its ability to support virus replication. *J Gen Virol*, **78 (Pt 11)**, 2913-2922.
5. Preston, C.M. (1979) Abnormal properties of an immediate early polypeptide in cells infected with the herpes simplex virus type 1 mutant tsK. *J Virol*, **32**, 357-369.
6. Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*, **60**, 2256-2268.
7. Holm, L. and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research*, **38**, W545-W549.
8. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, **372**, 774-797.

12.4 CAR Paper

Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis

Deepankar Gahloth[1,5], Mark S Dunstan[1,5], Daniela Quaglia[1,4], Evaldas Klumbys[1], Michael P Lockhart-Cairns[2,3] [iD], Andrew M Hill[1] [iD], Sasha R Derrington[1], Nigel S Scrutton[1], Nicholas J Turner[1] [iD] & David Leys[1]*

Carboxylic acid reductase (CAR) catalyzes the ATP- and NADPH-dependent reduction of carboxylic acids to the corresponding aldehydes. The enzyme is related to the nonribosomal peptide synthetases, consisting of an adenylation domain fused via a peptidyl carrier protein (PCP) to a reductase termination domain. Crystal structures of the CAR adenylation–PCP didomain demonstrate that large-scale domain motions occur between the adenylation and thiolation states. Crystal structures of the PCP–reductase didomain reveal that phosphopantetheine binding alters the orientation of a key Asp, resulting in a productive orientation of the bound nicotinamide. This ensures that further reduction of the aldehyde product does not occur. Combining crystallography with small-angle X-ray scattering (SAXS), we propose that molecular interactions between initiation and termination domains are limited to competing PCP docking sites. This theory is supported by the fact that (R)-pantetheine can support CAR activity for mixtures of the isolated domains. Our model suggests directions for further development of CAR as a biocatalyst.

The ATP- and NADPH-dependent reduction of carboxylic acids to the corresponding aldehydes is catalyzed by CAR enzymes[1–3]. These enzymes consist of an adenylation domain fused via a PCP to a reductase termination domain, and are related to the nonribosomal peptide synthetases (NRPS)[4–7]. CAR enzymes show substantial promise as green biocatalysts for the conversion of aromatic and short-chain carboxylic acids into the corresponding aldehydes[3]. Individual CARs have been shown to convert a wide range of substrates, featuring in applications that range from the conversion of long-chain fatty acids into fuel precursors[8–12] to the production of starting materials for cascade reactions that generate enantiomerically pure chiral building blocks[13] (**Supplementary Results**, **Supplementary Fig. 1**). Newly characterized CAR family members continue to expand this synthetic chemistry toolbox[14]. However, the unlocking of the catalytic potential of these enzymes is hampered by the shortage of mechanistic and structural insights. CAR represents a distant relative of the NRPS family and lacks any extension module (**Fig. 1a**). In fact, CAR consists of a substrate-activating adenylation domain that is more closely related in substrate specificity and sequence to the acyl-CoA synthetase members of the ANL superfamily of adenylating enzymes[15]. While the latter enzymes generate CoA thioester products, the CAR adenylation domain is fused to a PCP domain, and thus resembles the NRPS initiator (or adenylation) module with respect to the thioester product. The similarity of CAR to the modular NRPS enzymes extends to the inclusion of a terminator domain in CAR: an aldehyde-product-releasing reductase domain is fused to the acyl-intermediate carrier PCP. Whereas sequence similarity of the CAR adenylation domain with known structures is limited to ~20%, closely related reductase structures (~50% similarity to CAR) from NRPS enzymes are available[16,17]. However, these enzymes catalyze the progressive four-electron reduction of the PCP-bound acyl group to the corresponding alcohol. By contrast, CAR catalyzes a strictly two-electron reduction, releasing the corresponding aldehyde product (**Fig. 1a**).

To provide a detailed understanding of the CAR mechanism, we determined the crystal structure of individual CAR domains both with and without the PCP domain. A range of CARs were screened for crystallization, and structural data were obtained for enzymes from *Nocardia iowensis*, *Mycobacterium marinum* and *Segniliparus rugosus*. Combining the crystallographic structures with SAXS studies, we show that large-scale domain dynamics underpin catalysis in CAR. Furthermore, we reveal that docking of the phosphopantetheine group in the reductase active site leads to reorientation of the nicotinamide moiety of bound NADPH from a noncatalytic to a catalytically competent position. We propose that this change ensures that reduction does not proceed beyond the aldehyde product, and show that mutagenesis of a single Asp residue involved in the nicotinamide reorientation leads to modest formation of the four-electron-reduced alcohol product.

RESULTS

Structure of and substrate binding by the CAR A domain

To understand the key determinants underpinning substrate specificity and the mechanism of acyl-AMP formation, we determined the structure of the adenylation domains (A domains) of CAR from *N. iowensis* (CAR*ni*) and *S. rugosus* (CAR*sr*). In each case, the structure was obtained in complex with AMP, which remained tightly bound during purification (**Fig. 1b**). The AMP is bound at the A-domain center, establishing an extensive network of molecular contacts conserved across the ANL superfamily[15,18]. The most similar structures include the bacterial benzoate-CoA ligase[19] and the human medium-chain acyl-coenzyme A synthetase ACSM2A[20] (**Supplementary Fig. 2a**). Previous studies on these enzymes and other members of the ANL superfamily have revealed the presence of a mobile C-terminal domain (A$_{sub}$, residues 527–654) that
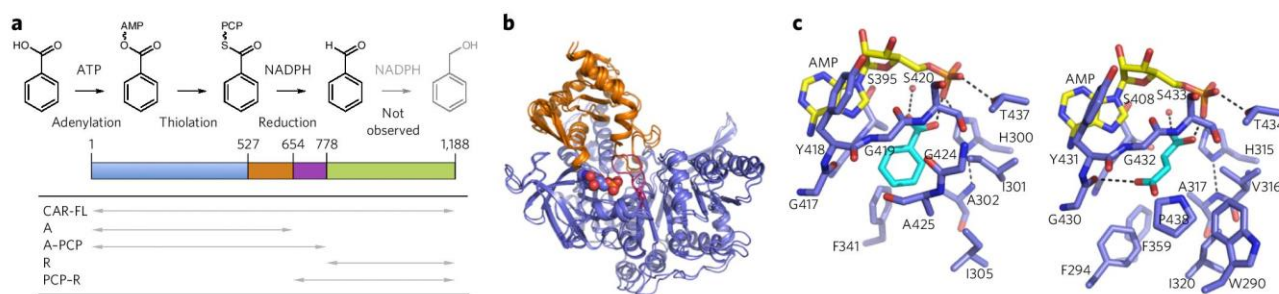
Figure 1 | Carboxylic acid reductase (CAR) is a modular enzyme. (**a**) Schematic overview of the CAR primary sequence, color coded according to individual domains. A general reaction, as catalyzed by CAR, is shown directly above. Various CAR fragments generated here are indicated by arrows. CAR-FL, full-length CAR. (**b**) An overlay of the CAR*sr* and CAR*ni* A-domain crystal structures, color coded as in **a**. (**c**) A side-by-side comparison of the CAR*ni* (left) and CAR*sr* (right) A-domain substrate-binding regions.

adopts distinct positions correlated to the specific reactions catalyzed. Domain motion has been inferred from structures using distinct nucleotide ligands, with the ATP-bound form representing the adenylation state of the enzyme and AMP or AMP-acyl-bound forms frequently found to be bound to the thiolation state[15]. For both CAR A-domain structures, the exact orientation of the A_{sub} domain is highly similar and, even though these contain AMP, is more akin to ANL structures that correspond to the adenylation state. The universally conserved Lys from ANL motif A10 (ref. 18) in the A_{sub} domain (K629 in CAR*sr*) makes contact with the nucleotide ligand (a hallmark of the adenylation state), while the conserved A8 Gly[18] (located close to the AMP-acyl substrate in the thiolation state; Gly532 in CAR*sr*) is directed away from the active site (**Supplementary Fig. 2a**). The P-loop that grips the terminal ATP phosphates is disordered in CAR*sr*, whereas it adopts a position wedged in between the A_{core} (residues 1–526) and the A_{sub} domains in CAR*ni*. In both CAR A-domain structures, the domain interface established between A_{sub} and the N-terminal A_{core} domain is more extensive than those of other ANL enzymes in the adenylation state. This suggests that the conformational equilibrium of the isolated CAR A_{sub} domain remains poised toward the adenylation state, apparently unaffected by the nature of the nucleotide ligand.

The acid substrate-binding site is identified by benzoic acid bound to CAR*ni* and an unidentified molecule (modeled as fumarate) present close to the AMP phosphate that was co-purified with CAR*sr* (**Fig. 1c**). Most of the relatively narrow substrate-binding pocket is lined by hydrophobic residues, with His300 (conserved in CAR enzymes; His315 in CAR*sr*) located close to both AMP phosphate and the substrate carboxylate. The active site volume of CAR*sr* is smaller than that in CAR*ni* due to the presence of Phe294 (Ser280 in CAR*ni*) and is distinct in shape due to insertion of Ala425 in CAR*ni*. The substrate specificity of these enzymes was determined by screening against a diverse carboxylic acid substrate panel including benzoic, heterocyclic, phenylacetic, phenylpropanoic and fatty acid substrates (**Supplementary Fig. 1**). *Para*- and *meta*-substituted benzoic acids, such as toluic acid, showed high levels of activity against several CAR enzymes. All CARs tested exhibited a poor tolerance for *ortho* substituents, presumably because of steric hindrance.

The CAR A–PCP didomain is a dynamic entity

To explore whether the A_{sub} conformation is affected by the presence of the PCP domain, and establish how the thioester linkage between the acyl-AMP and the PCP phosphopantetheine is formed, we determined the structure of the CAR*sr* adenylation–PCP region (A–PCP). CAR*sr* A–PCP could be crystallized in two distinct conformations that differ in the position of the A_{sub} and PCP domains. Although one A_{sub} domain conformation is similar to that previously observed for the single A domain (i.e., the adenylation state;

Fig. 2a,b), a second, distinct conformation was observed in a different crystal form (**Fig. 2c,d**). In the adenylation state, the PCP domain is positioned distantly from the A domain, with the A_{sub}–PCP linker region adopting an extended α-helical conformation. The PCP Ser702, which serves as the phosphopantetheine attachment site, is positioned 52 Å away from the bound AMP phosphate. By contrast, the relative orientation of both PCP and A_{sub} domains is altered in the second A–PCP crystal structure, with the Ser702–AMP phosphate distance dramatically shortened to 19 Å. This large reduction in distance is the cumulative effect of two distinct domain reorientations. The adenylation A_{sub} domain reoriented via a rotation of ~165 degrees at the A8–Lys528 hinge region, although the A_{sub} center of gravity remains largely in place. Both the position of the hinge and extent of rotation resemble the motion seen between the adenylation and thiolation conformations in other ANL family members[5,6,15] (**Fig. 2**; **Supplementary Fig. 2b**). The relative orientation of the PCP domain and the A_{sub} domain also changed, with an additional ~75 degree rotation at Ala651, leading to the more dramatic PCP domain reorientation. The cumulative effect of both the Lys528 and Ala651 rotations is a displacement of the PCP domain center of gravity by ~50 Å. The reorientation of both PCP and A_{sub} domains leads to a structure compatible with thiolation (**Fig. 2d**). Although phosphopantetheine is not present in our structure, an overlay with the related initiation module of NPRS LgrA[7] reveals a very similar PCP orientation relative to the A_{core} domain, with the phosphopantetheine linker accommodated by a narrow channel lined with conserved consensus sequence elements from the A_{sub} domain (**Supplementary Fig. 2b**).

In addition to the altered A_{sub}–A_{core} domain interaction of the thiolation conformation, a new domain interaction surface is established between the PCP domain and the A_{core} domain, overlapping in part with the adenylation A_{sub}–A_{core} domain interaction (**Supplementary Fig. 3**). The presence of this additional interaction is therefore likely to affect the adenylation–thiolation conformational equilibrium of the A_{sub} domain in comparison to the isolated A domain with the A–PCP didomain region. Biophysical parameters for the CAR*sr* A–PCP in solution were determined by SAXS, multi-angle light scattering (MALS) and area under the curve (AUC) analyses (**Supplementary Fig. 4**). Neither of the crystal structures obtained correlated well with the experimental data. Rigid-body modeling using both linker regions (Lys528 and Ala651) resulted in an ensemble of two models that accounted for the SAXS profile, one resembling the A–PCP thiolation state crystal structure and the other corresponding to an open conformation of the A_{sub} domain not observed in the crystal structures. These data are consistent with the expected flexibility of the A–PCP didomain. Whereas the observed position of the PCP domain in the adenylation state is influenced by crystal lattice contacts (a total of ten putative hydrogen bonding interactions can be observed between the PCP domain
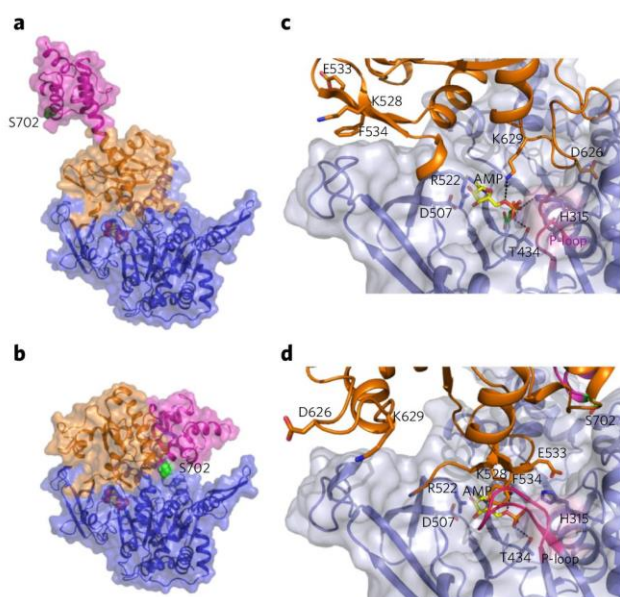
Figure 2 | Structure of the CAR*sr* A–PCP didomain reveals a dynamic entity. (**a**) The crystal structure of CAR*sr* A–PCP in the adenylation state. (**b**) The thiolation state crystal structure, color coded as in **Figure 1**. (**c,d**) A detailed view of the interactions established between the AMP-bound A$_{core}$ and the A$_{sub}$ domain for the adenylation state (**c**) and thiolation state (**d**).

and symmetry-related monomers), the conformation seen in the crystal structure likely represents one of a wider range of possible conformations the A–PCP didomain can adopt in solution.

CAR R-domain structures reveal an on–off equilibrium

We sought to complement our understanding of the CAR A–PCP didomain by determining the structure of the CAR reductase (R) domain. In particular, we wanted to determine how this domain ensures that further reduction of the aldehyde product does not occur. Crystal structures of the reductase domains from CAR*sr* and CAR*mm* (*M. marinum*) reveal that this region is highly similar in structure to the terminal reductase domains of other NPRS[16,17], albeit with a distinct orientation of the smaller substrate-binding domain (**Fig. 3a**). Furthermore, crystals could be readily obtained in a variety of crystal forms. For CAR*mm*, these crystal forms reveal that two distinct conformations of Asp984 (Asp998 in CAR*sr*) in the reductase active site can occur, corresponding to active and inactive forms of the reductase, respectively (**Fig. 3b**). In the active form, the nicotinamide moiety is ordered and placed adjacent to conserved residues Thr921 and Tyr956, which are proposed to form the oxyanion hole that assists in reduction of the thioester[21]. Asp984 is positioned pointing away from the nicotinamide, buried within the protein matrix and hydrogen bonding to Asp1034. This conformation is similar to that observed for the four-electron reductase from MxaA[17] (**Supplementary Fig. 5**). However, Asp984 adopts a distinct conformation in the majority of the CAR*mm* reductase monomers (similar to the Asp998 conformation observed for CAR*sr* reductase), leading to a disordered nicotinamide moiety and thus the inactive state. This disorder is a consequence of Asp984 and the Ser983 carbonyl group being located within the nicotinamide binding pocket. The motion of Asp984 between both conformations is concomitant with reorientation of the backbone of residues 983–985. Comparison of the various CAR*mm* reductase structures reveals that the conformation adopted by residues 983–985 appears to be linked to the position of the smaller substrate-binding domain. This suggests a possible means by which CAR ensures that reduction does not proceed

beyond the aldehyde product: binding of the substrate-thioester PCP affects the position of the substrate-binding domain and hence the conformational equilibrium of Asp984, which appears poised toward the inactive state in the absence of the PCP-acyl substrate. In contrast, the aldehyde product might lack sufficient binding affinity to affect this equilibrium. Recently studied polyketide synthase (PKS) enzymes containing reductase off-loading domains implicated in aldehyde (as opposed to alcohol) production also have an Asp residue at the corresponding position[22,23]. However, one of these residues further reduces the aldehyde product to alcohol in the absence of a suitable transaminase to the trap the transient aldehyde intermediate[22]. The conformational equilibrium between on and off states in the CAR R domain extends beyond Asp984 reorganization, including the backbone reorientation of residues 983–985. Hence, the presence of an Asp984-equivalent residue might not, in itself, be sufficient to determine the product scope. In the case of CAR*sr*, although mutation of the equivalent Asp998 to Gly does not appear to affect benzoic acid reduction rates, it does lead to formation of the modest levels of the alcohol product *in vitro* (**Supplementary Fig. 6**). Furthermore, in contrast to the wild-type (WT) CAR, the CAR*sr* D998G variant displays modest benzaldehyde reductase activity. This suggests that Asp998 is required to ensure strict two-electron reductase activity by CAR and that the R domain has little affinity for the isolated benzaldehyde (as opposed to the covalently linked benzoyl-phosphopantetheine moiety). The physiological substrate(s) of CAR is unknown, and it is possible that Asp998 is required to ensure that four-electron reduction does not occur for the corresponding aldehydes.

A phosphopantetheine-binding-induced activation mechanism

To verify our hypothesis, we determined the crystal structure of the CAR*sr* PCP–reductase (PCP–R) didomain. As observed for the isolated reductase domains, crystals could be readily obtained in the presence of NADPH, and a variety of crystal forms could be found. The PCP domain is clearly visible in the corresponding crystal structures, and adopts a conformation independent of crystal packing, presenting the first crystal structure for an intact PCP–R didomain (**Fig. 3c**). The PCP domain is docked onto the larger NADPH-binding domain, directly above the ribose 2′-phosphate binding pocket, with no direct contacts established to the smaller substrate-binding domain. This suggests that binding of the adenosine 2′,5′-diphosphate moiety of NADPH is likely to precede PCP docking, with reorganization of the nicotinamide group linked to phosphopantetheine binding. The PCP phosphopantetheine linker Ser702 is located 16 Å from the nicotinamide binding pocket. In the unmodified CAR*sr* PCP–R structure, the bound NADPH nicotinamide remains disordered, with Asp984 observed in the inactive conformation. In contrast, the crystal structure of the CAR*sr* PCP–R modified with phosphopantetheine reveals an active conformation (**Fig. 3d,e**). In the latter structure, the smaller substrate-binding domain has been reoriented slightly to establish contacts between the phosphopantetheine group and the Pro1013–Gln1015 region that, in turn, lead to reorientation of the Tyr966 loop region from the larger NADPH-binding domain with concomitant active site closure. A plausible model for the reductase acyl-PCP complex is made by extending the phosphopantetheine arm with a benzoyl moiety. This places the substrate acyl group directly above the *pro-S* hydrogen on the C4 nicotinamide, which is in agreement with observed kinetic isotope effects (KIE; **Supplementary Figs. 5 and 9a,b**). This model also reveals that few, if any, interactions are made between the R domain and the benzoyl moiety. The ~50-fold decrease in NADPH oxidation levels upon providing benzoyl-CoA as a substrate instead of benzoic acid and ATP (i.e., bypassing the A domain) reveals that the increase in effective concentration of the thioester substrate by covalent tethering is an important factor in achieving high enzyme activity (**Table 1**). However, no activity
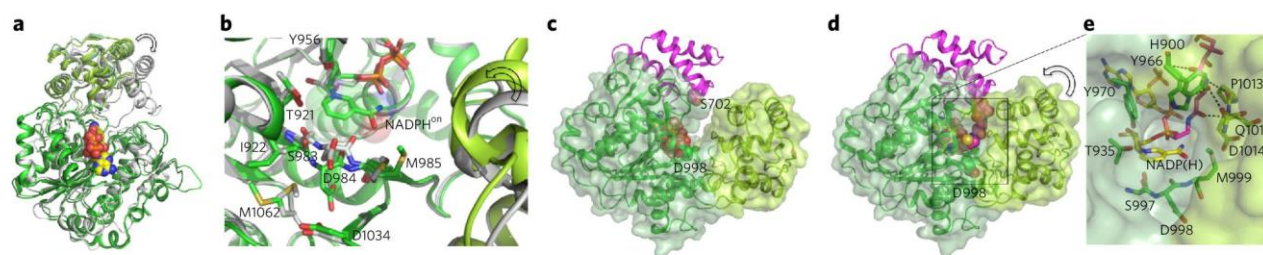
Figure 3 | Structure of CAR reductase and PCP–R regions. (**a**) An overlay of both CAR*mm* and CAR*sr* reductase domains (dark and light green, respectively) with the related MxaA reductase domain (gray). The position of the smaller substrate-binding domain is distinct for the CAR R domains, suggesting the possibility of domain motion, as indicated by the arrow. (**b**) An overlay of the active site of CAR*mm* reductase in the on (in light green carbons, with the NADPH nicotinamide visible, indicated by NADPH^{on}) and off (in gray carbons; nicotinamide disordered) active site conformations. The arrow indicates the domain motion of the smaller substrate-binding domain between both conformations. (**c,d**) The structures of unmodified (**c**) and phosphopantetheine-modified (**d**) CAR*sr* PCP–R didomain fragments, color coded as in **Figure 1**. (**e**) A detailed view of the CAR*sr* phosphopantetheine binding pocket.

could be observed with thiobenzoic acid, confirming that thioester substrate affinity is largely dependent on the phosphopantetheine linker region as opposed to the benzoyl moiety. This finding also provides a likely explanation for the modest reductase activity observed with benzaldehyde in the case of the CAR*sr* D998G variant. In addition, the free phosphopantetheine thiol group itself represents a steric hindrance to aldehyde binding directly above the nicotinamide. This is demonstrated both by the fact that the apo form of CAR*sr* D998G (i.e., unmodified with phosphopantetheine) displays higher benzaldehyde reduction activity compared to the holo-CAR*sr* D998G and by the fact that the presence of free phosphopantetheine inhibits benzaldehyde reduction observed for either of these variants (**Table 1**).

Mutation of residues involved in establishing contacts between the R domain and the phosphopantetheine group abolish CAR activity (**Supplementary Fig. 5d**). This strongly suggests that the presence of the substrate phosphopantetheine itself is required to induce Asp984–Ser985 reorientation and concomitant reductase activation, a possibility supported by the fact that benzoyl-CoA can act as a substrate for the isolated CAR*sr* R domain and R–PCP didomain. To prove that the (*R*)-pantetheine moiety itself is required for activation of the R domain, we tested for benzoic acid reductase activity of a CAR*sr* S702A mutant in the presence of (*R*)-pantetheine or a similarly sized thiol. Only the addition of (*R*)-pantetheine was able to rescue CAR activity for the S702A variant, confirming that free (*R*)-pantetheine can bypass the lack of PCP Ser702-bound phosphopantetheine group for the A domain, resulting in formation of a benzoyl-pantetheine thioester that can serve as a substrate for the R

domain. Biophysical parameters for the unmodified CAR*sr* PCP–R didomain in solution were determined by SAXS, MALS and AUC analyses, revealing that neither of the crystal structures obtained correlate well with the experimental data (**Supplementary Fig. 7**). Rigid-body modeling using Ser744 as a linker region revealed that an ensemble of multiple PCP–R conformations accounted for the SAXS profile obtained. This suggests a highly dynamic nature of the PCP domain in solution, which is in accordance with the relatively small interaction surface observed in the PCP–R crystal structure.

Modeling of the full-length CAR reveals a dynamic entity

Armed with the various crystal structures of the distinct CAR*sr* PCP-containing regions, we modeled the full-length CAR (**Fig. 4**). Whereas the A–PCP structure in the thiolation mode is clearly incompatible with the available PCP–R structures (the PCP cannot be simultaneously docked to both the respective partner domains), an overlay of the adenylation A–PCP conformation (i.e., PCP-off) with the modified PCP–R (i.e., PCP-on) reveals that only minor clashes occur between the adenylation and reductase domains (**Fig. 4a**). These clashes are easily avoided by minimal reorientation of the A$_{sub}$–PCP Ala651 hinge region. An overlay of the thiolation state A–PCP structure with the PCP–R SAXS-derived models reveals a very distinct model for the full-length enzyme (**Fig. 4b**).

To verify the dynamic nature of CAR, we determined the solution structure of CAR*sr* using SAXS (**Supplementary Fig. 8**). Neither of the models presented in **Figure 4a,b** account for the observed SAXS profile, and rigid-body modeling using four domains linked by the three identified hinge regions (K528, A651 and S744)

Table 1 | Substrate-dependent NADPH oxidation rates (min⁻¹) for CARsr WT and variants

Substrates added, in addition to ATP and NADPH	Full-length CAR	CAR S702A	Apo-CAR D998G	CAR D998G	R domain	PCP–R didomain	A + PCP–R didomain
Benzoic acid	278 ± 7	<0.001	<0.001	255 ± 9	<0.001	<0.001	<0.001
Benzoyl-CoA	11 ± 2	3.1 ± 0.4	4.3 ± 1.6	9.4 ± 1.0	1.3 ± 1.0	2.5 ± 0.2	1.4 ± 1.2
Benzoyl-CoA + (*R*)-pantetheine	ND	ND	2.7 ± 0.4	6.9 ± 0.7	ND	ND	ND
Thiobenzoate	<0.001	<0.001	ND	ND	<0.001	<0.001	<0.001
Benzaldehyde	<0.001	<0.001	0.15 ± 0.05	0.03 ± 0.02	ND	ND	ND
Benzaldehyde + (*R*)-pantetheine	ND	ND	<0.001	<0.001	ND	ND	ND
Benzoic acid + (*R*)-pantetheine	261 ± 4	6.8 ± 0.1	ND	ND	<0.001	<0.001	3.0 ± 0.8
Benzoic acid + decanethiol	281 ± 11	<0.001	ND	ND	<0.001	<0.001	<0.001
Benzoic acid + pentanethiol	238 ± 3	<0.001	ND	ND	<0.001	<0.001	<0.001

A range of substrates (with or without the presence of a thiol compound) were incubated with CAR*sr*, NADPH and ATP. Substrate NAPDH oxidation rates were highest for the wild-type enzyme with benzoic acid, but modest activity could be observed for both full-length and individual CAR domains when using benzoyl-CoA or in the presence of (*R*)-pantetheine. ND, not determined.
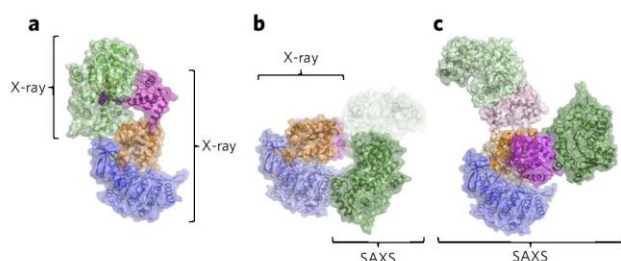
250

Figure 4 | Modeling of the full-length CAR structure. (**a**) A model for the CAR*sr* adenylation–reduction state, derived from an overlay of the PCP domains of the A–PCP (adenylation state) and PCP–R crystal structures, color coded as in **Figure 1**. (**b**) A model for the CAR thiolation state, derived by superimposition of the A–PCP thiolation state structure with the PCP–R SAXS models. Multiple conformations of the reductase domain are shown, as derived from the SAXS profile for the isolated PCP–R didomain (**Supplementary Fig. 7**). (**c**) The ensemble of structures that can account for the observed SAXS profile of the full-length enzyme (**Supplementary Fig. 8**). Although the individual models used can account for the observed SAXS profile of the full-length enzyme, this does not mean these are necessarily highly populated conformations.

Figure 5 | A dynamic model for CAR. A schematic representation of the conformational rearrangements that occur during the CAR enzymatic cycle. We propose that the CAR dynamic equilibrium randomly explores various conformations, some relevant with respect to catalysis and others used for ligand exchange, and is not poised toward either the adenylation–reduction or thiolation conformations. In addition to the large-scale reorientation of the individual domains, the reductase component displays a relatively small-scale conformational equilibrium that affects the active site region and is inherently poised toward the inactive state. Docking of the phosphopantetheine group in the reductase active site leads to substrate-induced active site remodeling and activation. It is possible that reduction might occur simultaneously with activation of the next substrate molecule in the adenylation domain. PPi, inorganic phosphate.

reveals that two distinct conformations, both representing an open conformation of the enzyme, can account for the observed profile (**Fig. 4c; Supplementary Fig. 8**). The addition of ATP and/or NADPH$_4$ (a reduced version of NADPH) did not substantially alter the scattering behavior of the sample. This therefore suggests that the dynamic equilibrium between the various conformations of the full-length enzyme is not poised toward either the adenylation–reduction or thiolation conformations, but likely randomly explores various conformations, some relevant with respect to catalysis and others applicable to ligand exchange. Substantial solvent viscosity effects on CAR catalytic efficiency support this hypothesis (**Supplementary Fig. 9c,d**).

Domain-exchanged CARs retain activity

While mixtures of isolated CAR domains do not display benzoic acid reductase activity, the addition of free (*R*)-pantetheine can support CAR activity in such mixtures (**Table 1**). This confirms that neither covalent linkage between individual domains nor the presence of a PCP domain is an absolute prerequisite for activity. Hence, this suggests that CAR enzyme activity is not inextricably linked to the exact nature of the covalent linkage between the individual CAR domains nor to the respective interdomain surface contacts. Hence, this robust dynamic nature of CAR suggests that domain exchange to produce new hybrid CARs should be feasible. Similar domain-exchange experiments carried out on NPRS systems have met with some success[24–27]. To test our hypothesis, we exchanged domains from CAR*ni* and CAR*mm* to create nmCAR and mnCAR hybrids. Interestingly, both hybrids showed activity toward a range of diverse substrates (**Supplementary Fig. 10**). Further studies will be required to determine the extent to which the respective A and R domains determine substrate specificity, but our present data suggest that the A domain is the key determinant.

DISCUSSION

The reduction of carboxylic acids to the corresponding aldehydes is a relatively simple reaction, but often suffers from the inadvertent production of alcohol byproducts through further reduction of the aldehyde product under the conditions used. The bacterial carboxylic acid reductase is able to catalyze the strictly two-electron reduction of a range of carboxylic acids using ATP and NADPH. CAR enzymes offer tremendous potential for future applications in organic synthesis, particularly with respect to generating aldehydes
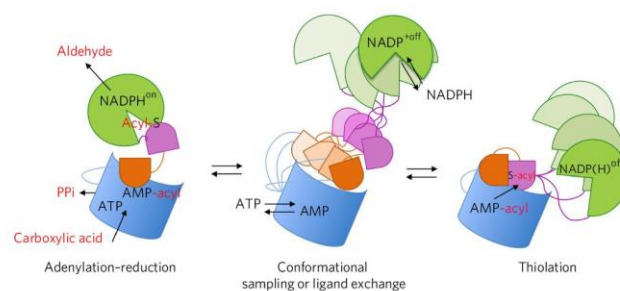
from carboxylic acids under mild reaction conditions. Moreover, as recently demonstrated[13], CARs can be combined with other enzymes (for example, transaminases and imine reductases) in cascade processes, thereby enabling the conversion of simple, inexpensive starting materials to products with greater functionality and complexity. Such cascades will increasingly require engineered CARs with broad substrate scope and activity for the synthesis of a wide range of target molecules.

Our crystallographic and solution characterization of CAR suggests an intricate mechanism that ensures reduction of the carboxylic acid substrate is restricted to the aldehyde level. Crystal structures of the isolated reductase domain reveal two distinct conformations of the active site region, only one of which appears to be catalytically competent (i.e., the 'on' state) by virtue of the fact the nicotinamide moiety is in close proximity to the conserved Thr and Tyr involved in catalysis. However, the on conformation is only apparent in a small subset of the available reductase domain monomer structures, with the majority adopting an 'off' state that has a disordered nicotinamide binding mode. The on–off conformational equilibrium thus appears poised toward the off state for the isolated reductase domain and is largely centered around the reorientation of a key Asp residue located close to the nicotinamide. Substrate binding, which in this case can be considered to be with the acyl-phosphopantetheine-PCP domain, can affect this equilibrium, leading to substrate-induced enzyme activation. The latter process has been observed in a wide range of enzymes, and often forms an integral part of those systems that contain inherently dangerous activities from the cellular perspective (i.e., proteases[28], kinases[29] and NAD$^+$-dependent ADP-ribosyltransferases[30]). Substrate-induced active site remodeling offers a suitable safeguard in such cases.

In the case of CAR, the structures of the PCP–R didomain in the presence and absence of the phosphopantetheine modification reveal that active site remodeling occurs as a direct consequence of an induced-fit phosphopantetheine docking in the reductase active site cleft and not in the PCP domain itself. This is supported by the fact that benzoyl-CoA can act as a substrate. Although the CAR R-domain key Asp residue that is affected in position by the induced fit of the pantetheine moiety is present in other terminal reductase domains[22,23], it remains unclear whether this represents a more general mechanism to control reduction activity.

In contrast to the intricate activation of the reductase domain, the communication and transfer of the activated acyl group appears to depend on a relatively simple and robust system consisting of a beads-on-a-string type of arrangement (**Fig. 5**). The overlay of the adenylation A–PCP conformation with the modified PCP–R structure presents a putative model for the full-length CAR*sr* structure when in the adenylation–reduction conformation (**Fig. 4a**). This model also suggests that reduction of the substrate–phosphopantetheine linkage might occur simultaneously with activation of the next substrate molecule in the adenylation domain, although a scenario whereby these reactions occur sequentially[6] remains equally likely. A putative model for the thiolation state of the full-length CAR is assembled by combining the thiolation A–PCP crystal structure (i.e., PCP-on) with the PCP–R SAXS models (i.e., PCP-off, **Fig. 4b**). This model reveals that a dramatic reorientation of the reductase domain with respect to the adenylation is required for the transition from the adenylation–reduction conformation to the thiolation conformation. Crucially, a comparison of the two models clearly suggests the absence of a long-lived interaction between the CAR terminal domains. The SAXS models of the full-length enzyme present a dynamic picture with the various domains sampling the conformational space available (**Fig. 4c**). Hence, the enzyme appears to function as a highly mobile entity that makes use of competing docking sites for the PCP domain on the A_{core} and reductase domains. This is consistent with the lack of ATP- and benzoic-acid-dependent NADPH consumption activity when adding either of the CAR A domains with the cognate PCP–R fragments (**Table 1**). The lack of interactions and/or sophisticated communication between the CAR terminal domains suggests that the CAR catalytic repertoire can be diversified by domain exchange with the CAR family, as we demonstrate here. A similarly robust design was recently uncovered for intermodular communication in hybrid PKS–NRPS[31], underpinning the engineering of polyketide–nonribosomal peptide interfaces[32]. By analogy, this suggests the possibility to explore new chemistry by fusion of CAR domains with appropriate NRPS components in the future. The fact that free (*R*)-pantetheine can be used to support the activity of isolated domains suggests that it might even be possible to bypass the need for covalent linkage between these components. This also confirms that the A domain can catalyze ATP-dependent formation of acyl-pantetheine in the absence of the PCP domain, reminiscent of the related acyl-CoA ligase activity[19]. Hence, although the isolated A and R domain are functional in the presence of (*R*)-pantetheine, the increase in effective concentration arising from covalent tethering to the PCP domain is a substantial contributor to the efficiency of the natural system.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

References

1. Gross, G.G. Formation and reduction of intermediate acyladenylate by aryl-aldehyde. NADP oxidoreductase from *Neurospora crassa*. *Eur. J. Biochem.* **31**, 585–592 (1972).
2. He, A., Li, T., Daniels, L., Fotheringham, I. & Rosazza, J.P.N. *Nocardia* sp. carboxylic acid reductase: cloning, expression, and characterization of a new aldehyde oxidoreductase family. *Appl. Environ. Microbiol.* **70**, 1874–1881 (2004).
3. Akhtar, M.K., Turner, N.J. & Jones, P.R. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc. Natl. Acad. Sci. USA* **110**, 87–92 (2013).
4. Strieker, M., Tanović, A. & Marahiel, M.A. Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234–240 (2010).
5. Gulick, A.M. Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases. *Curr. Opin. Chem. Biol.* **35**, 89–96 (2016).
6. Drake, E.J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).
7. Reimer, J.M., Aloise, M.N., Harrison, P.M. & Schmeing, T.M. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239–242 (2016).
8. Lennen, R.M. & Pfleger, B.F. Microbial production of fatty acid-derived fuels and chemicals. *Curr. Opin. Biotechnol.* **24**, 1044–1053 (2013).
9. Kallio, P., Pásztor, A., Thiel, K., Akhtar, M.K. & Jones, P.R. An engineered pathway for the biosynthesis of renewable propane. *Nat. Commun.* **5**, 4731 (2014).
10. Duan, Y. *et al.* Synthesis of α,β-unsaturated esters via a chemo-enzymatic chain elongation approach by combining carboxylic acid reduction and Wittig reaction. *Beilstein J. Org. Chem.* **11**, 2245–2251 (2015).
11. Xu, P., Qiao, K., Ahn, W.S. & Stephanopoulos, G. Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proc. Natl. Acad. Sci. USA* **113**, 10848–10853 (2016).
12. Zhou, Y.J. *et al.* Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat. Commun.* **7**, 11709 (2016).
13. France, S.P. *et al.* One pot cascade synthesis of mono- and di-substituted piperidines and pyrrolidines using carboxylic acid reductase (CAR), ω-transaminase (ω-TA) and imine reductase (IRED) biocatalysts. *ACS Catal.* **6**, 3753–3759 (2016).
14. Finnigan, W. *et al.* Characterization of carboxylic acid reductases as enzymes in the toolbox for synthetic chemistry. *ChemCatChem* **9**, 1005–1017 (2017).
15. Gulick, A.M. Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811–827 (2009).
16. Chhabra, A. *et al.* Nonprocessive [2 + 2]e- off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. USA* **109**, 5681–5686 (2012).
17. Barajas, J.F. *et al.* Comprehensive structural and biochemical analysis of the terminal myxalamid reductase domain for the engineered production of primary alcohols. *Chem. Biol.* **22**, 1018–1029 (2015).
18. Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651–2674 (1997).
19. Thornburg, C.K., Wortas-Strom, S., Nosrati, M., Geiger, J.H. & Walker, K.D. Kinetically and crystallographically guided mutations of a benzoate coa ligase (Bada) elucidate mechanism and expand substrate permissivity. *Biochemistry* **54**, 6230–6242 (2015).
20. Kochan, G., Pilka, E.S., von Delft, F., Oppermann, U. & Yue, W.W. Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. *J. Mol. Biol.* **388**, 997–1008 (2009).
21. Kavanagh, K.L., Jörnvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65**, 3895–3906 (2008).
22. Awodi, U.R., Ronan, J.L., Masschelein, J., de Los Santos, E.L.C. & Challis, G.L. Thioester reduction and aldehyde transamination are universal steps in actinobacterial polyketide alkaloid biosynthesis. *Chem. Sci.* **8**, 411–415 (2017).
23. Peng, H. *et al.* Deciphering piperidine formation in polyketide-derived indolizidines reveals a thioester reduction, transamination, and unusual imine reduction process. *ACS Chem. Biol.* **11**, 3278–3283 (2016).
24. Mootz, H.D., Schwarzer, D. & Marahiel, M.A. Construction of hybrid peptide synthetases by module and domain fusions. *Proc. Natl. Acad. Sci. USA* **97**, 5848–5853 (2000).
25. Hur, G.H., Vickery, C.R. & Burkart, M.D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **29**, 1074–1098 (2012).
26. Calcott, M.J. & Ackerley, D.F. Portability of the thiolation domain in recombinant pyoverdine non-ribosomal peptide synthetases. *BMC Microbiol.* **15**, 162 (2015).
27. Khosla, C., Herschlag, D., Cane, D.E. & Walsh, C.T. Assembly line polyketide synthases: mechanistic insights and unsolved problems. *Biochemistry* **53**, 2875–2883 (2014).
28. Merdanovic, M., Mönig, T., Ehrmann, M. & Kaiser, M. Diversity of allosteric regulation in proteases. *ACS Chem. Biol.* **8**, 19–26 (2013).
29. Kornev, A.P. & Taylor, S.S. Dynamics-driven allostery in protein kinases. *Trends Biochem. Sci.* **40**, 628–647 (2015).
30. Langelier, M.F. & Pascal, J.M. PARP-1 mechanism for coupling DNA damage detection to poly(ADP-ribose) synthesis. *Curr. Opin. Struct. Biol.* **23**, 134–143 (2013).
31. Dowling, D.P. *et al.* Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. USA* **113**, 12432–12437 (2016).

32. O'Connor, S.E., Walsh, C.T. & Liu, F. Biosynthesis of epothilone intermediates with alternate starter units: engineering polyketide-nonribosomal interfaces. *Angew. Chem. Int. Edn Engl.* **42**, 3917–3921 (2003).

Acknowledgments

Author contributions

M.S.D. cloned, expressed and purified various CAR enzymes (both truncations and full-length as well as hybrid forms) and obtained crystal structures for both A and R domains. D.G. cloned, expressed and purified various CAR enzymes and obtained crystals structures for the PCP-didomain constructs in addition to isolated R domain structures. D.G. and M.P.L.-C. performed SAXS data collection and modeling. E.K., D.Q., A.M.H. and S.R.D. performed kinetic data analysis. All authors discussed the results and participated in writing of the manuscript. N.J.T., N.S.S. and D.L. initiated and directed this research.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to D.L.

253

ONLINE METHODS

Cloning and molecular biology. All FL-CAR enzymes were cloned into pET28b using Infusion HD Cloning technology to give an N-terminal His-tag. The experiment was designed *in silico* using the SnapGene software and pET28-b vector linearized using the restrictions sites *NdeI* and *EcoRI*. Individual CAR domains and their boundaries were identified with the software MOTIF scan and PSI-blast searches. Sequence alignments with known NRPS modules were used to select the final domain boundaries for cloning. Truncated CAR constructs containing just N-terminal domains were designed and cloned with a N-terminal His-tag and C-terminal CAR domains with a C-terminal His-tag. Individual CAR domains were cloned using the same protocol adopted for FL-CARs except that *NcoI* was used instead of *NdeI* for N-terminal constructs. For CAR*sr* constructs, DNA corresponding to CAR*sr* full-length (FL), adenylation (A), adenylation–PCP (A–PCP), PCP-reductase (PCP–R) and reductase (R) was amplified by PCR from the *CARsr* gene. PCR products were cloned into the Ligation independent cloning site of pNIC28a-Bsa4 vector using the ligation-independent method (Infusion HD, Clontech). These constructs have N-terminal His-tag and tobacco etch virus (TEV) protease cleavage site followed by amino acid sequence. Sequence of all the constructs were verified by DNA sequencing (Eurofins). Point mutations in CAR*sr* were introduced by Q5 site-directed mutagenesis kit (NEB).

Protein expression and purification. Expression and purification of CAR*mm/ni* and *sr* enzymes and individual (di)domain. *E. coli* BL21 (DE3) were transformed with one or two plasmids: a pET28-b containing FL-CAR or containing the individual CAR domains, and where modification with phospho-pantetheine was required (for activity studies and/or crystallography), a second vector, pCDF1b-Sfp, containing the gene for the expression of Sfp: phosphopantetheine transferase from *Bacillus subtilis*. A single colony was then used to inoculate 5 mL of LB medium containing 50 μg/mL streptomycin (only where the pCDF1b-Sfp vector was present) and kanamycin, and grown overnight, shaking at 37 °C. For CAR*mm/ni* expression, the inoculum was transferred and grown to mid-log phase in 2 L baffled flasks, containing TB medium, induced with 0.4 mM IPTG and shaken for 24 h at 20 °C. Proteins were purified to homology using Nis60 Ni resin (Clontech), which was followed by size-exclusion chromatography with a prep grade 26/60 Hi-load Superdex S200 purification column. Final samples were eluted in 25 mM Tris pH 8.0 and 150 mM NaCL.

For expression and purification of the CAR*sr* constructs (FL, A, A–PCP, PCP–R and R), cultures were incubated at 37 °C to an optical density (OD_{600}) between 0.6 and 0.8, and protein expression was induced with 0.1 mM IPTG at 20 °C for 16 h. Cells were harvested by centrifugation and stored in −80 °C until further use. For purification, cells were resuspended in Ni-binding buffer (20 mM HEPES pH 7.4, 0.5 M NaCl, 10 mM Imidazole) supplemented with EDTA-free protease inhibitor cocktail tablet (Sigma). The cells were lysed by cell-disruptor followed by centrifugation at $48,000 \times g$ for 1 h. Cleared supernatant was loaded on to equilibrated Ni-NTA agarose resin (Qiagen) and final elution was done with 250 mM imidazole. Affinity His-tag was removed by TEV protease cleavage at 4 °C. Cleaved protein was passed back over the Ni-NTA agarose beads and flow-through was collected. Cleaved protein was further cleaned by ion-exchange chromatography using the ResQ (6 ml) column (GE Healthcare). Protein was eluted in a linear gradient of NaCl (0–1 M). Protein-containing fractions were pooled together and concentrated up to ~10 mg/ml with 10 K MWCO (Vivaspin) filtration unit.

***In vitro* modification of CAR*sr* PCP–R by Sfp.** CAR*sr* PCP–R didomain was modified by promiscuous phosphopantetheinyl transferase Sfp according to the method described previously[33]. In brief, 0.1 μM purified Sfp was mixed with 5 μM of PCP–R protein and 5 μM benzoyl-CoA (Sigma) in 50 mM HEPES pH 7.5 and 10 mM $MgCl_2$ in total 100 μl volume. This reaction mixture was incubated at room temperature minimum for 2 h. To remove Sfp, the mixture was separated using gel filtration.

Screening for CAR activity. The activity of FL-CAR toward a range of carboxylic acids was tested by following the reduction of NADPH at 340 nm over 1 min. Measurements were performed using a Tecan infinite M200 pro 96-well plate spectrophotometer fitted with an injector. Seven alternative CARs were tested using the following conditions: 0.15 mM NADPH 1 mM ATP, 10 mM $MgCl_2$, 5 mM substrate, 2.5% DMSO (for substrate solubility) in 50 mM Tris–HCl pH 7.5 at room temperature.

The reactions were tested in 96-well plates using a total volume of 200 μL per well. The amount of enzyme used for each reaction was constant (5 μL). The enzyme was pre-pipetted in the wells and the rest of the reaction mix injected in each well just before the measurement started. Each measurement ran for 1 min, and the slope was then calculated using the Magellan software considering the variation in absorbance between 3 s and 12 s, over 10 points of measure: this allowed for elimination of the mixing effect observed occasionally just after injection and allowed for measurement of the activity of the enzymes in the linear range. All measurements were corrected with an appropriate blank. All measurements were generally repeated in quadruplicate and at least in triplicate for each sample. The substrates tested are reported in **Supplementary Figure 1**.

Activity assays for individual N-terminal domains were carried out with the commercially available EnzChek assay kit and 1 mM ATP, 10 mM $MgCl_2$, 5 mM substrate and 50 mM Tris–HCl pH 7.5. Reactions were monitored for the conversion of 2-amino-6-mercapto-7-methylpurine riboside (MESG) by purine nucleoside phosphorylase (PNP) to ribose 1-phosphate and 2-amino-6-mercapto-7-methyl- purine at 360 nm over 15 min.

Substrate dependent NADPH oxidation activity of CAR*sr* (FL, ACP-R, R, A+ACP-R as well as variants) was monitored by NADPH consumption at 340 nm. Assays were carried out in 50 mM Tris–Cl pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM ATP, 1 mM NADPH, and 5mM of each benzoic acid, benzoyl-CoA or thiobenzoate. The effect of (*R*)-pantetheine, decanethiol and pentanethiol on substrate dependent NADPH oxidation was assessed by addition to 2 mM final concentration of these compounds. Activity assays using benzaldehyde as a substrate were carried out using a similar protocol, but with 15 mM of benzaldehyde as the final concentration.

Crystallization and structure determination. Crystals of A domain CAR*ni* (30 mg/ml) were obtained using the sitting-drop vapor-diffusion method and grown within 7 d at 4 °C in 0.12 M ethylene glycols 0.1 M Tris–Bicine pH 8.5 30% glycerol/PEG 4K. The structure of A domain of CAR*ni* was solved by single-wavelength anomalous (SAD) method. Experimental phases were obtained by soaking crystals with 1 M KI for 30 s before flash freezing in liquid N_2. A SAD diffraction data set was collected from a single flash-cooled crystal at Diamond light source (beamline IO4) and reflections merged and scaled with Xia2 (ref. 34). The A domain CAR*ni* was solved using the SHELX C/D/E software package[35]. A total of eight iodides were located (SHELX C), with sufficient phasing power to generate a heavy atom substructure (SHELX D), for phasing and density modification with SHELX E. Initial model building was done with ARP/wARP followed by iterative cycles of manual model building and refinement in COOT[36] and with Phenix.refine[37], respectively. Structure of the A domain CAR*ni* in complex with benzoic acid was obtained by soaking the crystals in a solution of mother liquor supplemented with 0.1 M benzoic acid before flash cooling.

Crystals of R domain CAR*mm* (20 mg/ml) were obtained by sitting-drop vapor diffusion in 0.2 M Na citrate tribasic dehydrate, 0.1 M Bis–tris propane pH 6.5 and 20% PEG 3350. Co-crystals of CAR*mm* R domain with $NADPH_4$ were obtained in 0.2 M sodium sulfate, 0.1 M Bis–Tris–propane pH 6.5, 20% PEG3350 (SG $C222_1$) and 0.2 M ammonium tartrate dibasic, 20% PEG3350 (SG $P2_1$). Crystals were cryo-protected either in PFO oil (Hampton research) or PEG 200 before flash freezing in liquid nitrogen. Structure of Red CAR*mm* was solved by molecular replacement using Phaser[35] and the NRP terminal reductase domain from *Mycobacterium smegmatis* (PDB: 4DQV).

Optimized crystals of CAR*sr* A domain (15 mg/ml) were obtained in 0.2 M calcium chloride hydrate and 20% PEG 3350. CAR*sr* A–PCP (15 mg/ml) (thiolation state) crystals were grown in 1.5 M lithium sulfate and 0.1 M HEPES pH 7.5. CAR*sr* A–PCP (adenylation–reduction state) crystals were obtained in 0.2 M sodium fluoride, 0.1 M Bis–Tris propane pH 6.5 and 20% PEG 3350. CAR*sr* PCP–R crystals were obtained in the condition of 0.1 M sodium malonate dibasic monohydrate, 0.1 M HEPES pH 7.0, 0.5% Jeffamine ED-2003. Sfp treated CAR*sr* PCP–R crystals were obtained in 0.1 M carboxylic

acids 0.1M, 0.1 M imidazole/MES monohydrate pH 6.5, 50% ethylene glycol/PEG8K of the Morpheus screen (Molecular Dimension, UK).

All the data sets were integrated and scaled using the program XDS[38]. Structure determination of adenylation domain was done by molecular replacement using the CAR*ni* A domain as a search model. Structures of A–PCP didomain were solved by molecular replacement[39], using the A domain CAR*sr* structure as a search model. Iterative cycles of manual building in Coot[36] and refinement in Refmac[40] were used to complete the models. Iterative cycles of manual building in Coot and refinement in Refmac were used to complete the models. CAR*sr* R domain structure was solved by molecular replacement using the CAR*mm* R domain structure. Structure of CAR*sr* PCP–R didomain structure was solved by molecular replacement with refined CAR*sr* R structure and carrier protein domain from the CAR*sr* A–PCP structure.

Validation of all the structures were done with Molprobity[41] and PDB-REDO[42] were integrated into the iterative rebuild and refinement procedure. The data collection and refinement statistics for all structures are summarized in the **Supplementary Table 1**.

Small-angle X-ray scattering. SAXS intensity data, $I(q)$ versus q, ($q = 4\pi . \sin 2\theta\lambda$), were collected using HPLC SAXS on beamline B21 at Diamond Light Source (Didcot, UK). 50 µL of CAR*sr* purified sample was loaded onto the Shodex KW-403 size exclusion column mounted on Agilent HPLC and the eluent was flowed through the SAXS beam at 0.15 mL/min; the buffer (25 mM Tris–Cl pH 8.0, 150 mM NaCl) used as the background was collected after one SEC column volume. SAXS data were collected at one-second intervals using a Pilatus 2M detector (Dectris, Switzerland) at a distance of 3.9 m and an X-ray wavelength of 1 Å. All the SAXS data sets were analyzed by ATSAS[43] and Scatter suite[44]. Crysol[45] and the FoXS web server[46,47] were used to assess the fitting of SAXS data with the corresponding models. For modeling of CAR*sr* FL, A–PCP and PCP–R proteins, 10,000 independent models were generated and analyzed using EOM[48] and the results cross-validated by MultiFoxs web server[46,47].

Multi-angle light scattering and analytical ultracentrifugation. Purified CAR*sr* protein samples were injected onto a Superdex-200 10/300 GL column (GE Healthcare) pre-equilibrated with 20 mM HEPES pH 7.4, 150 mM NaCl buffer for the MALS analysis. Light-scattering intensities of proteins were measured at different angles relative to the incident beam and data analysis was performed with ASTRA 6 software (Wyatt Technology Corp., CA, USA). Protein fractions from MALS were then used in sedimentation velocity experiments using XL-A ultracentrifuge (Beckman Instruments) at 50,000 × r.p.m. (18,200*g*) at 20 °C and scanning every 90 s respectively using a wavelength of 280 nm for a total of 200 scans. The sedimentation boundaries were analyzed using the program Sedfit v8.7 (ref. 49) and hydrodynamic radius (R_h) and frictional ratio (f/f_o) were calculated with Sednterp[50].

Stopped-flow experiments. NADPH was obtained from Sigma-Aldrich and *pro-R* and *pro-S* NADP[2]H were synthesized and characterized as described previously[51]. Stopped-flow studies were performed on an Applied Photophysics SX20 stopped-flow spectrometer. Experiments were conducted in 100 mM potassium phosphate buffer (pH 7.5). Before recording stopped-flow measurements CAR*mm* was activated using 10 mM MgCl₂, 2 mM α-methylcinnamic acid and 2 mM ATP at room temperature for 20 min. Reactions were initiated by mixing 0.5 µM NADPH (final concentration) with varied concentrations of activated enzyme, at 30 °C. For KIE measurements the same concentration of *pro-R* and *pro-S* NADP[2]H were mixed with the enzyme. To follow the reductive reaction, NADPH was excited at 340 nm, and emission changes were followed using a 400 nm cut-off filter.

Steady state kinetics. The steady-state turnover of CAR*mm* at different viscosities was determined at 30 °C on assay mixtures containing 10 mM MgCl₂, 0.1 µM CAR*mm*, 1 mM ATP, 200 µM NADPH and 3 mM α-methyl cinnamic acid in 100 mM potassium phosphate buffer (pH 7.5). Steady state NADPH oxidation rates were determined at 340 nm ($\varepsilon 340 = 6.22$ mM⁻¹ cm⁻¹) using Varian Cary 300 Bio UV-Visible Spectrophotometer. Glycerol solutions were prepared by weight and viscosity was calculated as described before[52]. ATP, NADPH and α-methylcinnamic acid concentrations were varied for

respective experiments to measure apparent K_M values for CAR*mm*, CAR*ni*, nmCAR and mnCAR.

CAR biotransformation and HPLC product detection. Reactions for analysis were carried out on a 500 µL scale in a 2 mL Eppendorf tube. Typically, biotransformation reactions contained 5 mM substrate, 0.25 mg nickel-purified enzyme, 10 mM ATP, 100 mM MgCl₂, 10 mM NADPH in 100 mM Tris–HCl pH 7.5. The reaction contained 6.25% v/v DMSO from the addition of the substrate. The reaction was incubated at 25 °C shaking at 170 r.p.m. for 24 h. Reactions were extracted by addition of 100 µL HPLC grade acetonitrile to 100 µL of sample, vortexed, centrifuged and filtered. Samples were then analyzed by HPLC. Samples to be analyzed by HPLC and product conversions calculated were extracted with acetonitrile containing the 1 mg/mL 1-phenylethanol as an external standard.

Reverse-phase HPLC was carried out using an Agilent 1200 Series system equipped with a G1322A degasser, G1311A quaternary pump, G1329A standard autosampler (ALS) and a G1315B diode-array detector (DAD). All HPLC analysis was carried out using a phenomenex HyperClone 5 µm ODS C18 120 Å LC column (250 × 4.6 mm). Samples were analyzed using a gradient method between two solvents, Solvent A the aqueous phase, HPLC grade H₂O (0.1% TFA) and solvent B LC–MS grade acetonitrile (0.1% TFA). The initial mobile phase was 90% A, 10% B, a linear gradient was then employed over 30 min to a ratio of 30% A, 70% B, this was then returned to 90% A, 10% B over a further 10 min. Samples were run at room temperature, and a sample injection volume of 10 µL, detection wavelength of 215 nm, and a flow rate of 1 mL/min were used. Conversions of products as a percentage were calculated using peak area integrations of products in ratio to the external standard 1-phenylethanol.

To calculate product conversion a calibration curve was performed for benzaldehyde and benzyl alcohol using the external standard 1-phenylethanol. Standards containing 1 mg/mL 1-phenylethanol and either 0.125 mg/mL, 0.25 mg/mL, 0.5 mg/mL and 1 mg/mL of benzaldehyde or benzyl alcohol were analyzed by HPLC using the method described. Conversions were calculated using the linear relationship determined between the known concentration of product standards and the ratio of peak areas of these products standards and the external standard.

HPLC grade water was purchased from Romil Ltd (Cambridge, UK), Acetonitrile CHROMASOLV LC–MS was purchased from Honeywell Riedel-de Haën (Bucharest, Romania). Benzoic acid, benzaldehyde, benzyl alcohol, 1-phenylethanol, DMSO and trifluoroacetic acid were purchased from Sigma (Dorset, UK).

Data availability. Coordinates and associated structure factors have been deposited with the PDB under accession codes 5MSC, 5MSD, 5MST, 5MSS, 5MSW, 5MSR, 5MSP, 5MSV, 5MSU and 5MSO.

33. Yin, J., Lin, A.J., Golan, D.E. & Walsh, C.T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **1**, 280–285 (2006).
34. Winter, G., Lobley, C.M. & Prince, S.M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
35. Sheldrick, G.M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
36. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
37. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
38. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
39. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
40. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
41. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

42. Joosten, R.P., Long, F., Murshudov, G.N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).

43. Petoukhov, M.V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).

44. Rambo, R.P. & Tainer, J.A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481 (2013).

45. Svergun, D., Barberato, C. & Koch, M.H.J. CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).

46. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).

47. Schneidman-Duhovny, D., Hammel, M., Tainer, J.A. & Sali, A. FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W1, W424–W429 (2016).

48. Tria, G., Mertens, H.D.T., Kachala, M. & Svergun, D.I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).

49. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).

50. Hayes, D., Laue, T. & Philo, J. Program Sednterp: Sedimentation Interpretation Program. (Alliance Protein Laboratories, Thousand Oaks, California, USA) (1995).

51. Pudney, C.R., Hay, S. & Scrutton, N.S. Practical aspects on the use of kinetic isotope effects as probes of flavoprotein enzyme mechanisms. *Methods Mol. Biol.* **1146**, 161–175 (2014).

52. Hay, S., Pudney, C.R., Sutcliffe, M.J. & Scrutton, N.S. Are environmentally coupled enzymatic hydrogen tunneling reactions influenced by changes in solution viscosity? *Angew. Chem. Int. Ed. Engl.* **47**, 537–540 (2008).

Supplementary information

Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis

Deepankar Gahloth[1]*, Mark S. Dunstan[1]*, Daniela Quaglia[1$], Evaldas Klumbys[1], Michael P. Lockhart-Cairns[2,3], Andrew M. Hill[1], Sasha R. Derrington[1], Nigel S. Scrutton[1], Nicholas J. Turner[1], David Leys[1]

[1]Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, 131 Princess Street, M1 7DN Manchester, UK
[2]Division of Cell Matrix Biology and Regenerative Medicine, School of Biological Sciences, University of Manchester, Oxford Road, M13 9PT Manchester, UK
[3]Diamond Light Source, Harwell Science & Innovation Campus, Didcot, OX11 ODE, Oxfordshire, UK
$present address: Département de chimie, Université de Montréal 2900, Boulevard Édouard-Montpetit, Montréal, Québec, Canada, H3T 1J4.

* equal contributions

Supplementary Results
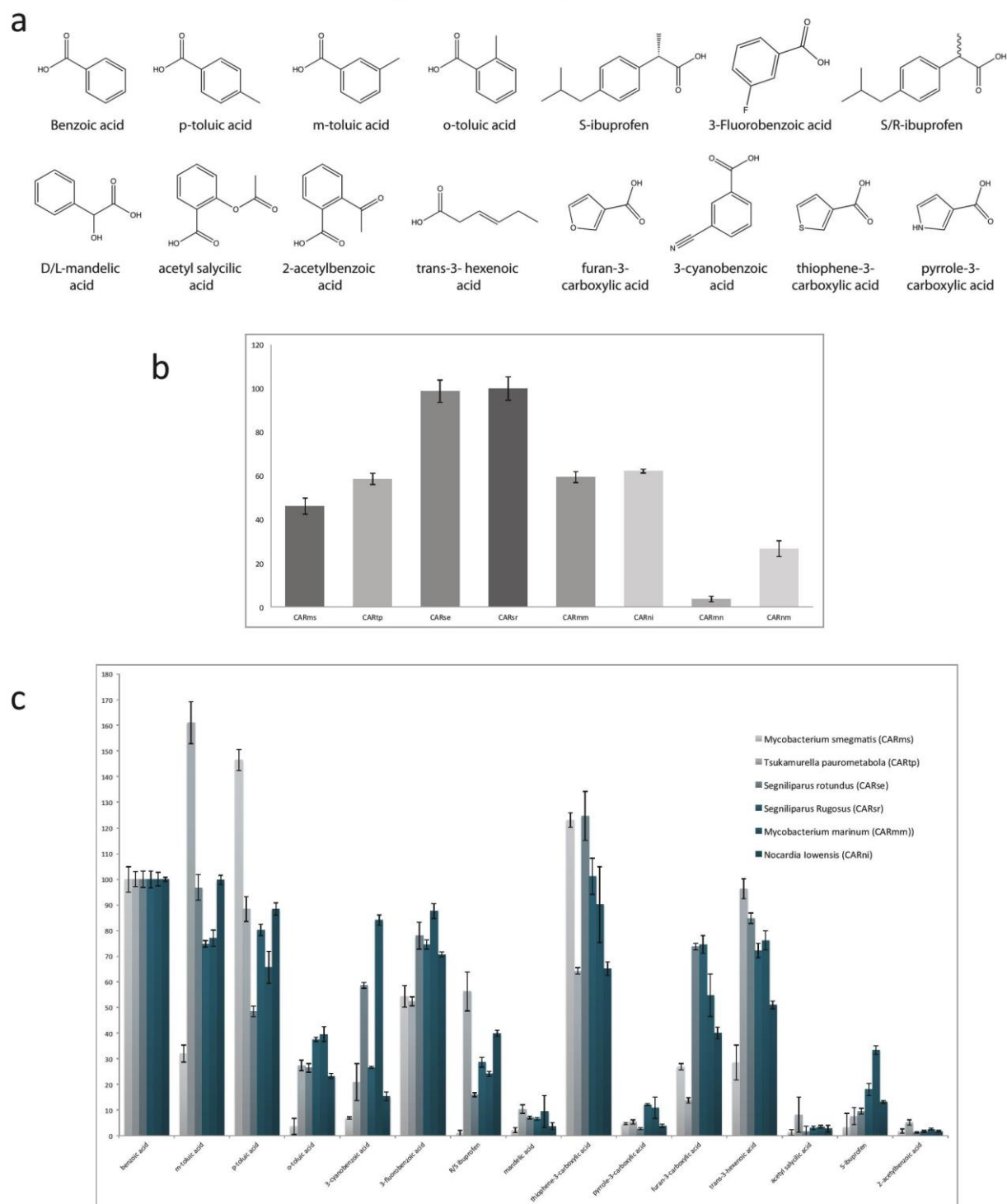
Suppl. Fig. 1. CAR activity screen. Panel **A** depicts the substrates tested in the CAR activity screen. Panel **B** shows the relative activities of CAR enzymes tested with benzoic acid. Panel **C** shows the relative activities of each enzyme against the substrate panel. On the y axis the percentage of activity, with 100% corresponding to the activity of each enzyme against benzoic acid (see panel B).
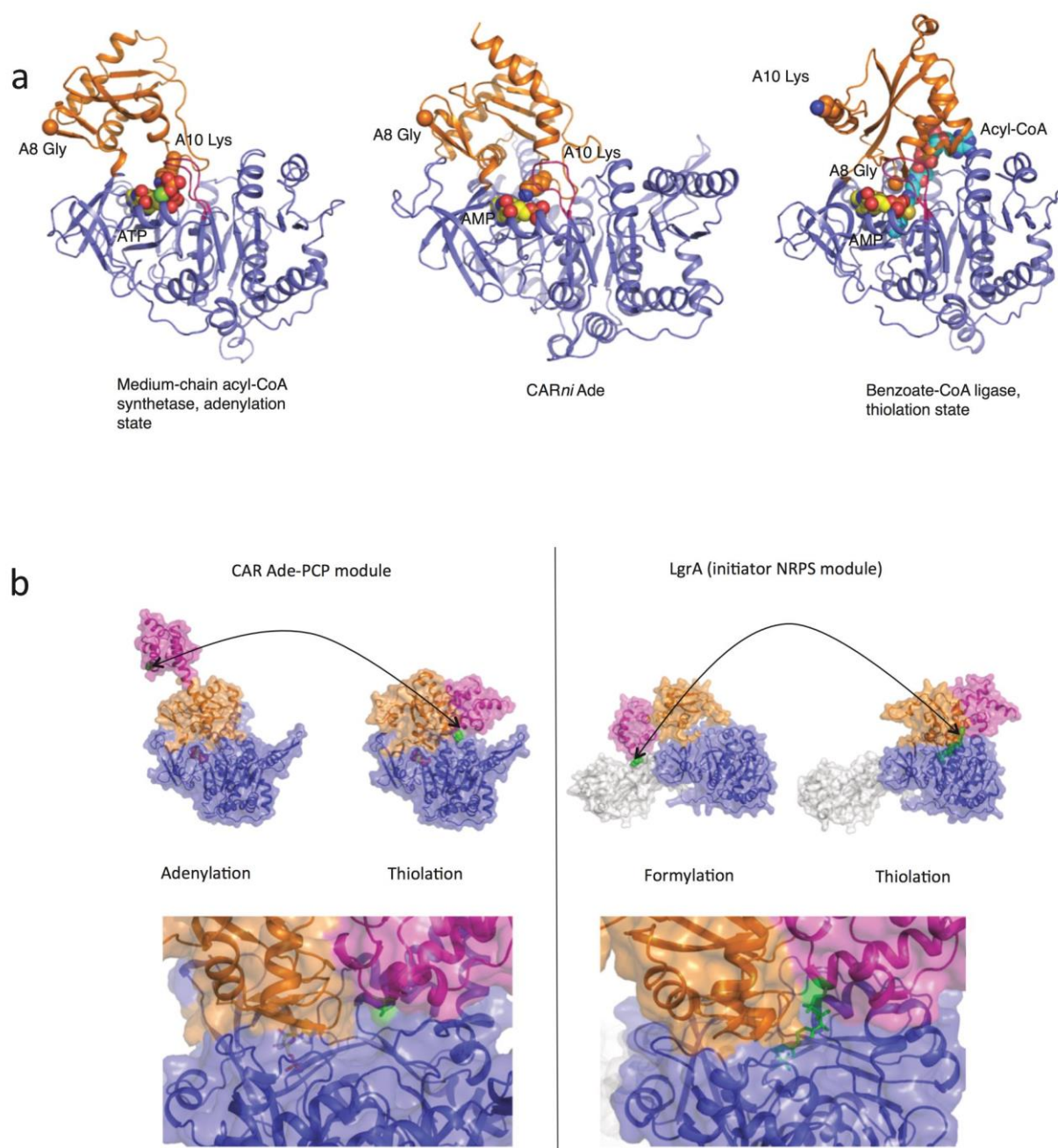
Suppl. Fig. 2. Comparison of the CAR A domain with related structures. Panel **A** shows a side by side comparison of CAR*ni* A (in the middle) with the human medium-chain Acyl-coenzyme A synthetase ACSM2A in the adenylation state ([20] PDB 3EQ6, Z score 32.8, rmsd 3.1 Å over 495 Calphas) to the left and the bacterial benzoate-CoA ligase in the thiolation state ([19] PDB 4ZJZ, Z score 35.2, rmsd 2.6 Å over 486 Calphas, to the right. Colour coding as Fig 1 and 2. The conserved A8 Gly and A10 Lys are highlighted by depiction as spheres. Panel **B** shows a side by side comparison of CAR*sr* A-PCP didomain (left panel) with the NRPS LgrA initiator module [7]. Large-scale reorientations of the PCP domain facilitated by reorientation of the A_{sub} domain are common to both enzymes. The respective thiolation conformations of both systems reveal a similar overall configuration and mode of PCP docking.
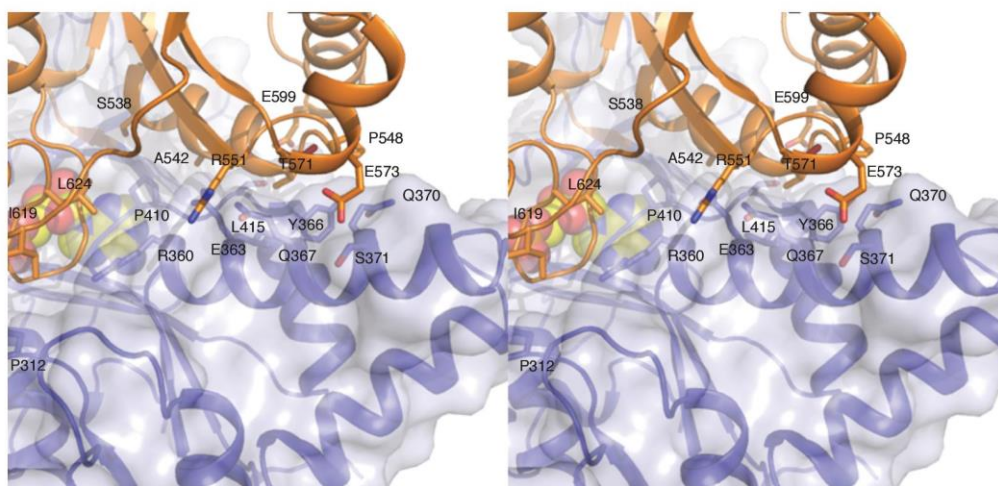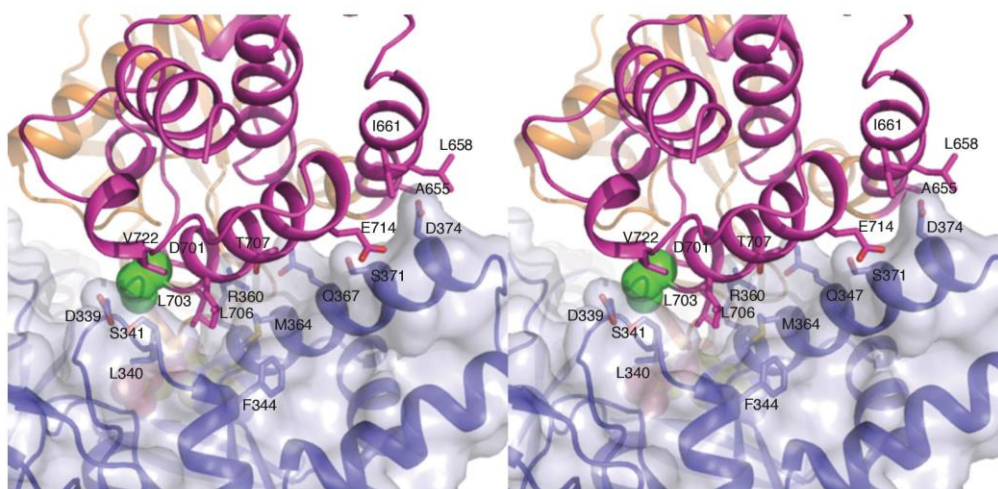
Suppl. Fig. 3. The CAR*sr* Ade$_{core}$-Ade$_{sub}$ and Ade$_{core}$-PCP interface. Panel **A** shows a stereoview of the CAR*sr* Ad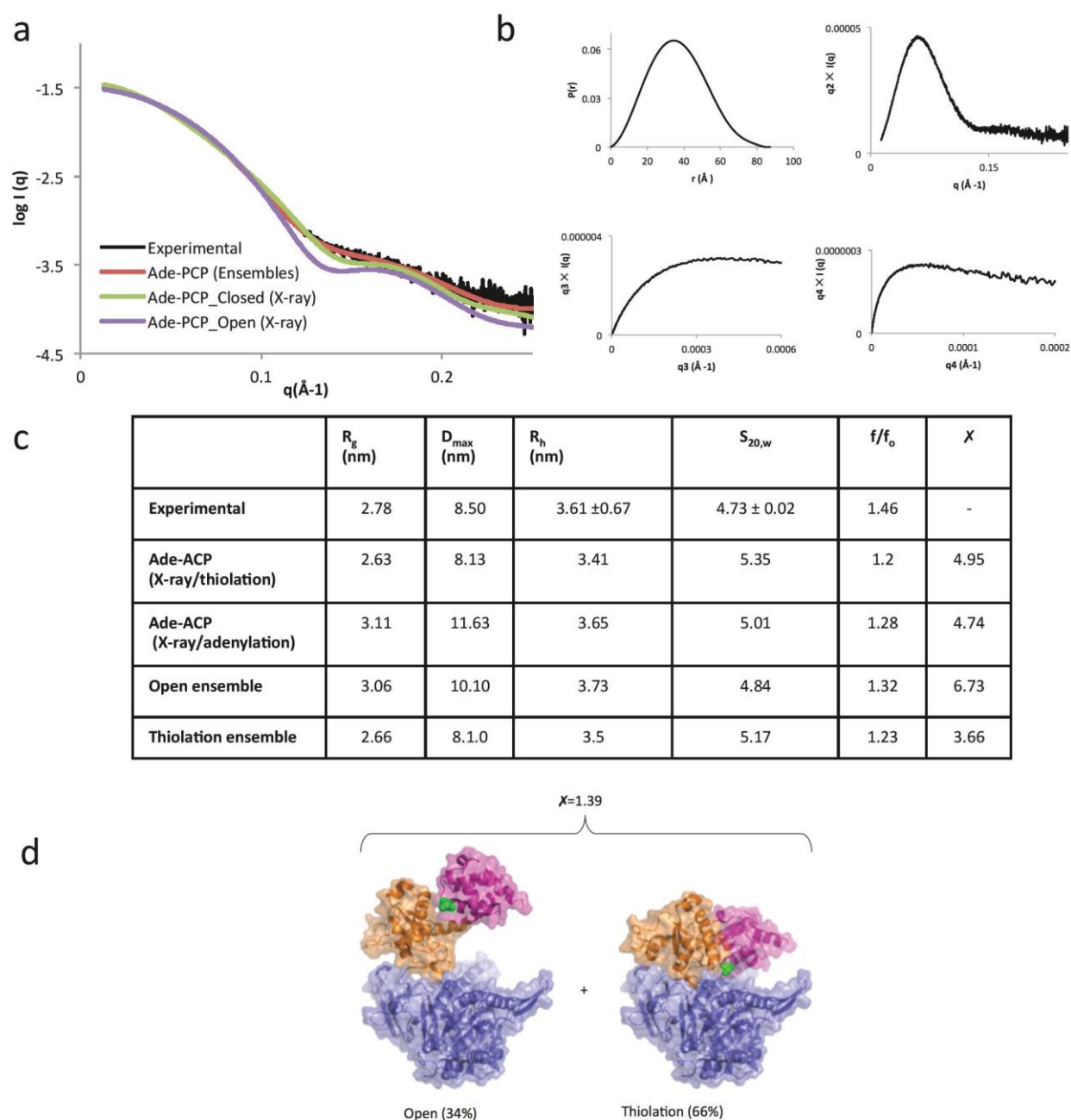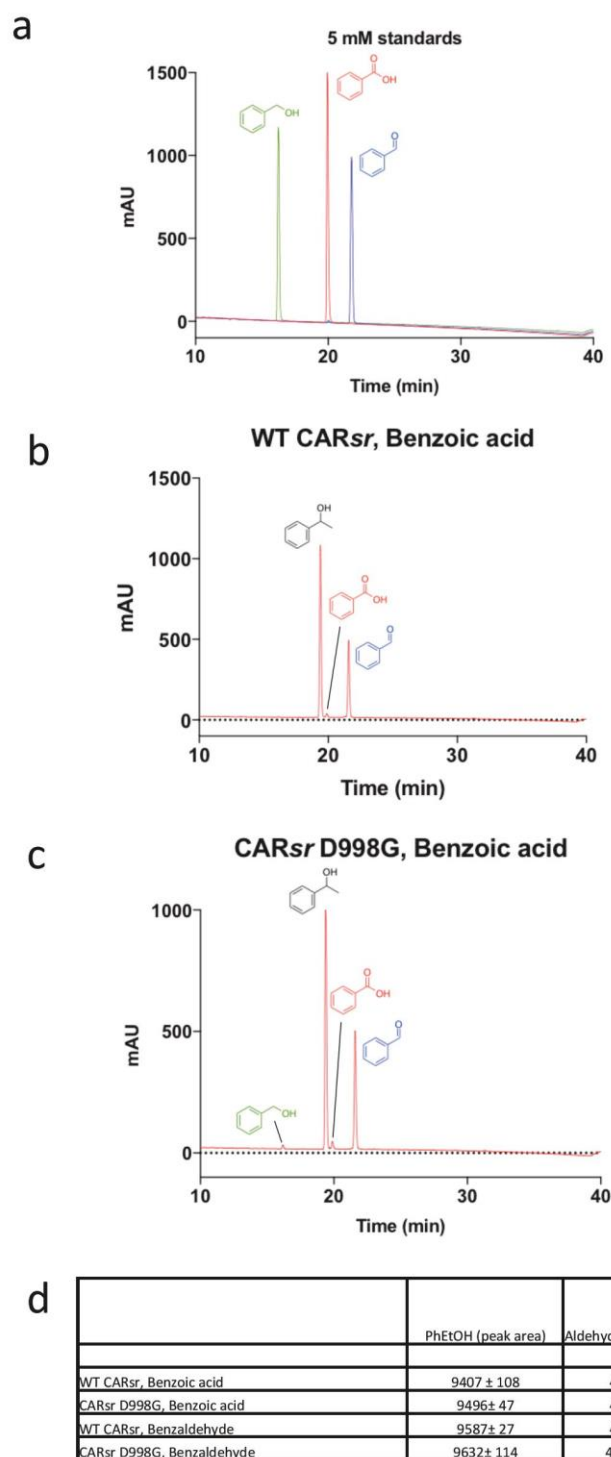e$_{core}$-Ade$_{sub}$ interface in the adenylation state, while panel **B** shows a stereoview of the competing Ade$_{core}$-PCP interface in the thiolation state. Colour coding as in Fig 1.

a

b

c

	R_g (nm)	D_{max} (nm)	R_h (nm)	$S_{20,w}$	f/f_o	χ
Experimental	2.78	8.50	3.61 ±0.67	4.73 ± 0.02	1.46	-
Ade-ACP (X-ray/thiolation)	2.63	8.13	3.41	5.35	1.2	4.95
Ade-ACP (X-ray/adenylation)	3.11	11.63	3.65	5.01	1.28	4.74
Open ensemble	3.06	10.10	3.73	4.84	1.32	6.73
Thiolation ensemble	2.66	8.1.0	3.5	5.17	1.23	3.66

d

χ=1.39

Open (34%)   +   Thiolation (66%)

Suppl. Fig. 4. Solution characterisation of the CAR*sr* A-PCP didomain. A) Scattering plot and fits of the theoretical scattering profiles for the selected ensembles and crystals structure of CAR*sr* A-PCP (open and closed). (Inset) Guinier plot of CAR*sr* A-PCP domain SAXS data. (**B**) Analysis of CAR*sr* A-PCP domain SAXS data (top left) Distance distribution plot (top right) Dimensionless kratky plot (bottom left) SIBYLS plot (top right) Porod-Debye plot. (**C**) Table showing the hydrodynamic and dimensional data for CAR*sr* A-PCP determined from AUC, MALLS, and SAXS, and parameters calculated with SoMo for the SAXS models and crystal structures. Rh, hydrodynamic radius; s20,w, sedimentation coefficient in water at 20 °C; f/f0, frictional coefficient; Rg, radius of gyration; Dmax, maximal linear dimension of the particles. (**D**) EOM ensembles of CAR*sr* A-PCP, which collectively fits with SAXS data with χ of 1.39.
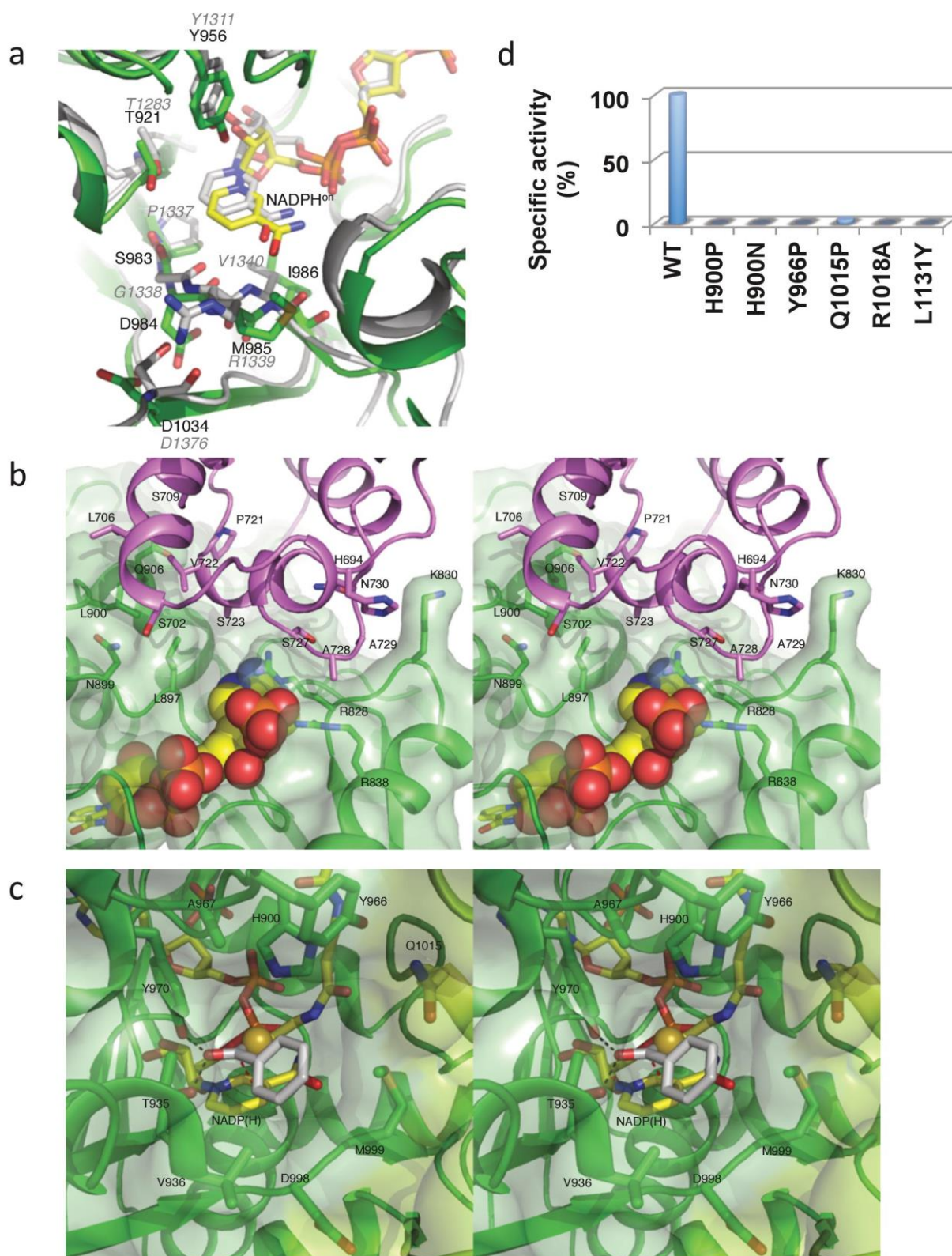
Suppl. Fig. 5. Crystal structure of the CAR*sr* PCP-R didomain. Panel **A** shows and overlay of the CAR*mm* reductase in the active conformation (in green) with the MxaA reductase structure (in grey). Panel **B** shows a stereoview of the reductase-PCP domain interface for the unmodified CAR*sr* PCP-R didomain. Panel **C** shows a stereoview of a model for the benzoyl-phoshopantetheine CAR*sr* complex. Panel **D** shows the effect of mutations in the phosphopantetheine binding site of CAR*sr* on NADPH oxidation rates in presence of benzoic acid.

a

5 mM standards

b

WT CAR*sr*, Benzoic acid

c

CAR*sr* D998G, Benzoic acid

d

	PhEtOH (peak area)	Aldehyde (peak area)	Alcohol (peak area)
WT CARsr, Benzoic acid	9407 ± 108	4473± 89	n/a
CARsr D998G, Benzoic acid	9496± 47	4490± 14	185 ± 3
WT CARsr, Benzaldehyde	9587± 27	4805± 19	n/a
CARsr D998G, Benzaldehyde	9632± 114	4510± 188	507 ± 35

Suppl. Fig. 6. Mutation of CARsr Asp 998 to Gly leads to alcohol production. Panel A shows an overlayed HPLC trace of standards for benzoic acid (in red), benzaldehyde (in blue) and benzylalcohol (green). Panel B and C show HPLC traces of biotransformation products from benzoic acid using CAR*sr* WT and CAR*sr* D998G respectively. Panel D shows observed peak area for various products (including the internal standard phenylethanol) from biotransformations of CAR*sr* Wt or D998G with benzoic acid or benzaldehyde (average taken from n=2). Modest amounts of alcohol are detected only in case of the CAR*sr* D998G mutant.
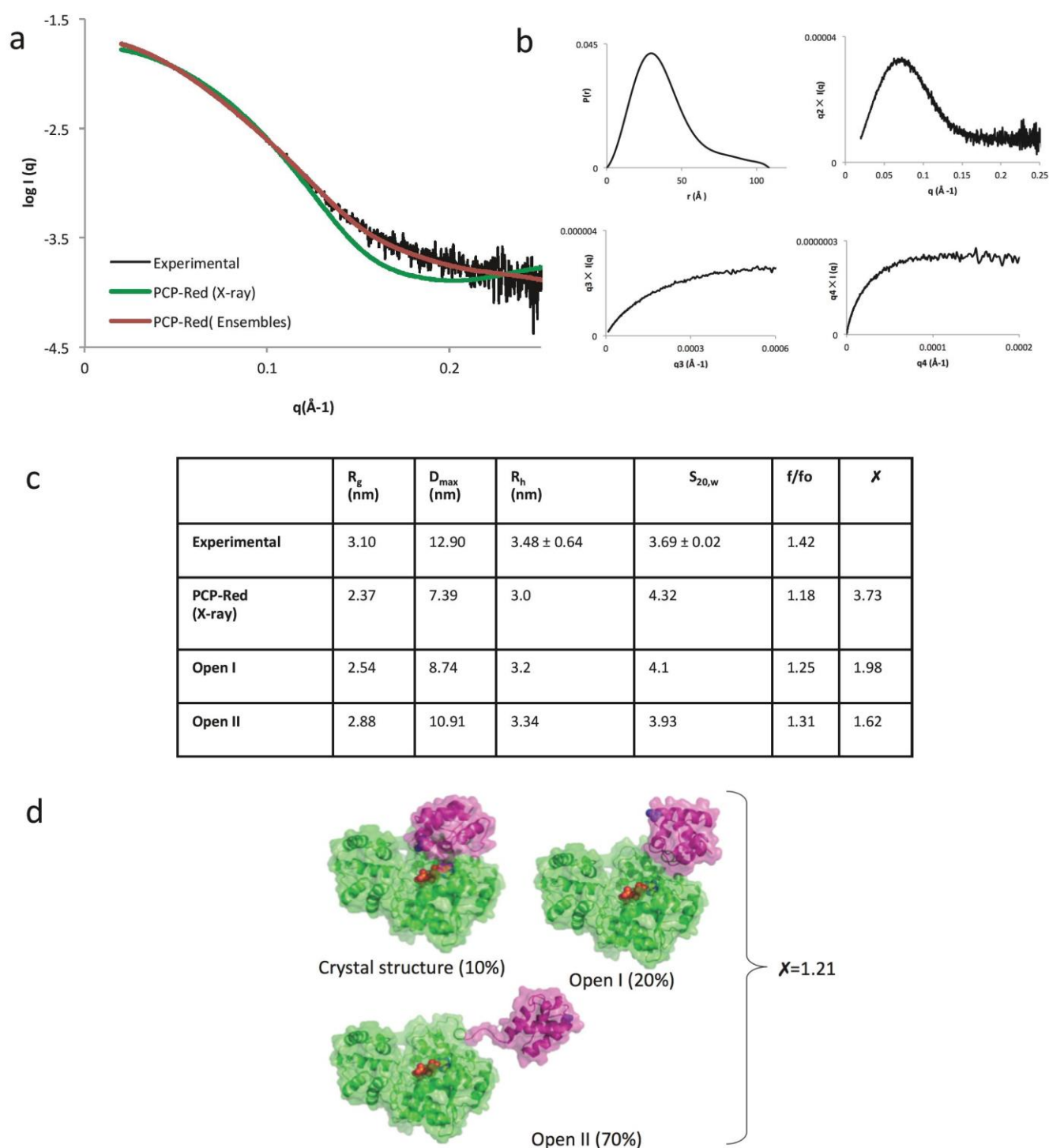
Suppl. Fig. 7. Solution characterisation of the CAR*sr* PCP-R didomain. (**A**) Scattering plot and fits of the theoretical scattering profiles for the selected ensembles and crystals structure of CAR*sr* PCP-R. (Inset) Guinier plot of CAR*sr* PCP-R didomain SAXS data. (**B**) Analysis of CAR*sr* PCP-R didomain SAXS data (top left) Distance distribution plot (top right) Dimensionless kratky plot (bottom left) SIBYLS plot (bottom right) Porod-Debye plot. (**C**) Table showing the hydrodynamic and dimensional data for CAR*sr* PCP-R determined from AUC, MALLS, and SAXS, and parameters calculated with SoMo for the SAXS models and crystal structures. Rh, hydrodynamic radius; s20,w, sedimentation coefficient in water at 20 °C; f/f0, frictional coefficient; Rg, radius of gyration; Dmax, maximal linear dimension of the particles. (**D**) EOM ensembles of CARsr PCP-R, which collectively fits with SAXS data with X of 1.21.

|	R_g (nm)	D_{max} (nm)	R_h (nm)	$S_{20,w}$	f/f_o	x
CAR-FL	4.47	14.60	4.60 ± 0.73	5.82 ± 0.05	1.56	
Open I	3.99	12.90	4.45	6.54	1.34	4.67
Open II	4.57	14.92	4.6	6.32	1.38	2.95
Adenylation/ Reduction model	4.03	13.85	4.39	6.64	1.32	3.00
Thiolation model	4.29	14.62	4.55	6.41	1.37	2.51
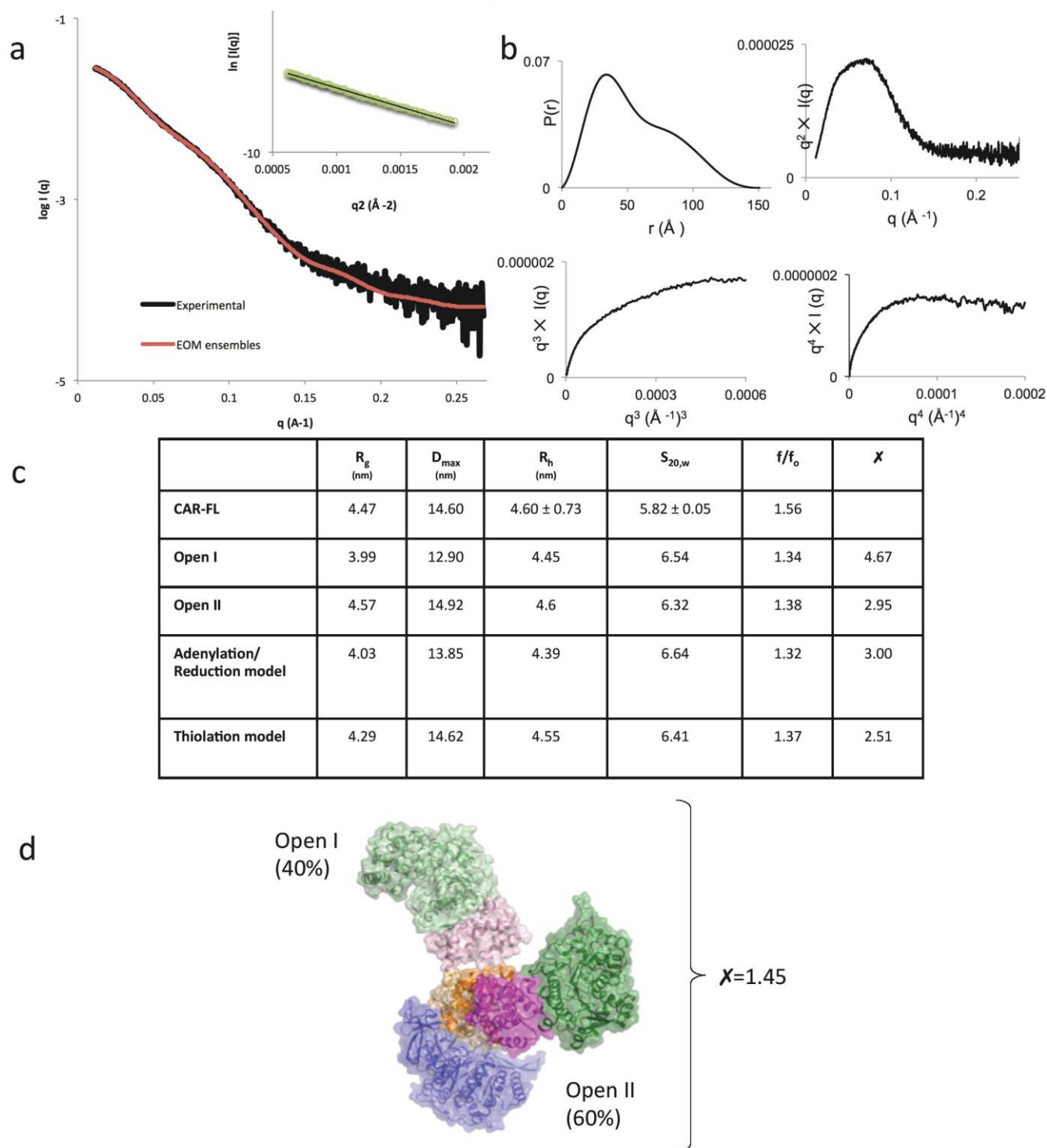
Suppl. Fig. 8. Solution characterisation of the full-length CAR*sr*. (A) Scattering plot and fits of the theoretical scattering profiles for the selected ensembles of CAR*sr* Full-length (FL). (Inset) Guinier plot of CAR*sr* FL SAXS data. **(B)** Analysis of CAR*sr* FL SAXS data (top left) Distance distribution plot (top right) Dimensionless kratky plot (bottom left) SIBYLS plot (bottom right) Porod-Debye plot. **(C)** Table showing the hydrodynamic and dimensional data for CAR*sr* FL determined from AUC, MALLS, and SAXS, and parameters calculated with SoMo for the SAXS and manually built models. Rh, hydrodynamic radius; s20,w, sedimentation coefficient in water at 20 °C; f/f0, frictional coefficient; Rg, radius of gyration; Dmax, maximal linear dimension of the particles. **(D)** EOM ensembles of CARsr FL, which collectively fits with SAXS data with x of 1.45.
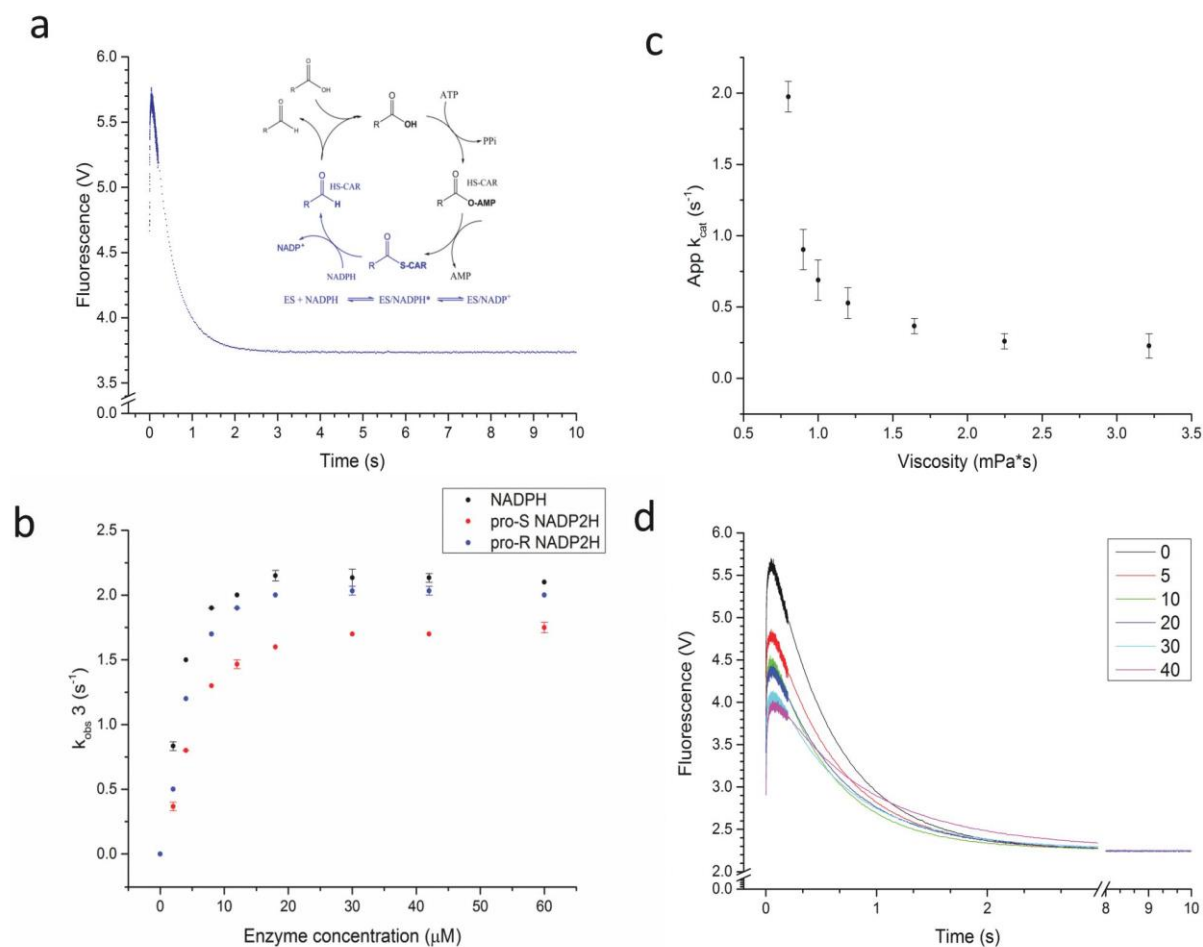
Suppl. Fig. 9. CAR*mm* kinetic characterization. A. An example stopped-flow fluorescence transient showing the reaction between CAR*mm* (60 μM final concentration) and NADPH (0.5 μM final concentration). The insert in A. shows the CAR*mm* catalytic cycle – highlighted in blue are the steps observed in the stopped-flow. B. Shows the observed rate constant associated with the reduction of CAR*mm* by NADPH (k_3) plotted against enzyme concentration (experiments performed in triplicate). C. Shows the effect of increasing the solvent viscosity on the steady-state apparent k_{cat} values (experiments performed in triplicate). D. Example transients showing the influence of glycerol on the binding and reduction of NADPH by CAR*mm*. All data in D. are normalised to the fluorescence signal recorded at 10 seconds. All experiments were performed in 100 mM potassium phosphate buffer (pH 7.5).

a

b
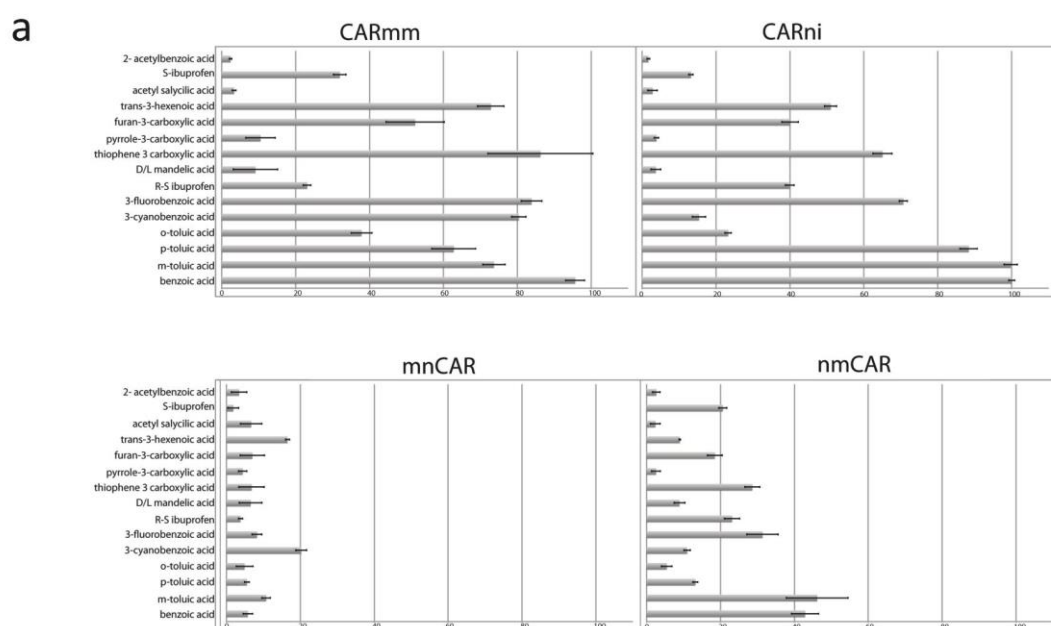
Enzyme	k_{cat} (s^{-1})	K_M α-methyl transcinnamic acid (mM)	K_I (mM)	K_M ATP (µM)	K_M NADPH (µM)
mnCAR	0.43 ± 0.01	0.027 ± 0.004	N/A	36.2 ± 6.1	15.7 ± 2.6
nmCAR	2.96 ± 0.13	0.236 ± 0.034	70 ± 37	39.5 ± 4.1	5.77 ± 0.65
CARmm	2.55 ± 0.10	0.102 ± 0.019	74 ± 38	37.8 ± 6.1	4.70 ± 0.54
CARni	4.09 ± 0.10	0.227 ± 0.019	48 ± 11	46.6 ± 2.4	13.1 ± 1.3

Suppl. Fig. 10. CAR hybrids activity screen. A. Comparison between the activity of *Mycobacterium marinum* (S.A. vs benzoic acid, 0.51 ± 0.014 U/mg), *Nocardia iowensis* (S.A. vs benzoic acid, 0.53 ± 0.004 U/mg), the mnCAR Hybrid (S.A. vs benzoic acid, 0.03 ± 0.007 U/mg) and the nmCAR Hybrid (S.A. vs benzoic acid, 0.23 ± 0.019 U/mg). Data was normalized to that of *Nocardia iowensis* against benzoic acid (100%). **B.** Steady state kinetic analysis of CARmm, CARni and their chimeras using α-methylcinnamic acid as a substrate. All reactions were performed at saturated conditions of substrates.

Suppl. Table 1. Crystallographic data collecton and refinement statistics.

	CARni Ade+AMP (PDB 5MSC)	CARni Ade AMP+Benzoic acid (PDB 5MSD)	CARsr Ade (PDB 5 MST)	CARsr Ade-PCP (Thiolation; PDB 5MSS)	CARsr Ade-PCP (Adenylation; PDB 5MSW)	CARsr PCR-Red (PDB 5MSR)	CARsr PCP-Red (PDB 5MSP)	CARsr PCP-Red (Modified; PDB 5MSV)	CARmm Red (PDB 5MSU)	CARmm Red (PDB 5MSO)
Data collection										
Wavelength (λ)	0.976	0.9686	0.9763	0.9763	0.9763	0.9763	0.9763	0.9763	0.9763	0.9763
Space group	*P1*	*P1*	*P2₁*	*C222₁*	*C2*	*P4₃*	*F222₁*	*P4₃*	*P2₁*	*C222₁*
Unit cells										
a, b, c (Å)	52.08, 54.25, 66.24	52.65, 54.5, 66.36	69.40, 90, 113.78	59.71, 127.70, 212.24	131.42, 65.33, 101.71	94.94, 94.94, 335.50	135.37, 137.68, 202.08	92.11, 92.11, 363.50	100.54, 58.95, 135.56	62.24, 136.16, 115.49
α, β, γ (°)	93.5, 98.8, 101.7	93.3, 97.7, 102.4	90, 92.84, 90	90, 90, 90	90, 110.8, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 105.92, 90	90, 90, 90
Res. (Å)	1.85 (8.27-1.85)	1.71 (1.75-1.71)	1.72 (1.78-1.72)	1.96 (2.03-1.96)	2.33 (2.41-2.33)	2.37 (2.46-2.37)	2.41 (2.5-2.41)	2.34 (2.42-2.34)	1.74 (1.80-1.74)	1.20 (1.24-1.20)
CC$_{1/2}$	1.0 (0.8)	1.0 (0.6)	1.0 (0.4)	1.0 (0.6)	1.0 (0.5)	1.0 (0.5)	1.0 (0.6)	1.0 (0.5)	1.0 (0.5)	1.0 (0.5)
R_{merge}	0.05	0.06	0.08	0.08	0.09	0.13	0.11	0.12	0.08	0.04
I/σI	12.1	9.1	12.8	10.2	10.9	6.4	8.0	11.3	11.7	18.0
Completeness	96.4 (87.2)	82.8 (81.9)	99 (99.5)	99 (99.5)	99 (98)	99 (99.5)	99 (99.5)	95 (92.3)	100 (100)	99 (100)
Refinement										
Res. (Å)	**1.85**	**1.71**	**1.72**	**1.96**	**2.33**	**2.37**	**2.41**	**2.34**	**1.74**	**1.20**
No. reflections	59592	63656	147461	58078	34371	11846	36214	120649	156520	143907
R_{work} / R_{free}	0.21/0.25	0.18/0.22	0.19/0.21	0.20/0.25	0.18/0.23	0.21/0.24	0.18/0.22	0.19/0.20	0.18/0.20	0.13/0.15
No. atoms	5325	5548	10824	5959	5747	16321	4025	16982	11730	3921
B-factors	28.4	29.8	21.37	22.21	49.62	63.61	69.11	30.03	30.93	22.32
R.m.s. deviations										
Bond lengths (Å)	0.009	0.008	0.078	0.012	0.015	0.009	0.017	0.009	0.01	0.012
Bond angles (°)	1.28	1.28	2.1	1.56	1.72	1.4	1.88	1.35	1.48	1.51

Each structure was determined from one crystal. *Values in parentheses are for highest-resolution shell.