The University of Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

Defining the mechanism of galectin-3-mediated TGF-β1 activation and its role in lung fibrosis.

Thesis for the degree of Doctor of Philosophy (PhD)

Submitted to the School of Medicine,

Division of Respiratory Medicine, University of Nottingham.

JESSICA FAE CALVER, BSc (Hons)

UNIVERSITY ID: 20205729

DR ALISON JOHN (PRINCIPAL SUPERVISOR), PROF GISLI JENKINS

DR DAVID SCOTT & DR AMANDA TATLER

Acknowledgements

First and foremost, I would like to thank and dedicate this work to my mum Dionne Calver. I do not think I will ever be able to put into words how truly grateful I am for her support throughout my PhD journey. She is the one person who has been there for me every single step of the way and I would not have been able to complete my PhD without her. She has helped me through the toughest of times and has celebrated by my side at the best of times. I would like to thank her for always pushing me to do my very best, for helping me relocate every 12 months over the past 4 years! (or as she would say 'going on tour again') and for her endless love. I would also like to sincerely thank my sister Eloise Calver for her emotional support and encouragement throughout my years of study. Thank you for always listening, for your advice and for checking in on my cells and proteins! I would like to give my warmest thanks to my family as a whole for their continuous support and understanding throughout my PhD, I feel very lucky to have you all in my life.

I would like to express my sincere appreciation to the Medical Research Council and Galecto Biotech for funding this research, along with the MRC IMPACT DTP for providing me with this opportunity. I would like to thank both of my supervisors Dr Alison John and Prof Gisli Jenkins (Imperial College London) for supporting my ideas, for helping with my professional development and for their guidance throughout, it has been a very unique PhD experience but I would not of had it any other way. Thank you both for motivating me and inspiring me with your enthusiasm for research. I would also like to extend my thanks to supervisors at the University of Nottingham

namely Dr David Scott and Dr Amanda Tatler and to placement supervisors namely Dr Robert Slack (Galecto Biotech) and Dr Stephen Carr (Research Complex at Harwell). I am especially grateful for the guidance, support and encouragement from Dr Robert Slack during my time at Galecto Biotech and throughout my thesis write up.

I would like to give special thanks to the people who have voluntarily taken on a supervisory role at some point during my PhD. I am forever grateful to Rochelle Edwards-Pritchard (University of Nottingham) for taking me under her wing when I first started my PhD, thank you for supervising me in the lab and for teaching me the techniques that have been fundamental throughout, I would not have made it through those first few months without you. An enormous thanks to Dr Gemma Harris (Research Complex at Harwell) for teaching me everything I know about SPR, thank you for always being patient with me and for helping troubleshoot experiments, you are the SPR queen! I would also like to give a huge thank you to Dr Ryan Lithgo for all of his guidance and support during my time at Harwell. Thank you for teaching me everything about proteins, for making me laugh and for all of the good memories. Thank you for continuing to support me even after leaving Harwell, I am grateful to have found a friend like you from my PhD.

Finally, I would like to thank my friend Amanda Greig for all of her support and encouragement during my thesis write up, it has been extremely difficult but it would have certainly been harder without you.

Thank you all for being a part of my PhD journey.

Declaration

I confirm that the work presented in this thesis is my own work, except where acknowledged. No material included in this thesis has been submitted for a previous academic award. The work presented in this thesis was carried out in accordance with University Guidelines and the Quality Manual.

Abstract

Introduction: Traction-dependent activation of the pro-fibrotic mediator transforming growth factor-β1 (TGF-β1), plays a critical role in idiopathic pulmonary fibrosis (IPF) pathogenesis. Galectin-3 potentiates TGF-β1 signaling to promote fibrogenesis and the efficacy of galectin inhibition in IPF patients is currently under investigation. Small molecule galectin-3 inhibitors block lysophosphatidic acid (LPA)-induced and TGF-β1-induced suppressor of mothers against decapentaplegic (SMAD) signaling in human lung fibroblasts (HLFs). However, the exact mechanism by which galectin-3 promotes fibrogenesis has not yet been defined. It was hypothesised that galectin-3 promotes TGF-β1 SMAD signaling and induces fibrogenesis by interacting directly with components of this signaling cascade in fibroblasts.

Methods: Full length galectin-3 protein was expressed and purified in-house (bacterial expression system). Surface plasmon resonance (SPR) was used to determine whether galectin-3 physically interacts with αv integrins, proteins constituting the large latent TGF-β1 complex (latency-associated peptide (LAP) and latent TGF-β binding protein 1 (LTBP1)) or the TGF-β1 receptor subunits (TGFβRI/ TGFβRII). SPR data was analysed in GraphPad Prism and on-off rate maps plotted in EVILFIT. The effects of protein deglycosylation or small molecule galectin inhibitors on binding was investigated and solution competition binding curves plotted in GraphPad Prism. Key binding data was validated *in vitro* by co-immunoprecipitation (Co-IP) and proximity ligation assay (PLA) in HLFs.

Results: By SPR galectin-3 bound to the αv integrins and TGFβRII subunit in a glycosylation-dependent manner, with enzymatic removal of glycans blocking the binding events. This galectin-3 binding was heterogeneous and not a 1:1 binding stoichiometry. Small molecule galectin-3 inhibitors blocked these binding interactions via the carbohydrate-recognition domain (CRD). No galectin-3 binding was detected to LAP, LTBP1 or the TGFβRI subunit. The binding of galectin-3 to the β1 integrin was validated in HLFs by Co-IP and PLA. Galectin-3 successfully co-immunoprecipitated with β1 integrin pulldown and inversely. Galectin-3 and the β1 integrin were also colocalised within 40 nm in IPF HLFs and galectin-3 inhibition prevented this colocalisation.

Conclusions: This work has encompassed cell, molecular and biophysical methods to define the mechanism of galectin-3-mediated TGF-β1 activation in fibroblasts and its potential role in IPF development. Galectin-3 bound to the αvβ1 integrin and TGFβRII subunit and this binding was blocked by galectin-3 inhibitors. Future work is required to validate the galectin-3-TGFβRII SPR binding data *in vitro*. It is hypothesised that upon binding, galectin-3 self-associates to form a lattice between the αvβ1 integrin and TGF-β1 receptor on adjacent cells to facilitate receptor clustering. Subsequently, when TGF-β1 is activated (traction-dependent), it is in close enough proximity to the TGF-β1 receptor to bind and signal, causing fibrogenesis. Understanding the precise role of galectin-3 in IPF pathogenesis may be critical for the continued development of more effective and selective treatments for IPF patients.

Table of Contents

Table of Figures

Table of Tables

Table of Abbreviations

ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
IIPs	Idiopathic interstitial pneumonias
UK	United Kingdom
GORD	Gastroesophageal reflux disease
HRCT	High-resolution computed tomography
UIP	Usual interstitial pneumonia
PaO_2	Partial pressure of oxygen
DLCO	Diffusing capacity for carbon monoxide
FVC	Forced vital capacity
NICE	National Institute for Health and Care Excellence
TGF-β1	Transforming growth factor-β1
BALF	Bronchoalveolar lavage fluid
HLF	Human lung fibroblast
TNF-α	Tumour necrosis factor-α
PDGF	Platelet-derived growth factor
IL-6	Interleukin 6
PDGFR	Platelet-derived growth factor receptor
FGFR	Fibroblast growth factor receptor
VEGFR	Vascular endothelial growth factor receptor
ATP	Adenosine triphosphate
MMPs	Matrix metalloproteinases
TIMPs	Tissue inhibitors of metalloproteinases
ECM	Extracellular matrix
AT1	Type 1 pulmonary alveolar epithelial cells
AT2	Type 2 pulmonary alveolar epithelial cells
AECs	Alveolar epithelial cells
FGF	Fibroblast growth factor
PNECs	Pulmonary neuroendocrine cells
Fas	Fas antigen
FasL	Fas ligand
HBECs	Human bronchial epithelial cells
CTGF	Connective tissue growth factor
ET-1	Endothelin 1
EMT	Epithelial-mesenchymal transition
α-SMA	α-smooth muscle actin
PTM	Post-translational modification
LOXL2	Lysyl oxidase-like 2
TG2	Transglutaminase 2
ER	Endoplasmic reticulum
SNPs	Single-nucleotide polymorphisms

GWAS	Genome-Wide Association Study
PCSK6	Proprotein convertase subtilisin/kexin type 6
PKN2	Serine/threonine-protein kinase N2
GPR157	G-protein coupled receptor 157
FAM13A	Family with sequence similarity 13 member A
FKBP5	FKBP prolyl isomerase 5
DEPTOR	DEP domain-containing mTOR-interacting protein
AKAP13	A-kinase anchoring protein 13
NPRL3	Nitrogen permease regulator-like 3
KIF15	Kinesin family member 15
SPDL1	Spindle apparatus coiled-coil protein 1
MAD1L1	Mitotic arrest deficient 1 like 1
KNL1	Kinetochore scaffold 1
STMN3	Stathmin-3
TERC	Telomerase RNA component
TERT	Telomerase reverse transcriptase
RTEL1	Regulator of telomere elongation helicase 1
DSP	Desmoplakin
DPP9	Dipeptidyl peptidase 9
MUC5B	Mucin-5B
FUT6	Fucosyltransferase 6
ATP11A	ATPase phospholipid transporting 11A
IVD	Isovaleryl-CoA dehydrogenase
ZKSCAN1	Zinc finger with KRAB and SCAN domains 1
MAPT	Microtubule associated protein tau
SFTPA1	Surfactant protein A1
SFTPA2	Surfactant protein A2
SFTPC	Surfactant protein C
ABCA3	ATP-binding cassette subfamily A member 3
PARN	Poly(A)-specific ribonuclease
DKC1	Dyskerin pseudouridine synthase 1
TINF2	TERF1 interacting nuclear factor 2
NAF1	Nuclear assembly factor 1 ribonucleoprotein
e-cigarettes	Electronic cigarettes
rRNA	Ribonucleic acid
AE-IPF	Acute exacerbation of IPF
EBV/ HHV-4	Epstein-Barr virus
CMV/ HHV-5	Cytomegalovirus
HHV-7	Human herpesvirus 7
HHV-8	Human herpesvirus 8
COVID-19	Coronavirus disease 2019
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SARS	Severe acute respiratory syndrome

MERS-CoV	Middle East respiratory syndrome coronavirus
CT	Computed tomography
aa	Amino acids
LAP	Latency associated peptide
LTBP1	Latent TGF-β binding protein 1
TGFβR	TGF-β1 receptor
SMAD	Suppressor of mothers against decapentaplegic
RhoA	Ras homolog family member A
ROCK	Rho-associated protein kinase
PI3K	Phosphoinositide 3-kinase
Akt	Protein kinase B
MAPK	Mitogen-activated protein kinase
ERK	Extracellular signal-regulated kinase
Wnt	Wingless/Int-1
Notch	Neurogenic locus notch homolog
SMURF	SMAD specific E3 ubiquitin protein ligase
MIDAS	Metal ion-dependent adhesion site
Ca^{2+}	Calcium
Mg^{2+}	Magnesium
Mn^{2+}	Manganese
RGD	Arginine-glycine-aspartic acid
scRNAseq	Single-cell RNA sequencing
LPA	Lysophosphatidic acid
PAR1	Protease-activated receptor 1
GTP	Guanosine triphosphate
WT	Wild-type
Fxa	Factor Xa
CCl_4	Carbon tetrachloride
mRNA	Messenger RNA
pSMAD3	Phosphorylated SMAD3
PCLS	Precision cut lung slices
COL1A1	Collagen type I alpha 1 chain
CRD	Carbohydrate-recognition domain
LacNAc	N-acetyllactosamine
RBCs	Red blood cells
GalNAc	N-acetylgalactosamine
Gal	D-galactose
MW	Molecular weight
kDa	Kilodalton
kb	Kilobase
PGAY	Proline-glycine-alanine-tyrosine rich
UUO	Unilateral ureteral obstruction
SAMs	Scar-associated macrophages

Ad-TGF-β1	Adenoviral TGF-β1
E-cadherin	Epithelial cadherin
siRNA	Small interfering RNA
TCF/LEF1	T-cell factor/lymphoid enhancer-binding factor 1
BAL	Bronchoalveolar lavage
PAI-1	Plasminogen activator inhibitor-1
CCL18	Chemokine (C-C motif) ligand 18
CHI3L1/ YKL-40	Chitinase-3-like protein 1
FDA	Food and Drug Administration
Å	Angstrom
PSCs	Pancreatic stellar cells
BDL	Bile duct ligation
TAA	Thioacetamide
HSCs	Hepatic stellate cells
2-DE	2D-electrophoresis
KCO	Carbon monoxide transfer coefficient
shRNA	Short hairpin RNA
HIF-1	HIF-1 transcription factor
FAK1	Focal adhesion kinase 1
PLA	Proximity ligation assay
iHBECs	Immortalised human bronchial epithelial cells
RCaH	Research Complex at Harwell
ICL	Imperial College London
ITC	Isothermal titration calorimetry
AUC	Analytical ultracentrifugation
SPR	Surface plasmon resonance
E. coli	*Escherichia coli*
CHO	Chinese hamster ovary
HEK293	Human embryonic kidney 293
LB	Luria-Bertani
TB	Terrific Broth
pH	Potential of Hydrogen
DNA	Deoxyribonucleic acid
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
dsDNA	Double-stranded DNA
ssDNA	Single-stranded DNA
GC	Guanine-cytosine
HF	High-fidelity
DMSO	Dimethyl sulfoxide
OPPF	Oxford Protein Production Facility
Tm	Melting temperature
dNTPs	Deoxyribonucleotide triphosphates

UV	Ultraviolet
rpm	Revolutions per minute
TAE	Tris-acetate-EDTA
SDS	Sodium dodecyl sulfate
NaOH	Sodium hydroxide
EDTA	Ethylenediaminetetraacetic acid
Dnases	Deoxyribonuclease
Rnase	Ribonuclease
A_{260}	260 nm
A_{280}	280 nm
IPTG	Isopropyl β-d-1-thiogalactopyranoside
Ni^{2+}	Nickel
HRV	Human rhinovirus
SEC	Size exclusion chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
DNase1	Deoxyribonuclease I
CV	Column volume
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
TCEP	Tris(2-carboxyethyl)phosphine
IMAC	Immobilised metal affinity chromatography
AUC	Analytical ultracentrifugation
SEC-MALS	Size exclusion chromatography – multi-angle light scattering
PEI	Polyethylenimine
GFP	Green fluorescent protein
Co^{2+}	Cobalt
CMA	Carboxymethyl aspartate
CMC	Critical micelle concentration
L-Arg	L-Arginine
L-Glu	L-Glutamine
MPL	Membrane Protein Laboratory
kHz	Kilohertz
FSEC	Fluorescence-detection size-exclusion chromatography
LDS	Lithium dodecyl sulfate
FL	Full length
TRUNC	Truncated
HeLa	Henrietta Lacks
EGFR	Epidermal growth factor receptor
Erk-P	Phosphorylated Erk
EGF	Epidermal growth factor
JAG1	Jagged canonical Notch ligand 1
DLL4	Delta like canonical Notch ligand 4
G3BP	Galectin-3 binding protein
PD-1	Programmed cell death protein 1

PD-L1	Programmed cell death ligand 1
pI	Isoelectric point
RU	Resonance units
K_d	Affinity
K_{on}	Association
K_{off}	Dissociation
$MnCl_2$	Manganese chloride
EDC	1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide
NHS	N-hydroxysuccinimide
Rmax	Analyte binding capacity
PNGase F	Peptide:N-glycosidase F
IC50	Inhibitory concentration 50
Mgat5	Alpha-1,6-Mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase
BHK	Baby hamster kidney cells
MAECs	Mouse aortic endothelial cells
HCLEs	Human corneal epithelial cells
ICC	Immunocytochemistry
RPECs	Retinal pigment epithelial cells
Co-IP	Co-immunoprecipitation
DMEM	Dulbecco's modified eagle’s medium
FBS	Foetal bovine serum
PBS	Phosphate-buffered saline
H	Heavy
L	Light
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
DSS	Disuccinimidyl suberate
Cu^{2+}	Cupric ions
Cu^{1+}	Cuprous ions
PVDF	Polyvinylidene difluoride
PBST	PBS with Tween 20
TBST	Tris-buffered saline with Tween 20
HRP	Horseradish peroxidase
PFA	Paraformaldehyde
DAPI	4',6-diamidino-2-phenylindole
FT	Flow through
MDCKs	Madin-Darby canine kidney cells
WHO	World Health Organisation
ACE2	Angiotensin converting enzyme 2
RBD	Receptor binding domain
ISG	Interferon-stimulated gene
IFN-I	Type 1 interferon

BSG	Basigin
NRP1	Neuropilin-1
GRP78	Glucose-regulated protein 78
TMPRSS2	Transmembrane serine protease 2
CTSL	Cathepsin L
IL-8	Interleukin 8
TMB	3,3',5,5'-Tetramethylbenzidine
H_2SO_4	Sulfuric acid
SD	Standard deviation
CDGs	Congenital disorders of glycosylation

Publications

Calver JF, Fabbri L, May J, *et al.* COVID-19 Lung Disease: Lessons Learned. COVID-19 in patients with chronic lung disease. Clinics in Chest Medicine.
10.1016/j.ccm.2022.11.013

Defining the mechanism of galectin-3-mediated TGF-β1 activation and its role in lung fibrosis. (Manuscript in preparation)

Abstracts

Calver J, Joseph C, John A.E, *et al.* 'The novel coronavirus SARS-CoV-2 binds RGD integrins and upregulates αvβ3 integrins in Covid-19 infected lungs.'
Accepted for the British Thoracic Society (BTS), February 2021 (Oral presentation).
http://dx.doi.org/10.1136/thorax-2020-BTSabstracts.37

Calver JF, Harris G, Lithgo RM, *et al.* 'Investigating the Pro-Fibrotic Effects of Galectins in IPF - A Potential Role for Glycan-Mediated Interactions with Integrins.'
Accepted for the British Association for Lung Research (BALR), June 2021 (Poster presentation).

Calver JF, Harris G, Lithgo RM, *et al.* 'Investigating the Pro-Fibrotic Effects of Galectins in IPF - A Potential Role for Glycan-Mediated Interactions with Integrins.'
Accepted for the British Thoracic Society (BTS), November 2021 (Oral presentation).
http://dx.doi.org/10.1136/thorax-2021-BTSabstracts.103
Won the BTS Abstract Award

Calver JF, Harris G, Lithgo RM, *et al.* 'Investigating the Pro-Fibrotic Effects of Galectins in Idiopathic Pulmonary Fibrosis'
Presented at the UoN Sue Watson Presentation Event, April 2022 (Oral presentation).
Awarded the 1st Place Sue Watson Prize for Best PhD Student Presentation

Calver JF, Harris G, Lithgo RM, *et al.* 'Investigating the Pro-Fibrotic Effects of Galectins in Idiopathic Pulmonary Fibrosis - A Potential Role for Glycan-Mediated Interactions with Integrins.'
Accepted for the American Thoracic Society (ATS) International Conference, May 2022 (Poster presentation).
https://doi.org/10.1164/ajrccm-conference.2022.205.1_MeetingAbstracts.A5330
Won the ATS Abstract Scholarship
Won the ATS International Conference Abstract Scholarship supported by Asthma and Lung UK

Calver JF, Harris G, Lithgo RM, *et al.* 'Galectin-3 binding to the αvβ1 integrin and TGFβRII extracellular domain is glycosylation-dependent'.
Accepted for the International Colloquium on Lung and Airway Fibrosis (ICLAF) International Conference, October 2022 (Poster presentation).
Co-chaired a session with Dr Kerri Johannson
Won the Action for Pulmonary Fibrosis Travel and Attendance Award

Calver JF, Harris G, Lithgo RM, *et al.* 'Galectin-3 Interacts Directly With Components of the TGF-β1 Signaling Pathway in Human Lung Fibroblasts'.
Accepted for the Keystone Symposia International Conference, March 2023 (Oral presentation and poster presentation).
Won the Keystone Symposia Future of Science Fund Scholarship

Coronavirus Disease 2019 Impact

During the coronavirus disease 2019 (COVID-19) pandemic I worked full time as part of the University of Nottingham COVID-19 Research Group whose aim was to investigate the pathogenesis, treatment and complications of lung disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus (lead by Prof Gisli Jenkins). Subsequently, this thesis also contains my studies as part of the group which aimed to investigate the potential of integrins and/or galectins acting as receptors or co-receptors for SARS-CoV-2 infection.

1. Introduction

1.1. Introduction to Idiopathic Pulmonary Fibrosis

1.1.1. Classification and Epidemiology

Interstitial lung diseases (ILDs) are a group of heterogenous disorders characterised by scarring of the lung tissue and inflammation (1). They have been classified into distinct subtypes using an integrated clinical, radiological and pathological approach. Idiopathic pulmonary fibrosis (IPF) falls under the 'major idiopathic interstitial pneumonias' (IIPs) subgroup and is defined as a chronic and progressive fibrosing interstitial pneumonia of unknown cause (2) (Table 1.1).

Table 1.1: Types of interstitial lung diseases

Interstitial Lung Disease Classification	Known Cause or Association Connective tissue & autoimmune Occupational & environmental causes Genetic Drug-induced
	Idiopathic Interstitial Pneumonias (IIPs) A. Major: **Idiopathic pulmonary fibrosis** Idiopathic non-specific interstitial pneumonia Respiratory bronchiolitis–interstitial lung disease Desquamative interstitial pneumonia Cryptogenic organising pneumonia Acute interstitial pneumonia B. Rare: Idiopathic lymphoid interstitial pneumonia Idiopathic pleuroparenchymal fibroelastosis C. Unclassified Idiopathic Interstitial Pneumonias
	Granulomatous Sarcoidosis Hypersensitivity pneumonitis Infections
	Other Lymphangioleiomyomatosis Histiocytosis X

IPF prevalence and incidence rates were estimated from an analysis of United Kingdom (UK) respiratory disease epidemiology covering 2004-2012 as part of the 'Respiratory Health of the Nation' project. It is the most common ILD subtype with >32,000 patients living with the disease in the UK and >6,000 newly diagnosed incidences annually (3). In the UK, its prevalence is approximately 50/100,000 with reports highest in Northern Ireland, north-west England, Scotland and Wales (3). There is worldwide variation in IPF mortality rates however they are steadily increasing globally. UK mortality rates are annually increasing by approximately 5% and it is causing ~7% of all UK respiratory deaths (4, 5). However, its clinical impact and disease burden is likely to be an under-estimation with only IPF as a primary cause of death being recorded and evaluated. Comorbidities include emphysema, pulmonary hypertension, lung cancer, sleep apnoea, cardiovascular conditions, gastroesophageal reflux disease (GORD), anxiety and depression.

1.1.2. Clinical Presentation and Diagnosis

Symptom onset is gradual with patient characteristics at diagnosis typically including dyspnea (shortness of breath), a nonproductive cough (dry), inspiratory crackles and finger clubbing (1). Additionally, characteristic abnormalities are identifiable on patient chest radiographs and high-resolution computed tomography (HRCT) scans with patients presenting with a usual interstitial pneumonia (UIP) pattern. This is defined by the presence of ground-glass opacities, honeycombing, reticulation and bronchiectasis, architectural distortion and inferior lobe volume changes may be apparent. This pattern is identified histologically by the presence of fibroblastic foci (composed of fibroblasts and myofibroblasts), fibrotic zones composed of dense collagen, cystic fibrotic airspaces and sometimes mild inflammation (6) (Figure 1.1).

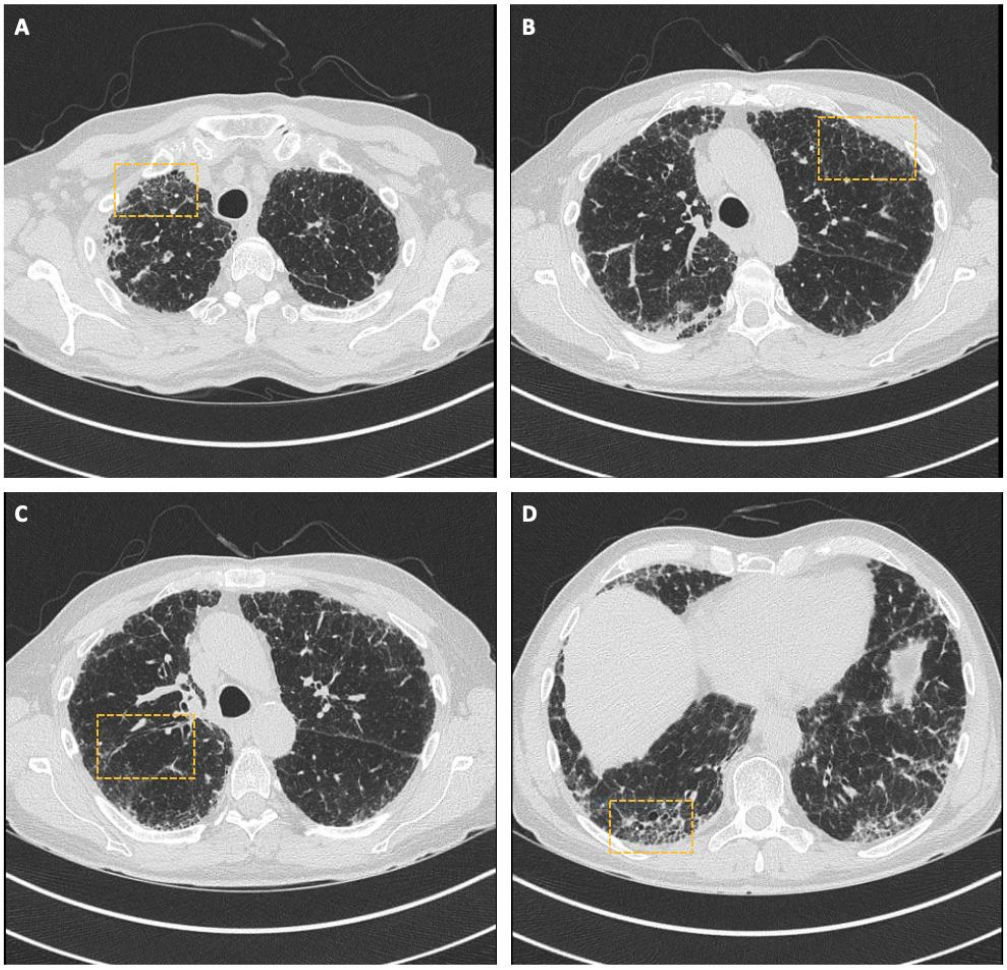


Figure 1.1: Usual interstitial pneumonia pattern
IPF patient HRCT images in the axial plane showing architectural distortion. ***(A)*** *Ground-glass opacities are sparse and discrete.* ***(B)*** *Reticulation evident by several linear opacities resembling a net or mesh.* ***(C)*** *Traction bronchiectasis caused by irreversible dilatation of the airway (bronchi and bronchioles) due to fibrosis of the surrounding lung parenchyma.* ***(D)*** *Honeycomb changes characterised by stacked cystic air spaces that reflect end-stage parenchymal destruction. HRCT images obtained through personal communication with Dr Laura Fabbri (Royal Brompton Hospital).*

For a definitive diagnosis, the UIP pattern must be present on HRCT with all other known ILD causes excluded (patients may or may not be subjected to surgical lung biopsy) (1). Abnormal pulmonary function indicative of restrictive/impaired gas exchange (reduced total lung capacity or vital capacity, abnormal partial pressure of oxygen (PaO_2) or diffusing capacity for carbon monoxide (DLCO)) is also essential for diagnosis (1).

IPF is more common in males and primarily affects older adults, an age > 50 years was previously included as a minor diagnosis criterion until later revised (7). The median survival of patients is 2.5-3.5 years from time of diagnosis, during this time patients tend to gradually decline yet there are periods of rapid decline in a minority of patients (1, 6). Currently, there is no proven pharmacological cure for IPF, subsequently patients experience an increase in respiratory symptoms with worsening pulmonary function tests and progressive fibrosis on HRCT (6). However, two orally administered antifibrotic medications, pirfenidone and nintedanib, are currently licensed for IPF management which slow the rate of lung scarring and progression, although an improvement in lung physiology or radiological abnormalities is rare (1, 8-10). Initially, only IPF patients with a forced vital capacity (FVC) between 50-80% were eligible for antifibrotic treatment in accordance with National Institute for Health and Care Excellence (NICE) guidelines (11, 12). However, it has recently been recommended that nintedanib is extended to IPF patients with a FVC of above 80% predicted (13). The FVC criteria for pirfenidone treatment is under review. Non-pharmacological interventions strongly recommended for management of IPF are limited to long-term oxygen therapy and lung transplantation (6).

1.1.3. Pharmacological Therapy

The mechanism of action of pirfenidone is not yet fully understood however it is suggested to have antifibrotic, anti-inflammatory and antioxidant effects. *In vivo* studies show pirfenidone to improve bleomycin-induced lung fibrosis quantified by a decrease in hydroxyproline levels (major component of collagen), lung collagen content and pro-collagen gene expression (14, 15).

Its anti-inflammatory effects have been explored in the same model in which pirfenidone suppressed the bleomycin-induced increase in transforming growth factor-β1 (TGF-β1) protein and the infiltration of innate (neutrophils and macrophages) and adaptive (lymphocytes) immune cells into the bronchoalveolar lavage fluid (BALF) (16). Additionally, pirfenidone suppresses the bleomycin-induced increase in TGF-β1 gene expression at the transcription level (17). Pirfenidone has also been shown to reduce human lung fibroblast (HLF) proliferation and myofibroblast differentiation (18). Its anti-inflammatory properties include inhibition of tumour necrosis factor-α (TNF-α), platelet-derived growth factor (PDGF) and interleukin 6 (IL-6) (19-22). Levels of lipid peroxidation and anti-/pro-oxidant enzymes have been used to assess its antioxidant effects (14, 15).

Nintedanib was initially identified as a strong inhibitor of the receptor tyrosine kinases associated with platelet-derived growth factor receptor (PDGFR), fibroblast growth factor receptor (FGFR) and vascular endothelial growth factor receptor (VEGFR) signaling in a tumor angiogenesis study (23). It blocks receptor phosphorylation by occupying the tyrosine kinase adenosine triphosphate (ATP)-binding pocket. This receptor tyrosine kinase inhibition contributes to the antifibrotic effects of nintedanib in IPF by inhibiting fibroblast proliferation, migration and preventing fibroblast to myofibroblast differentiation (24). Additionally, nintedanib reduces collagen secretion, modifies the secretion of matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs) which consequently affects extracellular matrix (ECM) composition (25, 26).

1.2. Current Understanding of Idiopathic Pulmonary Fibrosis Pathogenesis

1.2.1. Normal Wound Healing Response

Within a healthy, normal lung, the composition of the respiratory epithelium differs across the lower respiratory tract inclusive of the trachea, primary bronchi, bronchioles and alveoli. The outer alveolus epithelial wall is lined by type 1 pulmonary alveolar epithelial cells (AT1) and surfactant synthesising type 2 pulmonary alveolar epithelial cells (AT2), this acts as a physical barrier against infectious pathogens and harmful substances (27). Alveolar epithelial cells (AECs) are polarised, with their apical surface facing the lumen and basal surface in contact with the ECM. When alveolar epithelium injury occurs, physiological wound healing takes place in three distinct stages: Coagulation and inflammation, tissue formation and tissue remodeling. Immediately following injury, a local inflammatory response is initiated causing the recruitment of neutrophils and monocytes to the site of injury and innate alveolar macrophages activate the coagulation cascade (28). This, in association with fibrin, forms a platelet plug at the site of injury (29, 30). The activated platelets and wounded tissue release pro-inflammatory mediators and cytokines including TGF-β1, fibroblast growth factor (FGF) and PDGF which signal to both immune cells to promote resolution of inflammation and to structural cells to promote repair at the site of injury (31-33). Fibroblasts and myofibroblasts contract to reduce the size of the wound and secrete ECM proteins to form a temporary matrix at the site of injury. This is until AT2 cells extensively proliferate and migrate over the temporary matrix in the process of re-epithelialisation (34). Following this, the temporary matrix is degraded, TGF-β1 signaling reduced and ECM synthesis decreased as fibroblasts undergo apoptosis (35, 36). This facilitates resolution of injury for the physiological

basis of gas exchange. Yet, in IPF patients this post-injury repair is dysregulated and alveolar homeostasis is lost. Consequently, the lung epithelium is altered resulting in excessive ECM deposition and fibroblast proliferation (Figure 1.2).

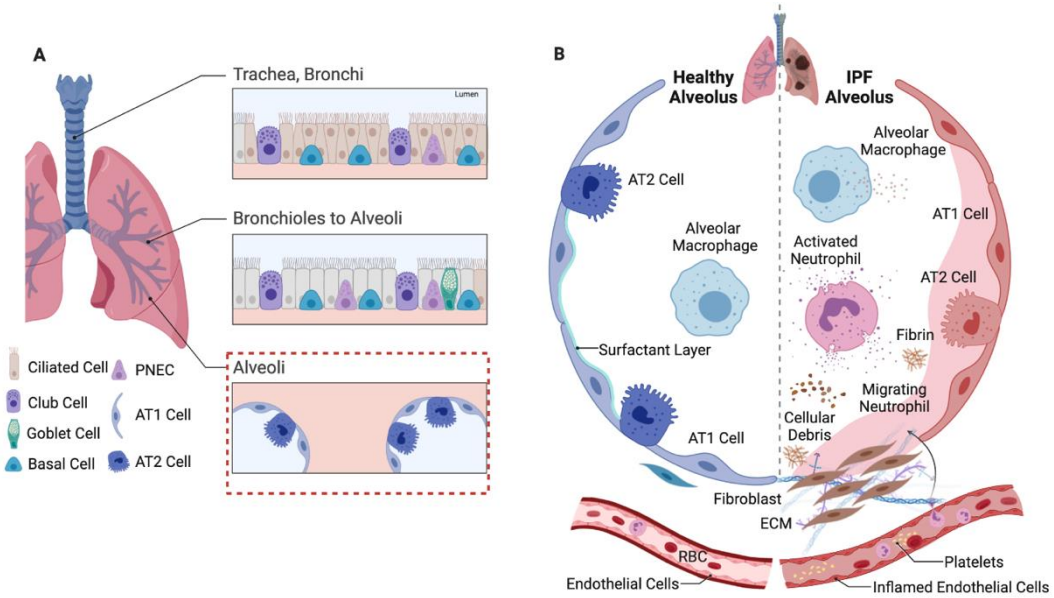


Figure 1.2: Anatomy of the lung and the respiratory epithelium
(A) *Diagram of the pseudostratified columnar epithelium composition along the lower respiratory tract. Cells of the tracheobronchial epithelium include club cells, basal cells and pulmonary neuroendocrine cells (PNECs). Mucin secreting goblet cells are found within the bronchioles through to the alveoli. AT1 and AT2 cells form the alveolar lining.* ***(B)*** *Comparison of histological and cellular interactions surrounding the alveoli in health vs IPF. The biological processes which influence manifestation of the disease phenotype are highlighted. (Created in BioRender.com)*

1.2.2. Wound Healing in Idiopathic Pulmonary Fibrosis Pathogenesis

IPF is believed to result from repeated injury to lung AECs in a genetically predisposed individual which leads to progressive fibrosis and homoeostatic imbalance due to dysfunctional wound repair. The 'epithelial injury hypothesis' proposes that compromised re-epithelialisation following epithelium injury enables interstitial cells to extensively proliferate and subsequently induce pulmonary fibrosis (37). How injury to the distal lung parenchyma induces apoptosis of the alveolar epithelium was first explored in the bleomycin model through the Fas antigen (Fas)/Fas ligand (FasL) apoptotic pathway (extrinsic pathway) (38, 39). Electron microscopy showed

apoptotic changes in AECs 14 days after bleomycin administration and necrotic AECs were also present (38). Although the relative importance of activation of apoptotic, necrotic and senescent pathways in the lung during development of IPF is still under investigation, epithelial cell senescence is increased in IPF lung tissue when compared to healthy controls and can be induced by TGF-β1 stimulation of human bronchial epithelial cells (HBECs) (40). The alveolar epithelium has been highlighted as the crucial 'IPF gatekeeper' as it is abnormal alveolar repair following injury which modulates the key downstream signaling pathways underlying lung fibrosis. Alternative naming for IPF has already been suggested such as 'epithelial-driven pulmonary fibrosis' based on current research findings (41).

Injury to the alveolar epithelium can typically result in alveolar collapse which subsequently exposes the underlying basement membrane (epithelial basal lamina) to damage (42). In the hamster bleomycin model, bleomycin administration disrupts the alveolar basement membrane with areas adjacent to sever injury appearing fragmented or folded, thickened and associated with extensive collagen deposition (43). This epithelial injury and basement membrane damage facilitates the clustering of fibroblasts in the alveolar airspaces where they proliferate and synthesise ECM components in attempt to repair the damage (44, 45). Lung fibroblasts utilise cell adhesion receptors (namely integrins) to mediate cell-ECM adhesion and migrate across the damaged basement membrane into the alveolar spaces. To do this they interact with the ECM via focal adhesions and fibrillar adhesions both of which are integrin-based structures (αvβ3 and α5β1, respectively) (46-49). These IPF fibroblasts

have been reported as being resistant to apoptosis meaning that post-injury repair is dysregulated and alveolar homeostasis is lost (50).

1.2.3. Fibroblastic Foci

Fibroblastic foci are a histopathological criteria for a definitive UIP pattern (6). Repeated injury to the alveolar epithelium causes growth factors, coagulants and chemotactic factors to be secreted which facilitate fibroblastic foci clustering at the injured site (51). These chemotactic factors include PDGF, TNF-α, TGF-β1, connective tissue growth factor (CTGF), endothelin 1 (ET-1), osteopontin and angiotensin (52-58). There are multiple myofibroblasts precursor cells and subsequently their origin is heterogenous. Most originate from resident tissue fibroblasts, but may also be bone marrow-derived (fibrocytes) or a result of epithelial-mesenchymal transition (EMT) (59-63). Pericytes and smooth muscle cells have also been hypothesised as a potential source of myofibroblasts (64).

Myofibroblast differentiation has been proposed as a two-stage model with two types of myofibroblasts identified showing distinct morphological characteristics (65). In response to mechanical tension, actin-containing stress fibres form in the cytoplasm of the quiescent fibroblasts altering the fibroblast functionality (66). These intermediate fibroblasts can generate contractile force and have been termed the "proto-myofibroblast", secreted factors promote modulation of these "proto-myofibroblast" into differentiated "myofibroblasts" (65). These are identifiable by expression of the cytoskeletal protein α-smooth muscle actin (α-SMA) and enhanced contractility (34, 67-70). Myofibroblasts are sometimes described as smooth muscle-like cells due to their acquired phenotype (67). Differentiation of the proto-

myofibroblast into a myofibroblast is dependent on TGF-β1 with α-SMA protein expression being TGF-β1-induced (71). Activated myofibroblasts are the primary effector cell in fibrosis and secrete ECM components inclusive of collagen, inflammatory and fibrogenic mediators and MMPs (72). After crossing the damaged basement membrane, myofibroblasts come into direct contact with the local collagen network causing subsequent contraction of the surrounding matrix and deposition of new collagen (65). Additionally, it has been suggested that myofibroblasts drive further fibrogenesis by inducing epithelial cell apoptosis and by preventing re-epithelialisation following injury (73, 74).

1.2.4. Extracellular Matrix ("Matrisome")

The ECM is a three-dimensional, non-cellular network within tissues that is in direct contact with the cells it surrounds. The lung ECM is critical for normal respiratory function and tissue structure, contributing both to the mechanical properties of the lung parenchyma and regulating a cells dynamic behaviour and cellular communication (75). The ECM functions as a solid-phase binding interface for secreted proteins therefore enabling their localisation in the ECM. There are two components of the ECM 1) the basement membrane composed of network-forming collagen IV and 2) the interstitial matrix composed of fibrillar collagens I and III (76-79). Adjacent collagen fibrils in the ECM are enzymatically crosslinked (covalent bond) to form fibers, this post-translational modification (PTM) is essential for the ECM structure and its mechanical properties (80). The ECM composition is regulated through the synthesis and breakdown of matrix components. ECM degradation is predominantly mediated by the MMP enzymes, however their action can be inhibited

by TIMPs to maintain homeostasis (81). Cells found within the ECM are either indigenous (reside in the matrix) or immigrant (migrate into the matrix).

In IPF, fibroblasts develop an invasive phenotype through direct contact with the ECM and mediate its remodeling through an excessive fibroproliferative response (82). This brings about an increase in collagen context which subsequently alters the type I to type III collagen ratio (44, 83, 84). Proteomic analysis has also shown matrisome-associated proteins to be enriched in IPF lung (85). Enzymes involved in ECM remodeling are also implicated in IPF pathogenesis. Interestingly, MMP levels are elevated in IPF and have been shown to promote rather than inhibit the fibrotic response to injury *in vivo* (86). Differences in the localisation and abundance of TIMPs 1-4 have also been shown in IPF lungs vs controls (87). However, only TIMP-1 (not 2 or 3) has been shown to be elevated in the lung following bleomycin-induced lung injury *in vivo* (88-90). There is less collagenolytic activity in an IPF patient lung caused by an increase in ECM stabilisation by covalent cross-linking of collagen (91, 92). Specifically, the cross-linking enzymes lysyl oxidase-like 2 (LOXL2) and transglutaminase 2 (TG2) are overexpressed in IPF which enhances pathological crosslinking and makes the ECM more resistant to proteolysis (93, 94). TG2 knockout mice have reduced fibrosis following bleomycin treatment upon histological assessment (94). Similarly, LOXL2 inhibition (antibody-mediated) decreases fibrillar collagen deposition following bleomycin challenge *in vivo* (93). The abnormally crosslinked ECM in IPF supports fibroblast growth through enhanced adhesion and proliferation (92, 93). Collectively, a resistant ECM and mechanical stress induces

profibrotic TGF-β1 activation thereby increasing lung stiffness in pulmonary fibrosis (95-97).

1.3. Idiopathic Pulmonary Fibrosis Pathophysiology

IPF has been proposed and accepted as a three-stage disease process: predisposition, activation and progression (41, 98). There are several non-modifiable and modifiable risk factors which predispose an individual to IPF development and identification of these has been informative of the disease pathophysiology. Repeated low-level injury to the alveolar epithelium of a genetically susceptible host causes endoplasmic reticulum (ER) stress-induced apoptosis of AT2 cells (42, 99). This is due to the production of misfolded and damaged proteins by the ER, a consequence of increased cell turnover demands (99). Activation occurs when alterations to the lung epithelium integrity causes the accumulation and activation of immune cells that subsequently release pro-fibrotic cytokines (100). This induces myofibroblast differentiation and ECM remodeling as previously described (Section 1.2.3 and Section 1.2.4). A change in ECM compliance also disrupts the endothelial cell-cell junctions causing vascular permeability (101). The modified ECM promotes IPF progression by supporting fibroblast growth and matrix deposition in a feed-forward loop of lung remodeling (41, 98). This relentless disease progression impairs gas exchange causing respiratory failure and IPF mortality. The evidence for each stage of the current pathogenesis model and the key molecules involved (TGF-β1, integrins and galectins) will be discussed in the following sections.

1.3.1. Predisposition

IPF is classified as either idiopathic or familial. If only one individual in a family has IPF it is described as sporadic (idiopathic), however if two or more cases of IPF are

present within the same family it is hereditary (familial) (102). Familial IPF is less common than the sporadic form with familial accounting for less than 5% of total IPF cases (6). But in both instances risk factors have been identified that may increase the likelihood of an individual developing the disease by contributing to the pathological wound healing response. As IPF is a chronic and progressive disease, age is highly associated with risk and survival is significantly related to age at presentation. Several hallmarks of aging are apparent in IPF lungs, specifically, altered intracellular communication, genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, mitochondrial dysfunction and cellular senescence (103). In addition, IPF prevalence is higher in men than women making sex a risk factor for the disease (104). Women have longer telomeres than men which may contribute towards the sex-related differences (105-107). Yet, higher sex hormone concentrations are protective against IPF onset and progression in both sexes, possibly by slowing telomere shortening (108). Most importantly, IPF is a polygenic disease and genetics predominately determine how susceptible a person is to developing the disease.

Genetic variants called single-nucleotide polymorphisms (SNPs) are present in the IPF genome and are determined by measuring the allele frequency between disease cases and controls. This is called a Genome-Wide Association Study (GWAS) in which clinical and/or biobank datasets can be applied. To date, 23 independent SNPs have been confirmed associated with IPF (109, 110). The genes identified are implicated in a range of biological functions and subsequently infer the pathological mechanisms underlying the disease. The common pathological consequences of these mutations

include decreased lung epithelium integrity, telomere shortening or increased cellular stress resulting in apoptosis (Table 1.2). More recently, there is increasing evidence of links between genetic variants (proprotein convertase subtilisin/kexin type 6 (PCSK6) and serine/threonine-protein kinase N2 (PKN2)) and different IPF phenotypes (111, 112).

Table 1.2: IPF genetic associations

Chromosome	Nearest Gene to SNP	Pathological Consequence
1	G-protein coupled receptor 157 (GPR157)	Cell signaling
4	Family with sequence similarity 13 member A (FAM13A)	
6	FKBP prolyl isomerase 5 (FKBP5)	
8	DEP domain-containing mTOR-interacting protein (DEPTOR)	
15	A-kinase anchoring protein 13 (AKAP13)	
16	Nitrogen permease regulator-like 3 (NPRL3)	
3	Kinesin family member 15 (KIF15)	Spindle assembly
5	Spindle apparatus coiled-coil protein 1 (SPDL1)	
7	Mitotic arrest deficient 1 like 1 (MAD1L1)	
15	Kinetochore scaffold 1 (KNL1)	
20	Stathmin-3 (STMN3)	
3	Telomerase RNA component (TERC)	DNA repair /Telomere maintenance
5	Telomerase reverse transcriptase (TERT)	
20	Regulator of telomere elongation helicase 1 (RTEL1)	
6	Desmoplakin (DSP)	Cell-cell adhesion
19	Dipeptidyl peptidase 9 (DPP9)	
11	Mucin-5B (MUC5B)	Host defence
19	Fucosyltransferase 6 (FUT6)	
13	ATPase phospholipid transporting 11A (ATP11A)	Mitochondrial dysfunction & metabolism
15	Isovaleryl-CoA dehydrogenase (IVD)	
7	Zinc finger with KRAB and SCAN domains 1 (ZKSCAN1)	Unknown
10	10q25.1	
17	Microtubule associated protein tau (MAPT)	

Adapted from Prof Louise Wain (University of Leicester) with permission (110, 113-117).

Associated genetic variants are classified as rare or common depending on their effect size (strength of the association) (118). Rare variants have a low allele frequency in the population but a high effect size, in contrast the allele frequency of common variants is higher but the effect size smaller. Rare variants are common in familial fibrosis and can be sub-classified based on their physiological gene function: alveolar health (surfactant protein A1 (SFTPA1), surfactant protein A2 (SFTPA2), surfactant protein C (SFTPC) and ATP-binding cassette subfamily A member 3 (ABCA3)) and telomere biology (TERT, TERC, RTEL1, poly(A)-specific ribonuclease (PARN), dyskerin pseudouridine synthase 1 (DKC1), TERF1 interacting nuclear factor 2 (TINF2) and nuclear assembly factor 1 ribonucleoprotein (NAF1)) (119). However, some of these rare telomere-related variants are also found in sporadic IPF (TERT, TERC, RTEL1 and PARN) (119). In sporadic IPF, several common variants each with a small effect contribute to the risk, however the MUC5B variant has a large effect size and is consequently associated with substantially higher susceptibility to disease (117, 120).

In addition to non-modifiable risk factors, occupational and environmental exposures have been associated with IPF susceptibility. Cigarette smoking has been identified as a risk factor for IPF, however the level of risk differs with smoking status (121, 122). Interestingly, the median survival time is extended in current smokers compared to former smokers (123). One possible explanation is that ongoing smokers have significantly lower interstitial cellularity and connective tissue deposition compared to former smokers which may reduce the risk of developing fibrosis (124). This has been termed “the healthy smoker effect” as smoking cessation may accelerate

progressive decline (125). Cigarette smoke has been shown to exacerbate the fibrotic response to bleomycin treatment *in vivo*. 6 weeks of exposure prior to bleomycin administration significantly increased fibroblast proliferation and lung collagen content, the MMP/TIMP balance and inflammatory cell profile was also adjusted (126). The impact of electronic cigarettes (e-cigarettes) and inhaled vapour on IPF risk is currently undetermined however it may present future new challenges as inhalation of occupational exposures (vapours, gas, dust or fumes) from metal, wood, silica and agricultural work can contribute to the IPF disease burden (127-137). A recent in-depth review and analysis of the occupational exposure data showed all except agricultural work to be significantly associated with IPF (138).

Microorganisms (viral, fungal and bacterial) have also been suggested as an independent IPF risk factor. Together with the discovery of IPF-associated variants in host defence genes (specifically mucociliary clearance of bacteria), a role for infection became evident from a combined immunosuppression therapy trial 'PANTHER-IPF' (NCT00650091) that resulted in severe adverse events and increased mortality in IPF patients (110, 117, 139). Bacterial 16S ribosomal ribonucleic acid (rRNA) sequencing has revealed that IPF patients have an increased bacterial load in BALF and an altered respiratory microbiome compared to control subjects (140, 141). In IPF patients, an increased bacterial load is associated with worse disease progression and a higher risk of mortality, but not with the extent of fibrosis (141, 142). Despite bacterial burden being associated with acute exacerbations of IPF (AE-IPF), routine treatment with antibiotic drugs is not supported as antibacterial therapy does not significantly

improve IPF patient mortality evident from the 'CleanUP IPF' (NCT02759120) and 'EME-TIPAC' trials (ISRCTN22201583) (143-145).

There is more substantial evidence to support a role for persistent or chronic viral infection in IPF development, with Epstein-Barr virus (EBV/ HHV-4), cytomegalovirus (CMV/ HHV-5), human herpesvirus 7 (HHV-7) and human herpesvirus 8 (HHV-8) all significantly associated with elevated IPF risk (146). However, whether viral infection is associated with AE-IPF or increases disease susceptibility via cytopathic effects or inducing epithelial injury is not fully understood (147, 148). Protein expression of the pro-fibrotic mediator TGF-β1 is elevated in epithelial cells from virus positive IPF lung sections when compared to virus negative (149). Additionally, It has been suggested that herpesvirus infection may exacerbate the fibrotic response by impacting telomere attrition and vascular remodeling (149, 150). The potential role of viral infection in progressive lung fibrosis may become clearer in the coming years due to long term follow up of large patient cohorts who have recovered from the coronavirus disease 2019 (COVID-19) severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. Betacoronavirus infections, specifically severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome coronavirus (MERS-CoV), are associated with lung abnormalities and lung fibrosis development (151, 152). An early report on patients hospitalised with SARS-CoV-2 infection similarly described ground glass opacities in all patients' lungs from computed tomography (CT) images (153). Discharged patients with SARS-CoV-2 had abnormal pulmonary function (diffusion capacity and lung volume) associated with disease severity (154). Although patient recovery from viral infection removes the damage

initiator, in a subset of patients the post-viral infection fibrosis persists and may progress further termed ‘post-COVID-19 fibrosis’; consequently, it is imperative that this cohort of survivors are rapidly identified and patients followed up (155, 156). It is evident from the ‘UKILD-Long COVID’ study (NCT05514522) that residual lung abnormalities are identifiable in a substantial proportion (11%) of patients discharged after COVID-19–related hospitalisation (157).

1.3.2. TGF-β1

Pro-fibrotic mediators play a major role in IPF activation and progression with TGF-β of critical importance (158). There are three mammalian isoforms of the TGF-β cytokine, TGF-β1, TGF-β2 and TGF-β3 and each is encoded by a different gene (159-161). All three isoforms are expressed in IPF lung tissue, although each isoform has a unique expression pattern (162). In lungs with early lesions of IPF TGF-β1 is primarily detected in alveolar macrophages, yet in lungs with advanced pulmonary fibrosis it is additionally localised to the epithelium and subepithelially matrix-associated (163). TGF-β1 protein expression in the matrix is concentrated to fibroblastic foci associated with sites of active fibrosis (164). By comparison, minimal TGF-β1 protein expression is detected in the non-fibrotic lung (54). TGF-β1 is a chemoattractant that augments its own effects by autoinduction as well as simultaneously inducing the expression of other fibrogenic cytokines (165). As TGF-β1 modulates both the inflammatory response and fibrogenesis, very tight regulation over its activation is required, however in IPF this tight regulation is lost and TGF-β1 is excessively activated. Mature TGF-β1 (112 aa) is secreted in complex with its prodomain (249 aa) as a biologically inactive precursor (Pro-TGF-β1, 361 aa) and is activated by furin protease cleavage (166, 167). The TGF-β1 prodomain named latency associated peptide (LAP) binds to

two TGF-β1 monomers (TGF-β1 homodimer) to form the 'small latent complex', this masks the active sites on TGF-β1 rendering it inactive (168). LAP targets the growth factor dimer for storage in close proximity to cells by directly binding to latent TGF-β binding protein 1 (LTBP1) in the ECM (169). Two monomers of the small latent complex are disulphide-bonded to a single molecule of LTBP1 (168). This 'large latent complex' directs localisation and anchorage of latent TGF-β1 to the ECM (Figure 1.3).

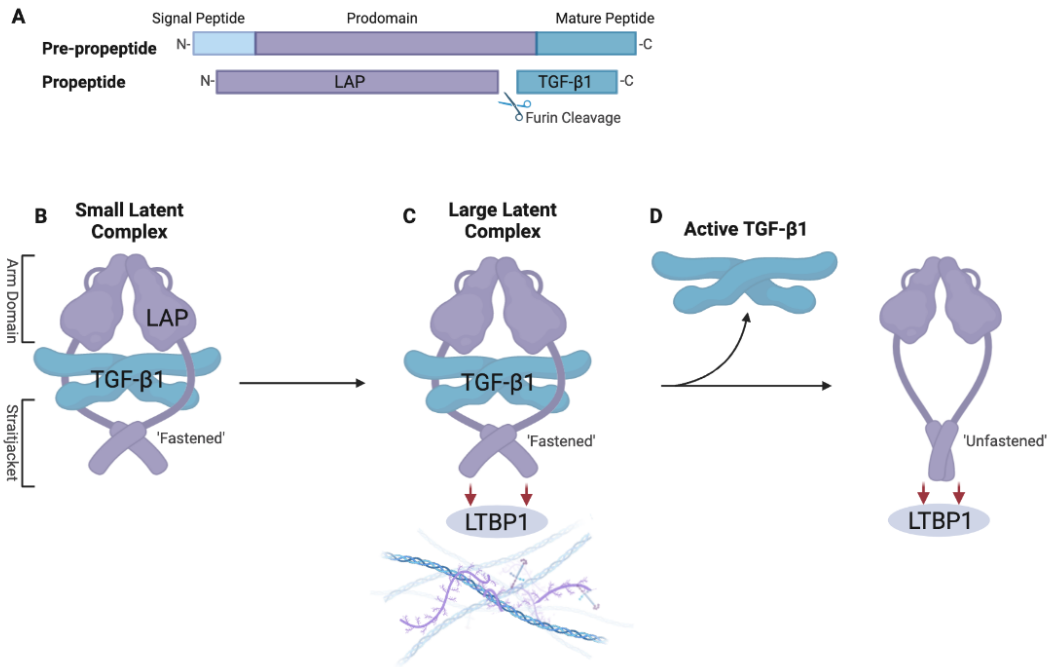


Figure 1.3: Structure of latent TGF-β1
Schematic diagram showing the synthesis, association and dissociation of the large latent TGF-β1 complex. ***(A)*** *Structure of the pro-TGF-β1 precursor.* ***(B)*** *TGF-β1 is encircled by the 'arm' and 'straightjacket' domains of LAP to form the inactive small latent TGF-β1 complex.* ***(C)*** *The binding of LTBP1 to the complex forms the large latent complex which localises the TGF-β1 homodimer to the ECM.* ***(D)*** *TGF-β1 becomes activated following its dissociation from LAP. (Created in BioRender.com)*

For TGF-β1 to become activated it must be released from the LAP domain of the complex. Activated TGF-β1 can then bind to its cell surface receptor and signal in an autocrine or paracrine manner. The TGF-β1 receptor (TGFβR) is a heterotetrameric structure composed of paired type I (TGFβRI) and type II (TGFβRII) receptors (170). Two TGF-β1 monomers can bind to the heterotetrameric receptor resulting in a 2:2:2

ternary complex (171). TGF-β1 ligand-binding to the TGFβRII subunit induces receptor oligmerisation and subsequent transphosphorylation of TGFβRI glycine-serine rich domain to activate its serine/ threonine kinase (172, 173). The suppressor of mothers against decapentaplegic (SMAD) cytoplasmic receptor regulated signal transducers SMAD2 and SMAD3 are subsequently phosphorylated to form a SMAD2/3 heterodimer (174-177). The co-mediator SMAD4 binds to the SMAD2/3 heterodimer to form a SMAD2/3/4 heterotrimeric complex (178, 179). SMAD2/3/4 translocates into the cell nucleus where it complexes with co-regulators to mediate gene expression (180, 181). TGF-β1 also signals via non-canonical pathways including ras homolog family member A (RhoA)/ rho-associated protein kinase (ROCK), phosphoinositide 3-kinase (PI3K)/ protein kinase B (Akt) to promote pro-survival signals, mitogen-activated protein kinase (MAPK)/ extracellular signal-regulated kinase (ERK) for cell proliferation, wingless/int-1 (Wnt) for tissue homeostasis and repair post-injury and neurogenic locus Notch homolog (Notch) pathways (182) (Figure 1.4).

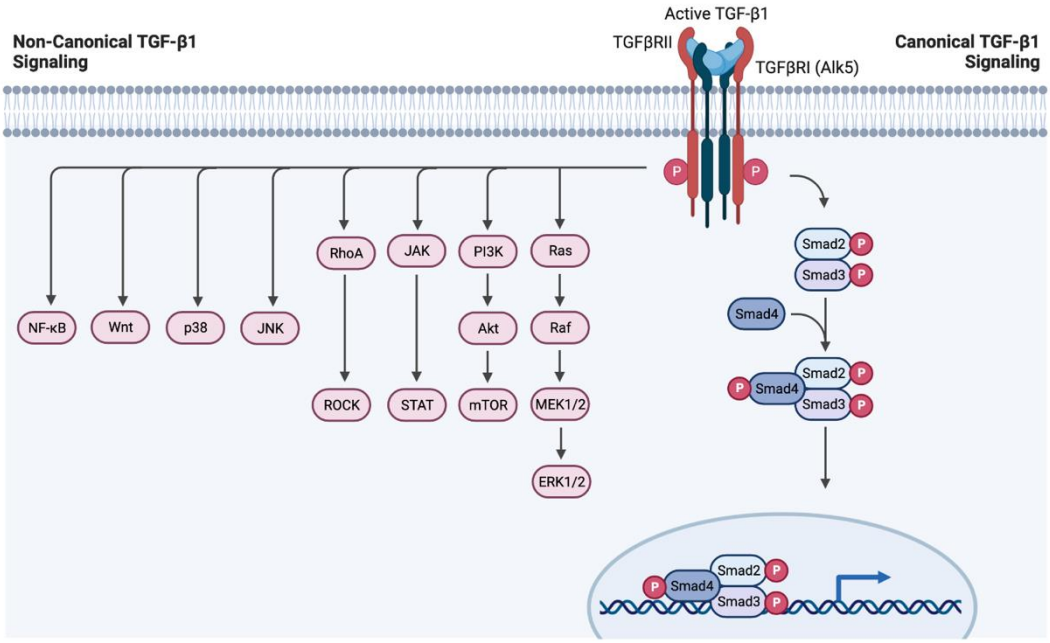


Figure 1.4: Overview of TGF-β1 signaling
Illustration of dimeric TGF-β1 signaling via both non-canonical and canonical TGF-β1 signaling pathways. Canonical SMAD signaling is initiated following the binding of active TGF-β1 to the heterotetrameric cell surface receptor. Phosphorylation events result in the formation of a SMAD2/3 heterodimer to which SMAD4 couples. This transcription factor complex subsequently translocates to the nucleus to regulate gene transcription. (Created in BioRender.com)

In health this TGF-β1 activation and signaling is tightly regulated through negative feedback mechanisms. This includes inhibitory SMADs which bind to the SMAD specific E3 ubiquitin protein ligase (SMURF) and target the TGF-βR for degradation (183-185). Yet in IPF, TGF-β1 activation serves as a positive feedback loop to activate more TGF-β1. It may do this by influencing cell surface integrin expression (186, 187).

1.3.3. Integrin-Mediated TGF-β1 Activation

Integrins are a family of heterodimeric receptors and are the integral transmembrane connection between the ECM and actin cytoskeleton (188, 189). Their structure is composed of alpha and beta subunits which are noncovalently bound, 18α and 8β subunits have been identified to date giving rise to 24 distinct mammalian integrins (190, 191). The two integrin subunits come together to form a globular “head”

(ectodomain) and two transmembrane "tails" (cytoplasmic) (192, 193). The transmembrane protein can signal bidirectionally namely inside-out (intracellular to extracellular) and outside-in signaling (extracellular to intracellular) (194). Integrins adopt three different major confirmational states which is regulated by divalent cation occupancy and corresponds to their ligand-binding affinity ('switchblade' model): bent with closed headpiece (inactive), extended with closed headpiece (primed) and extended with open headpiece (active) (195-197) (Figure 1.5).

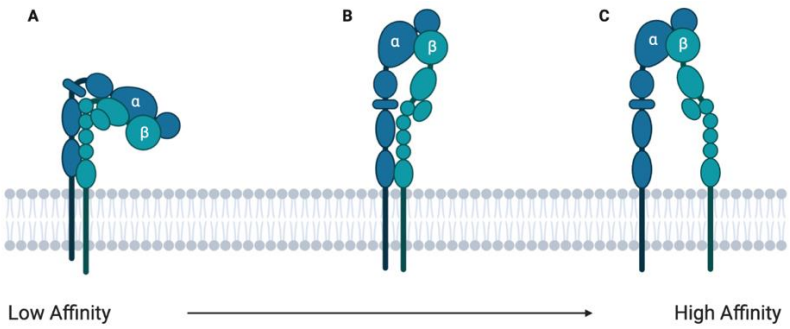


Figure 1.5: Integrin confirmation and binding affinity
Integrin confirmational state is determined by divalent cation occupancy and regulates ligand-binding activity. ***(A)*** *Bent confirmation (closed),* ***(B)*** *extended confirmation with closed headpiece and* ***(C)*** *extended confirmation with open headpiece (open).*

Divalent cations bind to the integrin metal ion-dependent adhesion site (MIDAS) motif which has metal specificity for calcium (Ca^{2+}), magnesium (Mg^{2+}) and manganese (Mn^{2+}) (198, 199). Integrins predominantly bind to ECM proteins however may also bind to cell surface or soluble ligands, eight integrins (α5β1, α8β1, αvβ1, αvβ3, αvβ5, αvβ6, αvβ8, and αIIbβ3) have an arginine-glycine-aspartic acid (RGD) tripeptide recognition site for binding RGD-containing proteins (194, 200-202).

Cell surface integrins can detect intracellular tension through their cytoplasmic tails which are in contact with cytoskeletal components such as talin and vinculin (203). In

physiological wound healing the actin cytoskeleton is reorganised promoting actin polymerisation and actomyosin contractility (204-206). The traction forces generated by this structural reorganisation are then conveyed to the ECM by inside-out integrin signaling. As previously described, inactive TGF-β1 is localised to the ECM (Section 1.3.2) and within the arm domain of TGF-β1 (LAP) there is an RGD sequence that RGD-binding integrins can recognise (168). Through this RGD binding, the arm domain of the TGF-β1 complex provides 'elastic resistance' to the mechanical strain e.g. contraction, resulting in the release and activation of TGF-β1 (95, 96). *In vitro* this traction-dependent TGF-β1 activation has been shown to occur via integrins: αvβ1, αvβ3, αvβ5 and αvβ6 (96, 207). Activated TGF-β1 then acts in an autocrine or paracrine manner to initiate canonical and non-canonical TGF-β1 signaling to regulate developmental processes, cell apoptosis, plasticity and migration of cells (Figure 1.6).

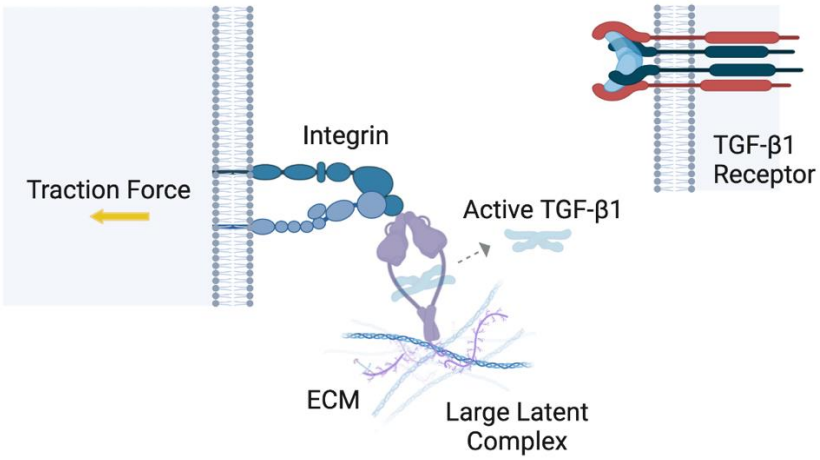


Figure 1.6: Overview of traction-dependent TGF-β1 activation
Schematic diagram illustrating how cytoskeletal changes are detected by the integrin which conveys these traction forces to the ECM causing release of TGF-β1 from the large latent complex. (Created in BioRender.com)

As TGF-β1 is produced by many cell types and is involved in several structural and signaling processes, systemic blockage of TGF-β1 has deleterious adverse effects.

Therapeutics which inhibit proteins known to be important for regulating TGF-β1 activity have consequently been developed with integrins being one of a number of targets under investigation. A more targeted approach to therapy may permit localised inhibition of TGF-β1 activation without the unwanted potential side-effects (208).

1.3.4. Integrin Expression in the Lung in Health and Disease

Generation of a single-cell atlas of pulmonary fibrosis by single-cell RNA sequencing (scRNAseq) has accelerated the assessment of gene expression profiles of individual cell types in health vs disease. Research by the Kaminski/ Rosas laboratory has enabled visualisation and comparison of the differential expression of a variety of integrins in lungs cells, these data are publicly available on the 'Idiopathic Pulmonary Fibrosis Cell Atlas' (209, 210). The β subunit genes (ITGβ1, ITGβ5 and ITGβ6) that encode the key fibroblast integrins (αvβ1 and αvβ5) and the epithelial-restricted integrin (αvβ6) are differentially expressed in control and IPF subjects. In addition, a population of transcriptionally distinct epithelial cells termed aberrant basaloid cells have been identified in IPF subjects which highly express ITGβ6 (209, 211) (Figure 1.7).

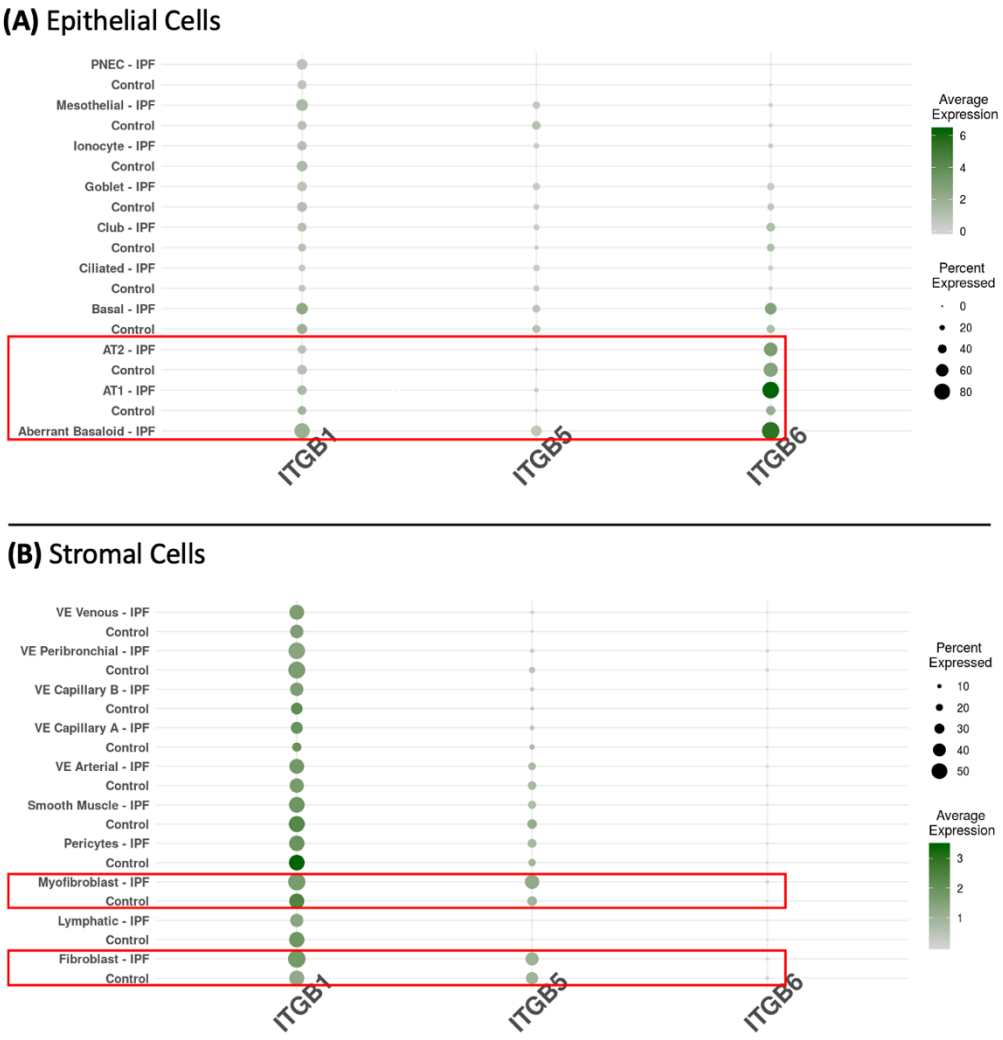


Figure 1.7: Differential expression of integrin genes
*Visualisation of integrin gene expression changes across **(A)** epithelial cells and **(B)** stromal cells generated by 'Batch Explorer' (209, 210). The size of the dot encodes the percentage of cells within a class, while the colour encodes the average expression level across all cells within a class.*

A number of studies have linked increased integrin expression to the development of lung fibrosis however the epithelial-restricted integrin αvβ6 is the most researched (212, 213). Protein expression of αvβ6 is low-undetectable in healthy pulmonary tissue, however gets upregulated in fibrotic regions of the lung in IPF patients (214). High levels of αvβ6 integrin on lung biopsy is significantly associated with a worse prognosis (215). *In vivo,* murine studies of bleomycin-induced lung injury revealed temporospatial upregulation of αvβ6 protein expression in areas of the lung

subsequently associated with development of fibrosis (207, 216). *In vitro*, TGF-β1 canonical signaling induces ITGβ6 gene expression and results in increased αvβ6 protein levels at the cell surface (187, 217). Therefore, in epithelial cells TGF-β1 activation can be regulated by two positive-feedback loops that amplify the fibrotic response 1) TGF-β1 autoinduction and 2) higher αvβ6 protein expression (207).

The critical importance of the αvβ6 integrin in the regulation of TGF-β1 activation was discovered by loss of or blockade of function in animal models of fibrosis. β6 knockout mice are protected against bleomycin-induced pulmonary fibrosis as evidenced by histological assessment and hydroxyproline quantification (207). Similarly, β6 knockout mice do not develop radiation-induced lung fibrosis as demonstrated by collagen staining (Masson's trichrome) and hydroxyproline context (218). Therapeutic blockade of αvβ6 function (anti-αvβ6 mAb) was first explored in an experimental model of renal fibrosis (219). However, the anti-αvβ6 mAb also protects against radiation-induced lung fibrosis and attenuates bleomycin-induced pulmonary fibrosis *in vivo* (214, 218). Blockade of αvβ6 in the bleomycin model decreases canonical TGF-β1 signaling as assessed by nuclear SMAD2/3 phosphorylation (214). Inversely SMAD3 knockout mice exposed to adenoviral TGF-β1 have reduced αvβ6 protein expression in the lung parenchyma (187). *In vitro* lysophosphatidic acid (LPA) and protease-activated receptor 1 (PAR1) agonist peptides induce αvβ6-mediated TGF-β1 activation in a guanosine triphosphate (GTP)-dependent (RhoA/ROCK) manner (95, 216). Activation of the coagulation cascade therefore contributes to lung injury and the anti-αvβ6 antibody blocks this TGF-β1 activation as evidenced by decreased SMAD2 phosphorylation (95, 216, 220). Furthermore following injection

of a PAR1-activating peptide, β6 knockout mice have reduced nuclear phospho-SMAD2 compared to wild-type (WT) mice (95).

Although it is damage to the alveolar epithelium that is believed to trigger IPF pathogenesis, it is the fibroblasts which predominately drive disease progression making them an attractive therapeutic target. Fibroblasts can express four different αv containing integrins: αvβ1, αvβ3, αvβ5 and αvβ8 and RGD-dependent binding of integrins to LAP was first shown for αvβ1 and αvβ5 (221, 222). Depletion of myofibroblast αv integrins using itgav^{flox}/^{flox};*Pdgfrb*-Cre (αv Cre) mice demonstrated their contribution to fibrotic diseases across multiple organs (kidney, liver and lung) (223). αv Cre mice are protected against bleomycin-induced pulmonary fibrosis as evidenced by collagen staining (Picrosirius red) and hydroxyproline context (223). Among the αv integrins, the αvβ5 integrin is localised to fibroblastic foci in IPF patients and αvβ5 blockade (anti-αvβ5 mAb) significantly decreases factor Xa (FXa)-induced myofibroblast differentiation and SMAD2 phosphorylation *in vitro* (224). However global or conditional loss of fibroblast integrins αvβ3, αvβ5 or αvβ8 is not protective against carbon tetrachloride (CCl_4)-induced hepatic fibrosis *in vivo* (223). Similarly dual αvβ3/αvβ5 knockout mice are not protected from bleomycin-induced lung fibrosis (225). As a result, αvβ1 was suggested to be the principle integrin responsible for adhesion to and activation of latent TGF-β1 in these cells (223).

The level of αvβ1 protein expression is significantly elevated in lung tissue from IPF patients compared to control subjects and in fibrotic regions of lung tissue from bleomycin treated mice (226). *In vitro* IPF patient-derived lung fibroblasts express the

αvβ1 integrin which mediates adhesion of these cells to LAP and traction-dependent TGF-β1 activity (221, 226). TGF-β1 has also been shown to elevate expression of the integrin β1 subunit at the messenger RNA (mRNA) and cell surface protein level suggesting that it too is subject to regulation via a positive feedback loop (217, 227). Inhibition of αvβ1 with a small-molecule inhibitor (compound c8) significantly reduces activation of TGF-β1 in both a model of bleomycin-induced lung fibrosis and CCl_4-induced hepatic fibrosis, as determined by the level of collagen deposition and phosphorylation of SMAD3 (221). Inhibiting several αv integrins may have an additive antifibrotic effect by targeting the activation of TGF-β1 in multiple cell types. The small molecule CWHM-12 blocks αv containing integrins and prevents the progression of established hepatic or pulmonary fibrosis *in vivo* (223). CWHM-12 significantly reduces the level of collagen staining (Picrosirius red) and hydroxyproline context in the established bleomycin model (223). Similarly, CWHM-12 significantly reduces collagen deposition in the CCl_4 model and the protein expression of both α-SMA and phosphorylated SMAD3 (pSMAD3) is reduced in CWHM-12 treated liver tissue (223). However, as αv integrins are required for homeostatic functions, more specific targeting of integrins overexpressed in fibrotic tissue is required.

A recent study using a dual selective inhibitor of αvβ6 and αvβ1 (PLN-74809) in precision cut lung slices (PCLS) prepared from IPF patient lung explants demonstrated that PLN-74809 significantly reduced the mRNA expression of collagen type I alpha 1 chain (COL1A1) and other fibrosis-related genes (226). Additionally, PCLS treated with PLN-74809 (7 days) had a significant reduction in SMAD2 phosphorylation. In bleomycin-challenged mice, PLN-74809 significantly reduced collagen deposition and

SMAD3 phosphorylation in a dose-dependent manner. As the inhibitor blocks both αvβ6 and αvβ1 integrin binding to LAP, it targets both the epithelial and fibrogenic response, respectively. Following this preclinical work, the efficacy and safety of PLN-74809 was evaluated in a phase 2a trial 'INTEGRIS-IPF' (NCT04396756) in which PLN-74809 was well tolerated and achieved statistically significant FVC increase in IPF patients (228).

1.4. Introduction to Galectins

1.4.1. Galectin Structure

Galectins are soluble proteins that bind to β-galactoside carbohydrates (229-231). 15 galectins have been discovered in mammals so far, with 12 of these galectin genes found in humans (232, 233). They are β-sandwich in structure and are composed of two anti-parallel β-sheets with the concave β-sheet being made up of 6 strands (S1-S6) and the convex being made up of 5 strands (F1-F5) (234). Within the concave side (S4-S6) there is a negatively charged groove/pocket in which glycan ligands can bind termed the carbohydrate-recognition domain (CRD) that is ~135 amino acids in length (235). The CRD is composed of five subsites (A-E) and the amino acids in subsites C-D are highly conserved among the family (232). It is the amino acid side chains of subsites C-D (histidine, asparagine, arginine, asparagine, glutamic acid, arginine) that are most important for glycan binding (232, 236). Subsite C is the most highly conserved and binds to β-galactoside whereas subsite D is less conserved but also contributes to the core binding site by accommodating the sugar residue (232). The CRD is the only functional domain, however galectins also have non-carbohydrate binding sites (237) (Figure 1.8).

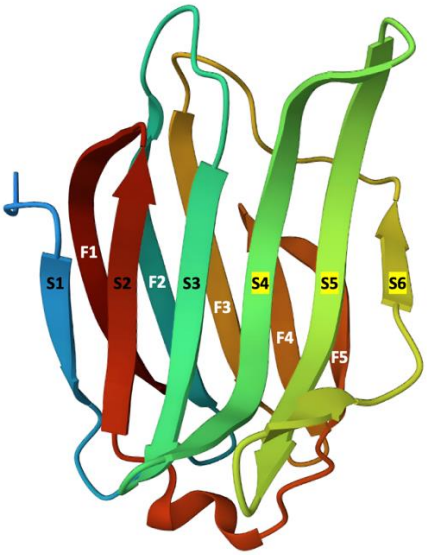


Figure 1.8: Structure of a galectin
Crystalised tertiary structure of a human galectin (PDB code: 5H9P) showing the S-sheets (S1-S6) and F-sheets (F1-F5). The S strands comprising the CRD are highlighted in yellow (S4-S6).

There are two different types of galectins: S-type lectins (sulfhydryl/thiol dependency) and C-type lectins (calcium dependency) (238). They have been classified into 3 groups based on their CRD: 1) prototypical galectins that contain two identical CRDs: galectin-1, 2, 5, 7, 10, 11, 13, 14, 15, chimeric galectins that contain a single CRD and a large amino terminal domain: galectin-3 and 3) tandem-repeat galectins containing two distinct CRDs: galectin-4, 6, 8, 9, 12 (239). Many of the galectin transcripts have alternative splicing therefore generating several different isoforms (233) (Figure 1.9).

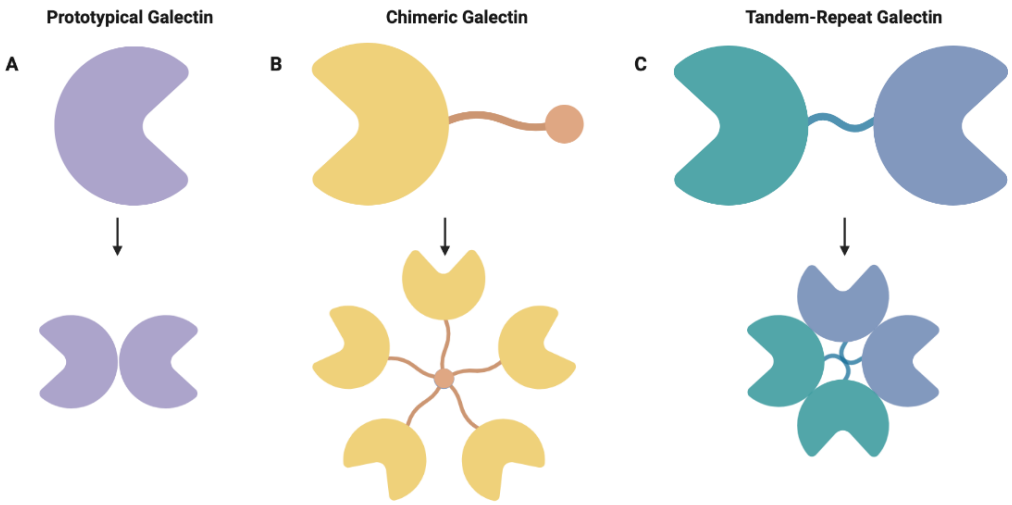


Figure 1.9: Galectin subtypes
Diagram depicting the three different forms of galectin structures: ***(A)*** *a prototypic galectin in both monomeric and dimeric form,* ***(B)*** *a galectin chimera capable of forming pentameric structures and* ***(C)*** *the structure of a tandem-repeat heterodimer in both monomeric and dimeric form. (Created in BioRender.com)*

1.4.2. Biosynthesis and Expression of Galectins

Galectins are synthesised on free ribosomes in the cytosol (unattached to the ER) (240). The newly synthesised proteins can then either remain in the cytoplasm, be transported to the nucleus or secreted from the cell via a non-classical secretion pathway (non-ER-golgi) (233, 241-243). Galectins do not have a signal sequence to direct the protein to the ER during synthesis, therefore it is thought that they are very important intracellularly but have acquired additional extracellular functions (244, 245). It has been suggested that galectins are not secreted conventionally as their carbohydrate-binding abilities would interrupt trafficking (235). Alternatively, perhaps the non-classical pathway allows the secretion of specific galectins in response to signals (235). Their secretion is generally stress-induced and different methods of secretion have become apparent depending on the galectin, cell type and its polarity (246).

1.4.3. Galectin Ligand-Binding and Function

Galectins are localised intracellularly and can signal in an autocrine fashion by binding to ligands within the cytoplasm, nucleus or mitochondria of the cell in which they were synthesised, or by binding to cell surface proteins on the same cell once secreted extracellularly (247). Alternatively, following their non-classical secretion galectins may signal in a paracrine fashion by binding to cell surface proteins on a neighbouring cell (248). Galectins can be involved in both carbohydrate-dependent protein-glycan interactions and carbohydrate-independent protein-protein interactions, although the latter tend to be less common and located intracellularly (247, 249, 250). Through their carbohydrate-dependent interactions, galectins can recognise and bind a wide range of proteins with suitable oligosaccharides rather than binding directly to specific individual receptors (251).

Galectin-carbohydrate screening arrays have been used to demonstrate the CRD glycan binding specificities, although these binding studies are not a true representation of the physiological context in which the glycoproteins are found (252). In total there are five types of glycosylation (N-, O-, C-linked, glypiation and phosphoglycosylation) with N-linked (NXS/T) and O-linked the most common. Generally, galectins have the highest binding affinity for N-linked glycans (236). The disaccharide N-acetyllactosamine [LacNAc] is the most basic unit required for galectin ligand-binding and it is found on every complex N-glycan, several types of O-linked glycans and on some glycolipids (253). The type of glycosylation and the extent of glycan branching present on a target ligand varies among glycoproteins and highly influences the galectin-ligand interactions. These glycan residues may also be further

modified (sialylation, sulfation and fucosylation) meaning that multiple glycoprotein isoforms exist with varying galectin binding affinity (236). For example, erythrocytes/ red blood cells (RBCs) express different oligosaccharides (N-acetylgalactosamine [GalNAc] or D-galactose [Gal]) on their cell surface which determine blood group (ABO) but also galectin binding affinity and sensitivity to haemagglutination (254, 255) (Figure 1.10).

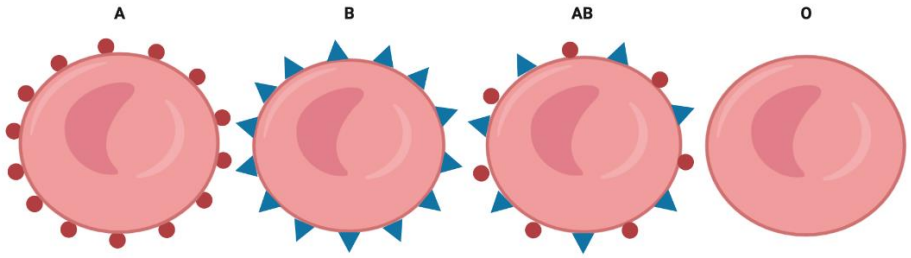


Figure 1.10: ABO blood group antigens
Blood group A has only A antigen [GalNAc] on the surface of RBCs and similarly only B antigen [Gal] for blood group B. AB blood group has both antigens present, but O group has no antigens on RBCs.

Although the direct binding of a galectin monomer to a single glycan residue is a low affinity interaction, multiple galectins may bind to a single sugar residue or multiple sugar residues present on a glycoprotein causing multivalent high-avidity reversible binding (256). These galectin-glycan interactions can be formed both in solution or at the cell surface to facilitate receptor clustering and lattice formation (257-259). This multivalency has been shown to occur through positive cooperativity whereby the binding of a single galectin recruits additional galectins to the glycoconjugate ligand (260). Galectins are pleiotropic proteins which participate in many context-dependent (galectin subtype, time and location) functions via their protein-glycan and protein-protein interactions. They are important for several biological processes including development, cell survival, the immune and inflammatory response and cell

communication (251). However, galectins have also been implicated in the pathogenesis of various diseases including fibrosis, cancer and cardiovascular disease (237). Galectin-3 and galectin-1 are the most studied members of the galectin family and will be discussed in the following sections (Section 1.4.4 and Section 1.4.5). IPF patients have distinct cell types with differential expression of the genes encoding galectin-3 (LGALS3) and galectin-1 (LGALS1) in comparison to healthy controls (Figure 1.11).

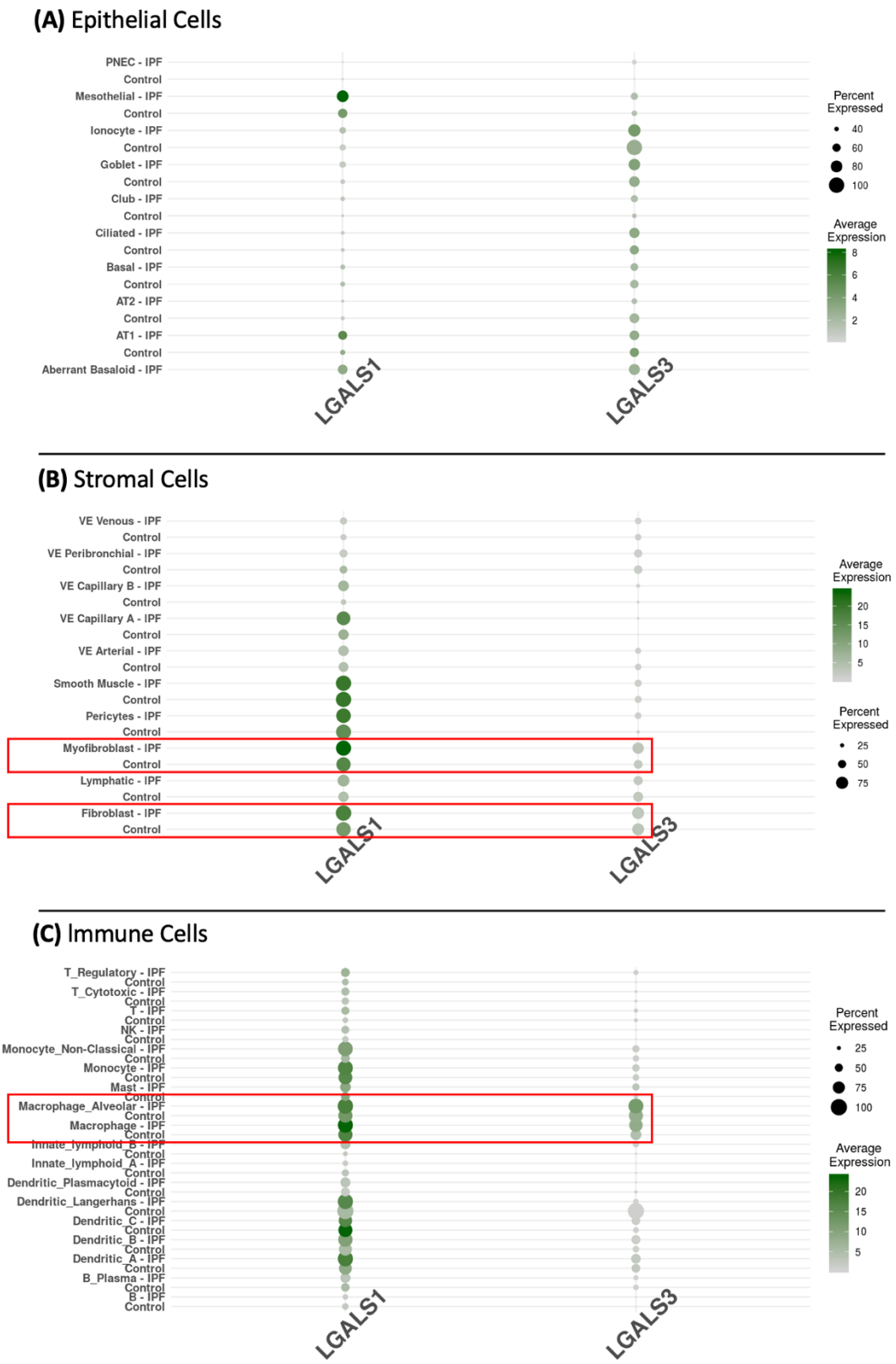


Figure 1.11: Differential expression of galectin genes
*Visualisation of galectin-1 and galectin-3 gene expression changes across **(A)** epithelial cells **(B)** stromal cells and **(C)** immune cells generated by 'Batch Explorer' (209, 210). The size of the dot encodes the percentage of cells within a class, while the colour encodes the average expression level across all cells within a class.*

1.4.4. Galectin-3

Galectin-3 is a 250 amino acid (MW: 27 kDa) chimeric galectin (261, 262). The protein is encoded by the LGALS3 gene mapped to chromosome 14 q21-22 (1.1 kb transcript) (235, 263). Galectin-3 has a C-terminal domain encoded by 3 exons (~130 aa) and an extended flexible N-terminal region encoded by 3 exons (~120 aa) (262, 264). Within the N-terminal domain there are two regions, a collagen like domain containing a proline-glycine-alanine-tyrosine rich (PGAY) repeat motif (7-14 repeats, each containing 8-11 amino acids) and a short end (235, 262, 264). Galectin-3 is expressed by various cell types (inflammatory cells, fibroblasts and epithelial cells) and is located intracellularly in both the cytoplasm and nucleus (249, 265). It is not a glycoprotein however, it can be phosphorylated at its N-terminal serine residues which has been linked to its nuclear export and ligand-binding capability (266, 267). The N-terminal domain of galectin-3 has also been implicated in its non-classical secretion (268). Galectin-3 is monomeric at moderate-high concentrations but can oligomerise through either the N-terminal or C-terminal domain (260, 269-271). It has also been reported to exist as a pentamer in the presence of its carbohydrate ligands (272).

1.4.4.1. Galectin-3 in Fibrosis

A role for galectin-3 in fibrosis was initially suggested in studies which reported abundant levels of galectin-3 protein in patient biopsies of liver cirrhosis and chronic pancreatitis (273, 274). Previous work has additionally shown galectin-3 to play a key role in the development of hepatic and renal fibrosis (275, 276). In the CCl_4-induced hepatic fibrosis model, galectin-3 is upregulated at both the gene and protein level with its expression temporospatial associated with fibrotic areas of the liver as evidenced by collagen staining (Picrosirius red) and colocalisation with α-SMA (275).

Galectin-3 knockout mice have significantly reduced collagen content and α-SMA present following CCl_4 treatment, although inflammatory cell recruitment (macrophage and neutrophil) and pro-inflammatory cytokine secretion (TNF-α) are unchanged in the absence of galectin-3 (275). Galectin-3 is also elevated in the unilateral ureteral obstruction (UUO) mouse model of renal fibrosis (276). Yet, galectin-3 knockout mice have significantly reduced collagen content (Picrosirius red) and α-SMA present after UUO despite no change in macrophage recruitment, IL-6 or TNF-α release (276). Within the lung there are resident alveolar macrophages and a recruited population of macrophages which are critical effector cells in wound healing. Galectin-3 expression is predominantly cytoplasmic in macrophages however its cell surface distribution is upregulated following irradiation-induced lung fibrosis (265). As with galectin-3 knockout mice, macrophage depletion in a UUO model significantly reduces myofibroblast activation and collagen deposition, thereby linking both macrophages and galectin-3 to the fibrotic phenotype (276). Macrophages within the fibrotic niche are reportedly monocyte-derived and have pro-fibrotic activity leading them to be termed 'scar-associated macrophages' (SAMs) (209, 211, 277-281). Identification of these SAMs by single-cell transcriptomic analysis and lineage tracing supersedes the view that alternatively activated macrophages are associated with enhanced fibrosis (282, 283).

Some of the most compelling evidence of a role for galectin-3 in the pathogenesis of IPF has been identified in mouse models of fibrotic lung disease. Administration of adenoviral TGF-β1 (Ad-TGF-β1) or intratracheal bleomycin to WT mice induces fibrotic changes in the lung and galectin-3 expression is localised within these fibrotic

areas (284). As in models of liver and kidney fibrosis, galectin-3 knockout mice are protected from the profibrotic effects of bleomycin or TGF-β1 overexpression as evidenced by decreased collagen context in the lung (Masson's trichrome and sircol assay) (284). TGF-β1 treatment induces myofibroblast differentiation and collagen synthesis in fibroblasts isolated from WT mice, but this is reduced by galectin-3 knockdown as determined by sircol assay, immunofluorescence and western blot analysis (284). Although galectin-3 is expressed by fibroblasts and can stimulate fibroblast cell proliferation, it is reportedly myeloid-derived galectin-3 which is the key contributor to driving fibrosis (261, 285, 286). The finding that AECs isolated from galectin-3 knockout mice show reduced EMT in response to TGF-β1 further implicates a role for galectin-3 in fibrosis (284). These AECs have maintained epithelial cadherin (E-cadherin) surface expression (epithelial marker) and reduced α-SMA (myofibroblast marker), therefore it has been proposed that EMT-myofibroblast activation is galectin-3-dependent (284). In addition, the galectin-3 receptor CD98 has recently been implicated in TGF-β1-induced EMT in A549 cells (lung carcinoma epithelial cells) (287). However, the EMT theory remains controversial with lineage tracing studies showing no evidence of EMT at the cellular or molecular level *in vivo* (288).

Galectin-3 is suggested to be involved in re-epithelialisation during wound healing with AECs from galectin-3 knockout mice having reduced TGF-β1-induced migration on the *in vitro* scratch-wound assay (284). Small interfering RNA (siRNA) knockdown of epithelial galectin-3 effects TGF-β1 receptor function and its downstream signaling in A549 cells. In addition, flow cytometric assessment detected a reduction in TGFβRII

cell surface expression by 66% thus suggesting that TGF-β1 receptor expression and subsequently TGF-β1 ligand-binding is in part regulated by galectin-3 (284). Although this had no effect on canonical TGF-β1 signaling as demonstrated by western blot, it did reduce TGF-β1-mediated β-catenin activation in a SMAD-independent manner. During active Wnt signaling the β-catenin destruction complex is inactive (by phosphorylation) meaning that β-catenin can translocate into the nucleus and induce T-cell factor/lymphoid enhancer-binding factor 1 (TCF/LEF1) transcriptional activity. However, during inactive Wnt signaling the destruction complex is activated causing cytoplasmic β-catenin proteolysis. The destruction complex is more active in AECs from galectin-3 knockout mice and consequently there is less TGF-β1-mediated β-catenin activation compared to the WT control (284). The potential importance of TGF-β1 signaling independent of SMAD has similarly been suggested in both the CCl_4 and UUO fibrosis models (275, 276). Although, this may represent a cell specific effect as TGF-β1 predominantly signals via SMAD and SMAD3-deficient mice are protected against bleomycin-induced pulmonary fibrosis (289).

Administration of Ad-TGF-β1 or intratracheal bleomycin *in vivo* elevates the concentration of galectin-3 in the BALF (284). Similarly, galectin-3 levels are significantly higher in the BALF and serum of IPF patients compared to controls (284, 290). This increase of galectin-3 in the BALF is at least partly derived from alveolar macrophages (290). UIP patient lung sections immunostained for galectin-3 show elevated galectin-3 expression within fibroblastic foci that is temporospatial associated with fibrosis (284). Pre-clinical research has already demonstrated that a galectin-3 inhibitor GB0139 (a thio-digalactoside) with some galectin-1 binding

affinity could reduce bleomycin-induced lung fibrosis *in vivo* as evidenced by a reduction in collagen deposition and has led to the compound being used in a clinical trial in IPF patients (284). The safety, tolerability, pharmacokinetics and pharmacodynamics of inhaled GB0139 were investigated in a phase 1/2a clinical trial (NCT02257177) and showed highly suitable for dosing IPF patients, with no significant treatment-related side effects (291). The time taken to reach maximum plasma concentration ranged from 0.6 to 3 hours and it had a plasma half-life of 8 hours (291). GB0139 reduced galectin-3 expression by bronchoalveolar lavage (BAL) alveolar macrophages and decreased plasma levels of biomarkers associated with IPF progression (PDGF-BB, plasminogen activator inhibitor-1 (PAI-1), galectin-3, chemokine (C-C motif) ligand 18 (CCL18) and chitinase-3-like protein 1 (CHI3L1/ YKL-40)) (291). The Food and Drug Administration (FDA) approved GB0139 as an investigational new drug which is now being used in an EU and US phase 1b/2a clinical trial 'GALACTIC-1' (NCT03832946).

1.4.5. Galectin-1

Galectin-1 is a 135 amino acid (MW: 15 kDa) prototypical galectin (244, 292). The cytogenic location of human LGALS1 is chromosome 22 q12-13, it has 4 exons and a transcript length of 0.6 kb (232, 235). Monomeric galectin-1 has a hydrophobic region (formed between the S1 strand and the F1 strand) and when these hydrophobic regions of monomers interact they form non-covalent galectin-1 homodimers (234, 293, 294). When galectin-1 dimerises the CRDs are ~40-50 Å apart and located at opposite ends of the dimer (293). Galectin-1 is differentially expressed by various cell types (immune cells and stromal cells) and is predominantly located intracellularly in both the cytoplasm and nucleus (209, 210, 247).

1.4.5.1. Galectin-1 in Fibrosis

A role for galectin-1 in fibrogenesis was initially suggested in chronic pancreatitis whereby galectin-1 positively correlated with the extent of pancreatic fibrosis (274). Activated pancreatic stellar cells (PSCs) have a myofibroblast-like phenotype and subsequently play a central role in this fibrogenesis (295). Galectin-1 mRNA and protein expression is significantly increased in activated PSCs and correlates with α-SMA protein expression *in vitro* (296). Previous work has similarly shown galectin-1 to play a key role in the development of hepatic fibrosis. Galectin-1 expression is increased in both the bile duct ligation (BDL) and thioacetamide (TAA)-induced hepatic fibrosis models (297). Hepatic stellate cells (HSCs) are the central effector cell in hepatic fibrosis and activated HSCs isolated from TTA treated rats show elevated galectin-1 at the mRNA and protein level (297, 298). Galectin-1 (intracellular and secreted) is differentially expressed in quiescent and activated HSCs *in vitro* as demonstrated by 2D-electrophoresis (2-DE) and mass spectrometry (299). Proteomic analysis similarly identified elevated galectin-1 (intracellular and secreted) in HSCs following CCl_4 treatment *in vivo* (299).

Using the same 2-DE proteomic approach it has been shown that galectin-1 is upregulated in whole lung tissue following irradiation-induced lung fibrosis *in vivo* (300). Similarly, galectin-1 protein expression is increased in the lung following bleomycin-induced lung injury as demonstrated by immunohistochemical staining and *ex vivo* western blot analysis (301). However, the galectin-1 inhibitor OTX008 has no detectable effect on the bleomycin-induced fibrotic phenotype upon histological assessment, although it should be considered that OTX008 is a low affinity allosteric

inhibitor (302). Galectin-1 has been suggested to have no, or minimal influence in this model (303). Comparison of BALF protein maps identified increased galectin-1 in IPF patient samples vs other ILDs (sarcoidosis and systemic sclerosis) and healthy controls (304-306). Significantly higher galectin-1 was detected in BAL from IPF former smokers than non-smokers and this correlated with number of packs/year among smokers (306). This increase in BAL galectin-1 concentration inversely correlated with DLCO and carbon monoxide transfer coefficient (KCO) in IPF (306). In contrast, serum galectin-1 concentrations are significantly higher in other fibrotic ILDs (sarcoidosis and non-specific interstitial pneumonia) and consequently may be a less clinically meaningful IPF biomarker (307).

Multiple *in vitro* studies have associated galectin-1 with myofibroblast differentiation and ECM production (301, 308). As previously described, myofibroblast differentiation is dependent on TGF-β1 (Section 1.2.3). Expectedly, TGF-β1 treatment of murine (NIH3T3) and human (IMR90) fibroblasts increases the protein expression of α-SMA, pro-collagen and fibronectin, yet galectin-1 levels simultaneously increase in a SMAD-independent manner (301). TGF-β1 treatment of murine fibroblasts elevates galectin-1 expression in the nuclear fraction of whole cell lysates and has been proposed to regulate fibrotic gene expression (301). TGF-β1 induces translocation of galectin-1 from the cytoplasm into the nucleus where it colocalises with SMAD2, subsequently the galectin-1-SMAD2 interaction has been hypothesised to initiate SMAD-mediated transcription and myofibroblast differentiation (301). Galectin-1 knockdown by short hairpin RNA (shRNA) has been shown to attenuate TGF-β1-induced murine fibroblast activation, as evidenced by a significant reduction

in α-SMA and pro-collagen at the protein level (301). Stimulation of human dermal fibroblasts with galectin-1 similarly increases the number of α-SMA positive myofibroblasts in culture and has an additive effect in combination with TGF-β1 (308). This effect was carbohydrate-dependent with an amino acid substitution (E71Q) in the galectin-1 CRD (subsites C-D) reducing its stimulatory activity (232, 308). Galectin-1 stimulated ECM production by these cells demonstrated by an increase in fibronectin staining and was reported to exhibit TGF-β1-like activity (308).

The potential role of galectin-1 in fibrosis has been investigated in the context of hypoxia-induced pulmonary fibrosis. As previously described (Section 1.2.2) abnormal alveolar repair following injury can typically result in alveolar collapse and this can cause local hypoxia as a secondary physiological insult (309). Using a novel bleomycin/hypoxia model it was shown that hypoxia as a secondary insult exacerbates the fibrotic response to bleomycin *in vivo* (302). Myofibroblast differentiation and collagen deposition is increased in the lung of bleomycin/hypoxia treated mice compared to bleomycin alone and this coincides with increased galectin-1 protein expression (302). Galectin-1 inhibition (OTX008) significantly reduced collagen context in the lung as determined by sircol assay, decreased airspace closure and improved lung function in these mice (302). Therefore suggesting that hypoxia signaling and potentially hypoxic lung injury exacerbates pulmonary fibrosis and lung remodeling via galectin-1.

Cellular adaption to hypoxia is mediated by oxygen-dependent regulation of the HIF-1 transcription factor (HIF-1α and HIF-1β heterodimer). HIF-1α levels are regulated

by oxygen-dependent degradation via the ubiquitin-dependent proteasome pathway. Consequently, under hypoxic conditions HIF-1α accumulates and can translocate into the nucleus to induce transcription of target genes. As expected, HIF-1α protein expression increases in lung epithelial cells (H441, AT2 cell-like) under hypoxic conditions leading to the expression of profibrotic genes (302). Hypoxia treatment of H441 cells also significantly increases the activation of focal adhesion kinase 1 (FAK1), a focal adhesion-associated kinase involved in cytoskeletal remodeling and myofibroblast differentiation (302). Increased FAK1 activity has been shown to induce galectin-1 expression at both the mRNA and protein level in a SMAD3-dependent manner and proximity ligation assay (PLA) has demonstrated galectin-1 and FAK1 colocalisation in H441 cells (302). Galectin-1 stimulation has also been shown to increase FAK1 activation *in vitro* (302). Therefore, a positive feedback loop may exist between FAK1 and galectin-1 that reciprocally enhances their activities under hypoxic conditions.

1.4.6. The Role of Galectin-3 in TGF-β1 Signaling

Unpublished data from Prof Gisli Jenkins group has demonstrated that TGF-β1 stimulation of IPF HLFs for 24 hours significantly increased the mRNA expression of galectin-3 compared to untreated cells. This response was not detected following stimulation of non-IPF HLFs or immortalised human bronchial epithelial cells (iHBECs) highlighting the potential importance of the relationship between galectin-3 and TGF-β1 signaling in fibroblasts derived from fibrotic lung tissue. Interestingly, TGF-β1 treatment did not induce galectin-3 secretion by HLFs (IPF or non-IPF) or iHBECs.

Further investigation of the role of galectin-3 in TGF-β1 signaling in fibroblasts and epithelial cells showed that the effects were cell type specific. Stimulating fibroblasts or epithelial cells with TGF-β1 induces canonical TGF-β1 signaling as demonstrated by increased SMAD2 phosphorylation. However, only in fibroblasts, both human and mouse embryonic derived cells, was it possible to block this increase in SMAD signaling with the galectin-3 inhibitor GB0139. HLFs stimulated with galectin-3 had a significant increase in TGF-β1 SMAD signaling, on par with cells stimulated with TGF-β1, and this increase was inhibited by GB0139. However galectin-3 stimulation had no effect on TGF-β1 canonical signaling in iHBECs, possibly due to the high expression of the galectin-3 receptor CD98 which may bind to galectin-3 and act as a decoy receptor.

LPA-induced TGF-β1 activation in HLFs is an integrin-dependent process, and incubation with a β1 small molecule inhibitor or anti-αvβ5 antibody blocks canonical TGF-β1 signaling. The discovery that LPA-mediated TGF-β1 activation was completely inhibited by galectin-3 inhibitors suggests that endogenous galectin-3 is essential for LPA-induced TGF-β1 activation. More specifically when comparing the effects of a cell permeable and cell-impermeable galectin-3 inhibitor, inhibition of extracellular galectin-3 was shown to reduce SMAD2 phosphorylation in a concentration-dependent manner. Some of this data has been previously reported in the form of abstracts (310-312).

1.5. Project Aim and Hypothesis

Small molecule galectin-3 inhibitors can inhibit both LPA-induced integrin-mediated TGF-β1 activation and TGF-β1-induced or galectin-3 induced SMAD signaling in HLFs, this suggests that galectin-3 is directly involved in TGF-β1 activation and signaling via its CRD. The overall aim of this thesis was to define how galectin-3 promotes TGF-β1 activation and signaling in lung fibroblasts. This was achieved by investigating whether galectin-3 physically interacts with αv integrins, components of the large latent complex and the TGF-β1 receptor.

It was hypothesised that galectin-3 may directly bind to αv integrins or components of the large latent complex and/or bind to the TGF-β1 receptor to promote TGF-β1 signaling in fibroblasts (Figure 1.12).

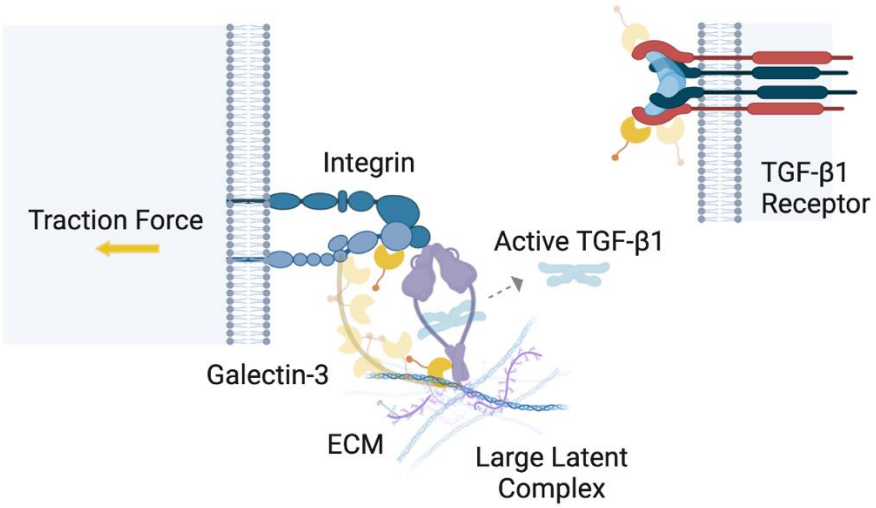


Figure 1.12: Proposed view of galectin-3 involvement in IPF development
Schematic diagram illustrating how galectin-3 may potentially modulate traction-dependent TGF-β1 activation and subsequently IPF development. Galectin-3 may bind to cell surface integrins or the large latent complex, or perhaps both and form a lattice between the two so that less traction force is required for TGF-β1 activation. Galectin-3 may potentially be involved in TGF-β1 signaling directly at the receptor level. Figure adapted from Parmar N. PhD Thesis (2019) with permission (313). (Created in BioRender.com)

2. Materials and Methods

2.1. Materials

Bacterial Cell Culture

Competent *E. coli* Top10 (IBA, 5-1600-020)
Lemo21 (DE3) competent E. coli (New England Biolabs, C2528J)
Luria Broth Granulated [Miller's LB Broth] (Melford, L24400)
Terrific Broth, Modified, Granulated (Melford, T15100)
Luria Agar Granulated [Miller's LB Agar] (Melford, L24022)
Carbenicillin (Disodium Salt) (Fisher Bioreagents, BP26485)
Chloramphenicol (Sigma-Aldrich, C0378)

Recombinant DNA Technology

Clonal Genes (Twist Bioscience, USA)
10mM Tris-Cl, pH 8.5 (Qiagen, 1014612)
Phusion High-Fidelity DNA Polymerase (Thermo Scientific, F530S)
dNTP Mix (Promega, U151A)
Primers (Sigma-Aldrich)
PureLink Quick Gel Extraction and PCR Purification Combo Kit (Invitrogen, K220001)
DpnI (New England Biolabs, R0176S)
CutSmart Buffer (10X) (New England Biolabs, B7204S)
ClonExpress II One Step Cloning Kit (Vazyme, C112)
Linearised Vectors (OPPF, RCaH)
QIAprep Spin Miniprep Kit (Qiagen, 27106)
Qiagen Plasmid Plus Midi Kit (Qiagen, 12943)
Qiagen Plasmid Giga Kit (Qiagen, 12191)

Recombinant Protein Expression & Purification

IPTG (isopropyl-β-D-thiogalactopyranoside) (VWR, 437145X)
HEPES (Sigma-Aldrich, H3375)
Sodium Chloride (VWR, 27800.360)
cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche, 5056489001)
Lysozyme from chicken egg white (Sigma-Aldrich, 62971)
Deoxyribonuclease I from bovine pancreas (Sigma-Aldrich, DN25)
HisTrap High Performance (Cytiva, 17-5248-01)
Imidazole (Sigma-Aldrich, I2399)
Amicon Ultra-15 Centrifugal Filter Unit (Millipore, UFC901024)
Spectra/Por 3 Dialysis Tubing 3.5 kD 45 mm 50ft (Repligen, 132724)
TCEP-HCl (Tris(2-carboxyethyl)phosphine hydrochloride) ≥99% (Strem Chemicals, 15-7400)
HRV 3C Protease (OPPF, RCaH)
Superdex 75 10/300 GL (Cytiva, 17517401)
Glycerol (Sigma-Aldrich, G7757)
Expi293 Expression Medium (Gibco, A1435101)
Trypan Blue Stain (0.4%) for use with the Countess Automated Cell Counter (Invitrogen, T10282)
Opti-MEM Reduced Serum Medium (Gibco, 31985070)
PEI-MAX (Polysciences, 24765)
Valproic acid sodium salt (Sigma-Aldrich, P4543)
Sodium propionate (Sigma-Aldrich, P1880)

D-(+)-Glucose solution (Sigma-Aldrich, G8769)
Detergents (OPPF, RCaH)
TALON Metal Affinity Resin (TakaRa, 635504)
L-Arginine monohydrochloride (Sigma-Aldrich, A5131)
L-Glutamic acid monosodium salt hydrate (Sigma-Aldrich, G1626)
Superdex 200 10/300 GL (Cytiva,17517501)
Superdex 75 Increase 10/300 GL (Cytiva, 29148721)

Mammalian Cell Culture

Dulbecco's Modified Eagle's Medium (DMEM) - high glucose (Sigma-Aldrich, D5671)
Foetal Bovine Serum, qualified (Gibco, 26140079)
L-Glutamine solution (Sigma-Aldrich, G7513)
Penicillin-Streptomycin (Sigma-Aldrich, P4458)
293T (ATCC, CRL-3216)
DMEM, high glucose, HEPES, no phenol red (Gibco, 21063029)
PBS, pH 7.4 (Gibco, 10010023)
Trypsin-EDTA solution (Sigma-Aldrich, T4049)
Trypan Blue solution (Sigma-Aldrich, 93595)
FuGENE-HD Transfection Reagent (Promega, E2311)
SB-525334 (Alk5 inhibitor) (Sigma-Aldrich, S8822)
Small molecule galectin inhibitors (Galecto Biotech, SBC)
Dimethyl sulfoxide (Sigma-Aldrich, D5879)
Oleoyl-L-α-lysophosphatidic acid sodium salt (Sigma-Aldrich, L7260)

Gel Electrophoresis

Agarose, DNA grade (100 bp - 23 kb) (VWR, 438792U)
TAE Buffer (Tris-acetate-EDTA) (50X) (Thermo Scientific, B49)
SYBR Safe DNA Gel Stain (Invitrogen, S33102)
Invitrogen E-Gel General Purpose Agarose Gels, 1.2% (Invitrogen, G501801)
1Kb Plus DNA ladder (Invitrogen, 10787018)
Gel Loading Dye, Purple (6X) (New England Biolabs, B7024S)
10% Mini-PROTEAN TGX Precast Protein Gels, 15-well, 15 µL (Bio-Rad, 4561036)
PageRuler Prestained Protein Ladder, 10 to 180 kDa (Thermo Scientific, 26616)
InstantBlue Coomassie Protein Stain (ISB1L) (Abcam, ab119211)
BenchMark Fluorescent Protein Standard (Invitrogen, LC5928)
NuPAGE LDS Sample Buffer (4X) (Invitrogen, NP0007)
Sodium dodecyl sulfate (Sigma-Aldrich, 436143)
Trizma base (Sigma-Aldrich, T1503)
Glycine (Sigma-Aldrich, G8898)
4X Bolt LDS Sample Buffer (Invitrogen, B0007)
10X Bolt Sample Reducing Agent (Invitrogen, B0009)
Amersham ECL Full-Range Rainbow (Cytiva, RPN800E)
PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa (ThermoFisher Scientific, 26619)
Bolt 4 to 12%, Bis-Tris, 1.0 mm, Mini Protein Gels (Invitrogen, NW04120BOX)
20X Bolt MOPS SDS Running Buffer (Invitrogen, B0001)

Western Blot

Pierce BCA Protein Assay Kit (Thermo Scientific, 23225)
iBlot 2 Transfer Stacks, PVDF, mini (Invitrogen, IB24002)
ATX Ponceau S red staining solution (Sigma-Aldrich, 09189)

Blotto Immunoanalytical Grade (Non-Fat Dry Milk) (Rockland Immunochemicals, B501-0500)
Phosphate buffered saline (Sigma-Aldrich, P4417)
Tween 20 (Sigma-Aldrich, P9416)
SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific, 34095)

Recombinant Proteins

Recombinant Human Galectin-1 (R&D Systems, 1152-GA)
Recombinant Human Galectin-3 (R&D Systems, 1154-GA)
Recombinant Human TGF-beta 1 Protein (R&D Systems, 240-B)
Recombinant human TGF beta Receptor I protein (Abcam, ab105908)
Recombinant Human TGF beta Receptor II protein (ab191920)
Recombinant Human LAP (TGF-beta 1) Protein (R&D Systems, 246-LP)
Recombinant Human LTBP1 protein (Abcam, ab114738)
Recombinant Human Integrin alpha V beta 1 Protein, CF (R&D Systems, 6579-AVB)
Recombinant Human Integrin alpha V beta 5 Protein, CF (R&D Systems, 2528-AV)
Recombinant Human Integrin alpha V beta 6 Protein, CF (R&D Systems, 3817-AV)
Recombinant Human Integrin alpha V beta 8 Protein, CF (R&D Systems, 4135-AV)
Recombinant Human Integrin alpha V beta 3 Protein, CF (R&D Systems, 3050-AV)
Recombinant Human Integrin alpha 2b beta 3 Protein, CF (R&D Systems, 7148-A2)
Recombinant Human Integrin alpha 5 beta 1 Protein, CF (R&D Systems, 3230-A5)
SARS-CoV-2 (2019-nCoV) Spike S1-mFc Recombinant Protein (Sino Biological, 40591-V05H1)
Recombinant Human ACE2 protein (Abcam, ab151852)

Antibodies

Goat anti-Mouse IgG Fc Cross-Adsorbed Secondary Antibody, HRP (Invitrogen, 31439)
Mouse anti-Human IgG1 Fc Secondary Antibody, HRP (Invitrogen, A-10648)
Anti-Integrin beta 1 antibody [12G10] (Abcam, ab30394)
Goat anti-Mouse IgG1 Cross-Adsorbed Secondary Antibody, Alexa Fluor 568 (Invitrogen, A-21124)
Negative control Mouse IgG1 (Dako, X0931)
Purified Rabbit IgG Isotype Standard (BD Biosciences, 550875)
Anti-Integrin beta 1 antibody [P5D2] (Abcam, ab24693)
Galectin 3 Polyclonal Antibody (Invitrogen, PA5-34819)
Human Integrin beta 1/CD29 Antibody (R&D Systems, MAB17781)
Mouse IgG1 Isotype Control (R&D Systems, MAB002)
Recombinant Anti-Galectin 3 antibody [EPR19244] (Abcam, ab209344)
Recombinant Rabbit IgG, monoclonal [EPR25A] (Abcam, ab172730)
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, HRP (Invitrogen, G-21040)

Immunocytochemistry

Phosphate buffered saline (Sigma-Aldrich, P4417)
Bovine Serum Albumin (Sigma-Aldrich, A7030)
Pierce 16% Formaldehyde (w/v), Methanol-free (Thermo Scientific, 28908)
Triton X-100 (Sigma-Aldrich, T8787)
DAPI (Roche, 10236276001)
Anti-Fade Fluorescence Mounting Medium - Aqueous, Fluoroshield (Abcam, ab104135)

Proximity Ligation Assay

Duolink In Situ Detection Reagents Red (Sigma-Aldrich, DUO92008)
Duolink InSitu Wash Buffers, Fluorescence (Sigma-Aldrich, DUO82049)

Duolink In Situ PLA Probe Anti-Mouse MINUS (Sigma-Aldrich, DUO92004)
Duolink In Situ PLA Probe Anti-Rabbit PLUS (Sigma-Aldrich, DUO92002)
4',6-diamidino-2-phenylindole (Sigma-Aldrich, 32670)
Fluoroshield (Sigma-Aldrich, F6182)
Fetal Bovin Serum (Gibco, A4766801)
Dimethyl sulfoxide (Sigma-Aldrich, D2650)
Paraformaldehyde (Sigma-Aldrich, P6148)
Phosphate buffered saline (Sigma-Aldrich, P4417)
Bovine Serum Albumin (Sigma-Aldrich, A7030)
Triton X-100 (Sigma-Aldrich, T8787)

Solid-Phase Binding Assay

Tris(Hydroxymethyl) Methylamine, Biological Buffer (Fisher Chemical, 10355910)
3KG Sodium Chloride, Certified AR for Analysis (Fisher Chemical, 10326390)
Manganese (II) chloride tetrahydrate (Sigma-Aldrich, M3634)
Bovine Serum Albumin (Sigma-Aldrich, A7030)
PBS (Sigma-Aldrich, P4417)
TMB Substrate Reagent Set (BD Biosciences, 555214)
Sulfuric acid (Sigma-Aldrich, 339741)
0.5M EDTA (Active Motif, 101665)

Co-Immunoprecipitation

Universal Magnetic Co-IP Kit (Active Motif, 54002)
Pierce Crosslink Immunoprecipitation Kit (Thermo Scientific, 26147)

Surface Plasmon Resonance

Series S Sensor Chip CM5 (Cytiva, BR100530)
Amine Coupling Kit (Cytiva, BR100050)
Acetate 4.0 (Cytiva, BR100349)
Acetate 4.5 (Cytiva, BR100350)
Acetate 5.0 (Cytiva, BR100351)
Slide-A-Lyzer MINI Dialysis Devices, 10K MWCO (Thermo Scientific, 69570)
HBS-P+ Buffer 10× (Cytiva, BR100671)
Manganese (II) chloride tetrahydrate (Sigma-Aldrich, M3634)
BIAnormalizing solution (70%) (Cytiva, 29207950)
Protein Deglycosylation Mix II (New England Biolabs, P6044)
PNGase F (New England Biolabs, P0704)

2.2. Methods

As this PhD project has spanned four different institutes: University of Nottingham, Research Complex at Harwell (RCaH), Imperial College London (ICL) and Galecto Biotech, the methods used throughout this thesis are chapter specific.

3. Method development: Recombinant protein expression and purification

3.1. Introduction

In order to investigate whether galectin-3 and/or galectin-1 physically interact with αv integrins, components of the large latent complex or the TGF-β1 receptor to regulate TGF-β1 activation and signaling in fibroblasts, it was determined that the most effective approach to investigating the physical interactions would be to utilise the range of structural and biophysical methods available at the RCaH. Analysis of the current literature and the data sheets for commercially available proteins of interest suggested that to improve the chances of success of the project, it would be necessary to express and purify the proteins in-house for reasons outlined below.

Firstly, in comparing the amino acid sequence of commercially available recombinant proteins for the binding partners of interest to the deposited protein sequence on UniProt it was apparent that most of the commercially available proteins were truncations and not full length proteins. As shown (Table 3.1), the recombinant protein sequences are not the complete protein amino acid sequences. As a consequence, the potential galectin binding site may not have been encoded in these truncated recombinant proteins or might not be present due to improper protein folding. In addition, some of the recombinant proteins available had an uncleavable fusion/affinity tag which can affect a proteins binding capability depending on its size and location (314). Secondly, the commercially available proteins were only available at a low concentration and quantity. As many of the biophysical methods used at the RCaH to assess protein interactions e.g. isothermal titration calorimetry (ITC)

stopped-flow, analytical ultracentrifugation (AUC) and surface plasmon resonance (SPR) require a large quantity of protein, purchasing all of the binding partners of interest to screen for potential protein interactions was not feasible in terms of cost.

Table 3.1: Recombinant human proteins commercially available for potential binding partners

UniProt Protein Sequence	Recombinant Protein Sequence	Supplier	Expression System	Protein Tag
αvβ1 (αv: 1048 aa, β1: 798 aa)	αv: F31 - V992 β1: Q21 - D728	R&D Systems (6597-AVB) 50 µg/ £430	CHO	αv: 6His
αvβ5 (αv: 1048 aa, β5: 799 aa)	αv: F31 - V992 β5: G24 - N719	R&D Systems (2528-AV) 50 µg/ £335	CHO	αv: 6His
αvβ6 (αv: 1048 aa, β6: 788 aa)	αv: F31 - V992 β6: G22 - N707	R&D Systems (3817-AV) 50 µg/ £337	CHO	-
LAP (390 aa)	L30 - R278	R&D Systems (246-LP) 25 µg/ £372	*S. frugiperda*	-
LTBP1 (1721 aa)	C729 - S826	Abcam (ab114738) 10 µg/ £565	Wheat germ	GST tag
TGFβRI (503 aa)	R80 - M503	Abcam (ab105908) 10 µg/ £360	Sf9 cells	GST tag
TGFβRII (566 aa)	T23 - D159	Abcam (ab191920) 10 µg/ £200	HEK293 cells	Fc tag

N.B. proteins listed were the most appropriate from those available in 2019, this is subject to change.

The final consideration when opting to invest time in expressing and purifying the binding partners in-house was that mammalian cells are the most native environment for proteins of human origin. Recombinant proteins generated in a mammalian expression system are synthesised with the PTMs to facilitate native protein folding and optimal bioactivity, which is highly desirable for protein interaction studies (315). As previously described (Section 1.4.1) galectins are a family of β-galactoside binding proteins that preferentially bind complex N-linked glycans (229-231, 236). Therefore, in order to investigate galectin interactions it was essential that any potential binding partners were produced in mammalian cells as it is the only expression system

capable of complex N-glycan synthesis (316). In comparison, insect cells primarily synthesise simple to intermediate N-glycans, in yeast N-glycans are predominantly of the high-mannose type whereas no glycosylation is observed if recombinant proteins are expressed in bacterial (*Escherichia coli* (*E. coli*)) or cell-free expression systems (316-319). As highlighted (Table 3.1), the commercially available integrins (αvβ1, αvβ5 and αvβ6) and TGFβRII were produced by a mammalian expression system, but LAP, LTBP1 and TGFβRI were only available from insect cell expression or a cell-free system.

There are several significant disadvantages to utilising a mammalian expression system. Growth is slow, with complex culture conditions that are costly when compared to alternative expression systems (bacterial, yeast, insect cells) (320). Chinese hamster ovary (CHO) and human embryonic kidney 293 (HEK293) cells are the most commonly used cell lines for mammalian protein expression but mammalian expression systems are also limited by the often low-moderate protein yield (320, 321).

In contrast to the binding partner proteins discussed above, human galectins are not glycoproteins and therefore do not require a mammalian expression system (245). They can instead be expressed in bacteria (*E. coli*) in which growth is rapid and of lower cost due to the simple culture conditions required (320, 322). Additionally, a high yield of recombinant protein can be obtained from the high cell density bacterial expression system (323). Although the commercially available galectins were full length proteins and produced by *E. coli* expression, they were of a low concentration

and quantity (Table 3.2). As mentioned above, many of the biophysical methods used at the RCaH to assess protein interactions require a large protein quantity and expectedly out of all the proteins, galectin-3 and galectin-1 were required in the largest quantity to screen all 7 protein interactions. The number of vials of commercially available recombinant galectin required to reach the protein concentration necessary for binding experiments was subsequently not feasible both in terms of cost and in the potential impact that batch-to-batch variation might have on the experiments. It was therefore necessary to generate an in-house galectin protein stock of high concentration.

Table 3.2: Recombinant human galectin proteins commercially available

Protein	Supplier	Sequence	Expression	Tag
Galectin-3 (250 aa)	R&D Systems (1154-GA) 50 µg/ £321	A2 - I250	*E. coli*	-
Galectin-1 (135 aa)	R&D Systems (1152-GA) 50 µg/ £244	A2 - D135	*E. coli*	-

N.B. proteins listed were the most appropriate from those available in 2019, this is subject to change.

Taking all of the above into account, recombinant protein production was performed at the RCaH for all proteins of interest. Selecting the correct expression system was critical in obtaining functional proteins for downstream binding experiments.

3.2. Aims

The overall aim of this chapter was to generate sufficient quantities of high quality protein to facilitate protein interaction studies. Subsequently, the aims of this chapter are as follows:

- To express and purify all potential binding partners of interest in a mammalian expression system
- To express and purify galectin-3 and galectin-1 in a bacterial expression system
- Perform protein characterisation techniques to assess the quality of purified proteins before future use

3.3. Methods

3.3.1. Bacterial Cultures

3.3.1.1. Bacterial Strains

Two different strains of competent *E. coli* cells namely Top10 cells and Lemo21 (DE3) cells were used for cloning and protein expression, respectively (Table 3.3).

Table 3.3: E. coli strains utilised for the production of recombinant human proteins

Strain	Genotype
Top10	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G
Lemo21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo(CamR) λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo = pACYC184-PrhaBAD-lysY

3.3.1.2. Bacterial Culture Medium

Luria-Bertani (LB) media was prepared by dissolving 25 g of luria broth granulated in 1L of ddH_20. Media was then sterilised by autoclaving at 121^oC for 15 minutes and stored at room temperature (used within 24 hours).

Terrific Broth (TB) media was prepared by dissolving 47.6 g of terrific broth granulated in 1L of ddH_20 and 4 mL glycerol. Media was then sterilised by autoclaving at 121^oC for 15 minutes and stored at room temperature (used within 24 hours).

LB agar was prepared by dissolving 37 g of luria agar granulated in 1L of ddH_20 and pH adjusted to 7.2. Media was then sterilised by autoclaving at 121^oC for 15 minutes and stored at room temperature.

3.3.2. Recombinant DNA Technology

3.3.2.1. Isolation of DNA (Gene)

Principle of the method: To clone genes of interest into expression vectors for protein expression and purification, the Twist Bioscience 'clonal gene' service was used to manufacture synthetic deoxyribonucleic acid (DNA) from gene sequences. The company uses a silicon-based DNA synthesis platform to deliver next-generation sequencing (NGS)-verified custom genes in a Twist cloning vector. The DNA can then be amplified by polymerase chain reaction (PCR) using primers that have ~20-30 nucleotides homologous to the gene and a ~15 nucleotide extension homologous to the linearised vector for cloning. There are 3 general steps to PCR: denaturation, annealing and extension. Firstly, the gene DNA template is heated to 94-98^{o}C which disrupts the hydrogen bonds between the double-stranded DNA (dsDNA) base pairs resulting in single-stranded DNA (ssDNA). Next, the reaction is heated to 50-70^{o}C which allows the PCR primers to anneal to the complementary ssDNA sequence for new strand synthesis. Lastly, the reaction is heated to 72^{o}C whereby DNA polymerase (magnesium-dependent) extends the primer sequence by adding complimentary nucleotides, it has 5'-3' DNA polymerase activity and 3'-5' exonuclease activity. For amplification of guanine-cytosine (GC)-rich regions, GC buffer was used (over high-fidelity (HF) buffer) as this improves the performance of DNA polymerase, although there is a higher error rate. Dimethyl sulfoxide (DMSO) was also added to the PCR reaction to reduced secondary structures which could inhibit polymerase progression.

Clonal gene synthesis: Genes of interest were synthesised by Twist Bioscience, USA. Gene sequences were obtained from UniProt and delivered as plasmid DNA in a pTwist amp high copy vector (pUC origin of replication derived from pMB1 plasmid, ampicillin resistance). Lyophilised Twist Bioscience DNA templates were resuspended in 10 mM tris chloride, pH 8.5 to a stock concentration of 50 ng/µL and stored at 4°C (Table 3.4).

Table 3.4: Genes synthesised and corresponding UniProt accession # assigned to the protein sequence

Gene (Human)	UniProt Accession #
LGALS1	P09382
LGALS3	P17931
ITGAV	P06756
ITGβ1	P05556
ITGβ5	P18084
ITGβ6	P18564
TGFβ1 (LAP)	P01137
LTBP1	Q14766
TGFβRI	P36897
TGFβRII	P37173

Gene insert amplification: To amplify the Twist vector insert sequence, PCR was performed using the Phusion High-Fidelity DNA Polymerase protocol (Thermo Scientific, F530S). A 20 µL reaction mix was prepared by combining the following components according to the manufacturers' recommendations (Table 3.5).

Table 3.5: Components used to prepare PCR reactions

Component	Volume/20 µL Reaction	Final Concentration
5X Phusion Buffer (HF or GC)	4 µL	1X
dNTPs (10 mM)	0.4 µL	200 µM each
DMSO (100%)	0.6 µL	3%
Phusion HF DNA Polymerase	0.2 µL	0.02 U/µL
Forward Primer (10 µM)	1 µL	0.5 µM
Reverse Primer (10 µM)	1 µL	0.5 µM
Template DNA (20 ng/µL)	1 µL	1 ng/µL
H_20	11.8 µL	-

PCR: PCR primers for in-fusion cloning were designed on the Oxford Protein Production Facility (OPPF) OPTIC server by inputting the gene (full length or truncated) and vector sequences. Primer melting temperature (Tm) was calculated on 'OligoCalc' from the primer sequence (324). The server has three Tm calculations: basic, salt-adjusted and nearest neighbor, salt-adjusted Tm calculations were used for PCR annealing. The extension time used was based on 30 seconds per 1000 base pairs but for products less than 1 kb a 45 second extension was used (PCR protocol 1) to ensure full extension. If PCR (or subsequent cloning) was unsuccessful using the salt-adjusted Tm, slowdown PCR was performed to optimise primer annealing (325). The slowdown PCR extension time was based on 15 seconds per 1000 base pairs (PCR protocol 2). Due to the size of LTBP1 it was synthesised as two fragments and recombined by overlap extension PCR (PCR protocol 3) (326).

PCR protocol 1:

Step 1:	98°C 30 seconds
Step 2 (30 cycles):	(Denaturing) 98°C 10 seconds
	(Annealing)__°C 30 seconds (temperature based on primer Tm)
	(Extension) 72°C __ seconds (time based on number of base pairs)
Step 3:	72°C 5 minutes
Step 4:	4°C hold

PCR protocol 2 (Slowdown PCR):

Step 1:	95°C 10 minutes
Step 2 (48 cycles):	(Denaturing) 95°C 30 seconds
	(Annealing) 70°C to 53°C 30 seconds
	(Extension) 72°C 1 minute
Step 3 (15 cycles):	(Denaturing) 95°C 30 seconds (temperature ramp 2.5°C/second)
	(Annealing) 58°C 30 seconds (temperature ramp 1.5°C/second)
	(Extension) 72°C 1 minute (maximum ramp rate)
Step 4:	Hold 4°C

PCR protocol 3 (Overlap Extension PCR):

Principle of the method: Due to the size of LTBP1, the gene was synthesised as two separate DNA fragments by Twist Bioscience. Firstly, the two primary fragments were amplified separately by PCR to generate two PCR products with overlapping sequences. Fragment 1 contained a 21 base pair sequence at the 3' end that overlapped with the 5' end of fragment 2. Following PCR amplification of both fragment 1 and fragment 2, the PCR products were gel extracted and a PCR clean-up performed before proceeding with overlap PCR. This was to purify the DNA and to minimise PCR reagents e.g. primers, deoxyribonucleotide triphosphates (dNTPs) and enzymes from step A (extension PCR) entering into step B (overlap PCR). The PCR clean-up kit uses silica matrices to purify DNA. DNA binds to the silica membrane in the presence of chaotropic salts (gel solubilisation buffer) and contaminants are subsequently removed by washing (ethanol wash buffer). The isolated DNA is then eluted under low salt conditions in alkaline elution buffer (10 mM Tris-HCl, pH 8.5). Next, the full length LTBP1 fusion gene was generated by overlap PCR, as this step is dependent on the overlapping sequences produced in step A (extension PCR) it does not require any primers. However, it is important that the primary fragments are present in equimolar amounts. Finally, the full length LTBP1 fusion product was amplified for cloning using the forward primer of fragment 1 and the reverse primer of fragment 2.

Step A (Extension PCR): PCR protocol 2 (slowdown PCR) was followed and both PCR products electrophoresed on a handcast agarose gel. The LTBP1 PCR products were gel extracted using the PureLink Quick Gel Extraction and PCR Purification Combo Kit

protocol (Invitrogen, K220001). Under ultraviolet (UV) light, the gel region containing the DNA fragment of interest was excised. This was transferred into an eppendorf and weighed. 3 volumes of gel solubilisation buffer (L3) was added to the gel slice. This was incubated at 50^{o}C for 10 minutes and the tube inverted every 3 minutes until the gel slice was completely dissolved. Once dissolved, the tube was incubated for an additional 5 minutes to ensure solubilisation. The soluble PCR product was then loaded onto a clean-up spin column and centrifuged for 1 minute at 13,000 rpm, the flow through was discarded. The column was washed with 500 µL of wash buffer (W1) and centrifuged for 1 minute at 13,000 rpm, the column was then spun again at full speed for 2-3 minutes to remove residual buffer. The column was transferred into a clean eppendorf tube for DNA elution. 50 µL of elution buffer (E5) was incubated on the membrane for 1 minute at room temperature prior to centrifugation for 1 minute at 13,000 rpm. The purified DNA was stored short-term at 4^{o}C. All centrifugation steps were performed on a Eppendorf Centrifuge 5424 R, 5424 rotor. The DNA concentration (ng/µL) in eluted samples was determined using a NanoDrop spectrophotometer (Table 3.6).

Table 3.6: Concentration of recovered DNA and volume of primary fragment required for ~100 ng/kb

Construct	DNA Concentration (ng/µL)	Volume (µL) for Overlap PCR (B)
LTBP1 Fragment 1 (M1-C900)	36.8	7.3 (270/36.8)
LTBP1 Fragment 2 (V1-E881)	38.3	6.8 (260/38.3)

Step B (Overlap PCR): After step A (extension PCR), overlap PCR was performed to generate the full length fusion gene from the two primary fragments (Table 3.7).

Table 3.7: Components used to prepare step B (overlap PCR) PCR reaction

Component	Volume/50µl Reaction	Final Concentration
5X Phusion Buffer (GC)	10 µL	1X
dNTPs (10 mM)	2 µL	400 µM each
DMSO (100%)	1.5 µL	3%
Phusion HF DNA Polymerase	0.5 µL	0.02 U/µL
LTBP1 Fragment 1 (100 ng/1kb)	7.3 µL	-
LTBP1 Fragment 2 (100 ng/1kb)	6.8 µL	-
H_20	21.9 µL	-

Step 1: 94°C 30 seconds
Step 2 (9 cycles): 94°C 15 seconds, 72°C (−0.5°C/cycle) 2 min 20 secs
Step 3 (5 cycles): 94°C 15 seconds, 67.5°C (−0.5°C/cycle) 15 seconds, 72°C 2 min 20 secs
Step 4: 72°C for 2 minutes
Step 5: Hold 10°C

Step C (Final Fusion Gene PCR Amplification): After step B (overlap PCR), final fusion gene PCR amplification was performed to amplify the full length fusion product for cloning (Table 3.8).

Table 3.8: Components used to prepare step C (final fusion PCR amplification) PCR reaction

Component	Volume/50 µL Reaction	Final Concentration
5X Phusion Buffer (GC)	10 µL	1X
dNTPs (10 mM)	2 µL	400 µM each
DMSO (100%)	1.5 µL	3%
Step B PCR reaction	4 µL	-
Phusion HF DNA Polymerase	0.5 µL	0.02 U/µL
LTBP1 Forward Primer (10 µM)	1 µL	0.2 µM
LTBP2 Reverse Primer (10 µM)	1 µL	0.2 µM
H_20	30 µL	-

Step 1: 94°C 30 seconds
Step 2 (17 cycles): 94°C 15 seconds, 67°C (−0.5°C/cycle) 15 seconds, 72°C 2 min 20 secs
Step 3 (23 cycles): 94°C 15 seconds, 59°C 15 seconds, 72°C 2 min 20 secs
Step 4: 72°C 2 minutes
Step 5: Hold 10°C

The PCR product at ~5 kb was gel-extracted from a handcast agarose gel and sent for sequencing. For Sanger sequencing 5 µL of PCR product (10 ng/µL) was required and 5 µL of each primer (3.2 pmol/µL).

All PCR conditions utilised are summarised below (Table 3.9).

Table 3.9: Designed PCR primers (using gene and vector sequences) and corresponding PCR conditions

Target	Vector	Forward Primer Sequence	Reverse Primer Sequence	PCR Protocol No.
Galectin-3 FL (M1-I250)	pOPINF	aagttctgtttcaggggcccgATGGCGGATAATTTTAGCTTACATGAC	atggtctagaaagctttaGATCATTGTATAGCTGGCCGAAGTG	2
	pOPINE	aggagatatacc**atg**GCGGATAATTTTAGCTTACATGACGC	gtgatggtgatgtttGATCATTGTATAGCTGGCCGAAGTG	1 (74°C, 45 seconds)
	pOPINEneo-GFP	aggagatatacc**atg**GCGGATAATTTTAGCTTACATGACGC	cagaacttccagtttGATCATTGTATAGCTGGCCGAAGTG	1 (74°C, 45 seconds)
	pOPINE-mCherry	aggagatatacc**atg**GCGGATAATTTTAGCTTACATGACGC	gtgatggtgatgtttGATCATTGTATAGCTGGCCGAAGTG	1 (64°C, 1 minute)
Galectin-1 FL (M1-D135)	pOPINF	aagttctgtttcaggggcccgATGGCGTGTGGCTTAGTAGCG	atggtctagaaagctttaATCGAACGCAACACATTTAATTTTGAAG	2
	pOPINE	aggagatatacc**atg**GCGTGTGGCTTAGTAGCGTCCAAC	gtgatggtgatgtttATCGAACGCAACACATTTAATTTTGAAG	1 (74°C, 45 seconds)
	pOPINEneo-GFP	aggagatatacc**atg**GCGTGTGGCTTAGTAGCGTCCAAC	cagaacttccagtttATCGAACGCAACACATTTAATTTTGAAG	1 (74°C, 45 seconds)
	pOPINE-mCherry	aggagatatacc**atg**GCGTGTGGCTTAGTAGCGTCCAAC	gtgatggtgatgtttATCGAACGCAACACATTTAATTTTGAAG	1 (64°C, 1 minute)
LAP FL (M1-S390)	pOPINE	aggagatatacc**atg**GCCTCCTTCCGGGTTGAGACTTTTA	gtgatggtgatgtttGGAACACTTGCAAGATCTCACAATC	2
	pOPINEneo-GFP	aggagatatacc**atg**GCCTCCTTCCGGGTTGAGACTTTTA	cagaacttccagtttGGAACACTTGCAAGATCTCACAATC	2
LAP TRUNC (L30-R278)	pOPINE	aggagatatacc**atg**TTGTCTACTTGCAAGACCATCGATATG	gtgatggtgatgtttGCGCCTATGGCGTGAGGACTGC	2
	pOPINEneo-GFP	aggagatatacc**atg**TTGTCTACTTGCAAGACCATCGATATG	cagaacttccagtttGCGCCTATGGCGTGAGGACTGC	2
LAP TRUNC (L30-A350)	pOPINE	aggagatatacc**atg**TTGTCTACTTGCAAGACCATCGATATG	gtgatggtgatgtttAGCCCCAGGGTTATGCTGATTATAG	2
LTBP1 Fragment 1 (M1-C900)	pOPINE	aggagatatacc**atg**GCTGGAGCATGGCTTCGGTG	TTCTGTGACCTGCGTGGGGGCGA**TCACCTGACTCGCGGAGCTGGT**	3
LTBP1 Fragment 2 (V841-E1721)	pOPINE	**ACCAGCTCCGCGAGTCAGGTG**ATCGCCCCCACGCAGGTCACAGAA	gtgatggtgatgtttCTCTAAATCGCTATCCTTTTCCAAGTTAAGGGC	3
αV FL (M1-T1048)	pOPINE	aggagatatacc**atg**GCCTTCCCACCACGGAGAAGAC	gtgatggtgatgtttAGTTTCAGAGTTCCCCTCCCCGT	2
	pOPINEneo-GFP	aggagatatacc**atg**GCCTTCCCACCACGGAGAAGAC	cagaacttccagtttAGTTTCAGAGTTCCCCTCCCCGT	2
αV TRUNC (F31-V992)	pOPINE	aggagatatacc**atg**TTTAATCTTGATGTTGACTCCCCAGC	gtgatggtgatgtttCACAGGCATTGGAGCAGGTTGAATC	1 (66°C, 1 minute 30 seconds)
	pOPINEneo-GFP	aggagatatacc**atg**TTTAATCTTGATGTTGACTCCCCAGC	cagaacttccagtttCACAGGCATTGGAGCAGGTTGAATC	2
β1 FL (M1-K798)	pOPINE	aggagatatacc**atg**AACCTTCAGCCTATCTTTTGGATCG	gtgatggtgatgtttTTGCCCTCGTACTTAGGATTGACC	2
	pOPINEneo-GFP	aggagatatacc**atg**AACCTTCAGCCTATCTTTTGGATCG	cagaacttccagtttTTGCCCTCGTACTTAGGATTGACC	2
β1 TRUNC (Q21-D728)	pOPINE	aggagatatacc**atg**CAGACCGATGAAAACAGATGTCTTAAG	gtgatggtgatgtttGTCCGGGCCAGTTGGGCATTCG	1 (66°C, 1 minute 30 seconds)
	pOPINEneo-GFP	aggagatatacc**atg**CAGACCGATGAAAACAGATGTCTTAAG	cagaacttccagtttGTCCGGGCCAGTTGGGCATTCG	1 (66°C, 1 minute 30 seconds)
β5 FL (M1-D799)	pOPINE	aggagatatacc**atg**GCCCCGCGCCCCAGCACCTCT	gtgatggtgatgtttGTCAACTGTCCCATTATAGGACTTG	1 (66°C, 1 minute 30 seconds)
	pOPINEneo-GFP	aggagatatacc**atg**GCCCCGCGCCCCAGCACCTCT	cagaacttccagtttGTCAACTGTCCCATTATAGGACTTG	2
β5 TRUNC (G24-N719)	pOPINE	aggagatatacc**atg**GGTCTCAATATTTGTACCTCCGGG	gtgatggtgatgtttGTTTGGCGTGTTACCGCATTCGG	1 (66°C, 1 minute 30 seconds)
	pOPINEneo-GFP	aggagatatacc**atg**GGTCTCAATATTTGTACCTCCGGG	cagaacttccagtttGTTTGGCGTGTTACCGCATTCGG	1 (66°C, 1 minute 30 seconds)
β6 FL (M1-C788)	pOPINE	aggagatatacc**atg**GGGATTGAGCTGCTCTGTCTCTTT	gtgatggtgatgtttACAATCAGTTGACAGGTCCACTTTC	2
	pOPINEneo-GFP	aggagatatacc**atg**GGGATTGAGCTGCTCTGTCTCTTT	cagaacttccagtttACAATCAGTTGACAGGTCCACTTTC	2
β6 TRUNC (G22-N707)	pOPINE	aggagatatacc**atg**GGATGTGCACTGGGAGGGGGCGG	gtgatggtgatgtttGTTTGGGGGTTTGGGACAGTCTTTC	1 (66°C, 1 minute 30 seconds)
	pOPINEneo-GFP	aggagatatacc**atg**GGATGTGCACTGGGAGGGGGCGG	cagaacttccagtttGTTTGGGGGTTTGGGACAGTCTTTC	2
TGFβR1 FL (M1-M503)	pOPINE	aggagatatacc**atg**GAGGCGGCCGTCGCCGCCC	gtgatggtgatgtttCATCTTAATGCCCTCCTGTTGTGAG	2
	pOPINEneo-GFP	aggagatatacc**atg**GAGGCGGCCGTCGCCGCCC	cagaacttccagtttCATCTTAATGCCCTCCTGTTGTGAG	2
TGFβR1 TRUNC (R80-M503)	pOPINE	aggagatatacc**atg**AGGGACAGGCCCTTCGTGTGTG	gtgatggtgatgtttCATCTTAATGCCCTCCTGTTGTGAG	2
	pOPINEneo-GFP	aggagatatacc**atg**AGGGACAGGCCCTTCGTGTGTG	cagaacttccagtttCATCTTAATGCCCTCCTGTTGTGAG	2
TGFβR2 FL (M1-K567)	pOPINE	aggagatatacc**atg**GGCGAGGACTCCTTCGGGGTC	gtgatggtgatgtttTTGGTGGTATTCAGGCTCCCGTC	2
	pOPINEneo-GFP	aggagatatacc**atg**GGCGAGGACTCCTTCGGGGTC	cagaacttccagtttTTGGTGGTATTCAGGCTCCCGTC	2
TGFβR2 TRUNC (R190-K567)	pOPINE	aggagatatacc**atg**AGAGTAAACAGACAGCAAAAGCTCTC	gtgatggtgatgtttTTGGTGGTATTCAGGCTCCCGTC	2
	pOPINEneo-GFP	aggagatatacc**atg**AGAGTAAACAGACAGCAAAAGCTCTC	cagaacttccagtttTTGGTGGTATTCAGGCTCCCGTC	2

Amplification verification: To determine if the insert sequence had amplified correctly, PCR products were electrophoresed on either a precast or handcast agarose gel. E-gel agarose gels (1.2% with ethidium bromide in-gel dye) were loaded with 2 µL of DNA ladder or PCR product, 4 µL of 6X gel loading dye and made to a 20 µL final volume with H_20. Gels were ran using the E-gel power system on the recommended program 'E-gel 0.8-2%', 26-32 minutes run time. Handcast 1% agarose (1X) tris-acetate-EDTA (TAE) gels, 50 mL or 250 mL, were made with SYBR safe DNA in-gel stain (1X final concentration) and left to set in a gel casting tray for 30 minutes. 50 mL and 250 mL gels were ran in 1X TAE running buffer at 70 V or 150 V, respectively for 1 hour at room temperature.

3.3.2.2. Insertion

Principle of the method: Before inserting the PCR amplified DNA into a vector of choice, template plasmid DNA was removed using DpnI. DpnI is a restriction enzyme that cleaves methylated DNA only, therefore selecting for the newly synthesised unmethylated DNA. Following DpnI digestion, the insert and vector with homologous ends were combined independently of DNA ligase using recombinase Exnase II. This significantly reduces vector self-ligation and improves recombination efficiency.

DpnI: Prior to inserting the amplified DNA product into a vector of choice, DpnI digestion was performed. For each PCR product a 5 µL DpnI reaction mix was prepared by combining the following components according to the manufacturers' recommendations (Table 3.10).

Table 3.10: Components used per DpnI reaction

Component	Volume/5µL Reaction
DpnI	0.25 µL
CutSmart Buffer (10X)	0.5 µL
H_20	4.25 µL

The DpnI reaction was incubated with PCR products at 37°C for 30 minutes in a thermocycler, followed by heat inactivation of DpnI at 85°C for 20 minutes. The DNA concentration (ng/µL) in DpnI digested samples was determined using a NanoDrop spectrophotometer.

Insertion of the amplified product into a vector: The Vazyme ClonExpress II One Step Cloning Kit (Vazyme, C112) was then used to insert the amplified DNA product into a linearised vector of choice. Vectors were designed by OPPF (327) (Table 3.11).

Table 3.11: Summary of the OPPF vector characteristics

Vector	Tag(s)	Cleavage Site	Selection
pOPINF	6His (N-terminal)	3C	Carbenicillin
pOPINE	6His (C-terminal)	-	Carbenicillin
pOPINEneo-GFP	8His (C-terminal), GFP	3C	Carbenicillin
pOPINE-mCherry	6His (C-terminal), mCherry	3C	Carbenicillin

A vector to insertion ratio of 1:2 was used for cloning, calculated as follows:

The optimal mass of vector (ng) = 0.2 x number of base pairs

The optimal mass of insert (ng) = 0.4 x number of base pairs

Recombination reactions were prepared (alongside controls) by combining reaction components according to the manufacturers' recommendations (Table 3.12). Reactions were incubated at 37°C for 30 minutes in a thermocycler and then immediately placed on ice.

Table 3.12: Components used per reaction to insert the amplified DNA product into a linearised vector

Component	Volume/Reaction
Linearized Vector (100 ng/µL)	1 µL
Insert (100 ng/µL)	2 µL
CE II Buffer (5X)	4 µL
Exnase II	2 µL
H_20	11 µL

3.3.2.3. Transformation and Selection

Principle of the method: After cloning, bacterial transformation was performed whereby the plasmid containing the gene of interest is taken up by competent bacterial cells (horizontal gene transfer). Bacterial cells are made permeable by subjecting them to heat shock treatment which creates a transient pore to facilitate plasmid entry (328). The cells are then incubated in media for a short period to recover and repair the damage. After transformation, the bacteria with plasmid uptake (containing an antibiotic resistance gene) are selected for on an antibiotic LB agar plate.

Bacterial transformation and selection: Cloned recombination reactions were used directly for the transformation of Competent *E. coli* Top10 Cells. Top10 cells were used as they allow stable replication of high-copy number plasmids. 1 µL of recombination product (100 ng/µL) was added to 20 µL of competent cells and gently mixed, this was incubated on ice for 30 minutes. Cells were then heat shocked at 42°C for 45 seconds followed by immediately chilling on ice for 2-3 minutes. 250 µL of LB media alone (without antibiotics) was then added to the cells and tubes incubated on a horizontal orbital plate shaker at 37°C, 850 rpm for 1 hour. 150 µL was then pipetted onto a 25 mL LB agar plate containing the selection antibiotic carbenicillin (50 µg/mL

final concentration) and left to air dry. This was scaled accordingly if using a 24-well plate for transformations. Plates were parafilm sealed and incubated upside down at 37°C for 16-18 hours. Following overnight culture several mono-colonies were picked from the transformation plate for colony PCR. Successful recombination was visualised by agarose gel electrophoresis as described (Section 3.3.2.1). For colony PCR the Phusion High-Fidelity DNA Polymerase protocol (F530S, ThermoFisher Scientific) was adapted and the 10 µL PCR product loaded onto a gel (Table 3.13).

Table 3.13: Components used to prepare colony PCR reactions

Component	Volume/10µL Reaction	Final Concentration
5X Phusion Buffer (HF or GC)	2 µL	1X
dNTPs (10 mM)	0.2 µL	200 µM each
DMSO (100%)	0.3 µL	3%
Phusion HF DNA Polymerase	0.1 µL	0.02 U/µL
Forward Primer (10 µM)	0.5 µL	0.5 µM
Reverse Primer (10 µM)	0.5 µL	0.5 µM
H_20	6.4 µL	-
Colony	1	-

3.3.2.4. DNA Analysis

Principle of the method: To extract plasmid DNA from bacterial cells alkaline lysis was used (329). Bacterial cells were lysed in a buffer containing sodium dodecyl sulfate (SDS) which solubilise the cell membrane and NaOH which helps to break down the cell wall. It also contained ethylenediaminetetraacetic acid (EDTA) for deoxyribonuclease (DNases) inhibition and ribonuclease (RNase) for RNA degradation. The lysate pH was then neutralised by the addition of potassium acetate and adjusted to high-salt-binding conditions. Denatured proteins, chromosomal ssDNA and cellular debris were then precipitated away from the soluble plasmid DNA by potassium dodecyl sulphate and centrifugation. The QIAprep Spin Miniprep Kit uses silica matrices to purify DNA. DNA binds to the silica

membrane in the presence of chaotropic salts at a neutral pH. The salts are then removed by washing (ethanol wash buffer) as DNA only elutes in low-salt buffer. The DNA was efficiently eluted under low salt conditions in alkaline elution buffer (10 mM Tris-Cl, pH 8.5).

DNA Isolation: Colonies with successful recombination were cultured overnight in 5 mL LB media with carbenicillin (50 µg/mL final concentration) at 37°C, 200 rpm for subsequent DNA purification. To isolate plasmid DNA the QIAprep Spin Miniprep Kit was used (Qiagen, 27106). Overnight cultures were centrifuged at 4500 rpm for 10 minutes and the supernatant discarded (Beckman coulter Allegra X-22R centrifuge, SX4250 rotor). Pelleted bacterial cells were then resuspended and lysed in buffer P1 (250 µL). An equal volume of buffer P2 was then added (250 µL). This was followed by the addition of buffer N3 (350 µL) and centrifugation at 13,000 rpm for 10 minutes. The supernatant (800 µL) was applied to a QIAprep spin column, the column was then centrifuged at 13,000 rpm for 1 minute and the flow through discarded. Next, the membrane was washed with buffer PE (750 µL) and centrifuged at 13,000 rpm for 1 minute. This centrifugation step was repeated to remove residual wash buffer. The column was then transferred into a clean eppendorf tube for DNA elution. 30 µL of buffer EB was incubated on the membrane at room temperature for 1 minute prior to centrifugation at 13,000 rpm for 1 minute. All centrifugation steps were performed on a Eppendorf Centrifuge 5424 R, 5424 rotor unless otherwise stated. The DNA concentration (ng/µL) in the eluted samples was determined using a NanoDrop spectrophotometer and nucleic acid purity assessed by absorbance readings at 260 nm and 280 nm (A_{260}/A_{280}). For DNA isolation on a larger scale

overnight cultures were scaled up to 50 mL and the Qiagen Plasmid Plus Midi Kit (Qiagen, 12943) or the Qiagen Plasmid Giga Kit (Qiagen, 12191) used.

DNA sanger sequencing: Sequence verification of the purified DNA was performed by Source Bioscience, UK. Results were viewable in SnapGene Viewer and the nucleotide sequencing obtained was translated into the amino acid sequence using Expasy Translate. Verification of final products was performed by comparison to the Twist protein sequences. For Sanger sequencing 5 µL of plasmid DNA (100 ng/µL) was required per forward and reverse read and 5 µL of each primer (3.2 pmol/µL). All OPPF vectors selected for cloning have a T7 forward promoter, therefore the Source Bioscience T7F sequencing primer was used. Designed reverse primers were posted alongside purified DNA samples for reverse readings.

3.3.3. Bacterial Protein Expression and Purification

3.3.3.1. Recombinant Protein Expression (E. coli)

Principle of the method: Regulation of the *E. coli* lac operon gene is well understood and is widely used to control bacterial expression of recombinant proteins. The lac operon encodes 3 genes involved in lactose metabolism (lacZ, lacY and lacA) however its transcription is inhibited by the lac repressor protein (encoded by the upstream lacI gene). In the absence of lactose, the lac repressor binds to the operator which represses the transcription of lac operon genes (basal level transcription). However, when lactose is present it enters the cell and binds to the lac repressor releasing it from the operator. As a result, the T7 RNA polymerase can bind to the T7 promoter and initiate transcription of lac genes. All OPPF vectors selected for cloning have a T7 forward promoter. Isopropyl β- d-1-thiogalactopyranoside (IPTG) is a non-degradable

molecular mimic of allolactose and was therefore used to induce *E. coli* lac operon activity and recombinant protein expression.

Expression in Lemo21 (DE3 Strain): For galectin protein expression, sequence-verified DNA was transformed into Lemo21 (DE3) competent *E. coli* cells as previously described (Section 3.3.2.3). Transformed plates were kept short-term at 4°C. Lemo21 cells were used as protein expression can be controlled by IPTG induction of the T7 RNA polymerase. Additionally, Lemo21 (DE3) cells were used to prevent 'leaky' background/ basal level expression of the target gene. Both carbenicillin and chloramphenicol were used as selection antibiotics. Following transformation, a single colony was cultured overnight in 15 mL LB media (carbenicillin 50 µg/mL and chloramphenicol 35 µg/mL final concentration) at 37°C, 200 rpm. 10 mL of the culture was then added to 1L of TB media containing carbenicillin and chloramphenicol (carbenicillin 50 µg/mL and chloramphenicol 35 µg/mL final concentration). Inoculated cultures were left to grow at 37°C, 210 rpm until an OD_{600nm} of 1.4. Cells were then induced with 1 mL of IPTG (1 mM final concentration) and left to grow overnight at 22.5°C, 210 rpm. Cells were harvested by centrifuging at 5000 rpm for 13 minutes, 4°C (Beckman Coulter Avanti J-26S XPI, JLA8.1 rotor). Supernatants were discarded and bacterial cell pellets stored at -80°C until protein purification.

3.3.3.2. Recombinant Protein Purification (E. coli)

Principle of the method: Nickel (Ni^{2+}) affinity chromatography can be used for the purification of His-tagged recombinant proteins as it has a high binding affinity for histidine residues. HisTrap HP columns are packed with resin containing cross-linked agarose beads (sepharose) coupled to a chelating group that is pre-charged with Ni^{2+}.

This allows His-tagged proteins to be easily purified however, endogenous proteins containing histidine clusters may also bind to the column which reduces its specificity. Imidazole competes with histidine for the metal-charged resin and is therefore added into the binding and wash buffer at a low concentration to minimise non-specific binding of host cell proteins. An imidazole gradient is used to elute the His-tagged recombinant protein from the column, the target protein is eluted at a higher imidazole concentration than contaminates due to the higher binding affinity. OPPF vectors selected for bacterial protein expression and purification have a Human Rhinovirus (HRV) 3C cleavage site which is used to remove the His-tag following the histidine-dependent purification step. The cleaved His-tag can then be separated away from the purified protein based on the same principle explained above. Lastly, size exclusion chromatography (SEC) can be performed to separate the target protein from any residual contaminants, whereby proteins are separated by size by filtration through a porous matrix of beads (larger proteins elute faster). The protein elution time is compared against the column calibration curve to determine MW.

Nickel purification (His-tag): Galectin cell pellets (1L expression) were resuspended in 150 mL lysis buffer; 50 mM HEPES (pH 7.8), 300 mM NaCl, cOmplete EDTA-free protease inhibitor cocktail (1 tablet per 100 mL), 0.1 mg/mL lysozyme and 10 µg/mL Deoxyribonuclease I (DNase1). Resuspended pellets were then filtered with a cell strainer and lysed with two passes at 28 kpsi using a cell disruptor (Constant Systems). The lysate was clarified by centrifuging at 21,000 rpm for 30 minutes, 4^{o}C (Beckman Coulter Avanti J-26S XPI, JA 25.5 rotor). The supernatant containing the soluble galectin was then purified on a 5 mL HisTrap HP column (Ni^{2+} sepharose). The column

was pre-equilibrated with 5 column volumes (CV) (25 mL) of buffer containing 50 mM HEPES (pH 7.8), 300 mM NaCl and 10 mM imidazole. Imidazole was added to the clarified lysate to a final concentration of 10 mM, this was then passed through the column using a peristatic pump set to a 1 mL/min flow rate. Following binding, the column was washed with 5 CV of wash buffer; 50 mM HEPES (pH 7.8), 300 mM NaCl and 20 mM imidazole. The lysate load, flow through and wash were all later subjected to sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Bound protein was eluted on an ÄKTA Explorer (Fast Protein Liquid Chromatography) using an imidazole gradient (50 mM HEPES (pH 7.8), 300 mM NaCl, start concentration: 0 mM, end concentration: 750 mM). Gradient elutions were carried out over 20 CV (100 mL) with 3 mL fractions collected. An elution profile was measured at UV 280 nm absorbance (mAu). Fractions were selected from the elution profile and ran by SDS-PAGE. Coomassie stain was used to visualise the proteins resolved by electrophoresis. Fractions containing the eluted protein were combined, concentrated (MWCO: 10 kDa) and dialysed overnight at 4^{o}C against 2L of 50 mM HEPES (pH 7.8), 300 mM NaCl, 0.3 mM tris(2-carboxyethyl)phosphine (TCEP). The His-tag was cleaved during dialysis by the addition of HRV 3C-protease at a 1:100 ratio 3C:galectin. A reverse immobilised metal affinity chromatography (IMAC) column (5 mL HisTrap HP column) was used to separate the cleaved target protein from the 3C-His tag, uncleaved galectin and 3C protease. The reverse IMAC column was pre-equilibrated with 5 CV (25 mL) of buffer containing 50 mM HEPES (pH 7.8), 300 mM NaCl and 10 mM imidazole. The dialysed protein sample was passed through the reverse IMAC column using a peristatic pump set to a 1 mL/min flow rate and the flow through collected. Following binding, the column was washed with 5 CV of wash

buffer; 50 mM HEPES (pH 7.8), 300 mM NaCl and 20 mM imidazole. Bound material (3C-His tag, protease and non-cleaved material) was eluted from the column with 3 CV of buffer containing 50 mM HEPES (pH 7.8), 300 mM NaCl and 375 mM imidazole. The dialysed protein sample, flow through, wash and elution were all later subjected to SDS-PAGE and Coomassie staining to verify cleavage. Fractions containing cleaved galectin were combined and concentrated (MWCO: 10 kDa) down to 750 µL to be further purified using SEC. A Superdex S75 10/300 column (optimum protein separation range 3 - 70 kDa) was pre-equilibrated with SEC buffer; 25 mM HEPES (pH 7.8) and 500 mM NaCl. The concentrated protein sample was then loaded onto the column and ran at a flow rate of 0.5 mL/min and 0.5 mL fractions collected over a total volume of 1.1 CV (26 mL). A protein elution profile was measured at UV 280 nm absorbance (mAu). Fractions selected from the 280 nm absorbance trace were ran by SDS-PAGE and subjected to Coomassie staining to determine protein size and purity. The quality of the protein was assessed using analytical ultracentrifugation (AUC) (Section 3.3.6.1) and Size Exclusion Chromatography – Multi-Angle Light Scattering (SEC-MALS) (Section 3.3.6.2). 10% glycerol was added to 0.5 mL protein fractions and then snap frozen to be stored at -80^{o}C until required.

3.3.4. Mammalian Protein Expression and Purification

3.3.4.1. Recombinant Protein Expression (Expi293 cells)

Principle of the method: Polyethylenimine (PEI) is a positively charged polymer that has been utilised as a DNA transfection reagent (330). PEImax condenses negatively charged DNA into positively charged particles which can subsequently bind to anionic cell surfaces. The DNA:PEI max complex is subsequently taken up by the cell via endocytosis and the DNA released into the cytoplasm. To enhance gene expression

post-transfection the media can be supplemented with valproic acid and sodium propionate (histone deacetylase inhibitors) and glucose (carbon and energy source) (331).

Expression in Expi293 cells: Expi293 cells were seeded at 4-5x10^5 cells/mL and cultured in expi293 expression media in a humidified incubator set to 37°C with 5% CO_2, 120 rpm. Cells were then seeded in 125 mL flasks at 1-1.2x10^6 cells/mL (in 30 mL of media) for transfection the following day. Transfection was performed with a cell count between 1.5-2x10^6 cells/mL and a viability of at least 95%. For a 30 mL expression, 3 mL of Opti-MEM reduced serum media was mixed with 160 µL of PEI max (40K) and 60 µg of plasmid DNA (2 µg/ml). After a 10-minute incubation at room temperature this was added to the suspension cells and the flask returned back to the incubator. After 7 hours, enhancers were added to the suspension cells: 500 µL valproic acid, 200 µL sodium propionate and 550 µL glucose, and cells then incubated in a humidified incubator set to 30°C with 5% CO_2, 120 rpm for 48 hours. After 48 hours, green fluorescent protein (GFP) fluorescence was assessed under the microscope (EVOS Imager) and cells harvested by centrifugation at 1200 rpm for 10 minutes (VWR Mega Star 1.6 R, TX-400 rotor). Supernatants were discarded and mammalian cell pellets stored at -80°C until protein purification. For mammalian expression on a smaller scale (3 mL) reagent volumes were scaled accordingly.

3.3.4.2. Recombinant Protein Purification (Expi293 cells)

Principle of the method: Cobalt (Co^{2+}) is effective for small-scale purification of His-tagged recombinant proteins (membrane or soluble proteins) as it has a higher specificity but lower affinity than Ni^{2+} (332). TALON resin has a carboxymethyl

aspartate (CMA) matrix pre-charged with Co^{2+}. As there is less non-specific binding of endogenous proteins, less imidazole is required when using TALON resin (Co^{2+}) for protein purification. Protein purification efficiency is affected by buffer pH and salt concentration and both are specific to the protein of interest. Under optimal buffer conditions the protein will be more stable as extreme changes in pH can cause protein inactivation or precipitation. Changes in salt concentration can affect protein structure by influencing the hydrophobic and electrostatic interactions. Surfactants (detergents) are commonly used in membrane protein purification to help dissociate the protein from the cell membrane without affecting protein structure. Detergents interact with the membrane phospholipids to form micelles, the minimal detergent concentration at which this is observed is called the critical micelle concentration (CMC). To ensure micelle formation during purification the detergent is in excess at typically 3 X CMC (333). Glycerol can also be used to aid purification of membrane proteins as it creates a more native environment for the protein (increased hydrophobicity). For soluble proteins, the addition of charged amino acids (L-Arginine (L-Arg) and L-Glutamine (L-Glu)) increases protein stability during purification by reducing protein aggregation and precipitation (334).

Small-scale protein purification: Small-scale mammalian protein purification (3 mL expression) was performed using the Membrane Protein Laboratory (MPL) protocol as outlined below (335). For membrane protein purification buffers included 10% glycerol and/or 3 X CMC detergent to aid solubilisation, this was not required for soluble proteins. Cell pellets stored at -80^{o}C were thawed on ice and each resuspended in 1 mL lysis buffer; 20 mM HEPES (pH 7.8), 500 mM NaCl, 10 mM

imidazole and cOmplete EDTA-free protease inhibitor cocktail (1 tablet per 100 mL) (+/- 10% glycerol). Resuspended pellets were kept on ice and lysed by sonication at 16 kHz frequency for 15 seconds followed by a 15 seconds pause for a 2 minute total duration. 950 µL of lysate was then transferred into 96-well deep well plate. For membrane proteins 95 µL of DDM detergent (10% stock concentration) was added to the lysate (1% final extraction concentration) and incubated on a shaker at 450 rpm for 1 hour, 4^{o}C. For soluble proteins this step was not required. The lysate was then clarified by centrifugation at 3878 rpm for 30 minutes, 4^{o}C and the supernatant transferred into a new 96-well deep well plate. 100 µL of 50% (w/v) TALON resin (Co^{2+}) pre-equilibrated with lysis buffer was added to the supernatant, sealed and incubated on a shaker at 450 rpm for 1 hour, 4^{o}C. After 1 hour the supernatant-TALON suspension was transferred into a 96-well filter plate and centrifuged on top of a 96-well plate to collect the supernatant flow through. 500 µL of wash buffer containing 20 mM HEPES (pH 7.8), 500 mM NaCl and 25 mM imidazole (+/- 10% glycerol, 3 X CMC detergent) was added to the filter plate and centrifuged at 1466 rpm for 3 minutes, 4^{o}C. This wash step was performed 3 times and residual wash buffer removed by repeating the last centrifugation step. After washing, 50 µL of elution buffer; 20 mM HEPES (pH 7.8), 500 mM NaCl and 500 mM imidazole (+/- 10% glycerol, 3 X CMC detergent) was added the plate, sealed and incubated on a shaker at 450 rpm for 20 minutes, 4^{o}C. Purified protein was then eluted into a 96-well plate by centrifuging at 1466 rpm for 3 minutes, 4^{o}C. All centrifugation steps were performed on a Beckman coulter Allegra X-15R centrifuge, SX4750 rotor. The size and purity of the purified protein was assessed by gel electrophoresis (Coomassie and in-gel fluorescence) (Section 3.3.5) and fluorescence-detection size-exclusion

chromatography (FSEC) (Section 3.3.4.3). In-gel fluorescence and FSEC were used to confirm that the purified GFP fusion proteins were of the correct MW and facilitated the identification of free GFP and/or degraded protein. Coomassie staining was used to primarily assess the sample purity.

Detergent screening (membrane proteins): Large-scale mammalian protein purification (30 mL expression) was required for the detergent screening of membrane proteins (MPL protocol) as outlined below (335). This was to determine in which detergent the protein was most stable. Cell pellets stored at -80^{o}C were thawed on ice and resuspended in 24 mL lysis buffer; 20 mM HEPES (pH 7.8), 500 mM NaCl, 10% glycerol, 10 mM imidazole and cOmplete EDTA-free protease inhibitor cocktail (1 tablet per 100 mL). The pellet was kept on ice and lysed by sonication at 16 kHz frequency for 30 seconds followed by a 30 seconds pause for a 2 minute total duration. The lysate was subjected to ultracentrifugation at 40,000 rpm for 1 hour, 4^{o}C to pellet the membrane protein fraction and the supernatant discarded (Beckman Coulter Optima L-100 XP Ultracentrifuge, Type 45 Ti rotor). To determine the optimal detergent conditions, the small-scale protein purification protocol described above was performed with 12 different detergents per pellet (Table 3.14).

Table 3.14: Detergents included in screening and their CMC (Information provided by MPL)

Test No.	Detergent	Detergent Abbreviation	CMC (mM)
1	n-Dodecyl-β-D-Maltopyranoside	DDM	0.17 (0.0087%)
2	n-Dodecyl-β-D-Maltopyranoside + Cholesterol hemisuccinate	DDM + CHS	0.17 (0.0087%)
3	n-Decyl-β-D-Maltopyranoside	DM	1.8 (0.087%)
4	n-Decyl-β-D-Maltopyranoside + Cholesterol hemisuccinate	DM + CHS	1.8 (0.087%)
5	n-Octyl-β-D-Glucopyranoside	OG	18-20 (0.53%)
6	Lauryl Maltose Neopentyl Glycol	LMNG	0.01 (0.001%)
7	Octyl Glucose Neopentyl Glycol + Cholesterol hemisuccinate	OGNG + CHS	1.02 (0.058%)
8	n-Dodecyl-N,N-Dimethylamine-N-Oxide	LDAO	1-2 (0.023%)
9	Octaethylene Glycol Monododecyl Ether	C12E8	0.09 (0.0048%)
10	Polyoxyethylene(9)dodecyl Ether	C12E9	0.05 (0.003%)
11	5-Cyclohexyl-1-Pentyl-β-D-Maltoside	CYMAL-5	2.4 (0.12%)
12	n-Dodecylphosphocholine	Fos Choline-12	1.5 (0.047%)

As this required ~11.4 mL of lysate per protein (950 µL x 12), the membrane pellet was resuspended in 12.5 mL of lysis buffer and 950 µL pipetted into 12 wells across a 96-well deep well plate. 95 µL of each detergent was then added to the 12 different wells and incubated on a shaker at 450 rpm for 1 hour, 4°C. The lysate was clarified by centrifugation at 3878 rpm for 30 minutes, 4°C and the supernatant transferred to a new 96-well deep well plate. 100 µL of 50% (w/v) TALON resin pre-equilibrated with lysis buffer was added to the supernatant, sealed and incubated on a shaker at 450 rpm for 1 hour, 4°C. The remainder of the protocol was as described above, elutions were analysed by gel electrophoresis (Section 3.3.5) and FSEC (Section 3.3.4.3).

Buffer screening (soluble proteins): To determine the optimal buffer conditions for a soluble protein, the small-scale protein purification protocol described above was adapted and performed with 16 different buffer conditions per pellet. As a different lysis buffer was required for each test, the cell pellet from the 30 mL expression was thawed on ice and initially resuspended in 1.7 mL standard lysis buffer; 20 mM HEPES (pH 7.8), 500 mM NaCl, 10 mM imidazole and cOmplete EDTA-free protease inhibitor

cocktail (1 tablet per 100 mL). 100 µL of lysate was pipetted into 16 wells across a 96-well deep well plate and then 850 µL of the test lysis buffer added to individual wells (Table 3.15).

Table 3.15: Composition of lysis buffers included in buffer screening

Test No.	Lysis Buffer	Buffer pH
1	20 mM HEPES, 200 mM NaCl, 10 mM Imidazole	6.5
2		7.5
3		8
4		8.5
5	20 mM HEPES, 300 mM NaCl, 10 mM Imidazole	6.5
6		7.5
7		8
8		8.5
9	20 mM HEPES, 500 mM NaCl, 10 mM Imidazole	6.5
10		7.5
11		8
12		8.5
13	20 mM HEPES, 300 mM NaCl, 10 mM Imidazole 50 mM L-Arg, 50 mM L-Glu	6.5
14		7.5
15		8
16		8.5

Resuspended pellets were kept on ice and lysed by sonication at 16 kHz frequency for 15 seconds followed by a 15 seconds pause for a 2 minute total duration. The lysate was then clarified by centrifugation at 3878 rpm for 30 minutes, 4^{o}C and the supernatant transferred to a new 96-well deep well plate. 100 µL of 50% (w/v) TALON resin pre-equilibrated with standard lysis buffer was added to the supernatant, sealed and incubated on a shaker at 450 rpm for 1 hour, 4^{o}C. After 1 hour the supernatant-TALON suspension was transferred into a 96-well filter plate and centrifuged on top of a 96-well plate to collect the supernatant flow through. 500 µL of the test wash buffer was added to individual wells and the filter plate centrifuged at 1466 rpm for 3 minutes, 4^{o}C (Table 3.16). This wash step was performed 3 times and residual wash buffer removed by repeating the last centrifugation step.

Table 3.16: Composition of wash buffers included in buffer screening

Test No.	Wash Buffer	Buffer pH
1	20 mM HEPES, 200 mM NaCl, 25 mM Imidazole	6.5
2		7.5
3		8
4		8.5
5	20 mM HEPES, 300 mM NaCl, 25 mM Imidazole	6.5
6		7.5
7		8
8		8.5
9	20 mM HEPES, 500 mM NaCl, 25 mM Imidazole	6.5
10		7.5
11		8
12		8.5
13	20 mM HEPES, 300 mM NaCl, 25 mM Imidazole 50 mM L-Arg, 50 mM L-Glu	6.5
14		7.5
15		8
16		8.5

After washing, 50 µL of test elution buffer was added to the plate, which was then sealed and incubated on a shaker at 450 rpm for 20 minutes, 4°C (Table 3.17). Purified protein was then eluted by centrifuging at 1466 rpm for 3 minutes, 4°C. The size and purity of the purified protein was assessed by gel electrophoresis. The most favourable detergent for the protein purification was determined by comparing the results from in-gel fluorescence, Coomassie staining and FSEC analysis.

Table 3.17: Composition of elution buffers included in buffer screening

Test No.	Elution Buffer	Buffer pH
1	20 mM HEPES, 200 mM NaCl, 500 mM Imidazole	6.5
2		7.5
3		8
4		8.5
5	20 mM HEPES, 300 mM NaCl, 500 mM Imidazole	6.5
6		7.5
7		8
8		8.5
9	20 mM HEPES, 500 mM NaCl, 500 mM Imidazole	6.5
10		7.5
11		8
12		8.5
13	20 mM HEPES, 300 mM NaCl, 500 mM Imidazole 50 mM L-Arg, 50 mM L-Glu	6.5
14		7.5
15		8
16		8.5

3.3.4.3. Fluorescence-Detection Size-Exclusion Chromatography

Principle of the method: FSEC is fluorescence detection coupled to SEC and is used to assess GFP-tagged proteins following purification. The proteins are monitored by fluorescence and an elution profile plotted to assess mono-dispersity, protein aggregation and free GFP in the sample. For membrane proteins it can be used to compare solubility of the different detergent-membrane protein complexes and purification efficiency. The same protein samples are ran by NativePAGE to visualise the GFP-tagged protein by in-gel fluorescence followed by Coomassie staining to assess purity.

FSEC: For FSEC analysis of GFP-tagged proteins, 35 µL of the elution was transferred into an autosampler vial. FSEC was performed on a Prominence HPLC system (Shimadzu) operated by Nadisha Gamage (RFI, Harwell Campus) using a Superdex 200 10/300 column pre-equilibrated with FSEC Buffer; 20 mM HEPES (pH 7.8), 500 mM NaCl, 3 X CMC detergent. FSEC was performed at a flow rate of 1 mL/min, over 1 CV (20 mL). A fluorescence profile was measured at 510 nm absorbance.

3.3.5. Protein Gel Electrophoresis

Protocol: Two types of gel electrophoresis were performed: SDS-PAGE and NativePAGE. Mini-PROTEAN TGX precast protein gels were loaded with 1 µL of protein ladder or 10 µL of prepared sample. Protein samples were prepared by mixing 15 µL of sample with 5 µL of 4X lithium dodecyl sulfate (LDS) loading dye. For SDS-PAGE, samples were heated to 95^{o}C for 5 mins and gels ran in 1X tris-glycine SDS running buffer at 300V for 14 minutes (room temperature). For NativePAGE, samples were not heated and gels ran in 1X tris-glycine running buffer at 150V for 1 hour, 4^{o}C.

3.3.6. Biophysical Characterisation

3.3.6.1. Analytical Ultracentrifugation

Principle of the method: AUC combines ultracentrifugation with optical monitoring systems using spectrophotometer (absorbance) and interference optics. It is a solution technique whereby a series of sequential scans are taken of a protein at set intervals to observe how the protein moves through solution in real time. Sedimentation velocity experiments are used to yield information on the sedimentation coefficient, frictional ratio and MW of a protein, which are determined by the protein mass, shape and homogeneity/ assembly state. The sedimentation coefficient is directly proportional to the protein mass (larger proteins have a larger sedimentation coefficient). The frictional ratio is also highly dependent on mass but is inversely proportional to the sedimentation coefficient (proteins with a higher sedimentation coefficient have a smaller frictional ratio). This is described by the Svedberg equation:

$$s = \left(\frac{m\,(1 - \bar{v}\rho_0)}{f}\right)$$

s = sedimentation coefficient
m = mass of the particle
$\bar{v}$ = partial specific volume
ρ_0 = density of the medium
f = frictional coefficient

However the relationship between mass, the sedimentation coefficient and frictional ratio is modified by the shape of the protein (spherical vs elongated shape alters drag force). AUC data can be analysed by continuous c(s) analysis to generate a sedimentation coefficient distribution plot (336).

AUC: AUC sedimentation velocity experiments were performed at 50,000 rpm, 20°C (Beckman Coulter Optima Analytical Ultracentrifuge, An-50Ti rotor). Centrifuge cells had standard two-sector epon centrepieces and sapphire windows. Purified galectin protein was diluted to 0.5, 0.25 and 0.125 mg/mL in SEC buffer (25 mM HEPES (pH 7.8), 500 mM NaCl) for AUC analysis. For each galectin concentration, diluted protein was loaded into one sector of a cell and SEC buffer into the other as a reference (396 µL of sample or 400 µL reference per sector). Data were recorded using the absorbance (280 nm) optical detection system. The density and viscosity of the buffer was measured experimentally using a DMA 5000M densitometer equipped with a Lovis 2000ME viscometer module. The partial specific volume of the protein was calculated using SEDFIT from the amino acid sequence. Data were processed using SEDFIT, fitting to the c(s) model. Figures were made using GUSSI. All AUC experiments were performed with the instrument operator, Dr Gemma Harris (RCaH, Harwell Campus).

3.3.6.2. Size Exclusion Chromatography – Multi-Angle Light Scattering

Principle of the method: SEC is a chromatography technique that separates proteins based on hydrodynamic radius by filtration through a porous matrix of beads (larger proteins elute faster). The elution time of the protein sample can be compared against the column calibration curve to determine the proteins MW from known MW standards. The main disadvantage of this method is that it assumes the protein sample has the same molecular conformation, density and column interactions as the calibration standards, with retention time being solely dependent on MW. Subsequently, SEC-MALS is alternatively used for a more accurate determination of MW independent of retention time. Briefly, a laser is directed at a protein sample in

solution, the light is then scattered by the molecules in solution and the intensity of that scattered light is measured at several different angles. Intensity of the scattered light is proportional to MW x concentration (g/L) x change in refractive index increment with concentration (commonly 0.185 mL/g). Therefore, MW can be determined by measuring the intensity of scattered light by a protein at a specific concentration with a refractive index increment value of 0.185 mL/g. As this is independent of shape, retention time and calibration standards, it gives a more accurate determination of MW and impurities are more easily identifiable.

SEC-MALS: SEC-MALS was performed on an ÄKTA Pure 25 System (Cytiva). The purified galectin protein was diluted to 3 mg/mL in SEC buffer (25 mM HEPES (pH 7.8), 500 mM NaCl). 100 µL was then loaded onto a Superdex 75 10/300 increase column pre-equilibrated with SEC buffer. The protein was eluted with 1 CV (24 mL) of SEC buffer at a flow rate of 0.5 mL/min. The eluting protein was monitored using a DAWN HELEOS-II 18-angle light scattering detector (Wyatt Technologies) equipped with a WyattQELS dynamic light scattering module, a U9-M UV/Vis detector (Cytiva), and an Optilab T-rEX refractive index monitor (Wyatt Technologies). UV absorbance (280 nm), refractive index and light scattering data were analysed using Astra v7 with a refractive index increment value of 0.185 mL/g. All SEC-MALS experiments were performed with the instrument operator, Dr Gemma Harris (RCaH, Harwell Campus).

3.4. Results

3.4.1. Cloning and transformation of target proteins

Full length and/or truncated constructs were designed for all proteins of interest (Table 3.18). The full length constructs were designed to contain the complete gene sequence as obtained from UniProt, this meant that the full length receptor subunits (integrins or TGFβR) were membrane bound. In contrast, the truncated constructs were designed to exclude the transmembrane region of the protein therefore making the receptor subunits soluble. This was to replicated the commercially available proteins but to also increase the chance of successfully purifying the binding partners as it can be difficult to solubilise and purify membrane proteins. From the full length or truncated amino acid sequence, the expected gene length (base pairs) was calculated by multiplying the total number of amino acids by 3. The predicted protein molecular weight was calculated using the Expasy ProtParam tool (337).

Table 3.18: Plasmid constructs designed for protein expression and purification

Designed Protein Construct	Base Pairs	MW (kDa)	MW + Fluorescent Tag (kDa)	Vectors
Galectin-3 FL (M1-I250)	750	26.15	53.15	pOPINF pOPINE pOPINEneo-GFP pOPINE-mCherry
Galectin-1 FL (M1-D135)	405	14.72	41.72	
LTBP1 Fragment 1 (M1-C900)	2,700	95.87	-	pOPINE
LTBP1 Fragment 2 (V841-E1721)	2,643	97.18	-	
LAP TRUNC (L30-A350)	960	36.91	-	
LAP FL (M1-S390)	1,170	44.34	71.34	pOPINE pOPINEneo-GFP
LAP TRUNC (L30-R278)	744	28.49	55.49	
αv FL (M1-T1048)	3,144	116.04	143.04	
αv TRUNC (F31-V992)	2,883	106.27	133.27	
β1 FL (M1-K798)	2,394	88.42	115.42	
β1 TRUNC (Q21-D728)	2,121	78.35	105.35	
β5 FL (M1-D799)	2,397	88.05	115.05	
β5 TRUNC (G24-N719)	2,085	76.52	103.52	
β6 FL (M1-C788)	2,364	85.94	112.94	
β6 TRUNC (G22-N707)	2,055	74.30	101.3	
TGFβRI FL (M1-M503)	1,509	55.96	82.96	
TGFβRI TRUNC (R80-M503)	1,269	47.78	74.78	
TGFβRII FL (M1-K567)	1,701	64.57	91.57	
TGFβRII TRUNC (R190-K567)	1,131	43.28	70.28	

Sequence of full length (FL) and truncated (TRUNC) constructs shown. Gene length in base pairs (bp) and protein size (daltons) indicated for each designed construct. The desired cloning vectors for each construct is shown.

To determine whether the gene sequences for the constructs described in (Table 3.18) had been successfully amplified using 'PCR protocol 1' (Section 3.3.2.1) from the synthetic DNA template, PCR products were subject to electrophoresis on 1.2% agarose gels. Amplified DNA for full length galectin-3 (~750 bp) and galectin-1 (~405 bp) are shown in (Figure 3.1A - B) and full length or truncated αv (~2,883 bp), β1 (~2,121 bp), β5 (~2,397 bp or ~2,085 bp) and β6 (~2,055 bp) in (Figure 3.1C - D). All of these constructs ran at the expected molecular weight indicated by the DNA ladder and were successfully cloned. Lanes that have not been annotated with a plasmid construct are PCR products that were not successfully amplified or cloned by 'PCR

protocol 1' evident from incorrect sequencing data and were therefore amplified by 'PCR protocol 2'.

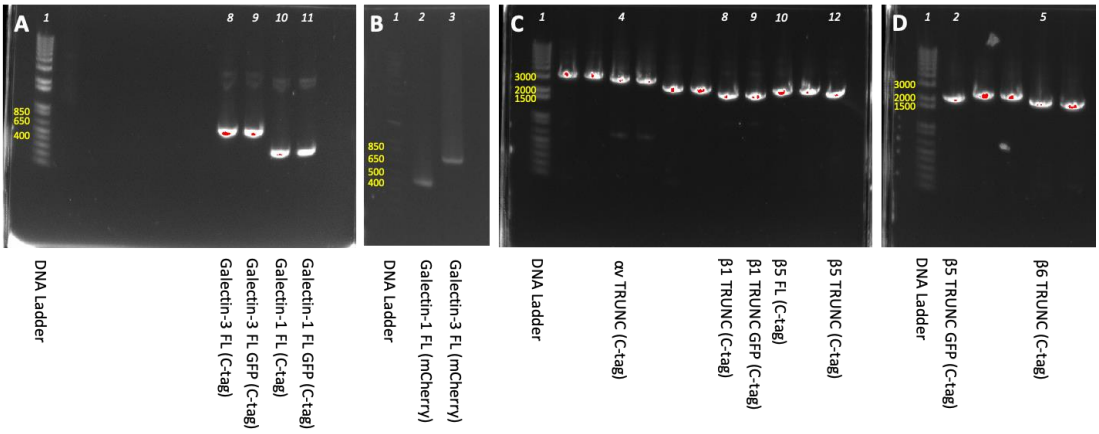


Figure 3.1: Confirmation of successful PCR amplification of clonal genes by 'PCR protocol 1' DNA separated by agarose gel electrophoresis and the target DNA fragment size estimated from the 1 Kb Plus DNA Ladder migration pattern. ***(A-B)*** *PCR analysis of galectin-3 full length and galectin-1 full length contructs* ***(C)*** *αv truncated, β1 truncated, β5 full length and β5 truncated constructs* ***(D)*** *and β5 truncated and β6 truncated constructs.*

A similar approach was used to amplify DNA and confirm the sizes of the remaining constructs, the PCR products amplified using PCR 'protocol 2' (Section 3.3.2.1) were subject to electrophoresis on 1% or 1.2% agarose gels. Amplified DNA for full length or truncated LAP (~1,170 bp, L30-R278 ~744 bp or L30-A350 ~960 bp), αv (~3,144 bp or ~2,883 bp), β1 (~2,394 bp), β5 (~2,397 bp), β6 (~2,364 bp or ~2,055 bp) and TGFβRI (~1,509 bp or ~1,269 bp) are shown in (Figure 3.2A) and full length or truncated TGFβRI (~1,269 bp) and TGFβRII (~1,701 bp or ~1,131 bp) in (Figure 3.2B). Amplified DNA for full length galectin-3 (~750 bp) and galectin-1 (~405 bp) are shown in (Figure 3.2C - D). All of these constructs ran at the expected molecular weight indicated by the DNA ladder and were successfully cloned. Lanes that have not been annotated with a plasmid construct are the PCR products of LTBP1 fragments that were later re-amplified and overlapped using 'PCR protocol 3'. The three construct highlighted in orange in (Figure 3.2A - B), namely β1 full length GFP, β6 full length and TGFβRII full

length GFP were amplified by PCR to the expected gene length (base pairs), but subsequent cloning was not successful evident from incorrect sequencing data.

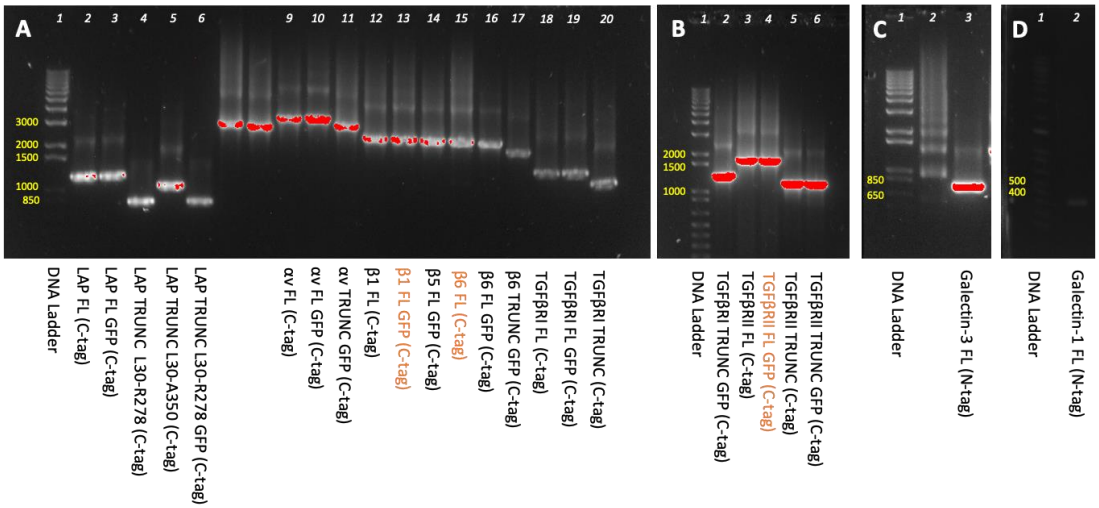


Figure 3.2: Confirmation of successful PCR amplification of clonal genes by 'PCR protocol 2' DNA separated by agarose gel electrophoresis and the target DNA fragment size estimated from the 1 Kb Plus DNA Ladder migration pattern. ***(A)*** *PCR analysis of LAP full length, LAP truncated (L30-R278), LAP truncated (L30-A350), αv full length, αv truncated, β1 full length, β5 full length, β6 full length, β6 truncated, TGFβRI full length and TGFβRI truncated constructs,* ***(B)*** *TGFβRI truncated, TGFβRII full length and TGFβRII truncated constructs,* ***(C)*** *galectin-3 full length and* ***(D)*** *galectin-1 full length.*

LTBP1 fragment 1 (~2,700 bp) and LTBP1 fragment 2 (~2,643 bp) were successfully amplified and recombined by overlap extension PCR using *'PCR protocol 3'* (Section 3.3.2.1). Agarose gels for the extension PCR (step A) (Figure 3.3A) and final fusion gene PCR amplification (step C) are shown (Figure 3.3B). Although the full length fusion gene was amplified by PCR to the expected gene length (~5,343 bp) based on the predicted base pairs, subsequent cloning was not successful evident from incorrect sequencing data.

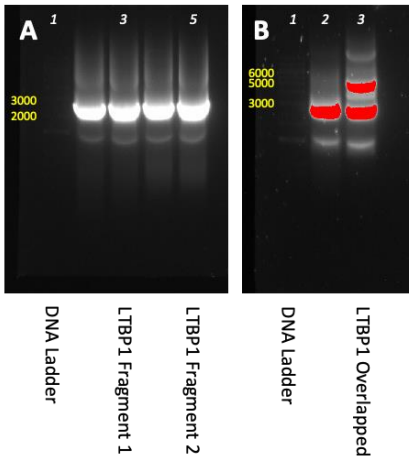


Figure 3.3: Confirmation of successful PCR amplification of LTBP1 by 'PCR protocol 3'
DNA separated by agarose gel electrophoresis and the target DNA fragment size estimated from the 1 Kb Plus DNA Ladder migration pattern. ***(A)*** *PCR analysis of LTBP1 fragment 1 and LTBP1 fragment 2 and* ***(B)*** *full length LTBP1 assembled via complementary DNA fragment overlap.*

PCR amplification, cloning and sequence verification was successful for 34 out of 38 designed constructs (Table 3.19) and these 34 constructs were taken forward for protein expression and purification trials. The 4 constructs that were not successfully sequence-verified namely β6 full length, LTBP1 full length, β1 full length GFP and TGFβRII full length GFP were not taken forward as because of time restraints it was not possible to troubleshoot the plasmid cloning experiments.

Table 3.19: Summary of constructs generated for expression studies

A

Target	FL (pOPINE)	FL (pOPINEneo-GFP)	TRUNC (pOPINE)	TRUNC (pOPINEneo-GFP)
αv	Yes	Yes	Yes	Yes
β1	Yes	-	Yes	Yes
β5	Yes	Yes	Yes	Yes
β6	-	Yes	Yes	Yes
LAP	Yes	Yes	Yes (x2 truncations)	Yes
TGFβRI	Yes	Yes	Yes	Yes
TGFβRII	Yes	-	Yes	Yes
LTBP1	-			

B

Target	FL (pOPINE)	FL (pOPINEneo-GFP)	FL (pOPINF)	FL (pOPINE-mCherry)
Galectin-3	Yes	Yes	Yes	Yes
Galectin-1	Yes	Yes	Yes	Yes

*Sequence-verified constructs to be taken forward into **(A)** mammalian or **(B)** bacterial protein expression denoted by 'Yes'. Constructs that were not sucessfully cloned and/or sequence-verified following PCR amplification denoted by '-'.*

3.4.2. Expression and purification of recombinant proteins from mammalian cells

As previously discussed (Section 3.1) a mammalian cell expression system was selected for the production of all recombinant glycoproteins. Experiments initially focused on the expression and purification of constructs in the pOPINEneo-GFP vector because the GFP tag facilitates tracking protein expression and optimising purification conditions. Addition of a GFP tag allows the protein expression to be monitored visually without needing to purify the protein and both in-gel fluorescence and FSEC analysis can be used alongside Coomassie staining to evaluate the purification efficiency.

3.4.2.1. Small-scale protein expression and purification

Recombinant protein expression was performed in the Expi293 expression system (3 mL scale) and small-scale protein purification performed as outlined in (Section 3.3.4.1 and Section 3.3.4.2). Protein expression and purification efficiency was

determined from the intensity of Coomassie staining and in-gel fluorescence results with each construct run in duplicate (Figure 3.4).

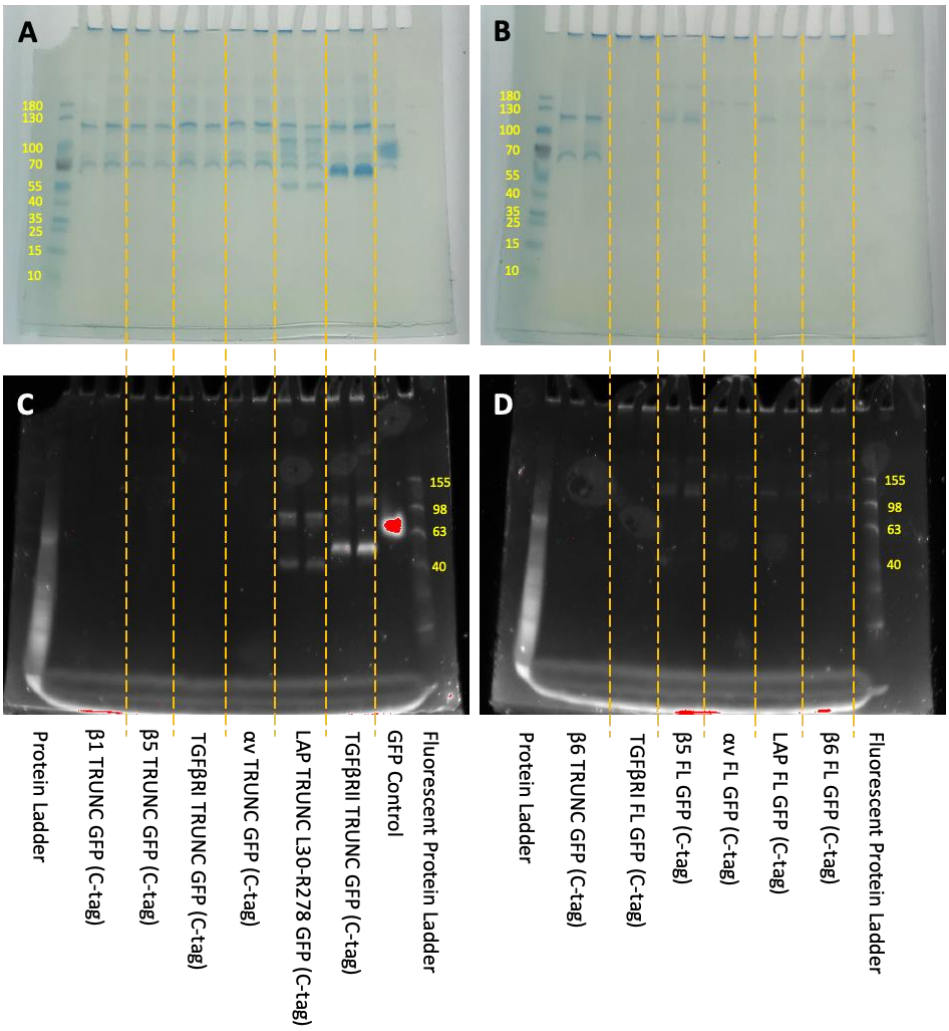


Figure 3.4: Gel electrophoresis of small-scale protein expression and purification
*Protein elutions visualised with **(A-B)** Coomassie stain or by **(C-D)** in-gel fluorescence. Target protein size estimated from the PageRuler Prestained Protein Ladder (10 to 180 kDa) migration pattern and BenchMark Fluorescent Protein Standard (11 to 155 kDa).*

The results suggested that expression and purification had been successful for only 3 of the constructs with a band at the expected protein size (kDa) shown by both methods for β5 full length GFP, αv full length GFP and β6 full length GFP. Results for LAP truncated GFP were inconclusive, a band at the correct size was present on the Coomassie gel but was slightly lower than expected by in-gel fluorescence. Results

for all other constructs suggested that the proteins were either expressed at very low levels or that the purification required optimisation. These results are summarised below (Table 3.20).

Table 3.20: Summary of small-scale mammalian expression and purification gel results

Construct	MW Including Tag (kDa)	Coomassie Staining	In-Gel Fluorescence
β1 TRUNC GFP	105.35	Band not present	Band not present
β5 TRUNC GFP	103.52	Band not present	Band not present
TGFβRI TRUNC GFP	74.78	**Band present**	Band not present
αv TRUNC GFP	133.27	Band not present	Band not present
LAP TRUNC GFP	55.49	**Band present**	**Inconclusive**
TGFβRII TRUNC GFP	70.28	Band not present	Band not present
β6 TRUNC GFP	101.3	Band not present	Band not present
TGFβRI FL GFP	82.96	Band not present	Band not present
β5 FL GFP	115.05	**Band present**	**Band present**
αv FL GFP	143.04	**Band present**	**Band present**
LAP FL GFP	71.34	Band not present	Band not present
β6 FL GFP	112.94	**Band present**	**Band present**

Protein elutions from the small-scale protein expression and purification were also subjected to FSEC analysis (Section 3.3.4.3) and the results showed the highest fluorescence intensity for the αv full length GFP, β5 full length GFP and β6 full length GFP constructs (Figure 3.5). This corresponds to what was visualised by gel electrophoresis (Figure 3.4).

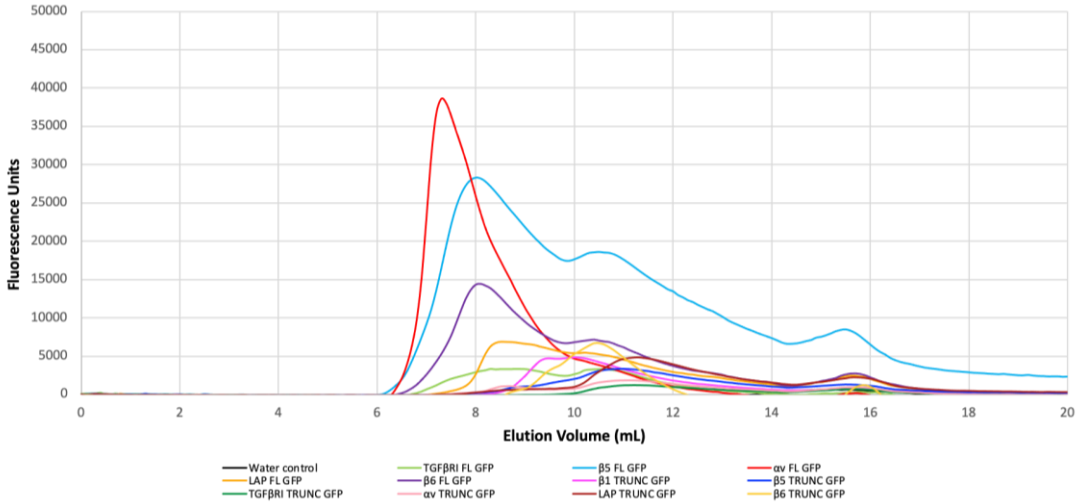


Figure 3.5: FSEC analysis of purified proteins following small-scale protein expression Protein elutions analysed by FSEC and a fluorescence profile measured at an absorbance of 510 nm. The retention time and subsequently elution volume is dependent on protein MW.

The FSEC trace for the TGFβRII truncated GFP construct showed a large peak at ~11 mL corresponding to free GFP and consequently this trace was removed from the fluorescence profile. The FSEC results for all other constructs implied a low protein expression and/or low purification efficiency (Figure 3.5). This corresponds to what was visualised by gel electrophoresis (Figure 3.4). Collectively, it was concluded from the small-scale purification results (Coomassie staining, in-gel fluorescence and FSEC) that 4 constructs, β5 full length GFP, αv full length GFP, β6 full length GFP (all membrane proteins) and LAP truncated GFP (soluble protein) could be taken forward for large-scale protein expression (30 mL scale) in order to further optimise the purification process.

3.4.2.2. Detergent screen for membrane proteins

For the 3 membrane proteins taken forward from the initial small-scale purification experiment, detergent screening was performed as previously outlined (Section 3.3.4.2). For the αv full length GFP protein, a band of the correct size (143.04 kDa) was detected by in-gel fluorescence (Figure 3.6A) and with Coomassie staining (Figure

3.6B) for all 12 detergents tested. FSEC analysis showed the highest fluorescence intensity in DM + CHS detergent (Figure 3.6C), subsequently this was selected as the detergent of choice for all future purifications.

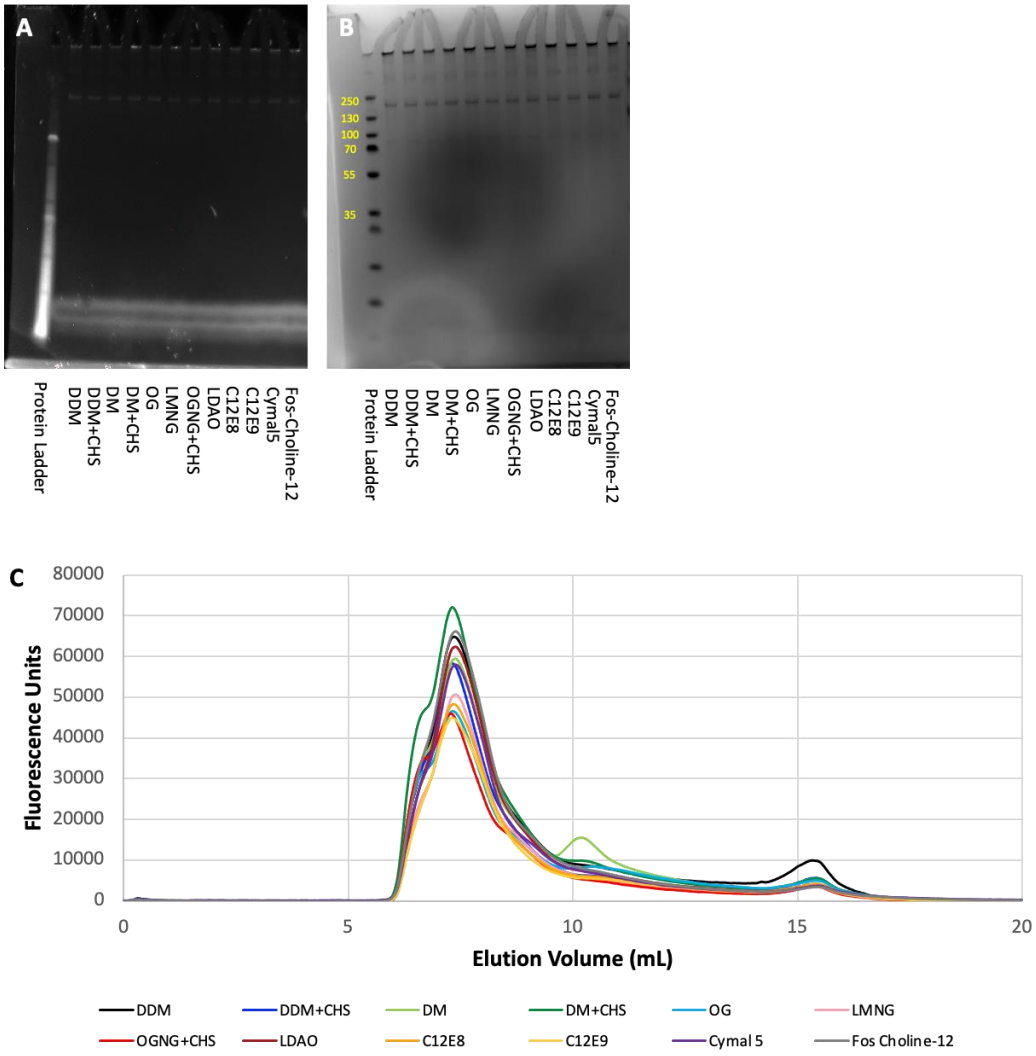


Figure 3.6: Detergent screen for the purification of αv full length GFP
*Protein elutions for αv full length GFP visualised by **(A)** in-gel fluorescence, **(B)** with Coomassie stain or **(C)** FSEC analysis with the fluorescence profile measured at absorbance 510 nm. Target protein size estimated from the PageRuler Prestained Protein Ladder (10 to 180 kDa) migration pattern.*

The most favourable detergent for the purification of β5 full length GFP was similarly determined by analysis of the in-gel fluorescence, Coomassie staining and FSEC results. For all 12 detergents tested, a band of the correct size (115.05 kDa) was detected by in-gel fluorescence (Figure 3.7A) and with Coomassie staining (Figure

3.7B). FSEC analysis showed a comparable fluorescence intensity in DDM + CHS and DM + CHS detergent, however a peak at ~11 mL was also present with DM + CHS which corresponds to degraded protein with a GFP tag (Figure 3.7C) and as a result DDM + CHS was selected as the detergent of choice for all future purifications of this protein.

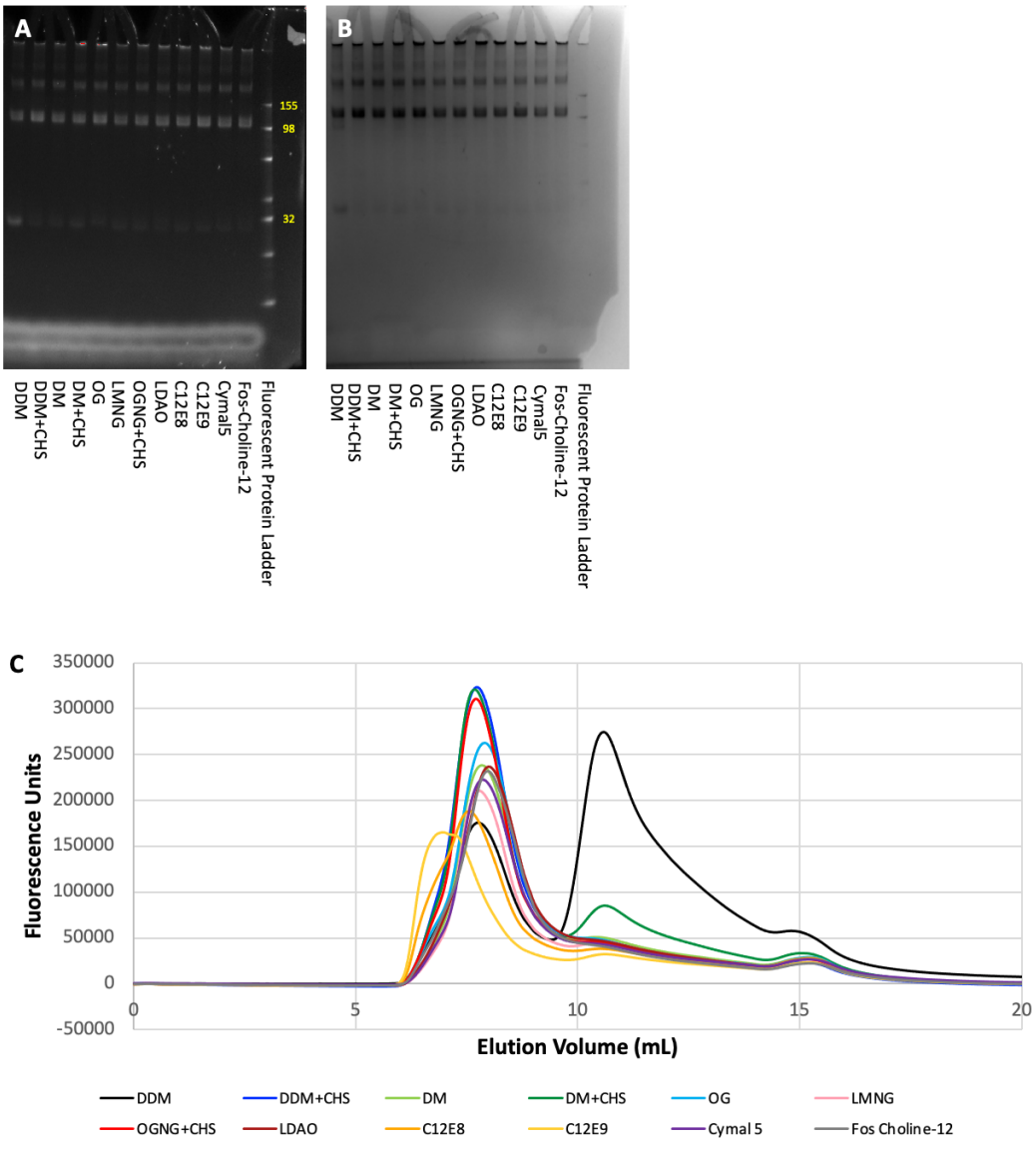


Figure 3.7: Detergent screen for the purification of β5 full length GFP
*Protein elutions for β5 full length GFP visualised by **(A)** in-gel fluorescence, **(B)** with Coomassie stain or **(C)** FSEC analysis with the fluorescence profile measured at 510 nm absorbance. Target protein size estimated from the BenchMark Fluorescent Protein Standard (11 to 155 kDa).*

For the purification of β6 full length GFP, a band of the correct size (112.94 kDa) was detected by in-gel fluorescence (Figure 3.8A) and with Coomassie staining (Figure 3.8B) for all 12 detergents tested. FSEC analysis showed a comparable fluorescence intensity in LDAO and Fos-Choline-12 detergent, however a peak at ~11 mL was also present with Fos-Choline-12 which corresponds to degraded protein with a GFP tag (Figure 3.8C) and subsequently LDAO was the detergent of choice for purification of this protein.

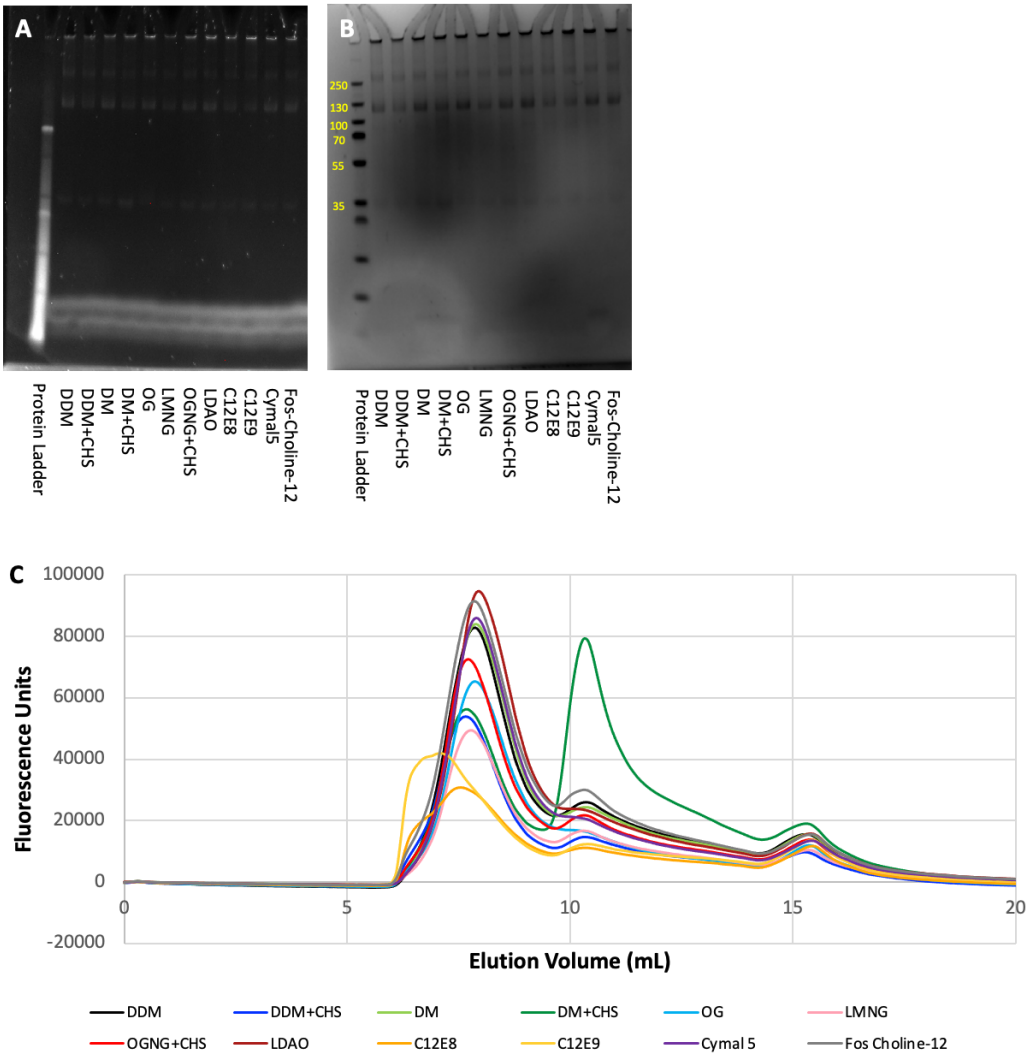


Figure 3.8: Detergent screen for the purification of β6 full length GFP
*Protein elutions for β6 full length GFP visualised by **(A)** in-gel fluorescence, **(B)** with Coomassie stain or **(C)** FSEC analysis with a fluorescence profile measured at 510 nm absorbance. Target protein size estimated from the PageRuler Prestained Protein Ladder (10 to 180 kDa) migration pattern.*

3.4.2.3. Buffer screen for soluble proteins

For the only soluble protein taken forward from the initial small-scale purification experiment namely LAP truncated GFP, buffer screening was performed as previously outlined (Section 3.3.4.2). As previously, the optimal buffer conditions for the purification of LAP truncated GFP were determined by analysis of the in-gel fluorescence (Figure 3.9A) and Coomassie staining (Figure 3.9B) results. A band of the correct size (55.49 kDa) was detected by in-gel fluorescence for two of the buffer conditions tested (lane 14 and 15) (Figure 3.9A). The in-gel fluorescence analysis indicated that the protein was generally unstable with a band corresponding to free GFP (27 kDa) present in all 16 buffer conditions. Analysis by Coomassie staining detected a band in only one of the buffer conditions (lane 14), although this band was of a size slightly lower than expected (Figure 3.9B). For future optimisation experiments using the LAP truncated GFP protein, the 300 mM NaCl, 50 mM Arg:Glu, 20 mM HEPES (pH 6.5) buffer was utilised.

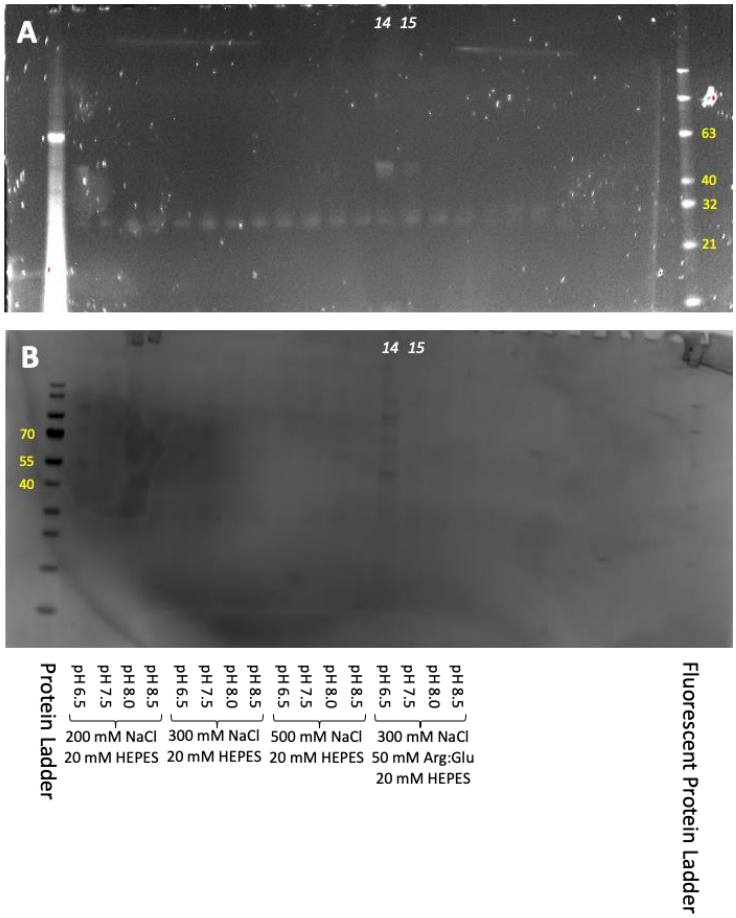


Figure 3.9: Buffer screen for the purification of LAP truncated GFP
*Protein elutions for LAP truncated GFP visualised by **(A)** in-gel fluorescence or **(B)** with Coomassie stain. Target protein size estimated from the PageRuler Prestained Protein Ladder (10 to 180 kDa) migration pattern and BenchMark Fluorescent Protein Standard (11 to 155 kDa).*

3.4.3. Expression and purification of recombinant proteins from bacterial cells

As previously discussed (Section 3.1) a bacterial cell expression system was selected for the production of recombinant galectin-3 and galectin-1. The pOPINF vector (6His N-terminal tag) was chosen for the galectin constructs, as unlike in the pOPINE vector, the his-tag can be cleaved from the protein during purification. This was preferable as the galectins were being produced for binding studies and a fusion/affinity tag can affect a proteins binding capability.

3.4.3.1. Recombinant human galectin-3 protein

Galectin-3 protein expression was performed in *E. coli* cells and subsequent protein purification performed using methods previously outlined (Section 3.3.3.1 and Section 3.3.3.2). Firstly, an imidazole gradient was used to elute the protein from a 5 mL Ni^{2+} HisTrap HP column. The ÄKTA results showed a broad peak in UV 280 nm absorbance between 20 - 40 mL elution volume corresponding to 150 - 450 mM imidazole and this represents the concentration range at which galectin-3 protein elution occurred (Figure 3.10A). The apex of the peak was at 40 mL elution volume corresponding to 300 mM imidazole. Gradient fractions were therefore selected from the elution profile and analysed by gel electrophoresis. Results showed the highest concentration of galectin-3 protein (27 kDa) was present in fractions A8 -> B1 (Figure 3.10B). Subsequently, these fractions were combined, concentrated and dialysed overnight with 3C-protease to cleave the His-tag from the protein. A reverse IMAC column was then used to separate the cleaved His-tag and 3C protease from the target protein. The gel electrophoresis results showed that galectin-3 (27 kDa) had predominately eluted in the load flow through and the wash, this was expected as the galectin had been cleaved from the His-tag that would bind to the Ni^{2+} (Figure 3.10C). However the intensity of the band in the elution suggest that not all of the galectin-3 had been cleaved from the His-tag. The 3C protease and contaminant proteins had been separated from the target protein and were present in the elution.

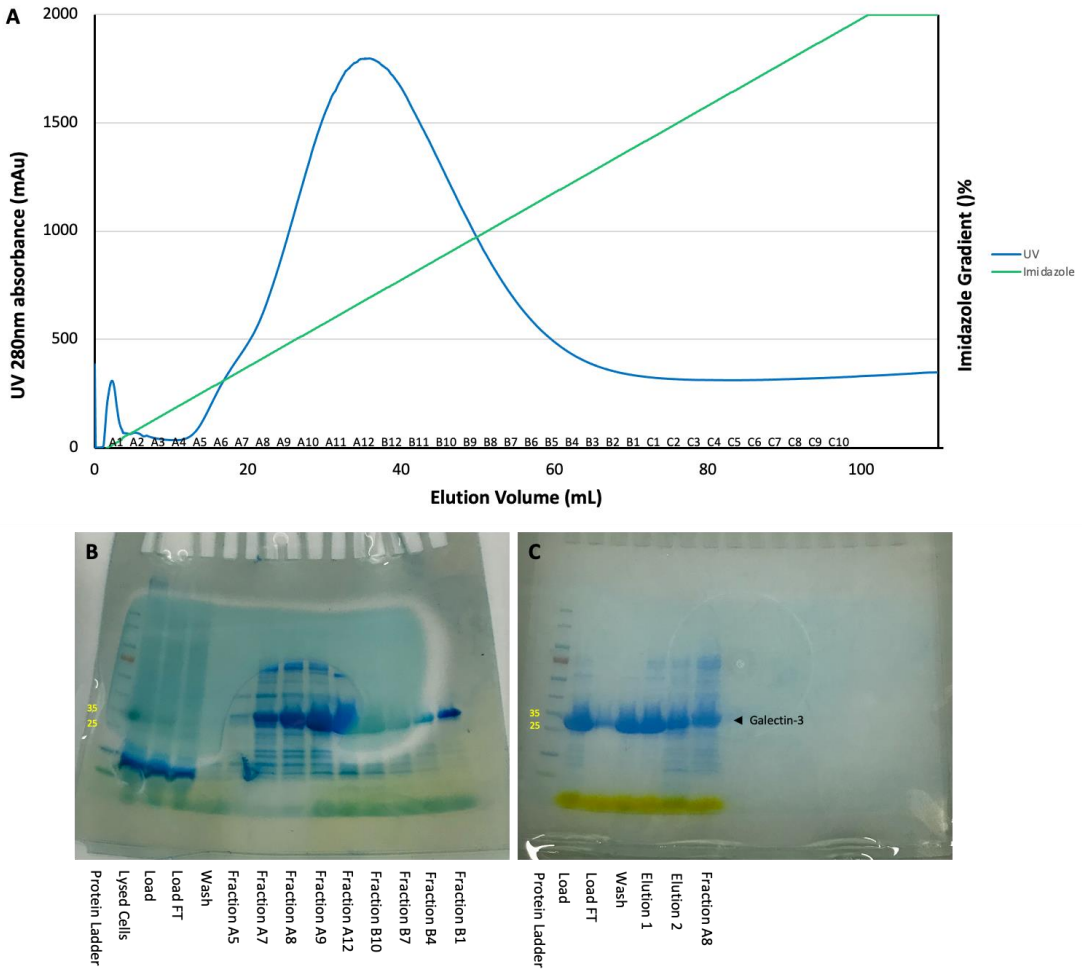


Figure 3.10: Ni^{2+} HisTrap HP column purification of galectin-3 (pOPINF vector)
***(A)** ÄKTA UV absorbance (280 nm) trace for galectin-3 elution in the presence of an imidazole gradient with a concentration range of 0 - 750 mM. **(B)** Fractions from the galectin-3 imidazole gradient elution analysed by SDS-PAGE to confirm the presence of galectin-3 in the expected fractions. Lysed cells, load, load flow through (FT) and wash steps loaded as controls. **(C)** Fractions collected following IMAC column purification analysed by SDS-PAGE to confirm 3C protease histidine cleavage from galectin-3.*

After reverse IMAC, the load flow through and wash were combined and concentrated down to 750 µL for SEC which is a final 'polishing step' in purification used to separate out any remaining trace impurities. The ÄKTA results showed a peak in UV 280 nm absorbance between elution volumes of 12 - 16 mL, this represents the range over which galectin-3 protein was eluted (Figure 3.11A).The small peak at ~10 mL elution volume was presumed to be a contaminant protein of a higher MW that had been separated out. Fractions were selected from the elution profile and ran by

gel electrophoresis, results showed very high concentrations of galectin-3 protein (27 kDa) present in fractions B3 -> B14 (Figure 3.11B). Given how heavily concentrated galectin-3 was and the band intensity for neat samples, fractions were diluted 1 in 10 in water and separated by gel electrophoresis (Figure 3.11C) which confirmed that the purified galectin-3 protein was highly pure with minimal contaminants.

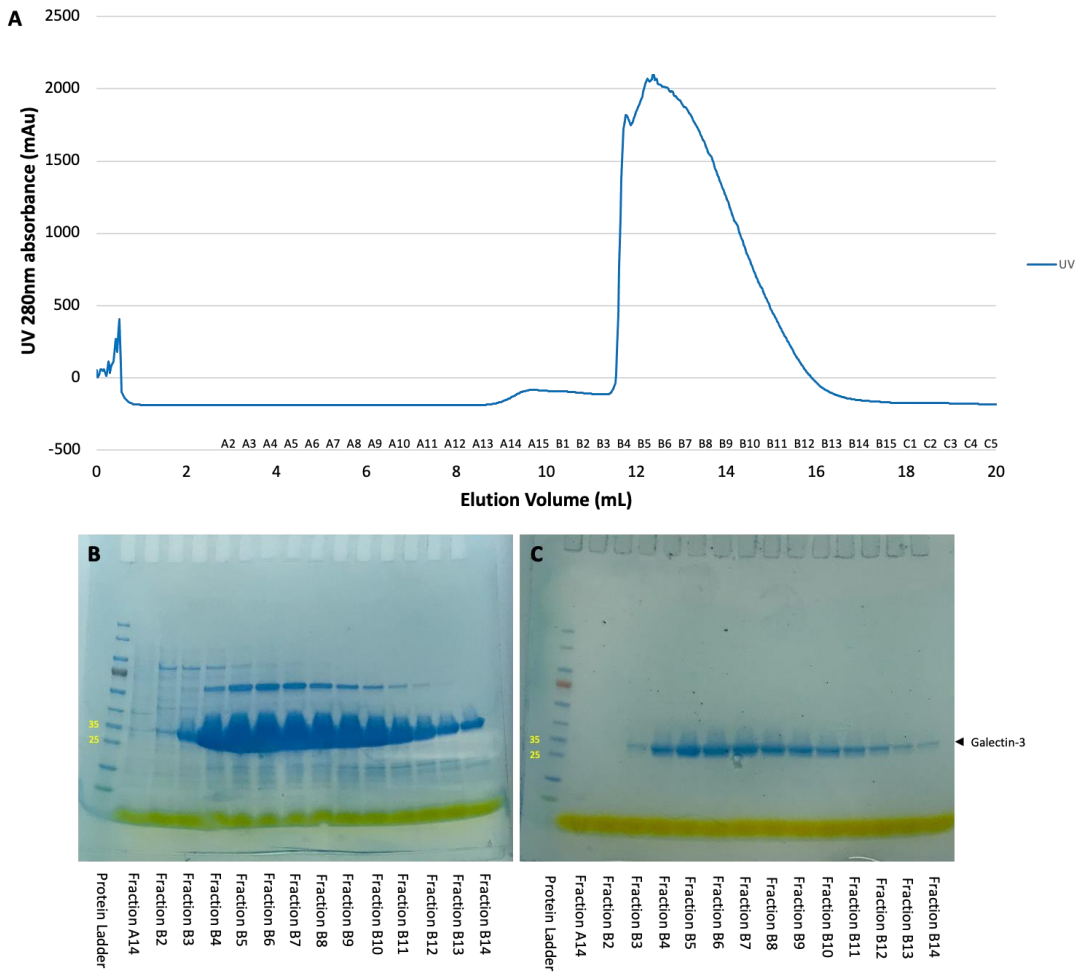
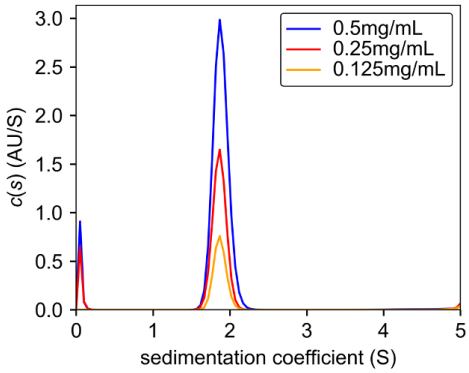


Figure 3.11: SEC column purification of galectin-3 (pOPINF vector)
***(A)** ÄKTA UV absorbance (280 nm) trace measured for galectin-3 following SEC. **(B)** Fractions from SEC analysed by SDS-PAGE to confirm galectin-3 protein purity. **(C)** SEC fractions diluted 1 in 10 and purity of galectin-3 protein assessed by SDS-PAGE.*

The quality of the purified protein was assessed by AUC (Section 3.3.6.1) a method that can be used to determine the protein mass and homogeneity/ assembly state.

The UV 280 nm absorbance data showed that the purified protein had a sedimentation coefficient value of approximately 1.9 S (Figure 3.12). The MW of the sedimented protein was calculated as 25 kDa (c(s) analysis). This is slightly lower than expected for galectin-3 (27 kDa), which will be discussed (Section 3.5), however the results suggested that the purified galectin-3 was a monomer in solution.



Concentration (mg/mL)	MW (Da)	Sedimentation Coefficient (S)	Frictional Ratio
0.5	24,597	1.882	1.423
0.25	24,383	1.864	1.428
0.125	25,108	1.871	1.451

Figure 3.12: Sedimentation coefficient distribution for purified galectin-3
The sedimentation coefficient value determined from absorbance data at three concentrations of galectin-3.

The quality of the purified protein was also assessed by SEC-MALS (Section 3.3.6.2) for comparison. The results showed a peak in UV 280 nm absorbance at an elution volume of 12 mL. This was expected for a protein the size of galectin-3 (monomer) (Figure 3.13). Astra v7 analysis determined that the MW of this protein was calculated as 23 kDa. This was similar to the MW estimated by AUC but is slightly lower than expected for a galectin-3 monomer (27 kDa).

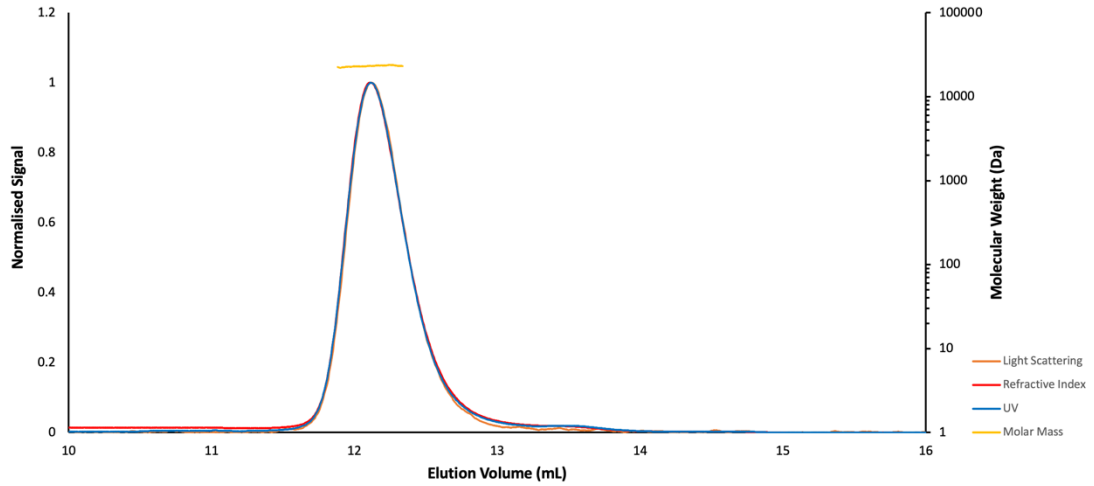


Figure 3.13: SEC-MALS data for purified galectin-3
UV 280 nm absorbance (blue), light scattering (orange), refractive index (red) and molar mass (yellow) data obtained by SEC-MALS.

3.4.3.2. Recombinant human galectin-1 protein

As with galectin-3, galectin-1 protein was expressed in *E. coli* cells and eluted from a 5 mL Ni^{2+} HisTrap HP column in the presence of an imidazole gradient. The ÄKTA results showed a broad peak in UV 280 nm absorbance between 10 - 40 mL elution volume corresponding to 75 - 300 mM imidazole concentrations at which galectin-1 protein elution was detected (Figure 3.14A). The apex of the peak was at 25 mL elution volume corresponding to 188 mM imidazole. The narrow peak at ~2 mL elution volume represents a contaminant protein that had been separated out. Gradient fractions were selected from the elution profile and assessed by gel electrophoresis. The results showed the highest concentration of galectin-1 protein (15 kDa) was present in fractions A5 -> B10, although contaminant proteins were also present (Figure 3.14B). Subsequently, these fractions were combined, concentrated and dialysed overnight with 3C-protease for His-tag cleavage and a reverse IMAC column was then used to separate the cleaved His-tag and 3C protease from the target protein. The gel electrophoresis results showed that galectin-1 (15 kDa) had predominately eluted in the load flow through and the wash, although contaminant

proteins of a higher MW were also eluted in the wash (Figure 3.14C). The 3C protease and contaminant proteins were present in the elution, as well as any galectin-1 that had not been cleaved from the His-tag.

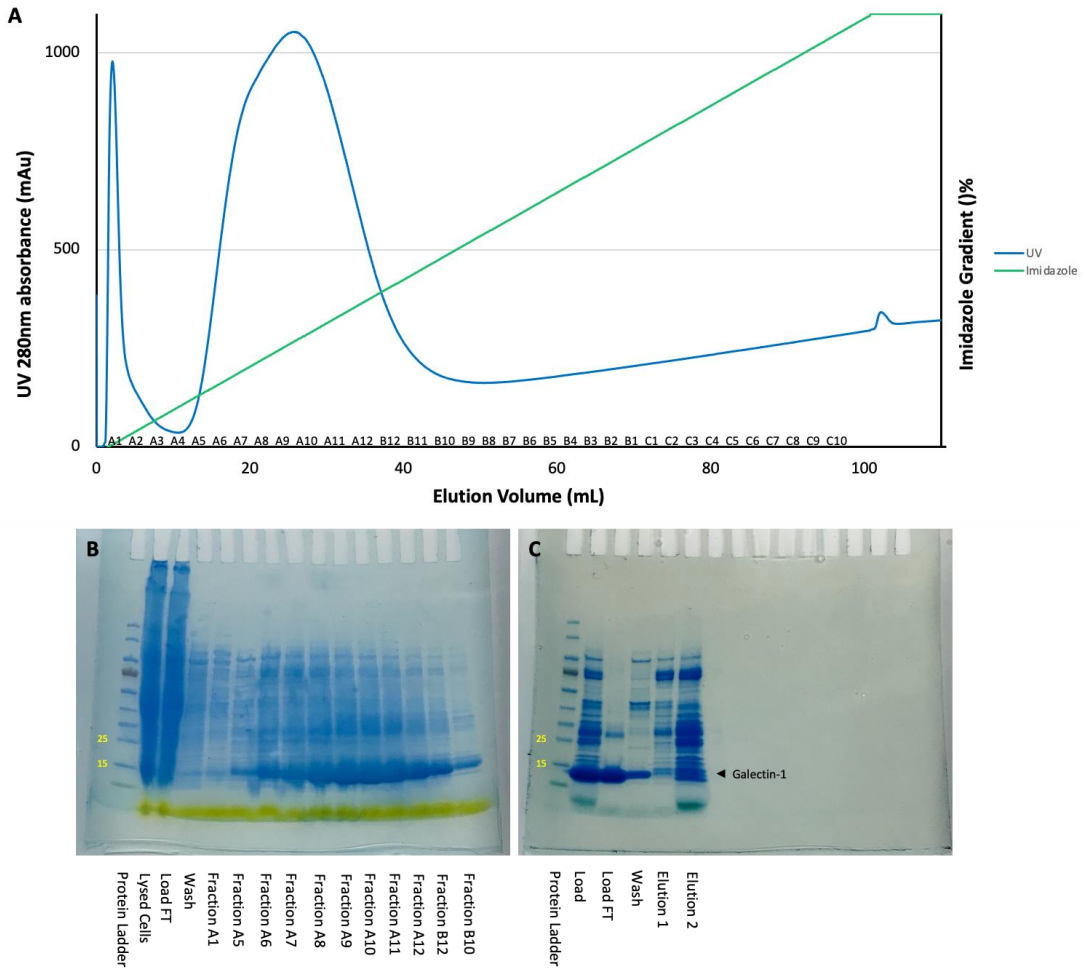


Figure 3.14: Ni^{2+} HisTrap HP column purification of galectin-1 (pOPINF vector)
***(A)** ÄKTA UV absorbance (280 nm) trace for galectin-1 elution in the presence of an imidazole gradient with a concentration range of 0 - 750 mM. **(B)** Fractions from the galectin-1 imidazole gradient elution analysed by SDS-PAGE to confirm the presence of galectin-1 in the expected fractions. Lysed cells, load flow through (FT) and wash steps loaded as controls. **(C)** Fractions collected following IMAC column purification analysed by SDS-PAGE to confirm 3C protease histidine cleavage from galectin-1.*

After reverse IMAC, the load flow through was concentrated down to 750 µL and SEC analysis performed. The ÄKTA results showed a peak in UV 280 nm absorbance between 13 - 17mL elution volume, which represented the galectin-1 protein elution (Figure 3.15A). The small peak eluting at ~9 mL elution volume was presumed to be

a contaminant protein or an aggregate that had been separated out. Fractions were selected from the elution profile and analysed by gel electrophoresis, with results showing very high concentrations of galectin-1 protein (15 kDa) present in fractions B9 -> C2 (Figure 3.15B). Samples were analysed neat and diluted 1 in 10 (Figure 3.15C) which confirmed that the purified galectin-1 protein was highly pure, although a band was also present at 30 kDa in fractions at the apex of the peak which were the fractions with the highest galectin-1 concentration, this was thought to be a galectin-1 dimer.

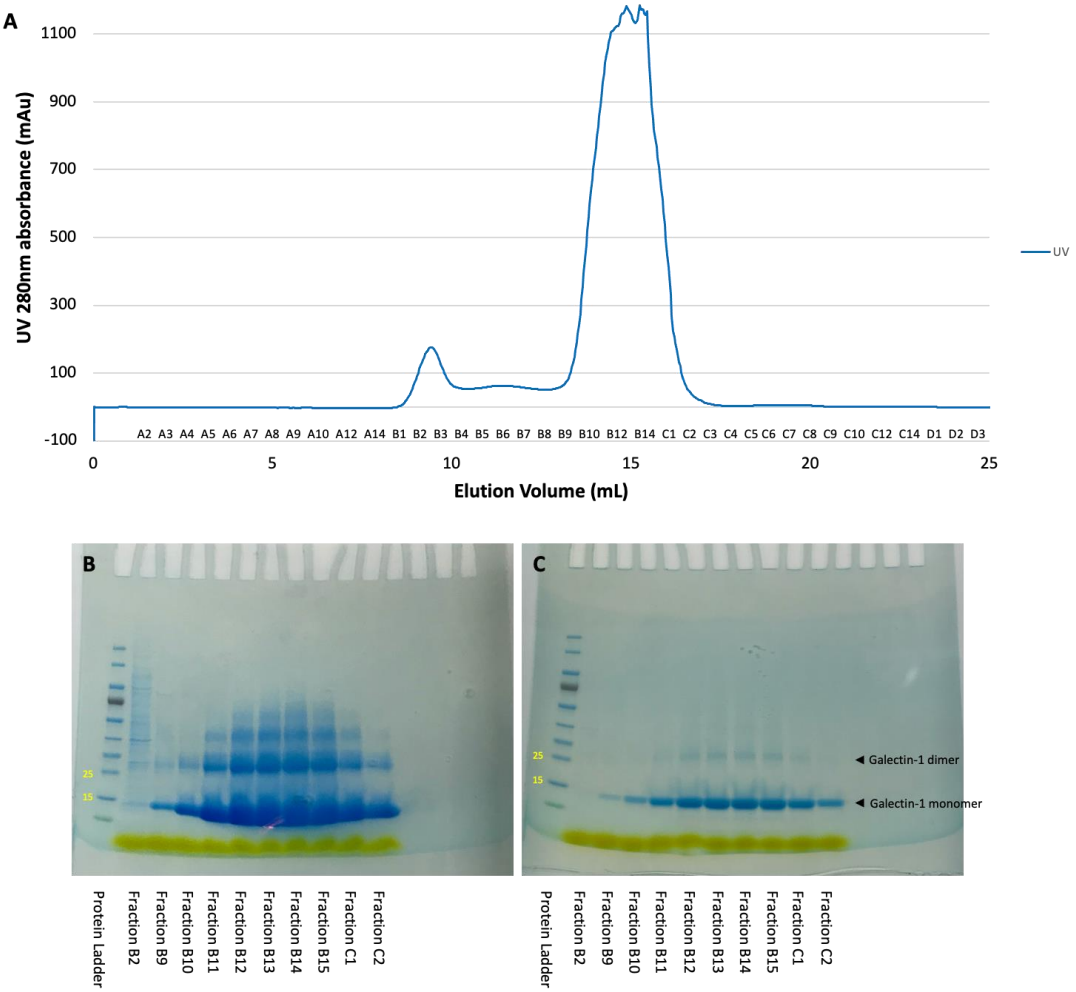
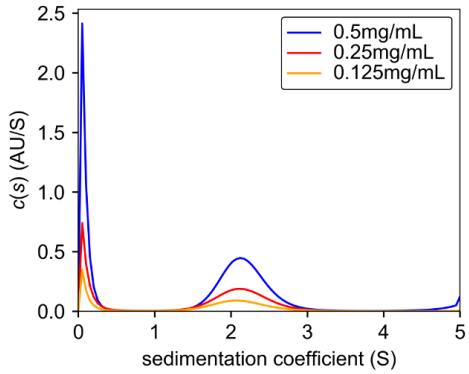


Figure 3.15: SEC column purification of galectin-1 (pOPINF vector)
***(A)** ÄKTA UV absorbance (280 nm) trace measured for galectin-1 following SEC. **(B)** Fractions from SEC analysed by SDS-PAGE to confirm galectin-1 protein purity. **(C)** SEC fractions diluted 1 in 10 and purity of galectin-1 protein assessed by SDS-PAGE.*

AUC analysis of the quality of the purified protein gave absorbance data that showed that the purified protein had a sedimentation coefficient value of approximately 2.2 S (Figure 3.16). The MW of the sedimented protein was calculated as 24 kDa (c(s) analysis). Compared with the AUC results for galectin-3 (Figure 3.12) the peak was quite broad here suggesting that purified galectin-1 was less homogenous. This corresponds to what was visualised by gel electrophoresis following SEC whereby both monomeric and dimeric galectin-1 were detected (Figure 3.15). However, the AUC results suggested that purified galectin-1 was predominantly dimeric in solution, despite the calculated MW being lower than expected for a dimer (30 kDa), which will be discussed (Section 3.5).



Concentration (mg/mL)	MW (Da)	Sedimentation Coefficient (S)	Frictional Ratio
0.5	24,015	2.174	1.204
0.25	23,355	2.142	1.200
0.125	23,472	2.139	1.206

Figure 3.16: Sedimentation coefficient distribution for purified galectin-1
The sedimentation coefficient value determined from absorbance data at three concentrations of galectin-1.

The quality of the purified protein was also assessed by SEC-MALS for comparison. The results showed a peak in UV 280 nm absorbance at an elution volume of 13 mL. This was expected for a protein the size of galectin-1 (dimer) (Figure 3.17). There was also a smaller peak in absorbance at an elution volume of 11.5 mL which was expected for a protein the size of monomeric galectin-1. The small peak observed at 14.5 mL elution volume was presumed to be a low level contaminant or galectin-1 in

an alternative assembly state. Astra v7 analysis determined that the MW of the protein at 13 mL was calculated as 24 kDa. This was similar to the MW estimated by AUC but is slightly lower than expected for a galectin-1 dimer (30 kDa).

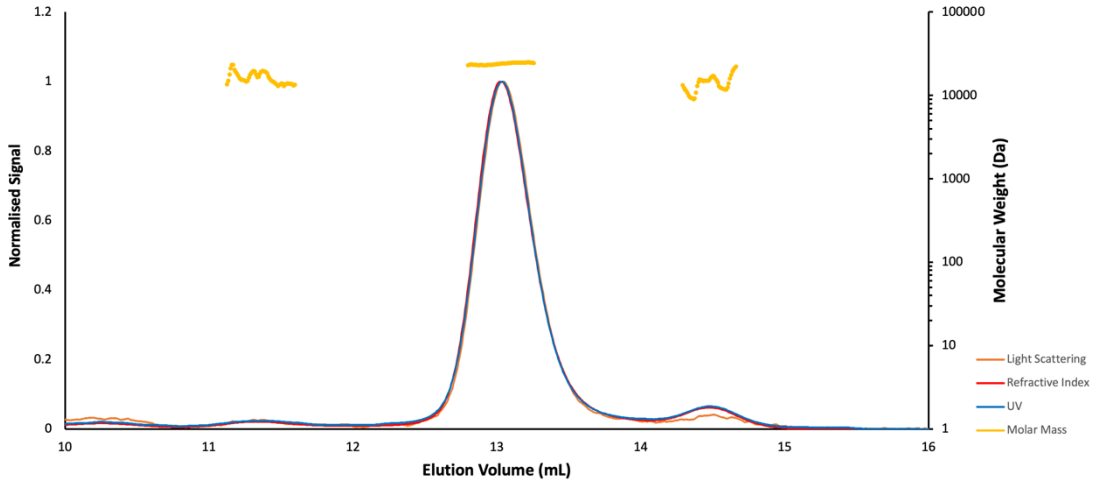


Figure 3.17: SEC-MALS data for purified galectin-1
UV 280 nm absorbance (blue), light scattering (orange), refractive index (red) and molar mass (yellow) data obtained by SEC-MALS.

3.5. Discussion

Due to the limitations of buying commercially available proteins for binding experiments, recombinant protein production was performed at the RCaH for all proteins of interest. PCR amplification and cloning was successful for 26 out of 30 constructs designed for mammalian protein expression (Section 3.4.1). Following this, a small-scale expression and purification trial was performed for GFP-tagged constructs with results most promising for β5 full length GFP, αv full length GFP, β6 full length GFP and LAP truncated GFP (Section 3.4.2.1). For all other constructs tested, results suggested that the expression and/or purification process required further optimisation but this was not possible due to time restraints. Detergent and buffer screens were utilised to identify the optimal purification conditions for β5 full length GFP, αv full length GFP, β6 full length GFP and LAP truncated GFP (Section 3.4.2.2 and Section 3.4.2.3), yet despite these advances one of the major issues encountered was low protein yield which made assessment of the quality of the purified protein difficult and the newly synthesised proteins did not meet the sample input requirement for protein interaction studies.

Monitoring GFP fluorescence in Expi293 cells 48 hours post-transfection showed a low abundance of recombinant protein (Appendix Figure 9.1) and indicated that the low protein yield obtained after purification was likely due to an insufficient level of initial protein expression. Suboptimal growth (culture) conditions is a common factor that can impact protein expression efficiency (338). For efficient expression, the rate of protein synthesis should match the secretory capacity of the host (339) and the conditions utilised during transfection and expression are important variables which

can affect this. The transfection and expression conditions used here were adapted from the MPL optimised protocol (331). However, it is important to consider that optimal conditions are protein dependent. Therefore, the factors to consider for future optimisation experiments are summarised below (Table 3.21).

Table 3.21: Parameters to optimise for the transient expression of a target proteins in Expi293 cells

Transfection	Expression
Flask and seeding density	Expression medium
Cell density and viability for transfection	Cell growth (temperature, shake speed, duration)
Transfection medium	
Transfection agent	Additives/ supplements (type, time of addition)
Plasmid DNA (amount and purity)	

Table adapted from article "Transient Transfection and Expression of Eukaryotic Membrane Proteins in Expi293F Cells and Their Screening on a Small Scale: Application for Structural Studies" with permission (331).

Upon reflection, the difficulties of mammalian cell-based protein expression and particularly the low level of protein expression observed, may perhaps explain why the commercially available proteins are of a low concentration and quantity. The commercial availability of only truncated constructs with fusion/affinity tags may be a further indication of the difficulties in purifying these proteins with the shortened and affinity tagged proteins being designed (340).

The work in this chapter was partly successful as several binding partners of interest were purified from the expression trials but were of low yield and consequently could not be used for protein interaction studies. Future work would therefore focus on optimising the growth (culture) conditions for mammalian protein expression, however it should be noted that this would be very costly in terms of time and reagents. In contrast to the above, expression and purification of galectin-3 and galectin-1 in a bacterial expression system was successful and subsequently, both

galectins were produced on a large-scale. It was demonstrated by gel electrophoresis, AUC and SEC-MALS that both purified proteins were highly pure with minimal contaminants (Section 3.4.3). Furthermore, the estimated MW calculated by AUC and SEC-MALS analysis was within an acceptable range to that of the expected MW. By both methods, the calculated MW was slightly lower than anticipated for monomeric galectin-3 (27 kDa) and dimeric galectin-1 (30 kDa) although there are several possible explanations for this. In AUC, the input values (buffer density, buffer viscosity, partial specific volume) as well as sample purity, influence the output values (frictional ratio, sedimentation coefficient and MW) and subsequently there is systematic error. The greatest difference between the calculated vs expected MW was observed for galectin-1. However the likely explanation for this is that the galectin-1 purified was a population of both monomeric and dimeric protein. In AUC c(s) analysis the frictional ratio is fitted for the distribution as a whole, not for each individual peak. Therefore if more than one peak is present the frictional ratio fitted is a population weighted average, this results in a broader peak and the MW calculated is less accurate. Similarly in SEC-MALS, the input value for the refractive index increment influences the output value (MW) and there is a % uncertainty in the MW value estimated (Appendix Table 9.2). Considering all of the above, there is a rational explanation for the lower MW calculated which provides confidence in the results generated.

Following the successful purification of galectin-3 and galectin-1, their biological activity was confirmed at Galecto Biotech (PhD industrial sponsor) by galectin-induced haemagglutination and apoptosis (data not shown). Both the purified

galectin-3 and galectin-1 were shown to agglutinate RBCs and induce apoptosis of Jurkat cells (immortalised T lymphocyte cells) in a concentration-dependent manner. The detection of biological activity in these purified proteins confirmed that both could be used for the protein binding studies designed to determine whether galectins directly interact with components of the TGF-β1 signaling pathway.

Summary: Recombinant human galectin-3 and galectin-1 were produced in large quantities by bacterial (*E. coli*) expression whereby both purified proteins were highly pure and subsequently showed biological activity. In contrast, attempts to express and purify the potential binding partners including αvβ1, αvβ5, αvβ6, LAP, LTBP1, TGFβRI and TGFβRII were less successful as all of the proteins produced by mammalian expression were of low abundance. Consequently, the quantity of purified protein was too low for protein characterisation or protein interaction studies.

4. Investigating the direct binding of galectin-3 and galectin-1 to components of the TGF-β1 signaling pathway by surface plasmon resonance

4.1. Introduction

As previously described, both galectin-3 and galectin-1 have been implicated in TGF-β1 signaling in IPF (Section 1.4). Current literature highlights the particular importance of galectin-3 in epithelial TGF-β1 signaling independent of SMAD (284). However, a role for galectin-3 in canonical TGF-β1 signaling has also been proposed in fibroblasts (Section 1.4.6). Briefly, LPA induces TGF-β1 activation and SMAD signaling via the integrins αvβ1 and αvβ5 in HLFs (310). This LPA-induced TGF-β1 activity is inhibited by GB0139 or galectin-3 inhibitors which target the galectin CRD (310). In particular, the cell-impermeable galectin-3 inhibitor GB0149 reduces TGF-β1 activation and downstream SMAD signaling in a concentration-dependent manner (313). This suggests that endogenous extracellular galectin-3 is essential for the initial activation of TGF-β1 in HLFs. Galectin-3 is also proposed to be involved in canonical TGF-β1 signaling directly at the receptor level as the galectin-3 inhibitor GB0139 prevents TGF-β1-induced SMAD2 phosphorylation (310). Taken together, it has been hypothesised that galectin-3 may be involved in TGF-β1 activation in fibroblasts via a direct interaction with αv integrins, components of the large latent complex or TGF-β1 receptor subunits.

Integrins are glycoproteins with glycosylation of both the α and β integrin subunits important for heterodimerisation, activation and biological function (341, 342). Although data is not available in fibroblasts, the current literature shows that galectins (specifically galectin-1, galectin-3 and galectin-8) are capable of regulating

integrin activity by interacting with their surface glycans. Galectin-3 regulates fibronectin polymerisation (fibrillogenesis) in mammary epithelial tumor cells and lamellipodia formation in corneal epithelial cells by binding to integrins α5β1 and α3β1, respectively (343-345). Exogenous galectin-3 has also been shown to regulate β1 integrin lateral mobility and clustering on the surface of Henrietta Lacks (HeLa) cells (346). Similarly, the integrins α4β1, α5β1 and α4β7 have been identified as galectin-1 binding partners (carbohydrate-dependent) involved in synapse formation (pre-B/stromal cells) (347). The β1 subunit is suitably glycosylated for protein-glycan binding as it carries highly branched N-glycans (β1,6-branched N-glycans) that are the preferred galectin ligand (348). To date, the glycosylation profiles of integrins α5β1, α3β1 and αvβ3 are the most well studied (349). However N-glycan sites have been reported or predicted for integrin subunits αv (13 sites), β1 (12 sites), β5 (8 sites) and β6 (9 sites) from the NXS/T consensus sequence (350, 351). There is currently no consensus sequence motif to enable the prediction and identification O-glycans.

Similar to integrins, the literature demonstrates that galectin-3 can interact with and regulate the activity of growth factor receptors. In mammary carcinoma cells, galectin-3 binds to both the epidermal growth factor receptor (EGFR) and TGF-β1 receptor (259). These interactions can be inhibited by N-glycan removal or competitive inhibition of the galectin-3 CRD. N-glycan removal decreases cytokine signaling evidenced by reduced phosphorylated Erk (Erk-P) and Smad2 nuclear translocation following epidermal growth factor (EGF) or TGF-β1 stimulation, respectively. In these studies nuclear translocation is re-established by restoring glycan levels or by blocking endocytosis. Galectin-3 binding to EGFR and the TGF-β1

receptor has subsequently been suggested to prevent receptor endocytosis thereby retaining their cell surface expression. Supporting this, knockdown of galectin-3 (siRNA-Gal-3) in A549 cells similarly effects surface TGF-β1 receptor expression and downstream non-canonical signaling (Wnt/β-catenin) (284). In contrast with αv integrins, the TGF-β1 receptor subunits TGFβRI and TGFβRII have only one and three N-glycan sites, respectively (259). However, glycosylation is evidently essential for the TGF-β1 receptor biological function with its cell surface transportation also being N-glycan mediated (352).

At present, the only evidence suggesting that galectin-3 directly interacts with LTBP1 is from solid-phase binding assays performed by Parmar, N (313). There is no additional evidence from the literature that suggests galectin-3 and/or galectin-1 can directly interact with components of the large latent complex (LTBP1 or LAP). However, 6 and 3 N-linked glycan sites have been reported for LTBP1 and LAP, respectively (353).

As previously described, SPR is a biophysical method which can be used to investigate protein interactions (Section 3.1). It has been demonstrated as being a suitable method for investigating galectin binding interactions inclusive of proteins and inhibitors with both galectin-1 and galectin-3 shown to be bound to Factor VIII in a reversible and dose-dependent manner (354). Similarly, galectin-3 bound to both jagged canonical Notch ligand 1 (JAG1) and delta like canonical Notch ligand 4 (DLL4) ligands which could be prevented by the presence of lactose (355). The binding of galectin-3 to galectin-3 binding protein (G3BP) and the effect of galectin-3 inhibitors

on binding has additionally been determined by SPR analysis (356). More recently, SPR was used to establish the binding capacity between galectin-3 and programmed cell death protein 1 (PD-1) or programmed cell death ligand 1 (PD-L1) (357). The optical technique has proven appropriate for detecting galectin binding interactions and can therefore be used to assess the binding of galectin-3 and/or galectin-1 to proteins in the TGF-β1 pathway.

4.2. Aims

Galectin-3 is involved in TGF-β1 activation and promotes TGF-β1-mediated downstream signaling pathways in fibroblasts, however the mechanism of action is not precisely defined. Subsequently, the aims of this chapter are as follows:

- To investigate whether galectin-3 and/or galectin-1 interact directly with αvβ1, αvβ5 and αvβ6 integrins
- To investigate whether galectin-3 and/or galectin-1 interact directly with the TGF-β1 receptor (TGFβRI/ TGFβRII)
- To determine whether galectin-3 and/or galectin-1 can directly bind to two components of the large latent TGF-β1 complex, LAP and LTBP1

4.3. Methods

4.3.1. Surface Plasmon Resonance

Principle of the method: SPR is a method used for the biophysical characterisation of molecular interactions. A biosensor chip with four individual channels (flow cells) is coated with a thin layer of gold and a dextran matrix onto which a ligand (protein of interest) can be covalently immobilised via amine coupling. As the dextran surface is negatively charged, the ligand is diluted in immobilisation buffer with a pH below the protein isoelectric point (pI), this facilitates electrostatic attraction to the sensor chip surface. Once covalently bound, an analyte (binding partner of interest) can be injected into the channel of the chip and is brought into close contact with the immobilised ligand. Detection of ligand-analyte interactions is based on the Kretschmann configuration whereby polarised light is directed through a glass prism at the glass-sample interface which gets reflected onto a detector. The angle at which the light is reflected is called the incident angle (or resonance angle) whereby light is absorbed by electrons in the gold surface layer causing them to resonate. If any analyte binds to the immobilised ligand there is a change in the mass near the gold surface which causes the incident angle to shift. This then gets detected by the detector and outputted as an SPR sensorgram. The interaction between the analyte and ligand is called the response and is measured in resonance units (RU) as a function of time. A sensorgram is a plot of the response against time which shows the progress of an interaction and gets displayed during the course of an analysis. To evaluate non-specific analyte binding, analyte is also injected into a reference channel with no ligand immobilised. Any binding to this channel then gets subtracted from the ligand-analyte reading. The ligand-analyte response is directly proportional

to the concentration of ligand immobilised on the surface and the MW of the analyte. If the sensorgram reaches steady state (equilibrium) during the sample injection then the affinity (equilibrium dissociation constant or K_d, units M) can be determined. If there is sufficient curvature during both association (K_{on}, units $M^{-1}s^{-1}$) and dissociation (K_{off}, units s^{-1}) phases then interaction kinetics can be determined. SPR Biacore evaluation software assumes a 1:1 stoichiometric ratio (ratio of analyte to ligand).

Wizard Application: SPR experiments were performed as a wizard run on the Biacore T200 (Cytiva) in HBS-P+ Buffer 1× (+/- 1 mM $MnCl_2$) running buffer. Prior to commencing experiments the instrument was successfully primed (three times) and underwent routine normalisation with BIAnormalizing solution (120 µL) to ensure a uniform signal during the assay. Protein immobilisation onto a Series S Sensor Chip CM5 was performed manually via amine coupling. The channel surface was activated by injection of a 1:1 pre-mixed solution of 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and 0.1 M N-hydroxysuccinimide (NHS) both prepared in Milli-Q water (injection: 10 µL/min, 7 minutes) to give reactive succinimide esters. Commercially available recombinant proteins (αvβ1, αvβ5, αvβ6, TGFβRI, TGFβRII, LAP and LTBP1) were then diluted in 10 mM sodium acetate immobilisation buffer (buffer pH protein-dependent) and immobilised onto individual channels (lysine ε-amino groups react with esters). The amount of immobilised ligand required was calculated using the formula:

$$\text{Immobilised ligand level (RU)} = \left(\frac{\text{Ligand MW}}{\text{Analyte MW}}\right) \times \text{Analyte binding capacity (Rmax)}$$

The analyte with the smallest MW (galectin-1) was used for the calculation and an analyte binding capacity (Rmax) of 100. Once the desired immobilised ligand level was reached (protein-dependent), excessive reactive groups were deactivated by injection of 1 M ethanolamine-HCL, pH 8.5 (ethanolamine) (injection: 10 µL/min, 7 minutes). When the signal was uniform, a series of 2-fold dilutions of galectin-3 or galectin-1 (dialysed against running buffer) were injected over the sensor chip surface at a 30 µL/min flow rate, 20°C, contact time: 120 seconds (dissociation time: protein-dependent). If required the sensor chip surface was regenerated with 5 mM EDTA between injections (protein-dependent). All sensorgrams were baseline-corrected using a reference flow cell with blank immobilisation prior to data analysis. Protein-dependent run settings are summarised below (Table 4.1).

Table 4.1: Protein-dependent SPR run settings

Protein	Immobilisation Buffer (pH)	Dissociation Time (seconds)	Regeneration
αv integrin	4.0	1200 (Glycosylated) 600 (Deglycosylated or compound)	EDTA
TGFβRI	5.0	300	N/A
TGFβRII	4.5		
LAP	5.0		
LTBP1	5.0		

Immobilisation buffer pH based on protein isoelectric point. Dissociation time and chip regeneration dependent on strength of binding interaction.

Binding data was analysed in GraphPad Prism v9, if steady-state was reached then the K_d value, units M (analyte concentration at which half of the ligand is occupied by the analyte at equilibrium) and Bmax (maximum specific binding) were determined by non-linear regression (binding saturation) - one-site specific binding. If steady-state was not reached then the minimum number of binding sites was estimated from the response at the highest analyte concentration tested (5000 nM). This was calculated using the formula (355):

$$\text{Minimum number of binding sites} = \left(\frac{\text{Maximum binding response}}{\text{Rmax}}\right)$$

For determination of interaction kinetics, on-off rate maps were plotted in EVILFIT from the sensorgram data (358). K_d values were not obtained from EVILFIT analysis, the on-off rate maps were plotted to solely demonstrate the size of the affinity distribution and the binding heterogenicity.

Protein Deglycosylation: To determine the effect of integrin (αvβ1, αvβ5, αvβ6) or TGFβRII deglycosylation on galectin binding, glycans were removed from recombinant proteins and confirmed via SDS-PAGE (Appendix Figure 9.2). For full deglycosylation (N- and O- linked glycan removal) of integrin proteins (αvβ1, αvβ5, αvβ6) and TGFβRII the Protein Deglycosylation Mix II (New England Biolabs, P6044) non-denaturing reaction conditions were adapted. 2.5 µL of both 10X deglycosylation mix buffer 1 and protein deglycosylation mix II were added to 20 µL of glycoprotein. After mixing, the reaction was incubated at 25°C for 30 minutes followed by a 16 hour incubation at 37°C. For partial deglycosylation (N-linked glycan removal) of integrin proteins (αvβ1, αvβ5, αvβ6) and TGFβRII the PNGase F (New England Biolabs, P0704) non-denaturing reaction conditions were adapted. 2 µL of glycobuffer 2 (10X) and 4 µL of peptide:N-glycosidase F (PNGase F) were added to 18 µL of glycoprotein. After mixing, the reaction was incubated at 37°C for 16 hours. Deglycosylated proteins were then immobilised as described above and the same galectin serial dilution injected.

Solution Competition Assay: Additionally, SPR was used to assess the effects of small molecule galectin inhibitors (GB0139, GB1107 or GB1490) on galectin binding to integrins (αvβ1, αvβ5, αvβ6) or TGFβRII. Glycosylated recombinant proteins were immobilised as described above and galectin inhibitors serially diluted 2-fold in running buffer containing 625 nM galectin across the dilution series. The serial dilution of galectin inhibitor (+ constant galectin concentration) was then injected across all flow paths at a 30 µL/min flow rate, 20^{o}C, contact time: 120 seconds (dissociation time: protein-dependent). The response values were normalised with respect to the highest RU response (0 nM compound). Competition binding curves were analysed in GraphPad Prism v9 and inhibitory concentration 50 (IC50) values (inhibitor concentration where the response (or binding) is reduced by half) determined by non-linear regression (binding saturation). The data was plotted on both a linear and log scale (data not shown) to determine which model to fit to the data. If the full curve was defined (GB0139 and GB1107) then the IC50 value was determined by specific binding with hill slope as the R^2 value was superior. If the curve was not defined (GB1490) then the IC50 value was determined by one-site specific binding.

4.4. Results

4.4.1. Galectin-3 physically interacts with αv integrins

Galectin-3 bound to the recombinant human integrins αvβ1 (Figure 4.1A), αvβ5 (Figure 4.1B) and αvβ6 (Figure 4.1C) in a concentration-dependent manner by SPR. The binding responses were 2-fold higher for αvβ1 and αvβ5 compared with αvβ6. The highest binding responses recorded for all three integrins were considerably higher than the theoretical maximum response (~135 Rmax) for analyte-ligand binding 1:1. It was not possible to saturate the immobilised ligand at the concentrations tested as evidenced by the lack of a plateau even at high concentrations. The binding to all three integrins was glycosylation-dependent with enzymatic removal of all N-linked and common O-linked glycans inhibiting the binding interaction (Figure 4.1A - C). Removal of N-linked glycans alone only partially inhibited the binding interaction. Galectin-3 binding data was analysed in EVILFIT for αvβ1 (Figure 4.1D), αvβ5 (Figure 4.1E) and αvβ6 (Figure 4.1F) to obtain a heat map of the association (K_{on}, units $M^{-1}s^{-1}$) and dissociation (K_{off}, units s^{-1}) distribution across the different analyte concentrations tested ($K_d = K_{off}/K_{on}$). The size of the affinity distribution (coloured circle) reflects the heterogenicity of the binding population (due to orders of magnitude covered) and the colour scalebar represents the abundance at each K_d/k_{off} combination (high abundance shown in red and low abundance in black). Results showed one heterogenous binding population (coloured circle) in which there are numerous binding events with a different K_d/k_{off} combination (abundance of each K_d/k_{off} combination represented by colour as described above). Together, these results showed that galectin-3 binding was heterogeneous for all three integrins. A second low abundance population was visible

for αvβ1 (Figure 4.1D) and αvβ6 (Figure 4.1F) (grey circle) which represented background noise only.

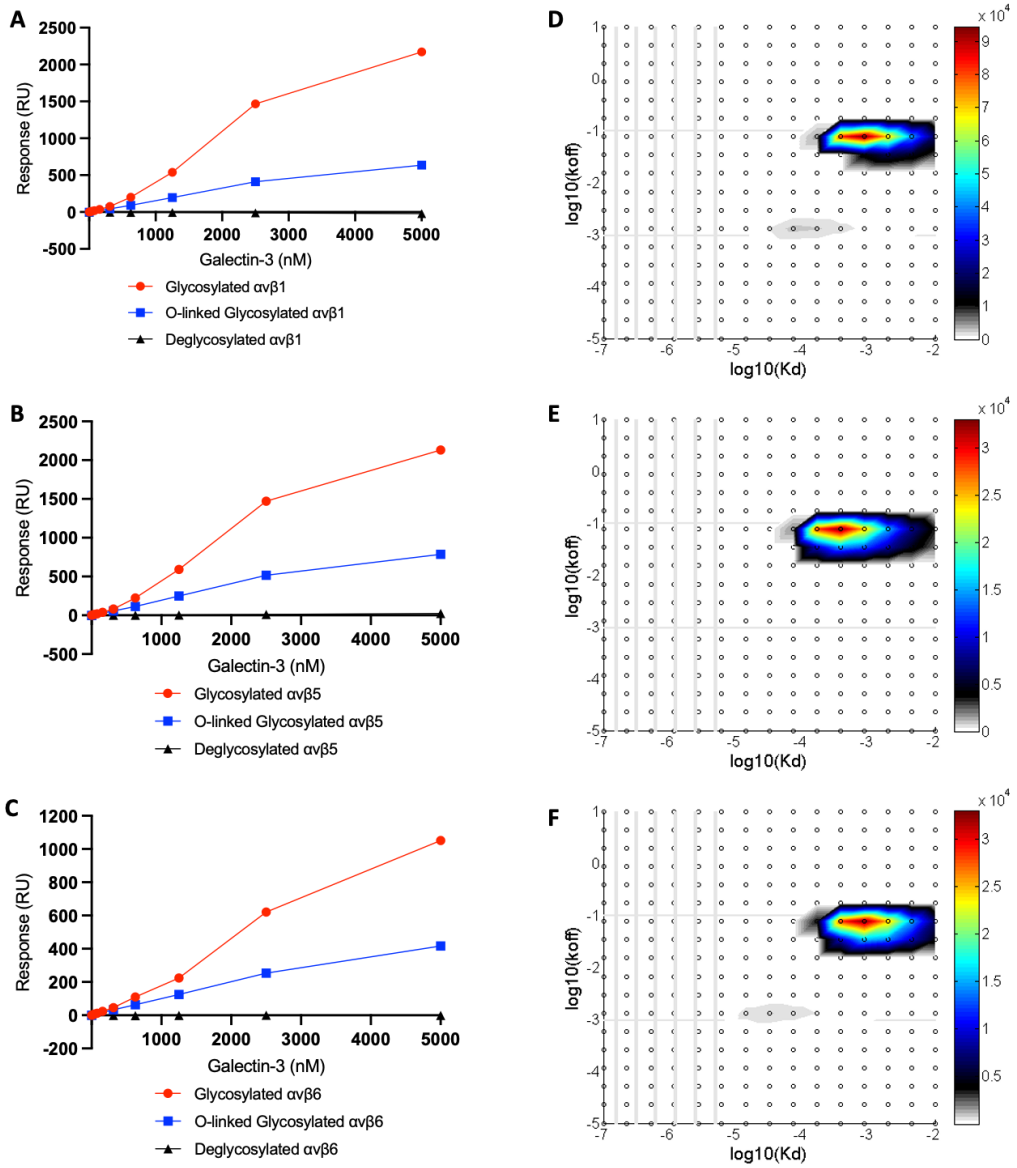


Figure 4.1: Galectin-3 binding to αv integrins by SPR
SPR data of soluble galectin-3 binding to glycosylated or deglycosylated αv integrins: ***(A)*** *αvβ1,* ***(B)*** *αvβ5 and* ***(C)*** *αvβ6 immobilised on the surface of a Series S sensor chip CM5 (~1000 RU). Sequential injections of galectin-3 protein (19.5 - 5000 nM) were performed, SPR signals were measured in response units (RU) and baseline subtracted. Binding response values plotted in GraphPad Prism with connecting line/curve shown.* ***(D-F)*** *Two-dimensional rate and affinity constant distributions plotted in EVILFIT from the fit of all experimental kinetic traces measured by SPR.*

4.4.2. Galectin-1 physically interacts with αv integrins

Galectin-1 also bound to the recombinant human integrins αvβ1 (Figure 4.2A), αvβ5 (Figure 4.2B) and αvβ6 (Figure 4.2C) by SPR and exhibited concentration-dependent binding. Again, the binding responses were considerably higher for αvβ1 and αvβ5 than αvβ6. The highest binding responses recorded for all three integrins were greater than the theoretical maximum response (~75 Rmax) for analyte-ligand binding 1:1. Consequently, the immobilised ligand was not saturated at the concentrations tested as evidenced by the lack of a plateau. Like galectin-3, the binding of galectin-1 to all three integrins was glycosylation-dependent with enzymatic removal of all N-linked and common O-linked glycans inhibiting binding (Figure 4.2A - C). Removal of N-linked glycans alone only partially inhibited the binding interaction as similarly observed for galectin-3. Galectin-1 binding data was analysed in EVILFIT for αvβ1 (Figure 4.2D), αvβ5 (Figure 4.2E) and αvβ6 (Figure 4.2F) and the results showed two distinct binding populations (Figure 4.2D - F). Within the predominant binding population (coloured circle) there are numerous binding events with a different K_d/k_{off} combination (abundance of each K_d/k_{off} combination represented by colour). There was also a second binding population (black circle) that is heterogenous (represented by size) but of low abundance (represented by colour). Collectively, these results showed that galectin-1 binding was heterogeneous for all three integrins.

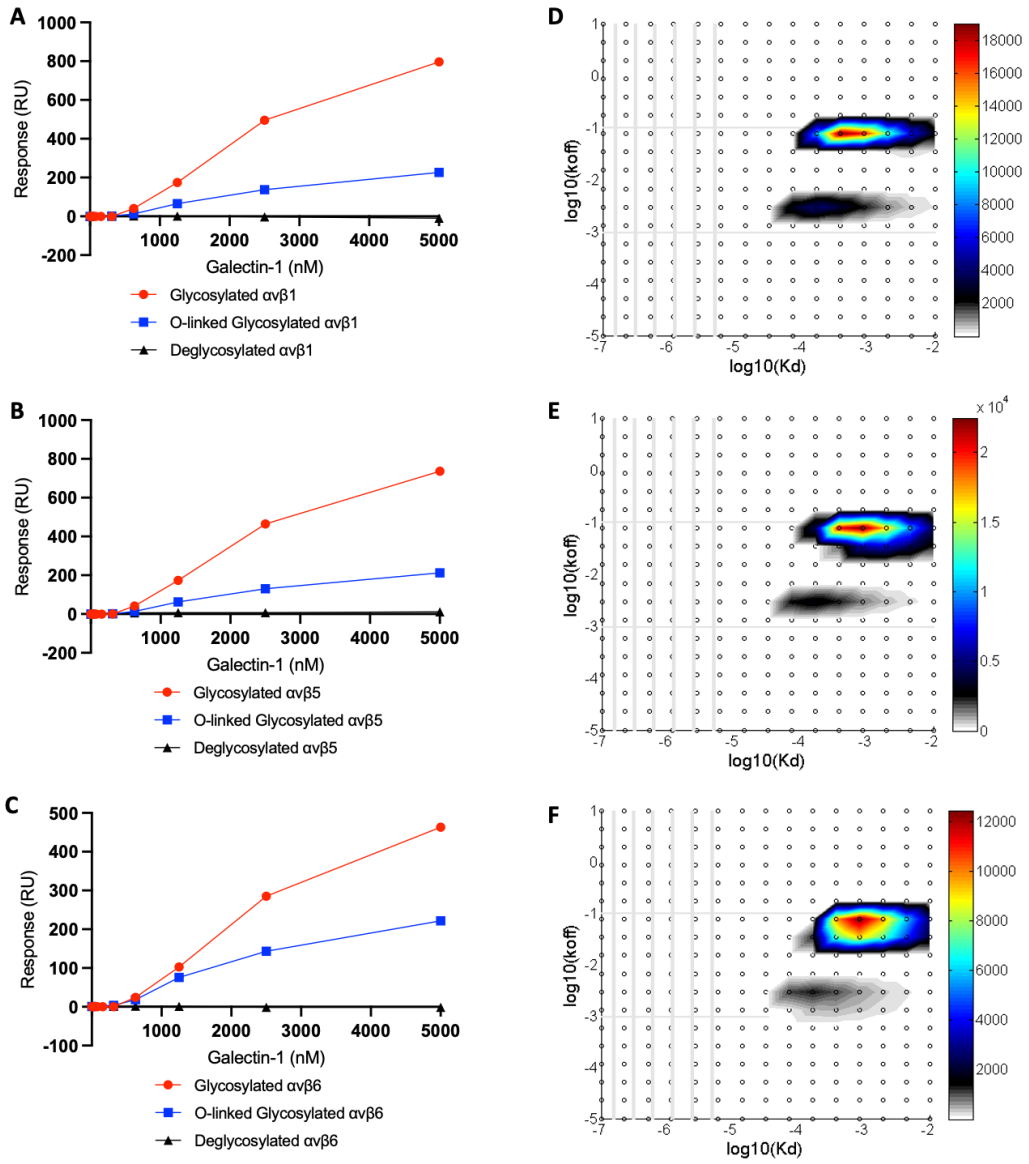


Figure 4.2: Galectin-1 binding to αv integrins by SPR
SPR data of soluble galectin-1 binding to glycosylated or deglycosylated αv integrins: ***(A)*** *αvβ1,* ***(B)*** *αvβ5 and* ***(C)*** *αvβ6 immobilised on the surface of a Series S sensor chip CM5 (~1000 RU). Sequential injections of galectin-1 protein (19.5 - 5000 nM) were performed, SPR signals were measured in RU and baseline subtracted. Binding response values plotted in GraphPad Prism with connecting line/curve shown.* ***(D-F)*** *Two-dimensional rate and affinity constant distributions plotted in EVILFIT from the fit of all experimental kinetic traces measured by SPR.*

4.4.3. Small molecule galectin inhibitors prevent galectin-3 binding to αv integrins

Solution competition binding assays showed that target inhibition of the galectin-3 CRD with the galectin-3 inhibitor GB1107 inhibited the binding of galectin-3 to αvβ1 (Figure 4.3A), αvβ5 (Figure 4.3B) and αvβ6 (Figure 4.3C). Concentration-dependent inhibition of galectin-integrin binding was demonstrated for all 3 αv integrins. At the

7500 nM concentration, galectin-3 was saturated with GB1107 (black line) and complete inhibition of binding was observed (Figure 4.3A - C). In comparison, galectin-3 was not completely saturated with the galectin-1 inhibitor GB1490 (red line) at the compound concentrations tested (Figure 4.3A - C). At the maximum compound concentration used (15000 nM), galectin-3 was ~70% saturated with GB1490. Like GB1107, close to complete binding inhibition was detected (~96%) with the higher affinity galectin-3 inhibitor GB0139 (blue line) but at a lower compound concentration (940 nM vs 1900 nM for GB0139 and GB1107, respectively) (Figure 4.3D - F). IC50 values determined by non-linear regression analysis were lower for GB0139 and GB1107 compared to GB1490 (Figure 4.3A - F). This was expected as GB0139 (K_d = 0.0023 µM) and GB1107 (K_d = 0.037 µM) are higher affinity galectin-3 inhibitors than GB1490 (K_d = 2.7 µM) (Appendix Table 9.1). Due to the galectin-3 concentration used in the assay (625 nM) the theoretical lower limit for IC50 is 312.5 nM, however the lower IC50 values obtained here are within 2-fold and were consistent across all 3 integrins. It is not possible to directly compare the IC50 values for GB1107 with GB0139 as both values are below the theoretical lower limit, these were also independent experiments performed with different batches of galectin-3. However the IC50 values do suggest that the concentration of active galectin-3 was less than 625 nM in both experiments.

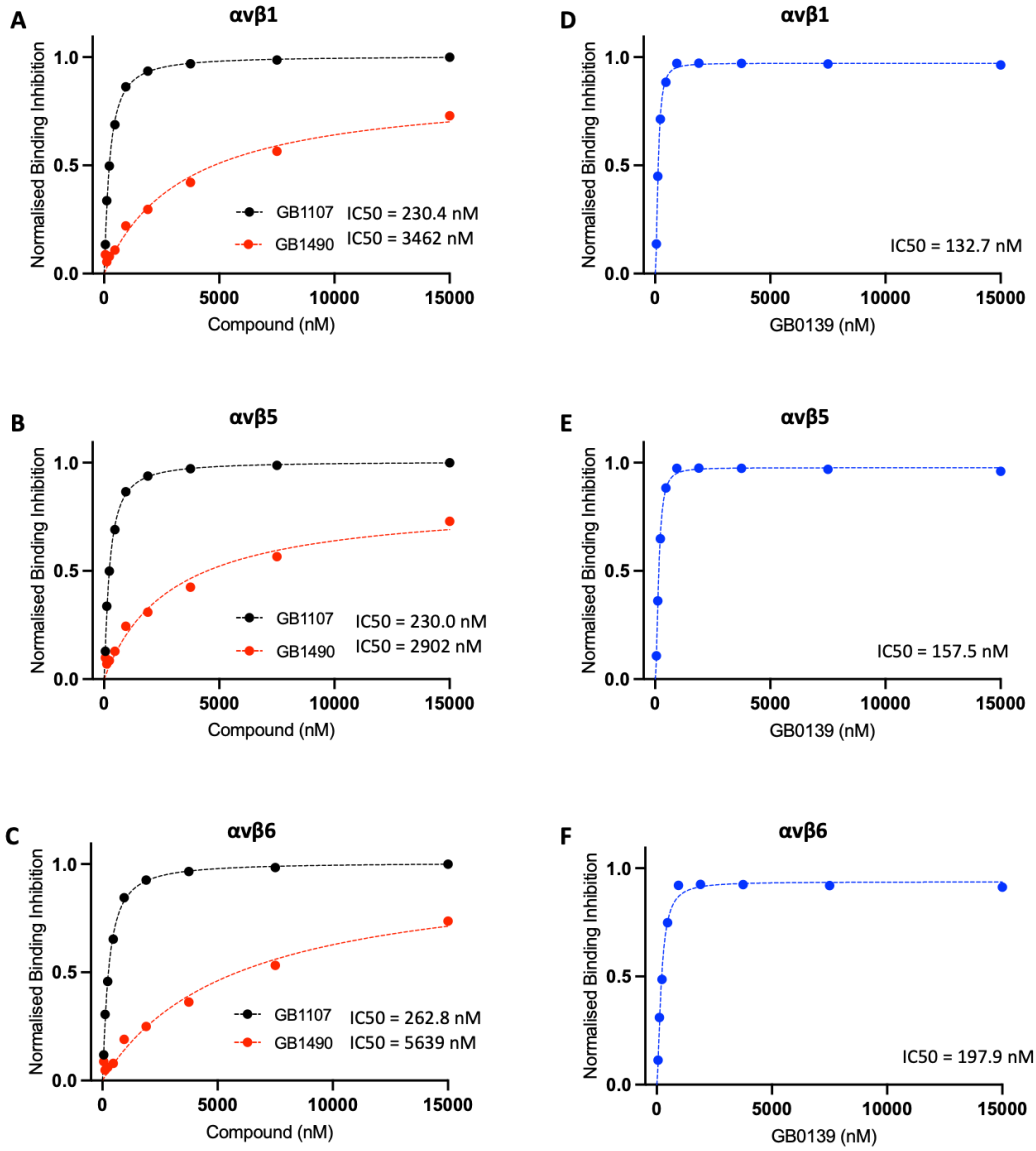


Figure 4.3: Competitive inhibition of galectin-3 binding to glycosylated αv integrins
Solution competition binding assays performed with the galectin-3 inhibitor GB1107 (black) or the galectin-1 inhibitor GB1490 (red) for αv integrins: ***(A)*** *αvβ1,* ***(B)*** *αvβ5 and* ***(C)*** *αvβ6 in the presence of galectin-3 at 625 nM. Solution competition binding assays were also performed with the higher affinity galectin-3 inhibitor GB0139 (blue) for αv integrins:* ***(D)*** *αvβ1,* ***(E)*** *αvβ5 and* ***(F)*** *αvβ6 in the presence of galectin-3 at 625 nM. Response values are normalised with respect to the highest binding response (0 nM compound) and competitive inhibition graphs plotted in GraphPad Prism. IC50 values were calculated by non-linear regression analysis (Specific binding with hill slope: GB1107 and GB0139, One-site specific binding: GB1490).*

4.4.4. Galectin-1 inhibition by GB0139 prevents binding to αv integrins

Inhibition of galectin-1 with GB0139 inhibited the binding of galectin-1 to αvβ1 (Figure 4.4A), αvβ5 (Figure 4.4B) and αvβ6 (Figure 4.4C) in a concentration-dependent manner as assessed by solution competition binding assays. At the 3750

nM concentration, galectin-1 was saturated with GB0139 and complete binding inhibition observed for all 3 αv integrins. IC50 values determined by non-linear regression analysis were similar across all three integrins. The GB0139 IC50 values were higher for galectin-1 than galectin-3 (Figure 4.3), this was expected as GB0139 has a higher affinity for galectin-3 (K_d = 0.0023 µM) than galectin-1 (K_d = 0.012 µM) (Appendix Table 9.1).

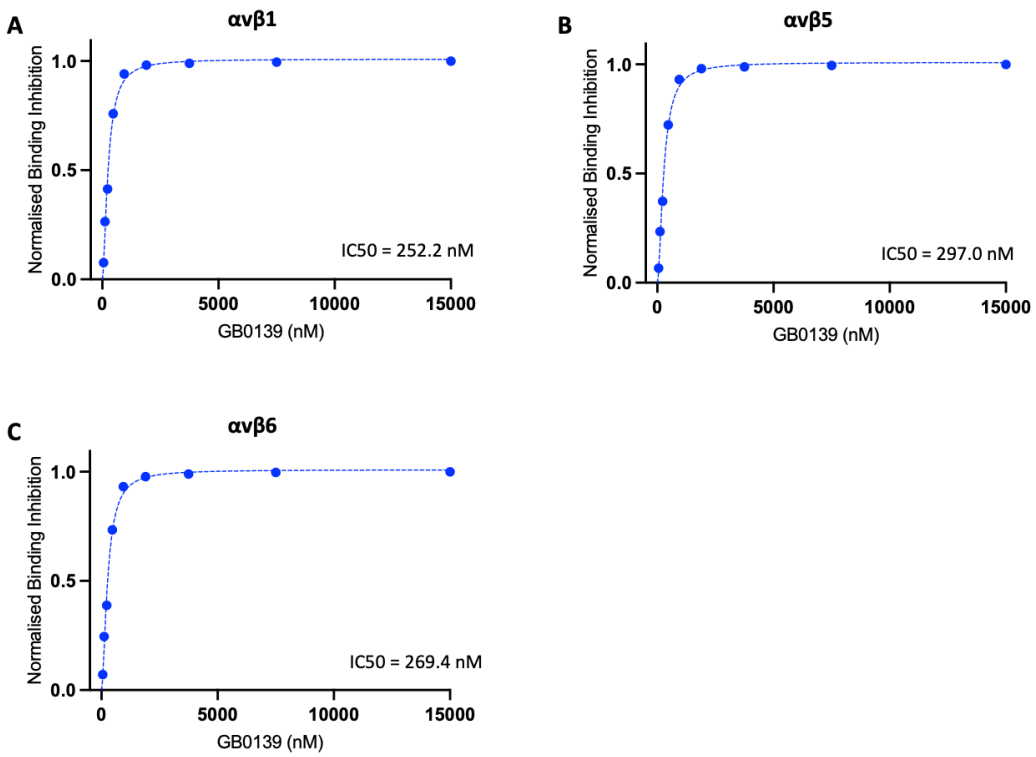


Figure 4.4: Competitive inhibition of galectin-1 binding to glycosylated αv integrins
Solution competition binding assays performed with the higher affinity galectin-3 inhibitor GB0139 (blue) for αv integrins: ***(A)*** *αvβ1,* ***(B)*** *αvβ5 and* ***(C)*** *αvβ6 in the presence of galectin-1 at 625 nM. Response values are normalised with respect to the highest RU response (0 nM compound) and competitive inhibition graphs plotted in GraphPad Prism. IC50 values were calculated by non-linear regression analysis (Specific binding with hill slope).*

4.4.5. Galectin-3 physically interacts with TGFβRII and binding is prevented by galectin inhibition

Galectin-3 bound to recombinant human TGFβRII protein in a concentration-dependent manner by SPR (Figure 4.5A). The highest binding response recorded was considerably higher than the theoretical maximum response (~251 Rmax) for analyte-

ligand binding 1:1. At the galectin-3 concentrations tested, saturation of the immobilised ligand was almost reached as evidenced by the plateauing of the curve at the highest concentrations. Subsequently, non-linear regression analysis was performed in GraphPad Prism for the glycosylated TGFβRII binding response, this yielded a K_d value of 5288 nM and a Bmax of 1473 RU. It is difficult to determine the biological relevance of this K_d value *in vivo* given that the binding interaction is not 1:1. From the Bmax (1473 RU) and the Rmax (~251) is was calculated (Bmax/Rmax) that a minimum of 6 individual galectin-3 (27 kDa) proteins are involved in the binding interaction. Galectin-3 binding to TGFβRII was glycosylation-dependent with enzymatic removal of all N-linked and common O-linked glycans inhibiting the binding interaction (Figure 4.5A). However, a similar decrease in the level of binding was observed after the removal of only N-linked glycans. EVILFIT results showed one heterogenous binding population (coloured circle) in which there are numerous binding events with a different K_d/k_{off} combination (abundance of each K_d/k_{off} combination represented by colour) (Figure 4.5B). A low abundance population was also visible (grey circle) which represented background noise only. Inhibition of the galectin-3 CRD with GB1107 inhibited galectin-3 binding to TGFβRII in a concentration-dependent manner (Figure 4.5C). At 15000 nM galectin-3 was saturated with GB1107 and complete binding inhibition observed. Like GB1107, close to complete binding inhibition was observed (~96%) with the galectin-1 inhibitor GB1490 (15000 nM). However, the curve for GB1490 is shifted to the right of GB1107 which demonstrates the difference in compound affinity for galectin-3, this is reflected in the IC50 values which are lower for GB1107 than GB1490. This was expected as GB1107 (K_d = 0.037 µM) is a higher affinity galectin-3 inhibitor than

GB1490 (K_d = 2.7 µM) (Appendix Table 9.1). Concentration-dependent inhibition was also observed in the presence of GB0139 with full saturation being reached (Figure 4.5D). The steepness of the curve demonstrates how GB0139 has high affinity for galectin-3 (K_d = 0.0023 µM) which is reflected in the IC50 value. Again it is not possible to directly compare the IC50 values for GB1107 with GB0139 as these were also independent experiments. However the IC50 values suggest that the concentration of active galectin-3 was higher than 625 nM.

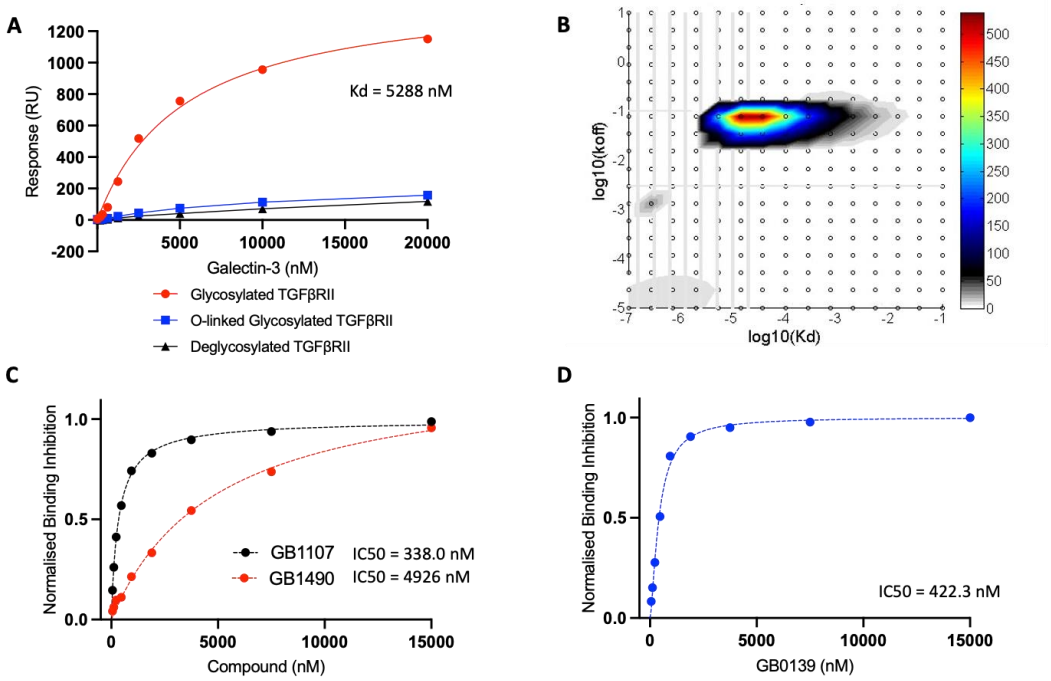


Figure 4.5: Galectin-3 binding to TGFβRII and competitive inhibition
(A) *SPR data of soluble galectin-3 binding to glycosylated TGFβRII immobilised on the surface of a Series S sensor chip CM5 (~400 RU). Sequential injections of galectin-3 protein (156.25 nM - 20 µM) were performed, SPR signals were measured in RU and baseline subtracted. Binding response values plotted in GraphPad Prism and non-linear regression analysis (Binding-Saturation, One Site-Specific Binding) performed for glycosylated TGFβRII. Connecting line/curve shown for O-linked glycosylated and deglycosylated TGFβRII.* ***(B)*** *Two-dimensional rate and affinity constant distributions plotted in EVILFIT from the fit of all experimental kinetic traces measured by SPR.* ***(C)*** *Solution competition binding assays were performed with the galectin-3 inhibitor GB1107 (black) or the galectin-1 inhibitor GB1490 (red) in the presence of galectin-3 at 625 nM.* ***(D)*** *Solution competition binding assays were also performed with the higher affinity galectin-3 inhibitor GB0139 (blue) in the presence of galectin-3 at 625 nM. Response values are normalised with respect to the highest RU response (0 nM compound) and competitive inhibition graphs plotted in GraphPad Prism. IC50 values were calculated by non-linear regression analysis (Specific binding with hill slope: GB1107 and GB0139, One-site specific binding: GB1490).*

4.4.6. Galectin-1 physically interacts with TGFβRII and binding is prevented by galectin inhibition

Galectin-1 bound to the recombinant human TGFβRII protein in a concentration-dependent manner by SPR (Figure 4.6A). The highest binding response recorded was higher than the theoretical maximum response (~140 Rmax) for analyte-ligand binding 1:1. It was not possible to saturate the immobilised ligand at the concentrations tested as evidenced by the lack of a plateau even at high concentrations. In contrast TGFβRII was saturated by galectin-3 at this analyte concentration range (Figure 4.5A), this shows that galectin-3 has a higher binding affinity for TGFβRII than galectin-1. Galectin-1 binding to TGFβRII was glycosylation-dependent with enzymatic removal of all N-linked and common O-linked glycans inhibiting the binding interaction (Figure 4.6A). Removal of N-linked glycans alone only partially inhibited the binding interaction. Galectin-1 binding data was similarly analysed in EVILFIT to obtain a heat map of the association (K_{on}, units $M^{-1}s^{-1}$) and dissociation (K_{off}, units s^{-1}) distribution across the different analyte concentrations tested ($K_d = K_{off}/K_{on}$). Here two binding populations are present but are only partially visible on the heatmap which is evidence that the K_d/k_{off} range for galectin-1 and TGFβRII is different compared to all of the above (Figure 4.6B). However the results still showed that within both populations there are numerous binding events with a different K_d/k_{off} combination (abundance of each K_d/k_{off} combination represented by colour). GB0139 (K_d = 0.012 µM) completely inhibited the binding of galectin-1 to TGFβRII in a concentration-dependent manner, however the IC50 value was below the theoretical lower limit (Figure 4.6C). This suggests that the concentration of active galectin-1 was less than 625 nM.

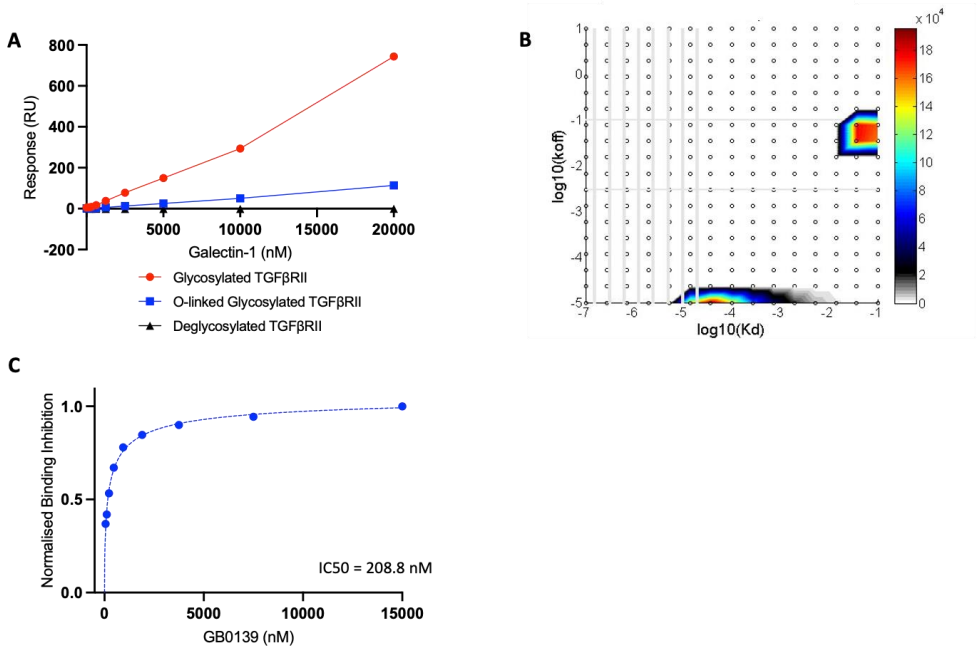


Figure 4.6: Galectin-1 binding to TGFβRII and competitive inhibition
***(A)** SPR data of soluble galectin-1 binding to glycosylated TGFβRII immobilised on the surface of a Series S sensor chip CM5 (~400 RU). Sequential injections of galectin-1 protein (156.25 nM - 20 µM) were performed, SPR signals were measured in RU and baseline subtracted. Binding response values plotted in GraphPad Prism with connecting line/curve shown. **(B)** Two-dimensional rate and affinity constant distributions plotted in EVILFIT from the fit of all experimental kinetic traces measured by SPR. **(C)** Solution competition binding assays were performed with the higher affinity galectin-3 inhibitor GB0139 (blue) in the presence of galectin-1 at 625 nM. Response values are normalised with respect to the highest RU response (0 nM compound) and competitive inhibition graphs plotted in GraphPad Prism. IC50 values were calculated by non-linear regression analysis (Specific binding with hill slope or one-site specific binding).*

4.4.7. Galectin binding to αv integrins and TGFβRII is not a 1:1 interaction

In all of the above binding experiments (excluding galectin-3 with TGFβRII) the immobilised ligand (αv integrins or TGFβRII) was not saturated with analyte (galectin-3 or galectin-1) at the concentrations tested. It was therefore not possible to yield an accurate K_d value or Bmax for these binding interactions. Subsequently, the minimum number of individual galectin-3 (27 kDa) or galectin-1 (15 kDa) proteins involved in the binding interaction was estimated from the response at 5000 nM of analyte. The results confirmed that galectin binding was not 1:1 and also demonstrated that the number of individual galectins was ligand-dependent (Table 4.2). However, the total

number of individual galectins in each binding interaction cannot be concluded from these experiments as saturation was not reached.

Table 4.2: Estimation of the number of individual galectins involved in ligand binding

Ligand	Analyte	Response (RU) at 5000 nM Analyte	No. of Individual Proteins
Galectin-3	αvβ1	2171.55	16
	αvβ5	2130.80	16
	αvβ6	1051.75	8
Galectin-1	αvβ1	796.15	11
	αvβ5	736.35	10
	αvβ6	463.25	7
	TGFβRII	743.9	6

4.4.8. Neither galectin-3 or galectin-1 physically interact with components of the large latent complex (LAP or LTBP1) or the TGFβRI subunit

No binding was detected for either galectin-3 or galectin-1 with immobilised LTBP1, LAP or TGFβRI at galectin concentrations up to a maximum of 5000 nM (Figure 4.7A-C). Small fluctuations in response were recorded across the analyte concentration series, these are commonly due to changes in refractive index and are unrelated to binding. Refractive index changes caused by buffer mis-match are most pronounced at high analyte concentrations which was detected at 5000 nM.

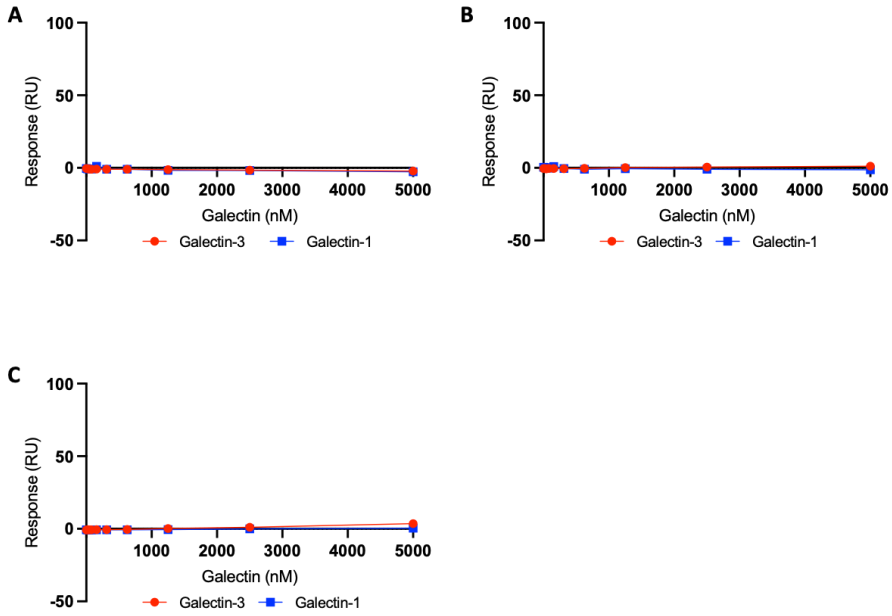


Figure 4.7: Negative SPR binding results
SPR data of galectin-3 or galectin-1 sequential injections (19.5-5000 nM) across a Series S sensor chip CM5 to assess binding to immobilised ***(A)*** *LTBP1,* ***(B)*** *LAP or* ***(C)*** *TGFβRI. SPR signals were measured in RU and baseline subtracted. Binding response values plotted in GraphPad Prism with connecting line/curve shown.*

4.5. Discussion

Galectins have been shown to bind to integrin glycosylation sites to regulate receptor activation and biological function (343-348). The overall aim of this study was to assess whether galectins can directly interact with different members of the complex responsible for integrin-mediated TGF-β1 activation. By SPR, full length galectin-3 and galectin-1 were found to bind to αvβ1 and αvβ5 which are integrins highly expressed on fibroblasts. Both galectins also bound to the epithelial-restricted integrin αvβ6. For both galectins the binding responses were higher for αvβ1 and αvβ5 compared to αvβ6. However, binding responses were higher for galectin-3 than galectin-1 across all three integrins. Despite both galectins having the highest affinity for poly-N-acetyllactosamine, galectin-1 binding requires a terminal β-galactose residue which is not required for galectin-3 binding and this may account for the difference in binding detected (233).

The binding responses recorded for both galectins across all three integrins were considerably higher than the theoretical maximum response for analyte-ligand binding 1:1. This demonstrates that multiple individual galectins are involved in the ligand interaction. Using the SPR data the minimum number of individual galectin-3 (27 kDa) proteins involved in the binding interactions was estimated (Section 4.4.7). As SPR measures a change in mass (concentration), it is not possible to accurately conclude if the total estimated number of galectin-3 protein is directly interacting with the immobilised ligand or if this number is as a result of oligomerisation. As galectin-3 can oligomerise through either the N-terminal or C-terminal domain this may have resulted in multivalency of carbohydrate-binding activity (whereby

multivalent galectin-3 binds to one or more cell surface glycans causing high avidity binding) (260, 269-272). Alternatively each galectin-3 protein could be binding to a single binding site on the integrin, although this is less likely given the estimated number of proteins and the size of galectin-3 when compared to the integrin receptor extracellular domain. It is therefore most probable that galectin-3 binds to one or more binding sites on the integrin causing subsequent self-association and lattice formation. Supporting this, galectin-3 has previously demonstrated positive cooperativity upon binding, whereby an increasing concentration of galectin-3 resulted in increased fractional binding to JAG1 and DLL4 by SPR (355). The authors used a N-terminally truncated galectin-3 (incapable of type-N self-association) to saturate the immobilised ligands and to determine the affinity of the monomeric interactions. This approach could have been used here (if the construct was available) to establish the affinity of galectin binding to αv integrins, although it may not have prevented type-C self-association. When evaluating the calculated number of galectin-1 proteins involved in the integrin interactions, it is important to considered that the galectin-1 protein produced was predominantly dimeric in solution (30 kDa) (Section 3.4.3.2). The estimated number of galectin-1 dimers involved in the binding interaction will therefore be half that of the minimum number of individual galectin-1 proteins (15 kDa) (Section 4.4.7). Like galectin-3, the oligomeric state of galectin-1 is particularly important in binding studies as the affinity of monomeric galectin-1 for immobilised ligands is significantly less than that of dimeric galectin-1 (359).

In a final attempt to measure binding affinity, SPR experiments were performed in the inverse setup in which galectin was immobilised to the chip and the binding

partner of interest injected across. However, no binding was detected during the test injections using this approach (data not shown). It has been suggested that as oligomerisation is often required for galectin function, the immobilisation process prevents a representative biological response (360). The negative binding responses observed using this set up are consistent with published data showing loss of galectin-3 activity following immobilisation (361). Collectively, from these experiments it was not possible to determine the stoichiometric ratio of galectin binding to αv integrins.

The binding of galectin-3 and galectin-1 to αv integrins by SPR was confirmed to be glycan-dependent. This is in agreement with the literature that states extracellular galectin interactions are commonly carbohydrate-dependent (247, 249, 250). Enzymatic removal of all integrin N-linked and common O-linked glycans inhibited the galectin-integrin binding interaction whilst only a partial reduction in binding was observed following the removal of N-linked glycans alone. From these experiments it was not possible to determine whether the partial binding observed following PNGase F treatment was due to O-linked glycan interactions or residual N-glycans. Although the efficiency of PNGase F treatment was visually confirmed by a shift in protein mobility by SDS-PAGE, this does not verify 100% N-glycan removal. Therefore, potentially some N-glycans remained on the integrin surface which the galectin could bind to. Alternatively, perhaps the O-linked glycans partially compensate for the higher affinity N-linked glycan interactions when no longer available or there is cooperative binding between the glycan subtypes. Despite N-glycans being the preferential galectin binding partner, there is evidence to suggest that O-glycans are still important for its biological function e.g. both galectin-1 and galectin-3-induced

apoptosis is regulated by O-glycan interactions (362). These findings are in agreement with the published literature whereby the carbohydrate dependency of galectin-integrin binding has been demonstrated using a number of techniques, other than SPR (343, 346). Knockdown of the glycosyltransferase Mgat5 (initiates branching on N-glycans) was previously used to determine the effect of deglycosylation on galectin-3 binding to α5β1 and its biological function in mammary carcinoma cells (343). Removal of sialic acids by neuraminidase enzymes has also been shown to increase the number of exposed glycans on the β1 integrin and alter galectin-3-mediated diffusivity (346).

Plotting the SPR data in EVILFIT as an association (kon) and dissociation (koff) distribution heatmap visually demonstrated that there is one heterogenous binding population for galectin-3 binding to αvβ1, αvβ5 and αvβ6. For galectin-1 two distinct but heterogenous binding populations are apparent on the EVILFIT heatmap which again could be explained by the dimeric state of galectin-1. A number of N-glycans have been reported or predicted from the NXS/T consensus sequence on the αv (13 sites), β1 (12 sites), β5 (8 sites) and β6 subunits (9 sites), several of which may be involved in the integrin-galectin interaction and therefore explain the binding heterogenicity (350, 351). However, it cannot be assumed that all reported/predicted sites are involved in the binding interaction as they could be inaccessible as a result of their distribution or the immobilisation process.

SPR competition experiments have previously been used to assess the inhibitory potency of galectin ligands relative to lactose (361). In the competition assays

described, asialofetuin was immobilised on the chip and galectin-1 or galectin-3 injected across the surface in the presence of inhibitory glycopeptides (361). Here, small molecule galectin inhibitors (GB1107, GB1490 and GB0139) which target the galectin CRD were shown to inhibit galectin-3 binding to αv integrins by SPR. All three compounds inhibited galectin-3 binding to αv integrins in a concentration-dependent manner and a difference in compound selectivity was demonstrated. As expected, the IC50 values were the lowest for GB0139 and GB1107, as these compounds have the highest binding affinity for galectin-3. GB1490 is a galectin-1 inhibitor with the lowest binding affinity for galectin-3 and subsequently had the highest IC50 values. GB0139 similarly inhibited galectin-1 binding to αv integrins in a concentration-dependent manner. However, the affinity of GB1107 and GB1490 for galectin-1 was too weak to determine IC50 values in these experiments (data not shown). The competitive inhibition data confirms that in the SPR experiments, galectin bound to the integrin glycan residues via the CRD. With carbohydrate ligand present in the CRD, it is perhaps more likely that any galectin-3 directly interacting with the integrin self-associates via its N-terminal (type-N self-association) and not the C-terminal (if self-association does occur at all) (271). However, galectin-3 self-association via the C-terminal (type-C self-association) cannot be ruled out based on these current data as the C-terminal domain of a galectin-3 could be binding to the C-terminal domain of another galectin-3 (independent of the CRD) resulting in the formation of oligomers (363).

Galectin-3 has previously been shown to be a binding partner of the TGF-β1 receptor with the TGFβRII subunit co-immunoprecipitating with galectin-3 pulldown (259).

Here, both galectin-3 and galectin-1 bound to TGFβRII by SPR, but the binding responses were higher for galectin-3 compared to galectin-1. As previously suggested, this could be because unlike galectin-3, galectin-1 requires a terminal β-galactose for binding (233). As a higher analyte concentration was used for the TGFβRII experiments than for the αv integrins, saturation of the immobilised ligand was almost reached with galectin-3. It was therefore possible to perform kinetic analysis on the data to yield a K_d value (5288 nM) and Bmax (1473 RU). The Bmax value obtained was higher than if the interaction was 1:1 which was supported by EVILFIT analysis showing a single heterogenous binding population. The minimum number of individual galectin-3 (27 kDa) proteins involved in the binding interaction was determined to be 6 from Bmax. Like with the integrins, it is not possible to conclude if the total number of galectin-3 protein is directly interacting with the immobilised TGFβRII or due to galectin-3 oligomerisation. By comparison, the immobilised ligand was not saturated with galectin-1 at the analyte concentrations tested so it was not possible to perform the same kinetic analysis. Instead, the likely number of galectin-1 binding sites was calculated (Section 4.4.7). The binding of galectin-3 and galectin-1 to TGFβRII was confirmed to be glycan-dependent. Enzymatic removal of all integrin N-linked and common O-linked glycans inhibited the binding interaction. Minimum galectin binding to TGFβRII was observed following the removal of all N-linked glycans. This is consistent with a previous study showing glycan-dependent binding of galectin-3 to the TGFβRII subunit by knockdown of Mgat5 in mammary carcinoma cells (259). Co-immunoprecipitation of the TGFβRII subunit with galectin-3 pulldown was only successful in Mgat5+/+ cells and binding was inhibited by lactose (259). The half-life of the TGF-β1 receptor was also reduced

in Mgat5-/- cells (259). Supporting this, galectin-3 has been proposed to mediate cell surface retention of the TGF-β1 receptor by binding to its poly-N-acetyllactosamine residues (284). The decrease in galectin binding observed following PNGase F treatment was greater for TGFβRII than the αv integrins. This data would suggest that TGFβRII N-linked glycosylation is essential for galectin binding to TGFβRII.

Small molecule galectin inhibitors (GB1107, GB1490 and GB0139) were shown to inhibit galectin-3 binding to TGFβRII by SPR in a concentration-dependent manner and a difference in compound selectivity was demonstrated. IC50 values were lowest for GB1107 and GB0139 compared to GB1490. GB0139 also inhibited galectin-1 binding to TGFβRII in a concentration-dependent manner. However, the affinity of GB1107 and GB1490 for galectin-1 was too weak to determine IC50 values in these experiments (data not shown). The competitive inhibition data confirms that in the SPR experiments, galectin bound to the TGFβRII glycan residues via the CRD.

There was no binding of galectin-3 or galectin-1 to LTBP1, LAP or TGFβRI detected by SPR in the current studies. Although there are a number of potential explanations for these findings beyond the conclusion that galectin-3 and galectin-1 do not bind to these proteins of interest. Firstly, it is important to consider that when performing solid-phase binding studies, the analyte binding site may be masked or the protein inactivated by the immobilisation process (364). Amine coupling is not homogeneous or oriented immobilisation due to the fact that most proteins contain several lysine ε-amino groups. For oriented immobilisation thiol coupling is often used however this involves site-specific introduction of a single thiol group into the protein. This capture

approach is only useful if the analyte binding site is known as orientation is determined by the location of the binding site. For example, this approach is used for site-oriented immobilisation of thiolated antibodies (365). Secondly, galectin binding to LTBP1, LAP or TGFβRI could be a lower affinity interaction and therefore a higher analyte concentration than already tested would be required to detect binding. Finally, when evaluating SPR data produced using recombinant proteins, the choice of expression system and protein amino acid sequence should also be considered. As galectin binding is mostly glycan-dependent, the expression system used to produce the recombinant proteins may affect binding interactions. The impact of the choice of expression system on binding has previously been demonstrated for both galectin-1 and galectin-3 binding to commercial Factor VIII expressed in either CHO or baby hamster kidney (BHK) cells (354). Galectin binding (%) was shown to be significantly higher to Factor VIII expressed in BHK cells than CHO cells, this was suggested to be because α1-3 galactosylation is expressed on BHK-derived Factor VIII which galectin-3 can bind. The commercially available LAP and TGFβRI proteins used for the current SPR experiments were produced in an insect expression system (baculovirus) which can synthesise simple to intermediate N-glycans. However, the only LTBP1 available was produced by cell-free protein expression (wheat germ) that is not capable of protein glycosylation. This may explain the negative binding by SPR which are inconsistent with the positive binding results obtained by solid-phase binding assay by Parmar, N. in which the same unglycosylated LTBP1 protein was used (313). By comparing the recombinant protein amino acid sequence to the UniProt accession # it is also evident that the proteins available and utilised in these studies were truncations, consequently it is possible that the analyte binding site may not have

been encoded. To verify the negative binding results these experiments would need to be performed using the full length protein (mammalian expression system) and by in solution binding techniques such as ITC, stopped-flow or AUC.

Summary: To summarise, both galectin-3 and galectin-1 bound to αv integrins (αvβ1, αvβ5 and αvβ6) and TGFβRII by SPR. However, binding responses were higher for galectin-3 than galectin-1 across all four proteins. In all experiments, analyte-ligand binding was heterogenous and not 1:1, although it was not possible to determine the stoichiometric ratio from the SPR data. Binding was glycan-dependent as deglycosylation (deglycosylation mix II or PNGase F) reduced the level of galectin binding to αv integrins and TGFβRII. Galectin binding was CRD-dependent as small molecule inhibitors which target the galectin CRD prevented the protein-glycan interactions in a concentration-dependent manner. Whether galectin-3 or galectin-1 binds to LTBP1, LAP or TGFβRI is inconclusive. As the integrin, the latent TGF-β1 complex and TGF-β1 receptor are required to be in close physical association for integrin-mediated TGF-β1 activation (168), it is hypothesised that galectins promote integrin-mediated TGF-β1 activation by facilitating the clustering of integrins and TGF-β1 receptors on their respective cell surfaces.

Limitations/ Future Work: Although these experiments were useful for investigating whether galectin-3 or galectin-1 interact with components of the TGF-β1 signaling pathway, a number of limitations do exist from the experimental design. Firstly, 100% N-glycan removal could not be confirmed by SDS-PAGE following PNGase F treatment (Appendix Figure 9.2), therefore the importance of O-linked glycans alone cannot be

concluded from these experiments. A more effective method of verifying complete N-glycan removal would be mass spectrometry, unfortunately this technique was not available at the RCaH. Secondly, as the immobilised ligand was not saturated at the analyte concentrations tested it was not possible to yield an accurate K_d value or Bmax. However, determining the interaction kinetics was not the aim of these experiments and beyond the scope of the 1-year placement at the RCaH, the experiments were solely designed to determine if the proteins could interact. Future work would therefore involve repeating these experiments using a higher analyte concentration so that complete kinetic analysis could be performed.

5. Investigating the direct binding of galectin-3 to αv integrins *in vitro*

5.1. Introduction

In Chapter 4, SPR was used for the high-throughput screening of galectin-3 and galectin-1 interactions with potential binding partners including αvβ1, αvβ5, αvβ6, LAP, LTBP1, TGFβRI and TGFβRII. This technique helped to short-list binding partners of interest confirming that both galectins were capable of binding to αv integrins (αvβ1, αvβ5 and αvβ6) and the TGFβRII subunit. However, SPR is somewhat of an artificial system that does not mimic the biological environment which is relevant for an interaction. It is therefore necessary to confirm that the key protein interactions detected by SPR are also present in cells using a range of cell biology techniques. The aim of this thesis is to define the mechanism of galectin-3-mediated TGF-β1 activation and its role in lung fibrosis and current evidence supports that an interaction between galectin-3 and αvβ1 in fibroblasts as likely most important for fibrogenesis (221, 284, 310). Experiments therefore focused on verify the binding of galectin-3 to αvβ1 in these cells.

The current literature indicates that galectin-3 is capable of binding to the β1 integrin subunit *in vitro* and may mediate integrin clustering, although interestingly this data has not been shown in fibroblasts. In mouse aortic endothelial cells (MAECs) galectin-3 and the α3 subunit co-immunoprecipitate with β1 integrin Ab pulldown, subsequently the α3β1 integrin was suggested to localise galectin-3 on the cell surface (366). Similarly, α3β1 was identified as a galectin-3 binding partner (alongside α6, αv and β4 integrins) in human corneal epithelial cells (HCLEs) by a combination of

affinity chromatography, mass spectrometry and immunostaining (345). Galectin-3 was therefore proposed to cross-link and cluster the α3β1 integrin on the cell surface to promote lamellipodia formation and cell migration. In support of this hypothesis, exogenous galectin-3 has been shown to increase lateral mobility and clustering of the β1 integrin in HeLa cells using single-particle tracking (346). The increase in lateral mobility was observed with full length exogenous galectin-3 but not if N-terminally truncated, therefore demonstrating that type-N self-association may be required for integrin clustering. How the increase in lateral mobility was only at a higher galectin-3 concentration supports this as galectin-3 oligomerisation is concentration-dependent (260, 270, 367, 368). By immunocytochemistry (ICC) staining, exogenous galectin-3 has similarly been shown to induce redistribution and clustering of the β1 integrin in retinal pigment epithelial cells (RPECs), with their interaction evident from co-immunoprecipitation (Co-IP) experiments (348). Inversely, blocking galectin-3 function (anti-galectin-3 mAb) has been shown to decrease the clustering of both α3β1 and α6β4 in keratinocytes (369).

Galectin-3 has also been proposed to play a role in acute inflammatory injury and chronic fibrosis responses via direct binding to the CD98: β1 integrin complex on epithelial cells with this complex being termed the 'gal-3-fibrosome' (287). These non-peer reviewed results show that in A549 cells the β1 integrin immunoprecipitates with galectin-3 Ab pulldown and the level of protein immunoprecipitated significantly increases with TGF-β1 stimulation. Additionally, colocalisation of galectin-3 and the β1 integrin was detected by PLA following TGF-β1 stimulation in an *ex vivo* human lung tissue model (non-IPF).

There are several methods commonly used to analyse protein interactions *in vitro* including Co-IP, affinity chromatography (pulldown assay), crosslinking and label transfer (370). However a limitation to several of these techniques is that proteins of interest require prior protein purification and labelling. Co-IP is generally considered to be the most physiological analysis method as the proteins are in their native conformation and interact with no external influence. In addition to protein interaction techniques, protein colocalisation methods are also often utilised to confirm that the proposed interacting proteins are in close proximity *in vitro* (371). A combination of these techniques would therefore provide physiologically relevant data on the binding of galectin-3 to the β1 integrin in fibroblasts.

5.2. Aims

Previously, galectin-3 binding to the αvβ1 integrin was detected via the biophysical technique SPR (Chapter 4), therefore to confirm that this interaction also occurs *in vitro* the following specific aim was addressed:

- To investigate the binding of galectin-3 to the αvβ1 integrin in primary lung fibroblasts using *in vitro* protein interaction techniques

5.3. Methods

5.3.1. Adherent Cell Culture

5.3.1.1. Primary Cells

Fibroblasts: IPF and Non-IPF HLFs were obtained from explanted human lung tissue post-lung biopsy with informed written consent. Cells were cultured in dulbecco's modified eagle's medium (DMEM) - high glucose supplemented with 10% foetal bovine serum (FBS), 4 mM L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin. FBS was not added to the culture media if serum free media was required. Cells were grown in a humidified incubator set to 37°C with 5% CO_2.

5.3.1.2. Immortalised Cells

HEK293T cells: HEK293T cells were cultured in DMEM media - 25 mM HEPES, high glucose, L-glutamine supplemented with 10% FBS, 100 U/mL penicillin and 100 µg/mL streptomycin. FBS was not added to the culture media if serum free media was required. Culture conditions were phenol red-free to decrease background fluorescence in images. Cells were grown in a humidified incubator set to 37°C with 5% CO_2.

5.3.1.3. Maintaining Cultured Cells

Passaging: When ~90% confluent cells were passaged. The media was aspirated, followed by a phosphate-buffered saline (PBS) wash and the addition of trypsin. Cells were incubated with trypsin for 2-5 minutes in the incubator (37°C with 5% CO_2). After 2-5 minutes, detachment of cells was visually confirmed under the microscope and the trypsin inactivated by adding an equal volume of FBS. Detached cells were pelleted by centrifuging for 5 minutes, 1500 rpm and the supernatant discarded (Fisherbrand GT2R Centrifuge, TX-400 rotor). The cell pellet was re-suspended in

fresh DMEM media and a cell count performed, cells were then seeded into a new flask at the desired cell density. To facilitate comparative biology experiments all cells were frozen down and seeded at a density that was consistent across the cell line, this was in attempt to ensure a consistent doubling population.

5.3.1.4. Cell Counting

Haemocytometer: Fibroblast cell counts were performed using a haemocytometer. 10 µL of cell suspension was mixed 1:1 with trypan blue and 10 µL of this pipetted onto the haemocytometer for cell counting under the microscope (10x magnification). The number of viable cells (trypan blue-negative) was counted in all four outer squares (16 squares/outer square). The number of viable cells/mL was calculated using the following formula:

$$\text{Number of viable cells/mL} = \left(\frac{\text{No. of cells counted}}{4}\right) \times 2 \times 10^4$$

The total number of cells was then calculated using the following formula:

$$\text{Total number of cells} = \text{Number of viable cells/mL} \times \text{Suspension volume}$$

Automated cell counter: HEK293T cell counts and % viability were determined using the TC20 automated cell counter (Bio-Rad). 10 µL of cell suspension was mixed 1:1 with trypan blue and 10 µL of this pipetted onto a dual chamber counting slide.

5.3.2. Co-Immunoprecipitation

Principle of the method: Co-IP is used to investigate protein interactions and can be performed using magnetic beads or agarose resin. Both methods work on the same principle however they are different in terms of binding capacity. In both instances

the protein lysate is firstly pre-cleared by incubating with beads or control agarose, this is to minimise non-specific binding during the immunoprecipitation. A primary antibody (pulldown antibody) against a known target protein is then incubated with the pre-cleared lysate. To pull the antigen-antibody complex out of solution Protein A/G magnetic beads or agarose resin are added. Protein A/G (Ig-binding protein from bacteria) binds to the Fc region of the primary/pulldown antibody which pulls the target antigen out of solution along with any potential binding partners. Any proteins that are non-specifically or weakly bound to the antigen-antibody-Protein A/G complex are removed by washing the resin or beads. Any non-specific protein binding not removed by washing are detected by performing the Co-IP in parallel with an isotype control primary antibody. The Co-IP samples are then subjected to SDS-PAGE gel electrophoresis and western blot. The primary antibody used for western blot is specific for the binding partner of interest that should have been pulled out by the antigen-antibody-Protein A/G complex if there was an interaction. The host species of the primary antibody used for western blot should be different to that of the Co-IP pulldown antibody to decrease background signal from antibody IgG heavy (H) and light (L) chains. Alternatively, the pulldown antibody can be coupled and crosslinked to the magnetic beads/resin to prevent it being present in the Co-IP elution.

5.3.2.1. Immunoprecipitation with Magnetic Beads

Protocol: Immunoprecipitation with magnetic beads was performed using the Universal Magnetic Co-IP Kit (Active Motif, 54002) as per manufacturer protocol (Whole Cell Extraction). Non-IPF HLFs were cultured in a T175 flask at p6. If cells were stimulated, they were grown until ~90% confluent, growth arrested for 24 hours and then 50 µM LPA (4 hours) or 2 ng/mL TGF-β1 (2 hours) added. If cells were untreated,

they were grown until ~100% confluent before harvesting. After growing and treating cells as required, the media was aspirated and the adherent cells washed with 5.5 mL of ice-cold PBS/inhibitor buffer. 4 mL of fresh ice-cold PBS/inhibitor buffer was then added to the flask and the cells detached by scraping. Detached cells were pelleted by centrifuging for 5 minutes, 1500 rpm at 4^{o}C and the supernatant discarded (Fisherbrand GT2R Centrifuge, TX-400 rotor). Cell pellets were stored at -80^{o}C and thawed on ice when required. Once thawed, cell pellets were resuspended in 130 µL of whole-cell lysis buffer (+/- IGEPAL CA-630 detergent) and incubated on ice for 30 minutes with vortexing every 5 minutes. After 30 minutes the lysate was clarified by centrifuging at 14,000 rpm for 10 minutes, 4^{o}C. To quantify total protein from cell lysates the bicinchoninic acid (BCA) assay (Section 5.3.3.1) was used with bovine serum albumin (BSA) standards prepared in Co-IP whole-cell lysis buffer and samples prepared at 3 different dilutions (neat, 1:5, 1:10). All of the following Co-IP steps were performed in duplicate (IgG control IP and target protein IP). To minimise non-specific binding, a pre-clear of the protein lysate was performed by incubating the lysate (experiment specific concentration) with 12.5 µL of Protein G magnetic beads and incubating on a rotary shaker for 1 hour at 4^{o}C. The magnetic beads were then pelleted using a magnetic stand and discarded. The antibody/extract mixture was prepared by combining the pre-cleared protein lysate with 1 µg of pulldown antibody and made up to a 500 µL final volume in Co-IP buffer. This was incubated on a rotary shaker for 4 hours at 4^{o}C. 25 µL of fresh Protein G magnetic beads were then added to the mixture and incubated on a rotary shaker overnight at 4^{o}C. The next day, the magnetic beads were pelleted and the supernatant (flow through) collected to be later subjected to gel electrophoresis. The beads were then washed by resuspending

in 500 µL of Co-IP buffer followed by pelleting the beads and collecting the supernatant (wash) for gel electrophoresis. This was performed 4 times in total and after collecting the final wash, the pelleted beads were resuspended in 20 µL of 2X sample buffer; Bolt LDS sample buffer (2X final concentration), +/- Bolt sample reducing agent (2X final concentration) and made up to 20 µL with H_20. 1 µL of Co-IP control steps (input, flow through, wash) was made to a 20 µL final volume in 2X sample buffer. All samples were heated to 95°C for 5 minutes and 20 µL/sample loaded onto a gel. The gel was ran by SDS-PAGE and western blot performed as described (Section 5.3.3.2). All centrifugation steps were performed on a Fisher Scientific accuSpin Microcentrifuge 17R, 24 x 1.5/2.0ml rotor (Thermo Scientific, 75003424) unless otherwise stated. Antibodies used for Co-IP and subsequent western blot are later summarised in a table (Section 5.3.2.3). Buffer components were supplied with the kit and buffers prepared according to the manufacturers' recommendations (Table 5.1).

Table 5.1: Components used to prepare Co-IP buffers

Buffer	Component	Volume
PBS/Inhibitor (10 mL)	PBS (10X)	1 mL
	H_20	8.4 mL
	Phosphatase Inhibitors	500 µL
	Deacetylase Inhibitor	100 µL
Whole-Cell Lysis Buffer (130 µL)	Whole-Cell Lysis Buffer	120.9 µL
	Phosphatase Inhibitors	6.5 µL
	Deacetylase Inhibitor	1.3 µL
	Protease Inhibitor Cocktail	0.65 µL
	PMSF (100 mM)	0.65 µL
	IGEPAL CA-630 (detergent)	Experiment specific
Co-IP Buffer (2.5 mL)	Co-IP/Wash Buffer	2.325 mL
	Phosphatase Inhibitors	125 µL
	Deacetylase Inhibitor	25 µL
	Protease Inhibitor Cocktail	12.5 µL
	PMSF (100 mM)	12.5 µL

5.3.2.2. *Immunoprecipitation with Beaded Agarose*

Protocol: Immunoprecipitation with agarose resin was performed by adapting the Pierce Crosslink Immunoprecipitation Kit (Thermo Scientific, 26147) manufacturer protocol. For Co-IP experiments, Non-IPF HLFs were cultured in a T175 flask at p6 until ~100% confluent and the cells harvested. The media was aspirated and the adherent cells washed with ice-cold 1X PBS, 5 mL of fresh PBS was then added to the flask and the cells detached by scraping. This was performed multiple times until most of the cells were detached which was visually confirmed under the microscope. Detached cells were pelleted by centrifuging for 5 minutes, 1500 rpm at 4^{o}C and the PBS discarded (Fisherbrand GT2R Centrifuge, TX-400 rotor). Cell pellets were stored at -80^{o}C and thawed on ice when required. Thawed pellets were weighed, resuspended in ice-cold IP lysis/wash buffer (300 µL/50 mg wet pellet) and incubated on ice for 30 minutes with vortexing every 5 minutes. After 30 minutes the lysate was clarified by centrifuging at 14,000 rpm for 10 minutes, 4^{o}C. To quantify total protein from cell lysates BCA was used (Section 5.3.3.1) with BSA standards prepared in IP lysis/wash buffer and samples prepared at 3 different dilutions (neat, 1:5, 1:10). All of the following Co-IP steps were performed in duplicate (IgG control IP and target protein IP) and centrifugation steps performed at 700 rpm for 1 minute. Firstly, 10 µg of pulldown antibody was diluted in 1X coupling buffer (100 µL volume) and coupled to protein A/G plus agarose (20 µL of resin suspension, pre-equilibrated with 1X coupling buffer) by incubating the column on a roller mixer for 1 hour at room temperature. After 1 hour the antibody coupled resin was washed 3 times in 1X coupling buffer (300 µL volume) by centrifugation. Next the bound antibody was crosslinked to the resin by incubating with 50 µL of crosslinking buffer (H_20, 1X

coupling buffer, 0.45 mM DSS) on a roller mixer for 1 hour at room temperature. The crosslinking reaction was stopped and non-crosslinked antibody removed by washing the resin 3 times in elution buffer (100 µL), each elution buffer wash was followed by a 200 µL wash in IP lysis/wash buffer. Before applying the protein lysate to the antibody coupled/crosslinked resin, a pre-clear of the protein lysate was performed by incubating 650 µg of lysate/column with control agarose resin slurry (52 µL of resin suspension, pre-equilibrated with 1X coupling buffer) on a rotary shaker for 1 hour at 4°C. After 1 hour the lysate was separated away from the control agarose by centrifugation and the resin discarded. The pre-cleared lysate was then incubated with the antibody coupled/crosslinked resin in a column overnight at 4°C. The next day, the column with lysate/resin was centrifuged and the remaining lysate (flow through) collected to be later subjected to gel electrophoresis. The resin was then washed 3 times in IP lysis/wash buffer (200 µL) by centrifugation, the wash was also collected for gel electrophoresis. In preparation for protein elution (60 µL total elution) the resin was washed in 1X conditioning buffer (100 µL). 10 µL of elution buffer was added to the resin and centrifuged immediately, another 50 µL of elution buffer was then incubated on the resin for 5 minutes before centrifuging again. Once the total elution was collected, all samples were prepared for gel electrophoresis. 28 µL of Co-IP elution was mixed with sample buffer (1X final concentration), Bolt sample reducing agent (1X final concentration) and made up to 40 µL with H_20. 1 µL of Co-IP control steps (input, flow through, wash) was mixed with sample buffer (1X final concentration), Bolt sample reducing agent (1X final concentration) and made up to 30 µL with H_20. Samples were heated to 95°C for 5 minutes and the total 40 µL of Co-IP sample loaded onto a gel alongside 10 µL of control steps. The gel was ran by SDS-

PAGE and western blot performed as described (Section 5.3.3.2). All centrifugation steps were performed on a Fisher Scientific accuSpin Microcentrifuge 17R, 24 x 1.5/2.0ml rotor (Thermo Scientific, 75003424) unless otherwise stated. Antibodies used for Co-IP and subsequent western blot are summarised below (Section 5.3.2.3).

5.3.2.3. Co-Immunoprecipitation Antibodies

The same antibodies were used when immunoprecipitating with magnetic beads or beaded agarose (Table 5.2).

Table 5.2: Antibodies used for Co-IP and subsequent western blot experiments

	Pulldown Antibody	Primary Antibody	Secondary Antibody
β1 integrin pulldown, galectin-3 Co-IP	Mouse Human Integrin beta 1/CD29 mAb (R&D, Clone No. P5D2) Mouse IgG1 Isotype Control mAb (R&D, Clone No. 11711)	Rabbit Anti-Galectin-3 mAb (Abcam, Clone No. EPR19244) Final concentration: 0.5 µg/mL	Goat Anti-Rabbit Immunoglobulins/HRP pAb (Dako, P0448) Final concentration: 0.1 µg/mL
Galectin-3 pulldown, β1 integrin Co-IP	Rabbit Anti-Galectin-3 mAb (Abcam, Clone No. EPR19244) Rabbit IgG Isotype Control mAb (Abcam, Clone No. EPR25A)	Mouse Anti-Integrin beta 1 mAb (Abcam, Clone No. 12G10) Final concentration: 1 µg/mL	Goat anti-Mouse IgG/ HRP pAb (Invitrogen, G21040) Final concentration: 0.2 µg/mL

Western blot antibody concentration was unchanged, the pulldown antibody final concentration was kit specific (1 µg with magnetic beads and 10 µg with agarose resin).

5.3.3. Protein Quantification and Detection

5.3.3.1. Bicinchoninic Acid Assay

Principle of the method: Cupric ions (Cu^{2+}) are reduced to cuprous ions (Cu^{1+}) by protein in an alkaline environment. The BCA assay is used to determine protein concentration by measuring the amount of reduced cuprous ions that have reacted

with BCA to form a purple water soluble complex. The complex has maximum absorbance at 562 nm which increases linearly with protein concentrations.

Protocol: Total protein from cell lysates was quantified using the Pierce BCA Protein Assay Kit (Thermo Scientific, 23225) with reference to BSA standards (9-point dilution). 5 µL/well of lysate or standard was pipetted into a 96-well plate. 95 µL of working reagent (50:1, reagent A to B) was then added to all wells and mixed thoroughly on a plate shaker for 30 seconds. The plate was sealed and placed in a humidified incubator set to 37°C with 5% CO_2 for 30 minutes. Absorbance was measured at 562 nm on a plate reader (Synergy HT, BioTek) with Gen5 software and the concentration (µg/mL) of each sample calculated in excel from the BSA standard curve.

5.3.3.2. Western Blot

Principle of the method: Western blot is an immunoassay used to detect a specific protein of interest in a protein lysate. Firstly, proteins within a lysate are separated by MW using gel electrophoresis. Proteins denatured by SDS carry a negative charge and therefore undergo electrophoretic migration towards the positively charged electrode (anode). Once separated the proteins on the gel are transferred onto a blotting membrane using the same principle, the negatively charged proteins are transferred onto the membrane by electrophoretic transfer. Polyvinylidene difluoride (PVDF) and nitrocellulose membranes are commonly used which bind to proteins by hydrophobic interaction. Protein transfer can be verified using Ponceau S, a reversible stain that binds positively charged amino groups and non-polar regions. After successful transfer the membrane is incubated with a protein blocking

agent (milk or BSA) that contains a high amount of protein to prevent non-specific binding of antibodies to the rest of the membrane, this enhances the antibody specificity. After blocking the membrane is incubated with a monoclonal primary antibody (high homogeneity) specific for the protein of interest (antigen). Any non-specifically bound antibody is removed by washing the membrane in a non-ionic, non-denaturing detergent (PBS with Tween 20 (PBST) or tris-buffered saline with Tween 20 (TBST)) to disrupt the hydrophobic interaction. The membrane is then incubated with a polyclonal secondary antibody (signal amplification) which is selected in accordance to the primary antibody isotype and host species. The secondary antibody can be labelled with an enzyme e.g. conjugated to horseradish peroxidase (HRP) for chemiluminescent detection. Lastly, following wash steps the blot is incubated with HRP substrate (luminol) which becomes oxidised to produce 3-aminophthalate, this emits light at 428 nm and can be captured on an imaging system. Detection of the correct size protein is verified by reference to MW standards.

Protocol: Bolt 4-12% Bis-Tris precast protein gels were used for protein separation. Protein samples were prepared and loaded onto the gel as described in the specific Co-IP methods (Section 5.3.2) alongside 5 µL of protein ladder. Gels were ran in a Bolt mini gel tank in 20X Bolt MOPS SDS running buffer (1X final concentration) at 200V for 32 minutes (room temperature). Proteins separated by gel electrophoresis were transferred onto a PVDF membrane using the iBlot 2 gel transfer device, ran at 200V for 7 minutes. Ponceau S staining was used to confirm that transfer was successful and the stain removed by washing in ddH_20. The membrane was blocked for 2 hours

with 5% milk in PBST (1X PBS, 0.05% tween 20). After 2 hours the membrane was washed in PBST (2 x 5 minute washes) and incubated with 5 mL primary antibody overnight at 4^{o}C. The next day, the membrane was washed in PBST (3 x 5 minute washes) and incubated with 10 mL of HRP-conjugated secondary antibody for 1 hour at room temperature. All primary and secondary antibodies were diluted in 5% milk in PBST. For chemiluminescent detection the membrane was incubated with 2 mL of western blot substrate (1:1, substrate A to B) for 5 minutes. Images were taken using the LiCor Odyssey Fc (Chemi channel 2 minutes, 700 channel 1 minute) and blots analysed on Image Studio Lite v5.2. The antibodies used for western blot following Co-IP are previously described (Section 5.3.2.3).

5.3.3.3. Immunocytochemistry

Principle of the method: Antibodies are used to visualise protein expression by cells (immunocytochemistry) or within tissue (immunohistochemistry). The sample is firstly preserved by fixation, there are two common fixative methods used: aldehyde based cross-linking (paraformaldehyde (PFA)) or alcohol based organic solvent fixation (methanol). PFA creates covalent cross-links between lysine residues thereby retaining protein structure, however the fixation process is longer and subsequently proteins may re-localise before complete fixation. By comparison, methanol works by dehydration and is therefore faster but the precipitated proteins are structurally altered (not chemically altered so retain antigenicity). After fixing, cell permeabilisation can be performed so that antibodies can interact with intracellular proteins. Samples are then incubated with a blocking agent to prevent non-specific binding of antibodies. Next, samples are incubated with a monoclonal primary antibody which recognises the protein of interest (antigen), the part of the protein

that is detected by the antibody is called the epitope. A polyclonal secondary antibody is then added that is designed to recognise and bind to the primary antibody. The secondary antibody is conjugated to a flurophore for visualisation by microscopy, this is called indirect detection. Indirect detection enables signal amplification (multiple secondary antibodies can bind to a single primary antibody) and is therefore more sensitive than direct detection with a flurophore-conjugated primary antibody. 4',6-diamidino-2-phenylindole (DAPI) is used as a nuclear counterstain to enable localisation of the protein of interest.

Protocol: Non-IPF HLFs at p6 were cultured in 8-well chamber slides for ICC. When ~80% confluent the media was aspirated, the cells washed in PBS and then fixed in 4% PFA (100 µL/well) for 20 minutes at room temperature. Cells were then washed 3 times with PBS and permeabilisation buffer (1X PBS pH7.4, 0.5% triton, 1% BSA) added (100 µL/well) for 20 minutes at room temperature. After 20 minutes, cells were washed as previously described and blocking buffer (1X PBS pH 7.4, 2% BSA) added (100 µL/well) for 1 hour at room temperature. After 1 hour the excess block was removed, a PBS wash performed and the plastic chamber removed from the slide. A hydrophobic barrier was created between all wells using a PAP pen. A mouse anti-β1 integrin primary antibody (Abcam, Clone No. 12G10) was diluted in blocking buffer (5 µg/mL final concentration) and 50 µL added to wells where required, 50 µL of blocking buffer alone was added to control wells. Cells were incubated with primary antibody overnight at 4^{o}C. The next day, cells were washed 3 times with PBS and then incubated with 50 µL of goat anti-mouse IgG1 secondary antibody, AF568 (Invitrogen, A-21124) diluted in blocking buffer (4 µg/mL final concentration) for 1

hour at room temperature. Cells were washed as previously described and then stained with DAPI diluted in PBS (1 µg/mL final concentration) for 20 minutes at 4°C. Anti-fade fluorescence mounting medium and a coverslip was added to the slide before imaging. ICC experiments were performed with Dr Kevin Liu (Imperial College London) and all confocal images taken on a SP5 microscope (Leica) by Dr Maria Zarcone (Imperial Collage London).

5.3.4. DNA Transfection

Principle of the method: HEK293T cells are commonly used for transient gene expression as they have a high transfection efficiency. FuGENE-HD is a highly efficient non-liposomal transfection reagent with low toxicity. The reagent complexes with nucleic acids and interacts with the cell membrane to transport the DNA inside the cell.

Protocol: HEK293T cells were cultured in a T25 flask and when confluent the media was aspirated, followed by a PBS wash and the addition of trypsin. Cells were incubated with 0.05% trypsin for 2 minutes in the incubator. After 2 minutes the trypsin was inactivated by adding an equal volume of fresh media. Detached cells were pelleted by centrifuging for 3 minutes at 1026 rpm and the supernatant discarded (VWR Mega Star 1.6 R, TX-400 rotor). The cell pellet was re-suspended in fresh DMEM media and a cell count performed. Cells were then seeded into an 8-well chamber slide at a cell density of 25000 cells/well and returned to the incubator for 7 hours prior to transfection. Cells were transfected with FuGENE-HD transfection reagent as per manufacturer protocol (Promega, E2311). 200 ng of isolated plasmid DNA (Table 5.3) was mixed with 10 µL of serum-free DMEM media, 0.5 µL of FuGENE

was then added to this. After a 10-minute incubation at room temperature the fuGENE-DNA mixture was pipetted into the appropriate wells. Cells were incubated at 37°C with 5% CO_2 for 36 hours before live-cell imaging. All DNA transfection experiments were performed with Dr Siva Ramadurai (RFI, Harwell Campus) and confocal images taken on a DMi8 microscope (Leica) by Dr Siva Ramadurai due to COVID restrictions.

Table 5.3: Plasmid-encoded proteins (αv integrins and galectins) with fusion tags

Plasmid DNA	Tag	Generated By
Galectin-1	mCherry	Jessica Calver
Galectin-3	mCherry	
αvβ5	GFP	
αvβ6	GFP	

5.3.5. Proximity Ligation Assay

Principle of the method: PLA can be used for *in-situ* detection of endogenous protein interactions. Firstly, two primary antibodies raised in different species (e.g. mouse and rabbit) are incubated on fixed cells which recognise their target antigen on the two different proteins of interest (if expressed). Two oligonucleotide-labelled secondary antibodies (PLA probes) that are against the host species of the primary antibodies are then added which will bind to the corresponding primary antibody-protein complex. If the PLA probes (and proteins of interest) are within 40 nm of each other, the two connector DNA strands will hybridise. This is then ligated into circular DNA by ligase and the resulting closed circle DNA template becomes amplified by DNA polymerase (rolling-circle amplification). Lastly, detection oligos coupled to fluorochromes hybridise to complementary sequences within the amplicon. This generates a fluorescent signal (for a single event) which can be visualised by microscopy.

Protocol: All PLA experiments were performed at the University of Leicester with Dr Panayiota Stylianou as a manuscript collaboration with Prof Bibek Gooptu. For PLA experiments, IPF and Non-IPF HLFs (at p3 or p4) were cultured in 8-well chamber slides and when ~80% confluent, cells were growth arrested in growth factor free media for 24 hours. After serum starvation cells were then stimulated with 2 ng/mL of TGF-β1 for 24 hours (400 µL/well) or untreated where required. Cells were then pre-treated with either 0.1, 1 or 10 µM of galectin inhibitor (GB0139) or an equivalent volume of DMSO (control) for 20 minutes (200 µL/well). Prior to beginning the Duolink PLA In-Situ Fluorescence protocol (Sigma-Aldrich, DUO92008) cells were fixed with 4% PFA (100 µL/well) for 10 minutes at room temperature. 100 µL of blocking solution (1X PBS pH 7.4, 10% FBS) was then added to each well and incubated for 2 hours at 37°C. After 2 hours, excess block was removed and cells incubated with primary antibody or isotype control antibody overnight at 4°C (100 µL/well). All antibodies used for PLA were diluted to a 5 µg/mL final concentration in antibody diluent (1x PBS pH 7.4, 2% FBS, 1% BSA) (Table 5.4).

Table 5.4: Antibodies used for PLA experiments

Isotype Control Antibodies	PLA Antibodies
Negative control mouse IgG1 (Dako, Clone No. DAK-GO1)	Mouse anti-Integrin beta 1 (Abcam, Clone No. P5D2)
Rabbit isotype control (BD Biosciences, Clone No. Poly1281)	Rabbit anti-galectin-3 (Invitrogen, PA5-34819)

The next day, excess antibody was removed and cells washed 2 times with 1X wash buffer A for 5 minutes. PLA probe solution containing anti-rabbit PLUS and anti-mouse MINUS probes diluted 5 in antibody diluent (1X final concentration) was then added (40 µL/well) to slides and incubated at 37°C for 2 hours. Excess PLA solution

was removed and cells washed 2 times with 1X wash buffer A for 5 minutes. Ligase (1 U/µL) was diluted in 1X ligation buffer (1:40) and incubated on slides (40 µL/well) for 30 minutes at 37^{o}C. Excess ligase solution was removed and cells washed 2 times with 1X wash buffer A for 5 minutes. After washing, polymerase (10 U/µL) diluted in 1X amplification buffer (1:80) was incubated on slides (40 µL/well) for 2 hours at 37^{o}C. Slides were then washed twice in 1X wash buffer B for 10 minutes and once in 0.01X wash buffer B for 1 minute. DAPI solution was diluted in antibody diluent (2 µg/mL final concentration) and 100 µL added to each well for 2 minutes. Lastly, cells were washed 3 times in 1X PBS pH 7.4 and mounted with fluoroshield for imaging. All confocal images were taken on a LSM 980 Airyscan 2 microscope (Zeiss) by Dr Panayiota Stylianou (University of Leicester).

5.4. Results

5.4.1. Galectin-3 physically interacts with the β1 integrin in HLFs

Co-IP was used to confirm the key SPR binding data in a cell-based system and to determine whether galectin-3 is a β1 integrin binding partner in HLFs. Magnetic beads coupled to anti-β1 integrin antibody were used for initial Co-IP experiments in untreated HLFs (Figure 5.1A) or following stimulation with LPA (50 µM) (Figure 5.1A) or TGF-β1 (2 ng/mL) (Figure 5.1B). Western blot results failed to detect galectin-3 (27 kDa) in association with the β1 integrin for all 3 culture conditions, this is evident by the lack of band at 27 kDa in Figure 5.1A lane 4 and lane 8 and in Figure 5.1B lane 6. However, galectin-3 was detected in all input and flow through controls. In all IP reactions both IgG H and L chains were detected by western blot (H chain: 50 kDa and L chain: 25 kDa). These data confirmed that galectin-3 protein was expressed by both untreated and stimulated HLFs but suggested that either: 1) the β1 integrin was not expressed by the fibroblasts, 2) the β1 integrin Ab pulldown was unsuccessful or 3) the proteins did not interact *in vitro.*

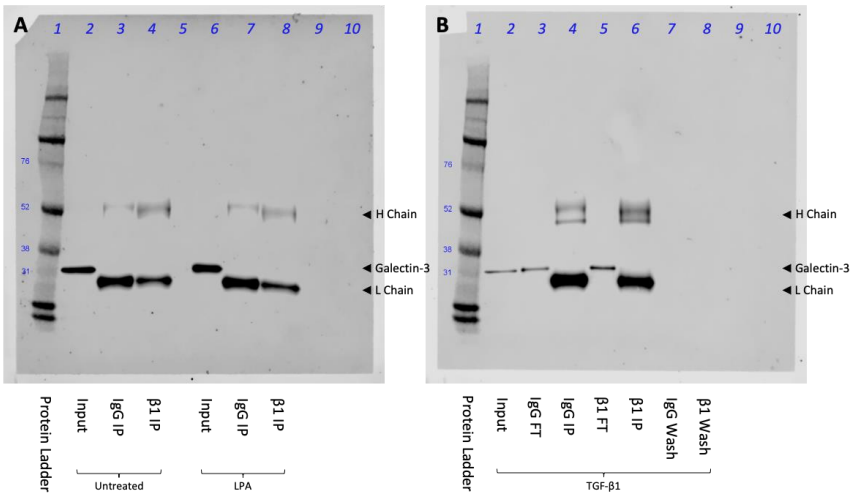


Figure 5.1: Western blot detection of galectin-3 following β1 integrin Ab pulldown in HLFs Protein lysates (250 µg/ IP reaction) from HLFs p6 (N=1), ***(A)*** *with or without prior stimulation with 50 µM LPA for 4 hours or* ***(B)*** *2 ng/mL TGF-β1 for 2 hours, were immunoprecipitated with an anti-β1 integrin antibody (1 µg/ IP reaction) and immunoblotted for galectin-3. Co-IP input, flow through (FT) and wash steps loaded as controls. Proteins separated by reducing SDS-PAGE and target protein size estimated from the Amersham ECL Full-Range Rainbow Markers (12 to 225 kDa) migration pattern.*

Following the unexpected negative Co-IP results, ICC was performed to determine whether untreated HLFs express the β1 integrin at p6. Staining confirmed that the β1 integrin protein was expressed by the cells identified by red signal (AF568) on the cell surface (Figure 5.2). These data therefore suggested two potential explanations for the failure to detect an interaction between galectin-3 and the β1 integrin, either 1) the β1 integrin was expressed but was not successfully immunoprecipitated during Co-IP or 2) the protein was expressed and immunoprecipitated but did not interact with galectin-3.

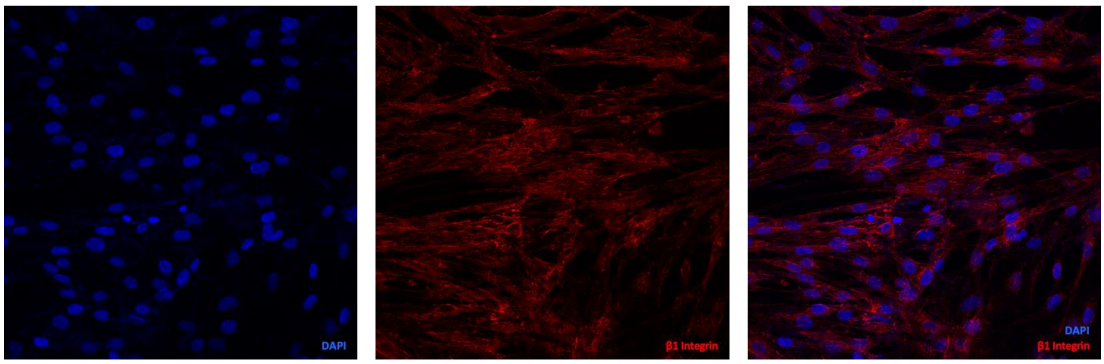


Figure 5.2: Immunofluorescence staining of the β1 integrin in HLFs
Representative confocal microscopy images (20x magnification) showing expression of the β1 integrin (red) and DAPI counterstaining (blue) in HLFs at p6 (N=1). Cells probed with an anti-β1 integrin primary antibody (5 µg/mL) and a AF568-conjugated secondary antibody (4 µg/mL).

In order to address the first possibility, a series of experiments were performed in an attempt to optimise the standard Co-IP protocol to improve the chances of detecting galectin-3 and β1 integrin interactions. As integrins are transmembrane receptors, the β1 integrin must be solubilised in the Co-IP lysis buffer for successful Ab pulldown. As previously described (Section 3.3.4.2) detergents are commonly used to aid solubilisation of membrane proteins. Therefore, a non-ionic, non-denaturing detergent (IGEPAL CA-630) was added to the Co-IP lysis buffer at a range of concentrations in an attempt to improve solubilisation of the β1 integrin with the results evaluated by western blot. After IP a band was detected at the expected protein size (130 kDa) for β1 integrin in the sample prepared in lysis buffer containing 1.5% detergent (lane 7) (Figure 5.3A). The results suggested that only a small fraction of the β1 integrin protein was soluble as no clear band was detected in control lanes (input, flow through and wash steps). Though this was not a problem given that a sufficient level of β1 integrin had immunoprecipitated and was detected in the IP reaction. In the sample prepared in lysis buffer containing 1% detergent (lane 4) a faint band was detected at 130 kDa which suggested that 1% detergent also increased

protein solubility, but not to the same extent as 1.5% detergent (Figure 5.3A). Following this, the Co-IP was repeated with a 1.5% detergent lysis buffer and samples separated by SDS-PAGE in reducing or non-reducing conditions. In reducing SDS-PAGE, a clear band the size of galectin-3 (27 kDa) was detected by immunoblotting after β1 integrin Ab pulldown (lane 4), no band the size of galectin-3 was detected in non-reducing SDS-PAGE (lane 7) (Figure 5.3B). Optimal detection of galectin-3 and the β1 integrin was with samples prepared in a lysis buffer containing 1.5% detergent and analysed under reducing conditions by SDS-PAGE, therefore these condition were used for further Co-IP experiments in untreated HLFs.

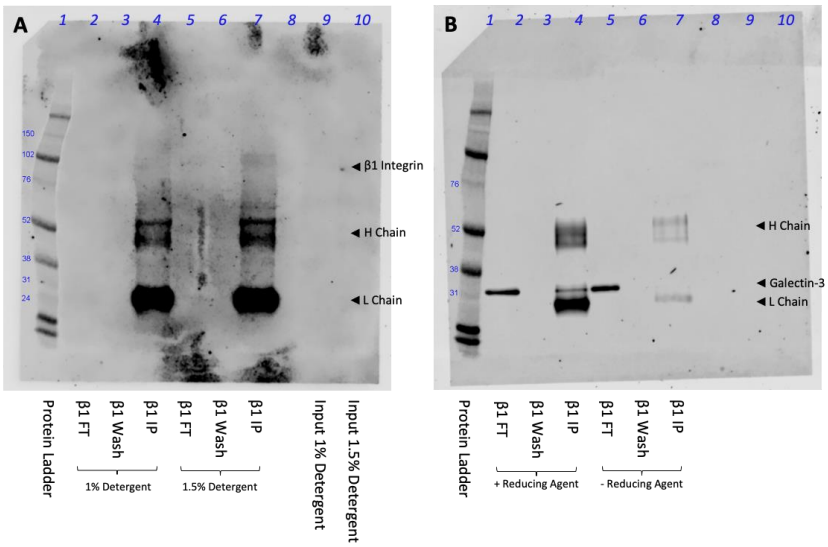


Figure 5.3: Optimisation of β1 integrin immunoprecipitation and galectin-3 immunoblotting ***(A)*** *Detergent (IGEPAL CA-630) was added to Co-IP lysis buffer at a 1% or 1.5% final concentration. Protein lysates (500 µg/ IP reaction) from untreated HLFs p6 (N=1) were immunoprecipitated with an anti-β1 integrin (1 µg/ IP reaction) antibody and immunoblotted for β1 integrin.* ***(B)*** *Untreated HLFs p6 (N=1) were lysed in 1.5% detergent lysis buffer and protein lysates (500 µg/ IP reaction) immunoprecipitated with an anti-β1 integrin antibody. Co-IP samples were subjected to SDS-PAGE in reducing or non-reducing conditions and immunoblotted for galectin-3. Co-IP input, flow through (FT) and wash steps loaded as controls. Proteins separated by reducing SDS-PAGE unless otherwise stated and target protein size estimated from the Amersham ECL Full-Range Rainbow Markers (12 to 225 kDa) migration pattern.*

Co-IP was performed using these optimised conditions to assess the galectin-3 and β1 integrin interaction alongside an IgG control IP. A faint band the size of galectin-3 (27 kDa) was detected after β1 integrin Ab pulldown (lane 6) and was not detected in the IgG control IP (lane 10) which suggested that there was no/minimal non-specific binding to the magnetic beads (Figure 5.4A). Galectin-3 was detected in the input (lane 2), in the β1 integrin IP flow through (lane 4) and in the IgG IP flow through (lane 8). This showed that galectin-3 was present in the lysate incubated with the magnetic beads (Protein G) coupled to anti-IgG_1 isotype control antibody and subsequently validated the negative IgG IP results. The blot was then re-probed for the β1 integrin (130 kDa) and as expected a band was detected in the β1 integrin IP (lane 6) but not in the IgG control IP (lane 10) (Figure 5.4B). Again β1 integrin protein was not detected in control lanes (input, flow through and wash steps) due to low solubility.

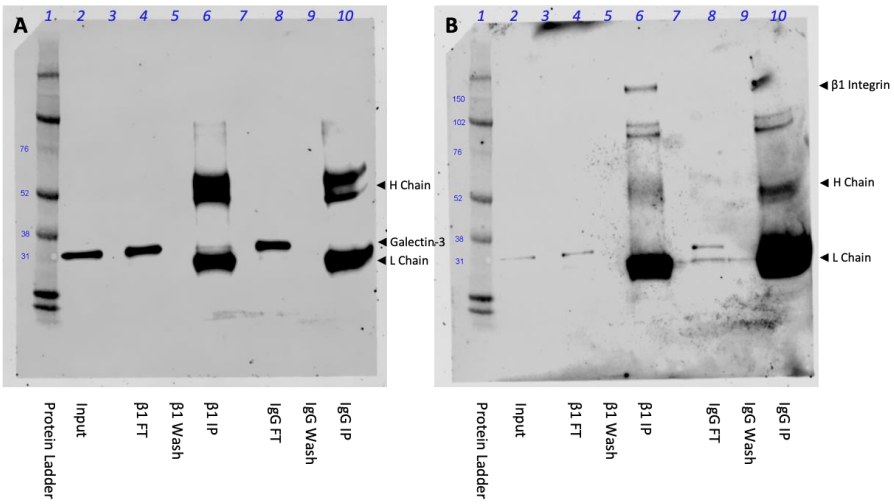


Figure 5.4: Co-immunoprecipitation of galectin-3 with β1 integrin Ab pulldown in HLFs
*Protein lysates (500 µg/ IP reaction) from untreated HLFs p6 (N=1) were immunoprecipitated with an anti-β1 integrin antibody (1 µg/ IP reaction) and immunoblotted for **(A)** galectin-3 or **(B)** β1 integrin. Co-IP input, flow through (FT) and wash steps loaded as controls. Proteins separated by reducing SDS-PAGE and target protein size estimated from the Amersham ECL Full-Range Rainbow Markers (12 to 225 kDa) migration pattern.*

Although galectin-3 successfully co-immunoprecipitated with the β1 integrin, the galectin-3 band (27 kDa) was extremely close to that of the IgG L chain (25 kDa). Therefore, further optimisation was attempted in order to remove/minimise the intensity of the IgG heavy and light chain. To achieve this, Co-IP experiments were performed with agarose beads and the pulldown antibody coupled/crosslinked to the resin. Covalently linking the pulldown antibody to the resin prevents the antibody fragments being eluted from the beads with the binding partner of interest. Using this approach, galectin-3 co-immunoprecipitated with β1 integrin Ab pulldown (lane 3) and neither IgG chain was detectable by immunoblotting (Figure 5.5A). In addition, no galectin-3 band was detectable in the IgG control IP (lane 2). Galectin-3 was detected in all input and flow through controls. As a final confirmation of the galectin-3 and β1 integrin interaction, the Co-IP was then performed in the opposite direction. The β1 integrin (130 kDa) successfully co-immunoprecipitated with galectin-3 Ab pulldown (lane 3), although a faint band was detected in the IgG control IP (lane 2) which indicated a low level of non-specific binding (Figure 5.5B). The β1 integrin was detected in all input and flow through controls.

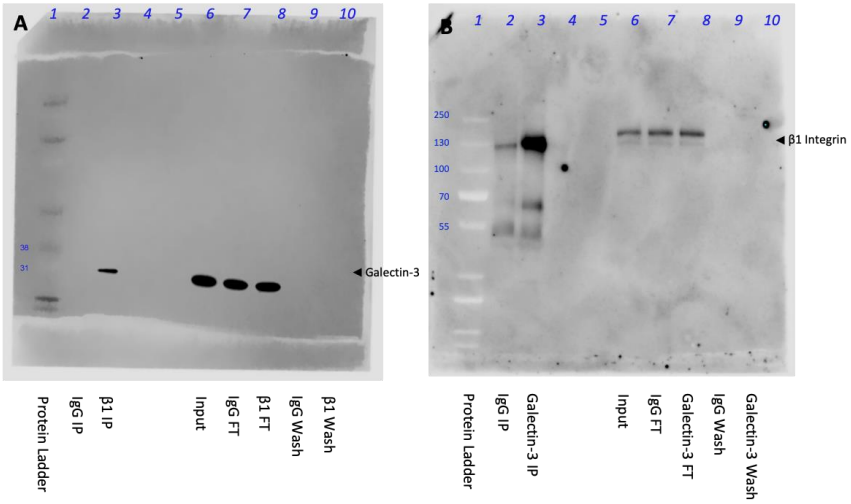


Figure 5.5: Galectin-3 binding to the β1 integrin by Co-IP
Protein lysates (650 µg/ IP reaction) from untreated HLFs p6 (N=3) were immunoprecipitated with ***(A)*** *an anti-β1 integrin antibody (10 µg/ IP reaction) and immunoblotted for galectin-3 or* ***(B)*** *immunoprecipitated with an anti-galectin-3 antibody (10 µg/ IP reaction) and immunoblotted for the β1 integrin. Co-IP input, flow through (FT) and wash steps loaded as controls. Proteins separated by reducing SDS-PAGE and target protein size estimated from the Amersham ECL Full-Range Rainbow Markers (12 to 225 kDa) migration pattern (left) or the PageRuler Plus Prestained Protein Ladder (10 to 250 kDa) (right).*

5.4.2. Localisation of galectin-3 and αv integrins in HEK293T cells

To assess the distribution of galectin-3 and αv integrins *in vitro*, preliminary experiments were performed in HEK293T cells which were co-transfected with plasmid DNA for galectin-3 and αvβ5 or αvβ6 (Figure 5.6). This approach was taken as a result of having no access to primary lung fibroblasts whilst at the RCaH and experiments were performed for proof of concept only given that cloning for the β1 full length GFP construct was unsuccessful (Section 3.4.1). Results showed that αv integrins indicated by green signal (GFP) were predominantly expressed at the cell surface, with low levels also localised to the cytoplasm. In contrast, galectin-3 was primarily expressed in the nucleus and cytoplasm indicated by red signal (mCherry). No galectin-3 was detectable at the cell surface or extracellularly which suggested that it had not been secreted by the cells, consequently it was not possible to assess

colocalisation of extracellular galectin-3 and αv integrins at the cell surface using this method.

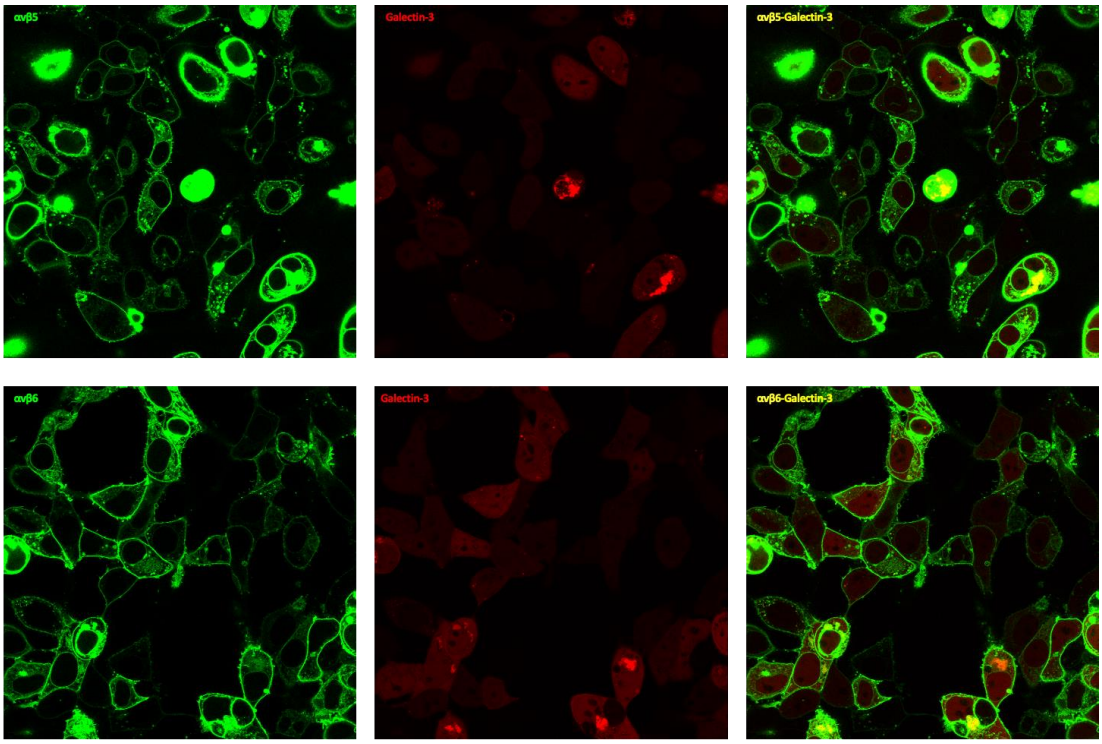


Figure 5.6: Distribution of αv integrins and galectin-3 by live cell imaging
Representative confocal microscopy images (63x magnification) showing the localisation of galectin-3 and αv integrins in HEK293T cells 36 hours post-transfection. Cells were transfected with 200 ng of isolated plasmid DNA for galectin-3 (mCherry-tagged) and αvβ5 or αvβ6 (GFP-tagged).

5.4.3. Galectin-3 and the β1 integrin are colocalised in IPF HLFs

Due to the drawbacks of transient protein expression, PLA experiments were performed in collaboration with Prof Bibek Gooptu and Dr Panayiota Stylianou (University of Leicester) to specifically assess colocalisation of galectin-3 and the β1 integrin in primary lung fibroblasts. No colocalisation was detected in untreated non-IPF HLFs indicated by lack of red PLA signal (Figure 5.7A). Yet following stimulation with TGF-β1 (2 ng/mL) a low level of colocalisation was detected as indicated by red signal (Figure 5.7B).

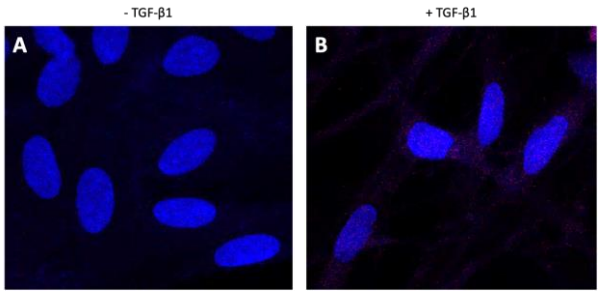


*Figure 5.7: Colocalisation of galectin-3 and the β1 integrin in non-IPF HLFs by in situ PLA Representative confocal microscopy images (63x magnification) showing PLA of galectin-3 and the β1 integrin in non-IPF HLFs p3-4 (N=3) in the **(A)** absence or **(B)** presence of TGF-β1 stimulation (2 ng/mL TGF-β1 for 24 hours). Cells probed with a mouse anti-β1 integrin primary antibody (5 µg/mL) and a rabbit anti-galectin-3 primary antibody (5 µg/mL) followed by anti-rabbit PLUS and anti-mouse MINUS probes. Colocalisation of galectin-3 and the β1 integrin ≤40 nm indicated by red fluorescence with DAPI counterstaining (blue).*

In contrast, when colocalisation of galectin-3 and the β1 integrin was assessed in IPF HLFs, a clear red signal was detected which indicated that the two proteins were colocalised on the cell surface ≤40 nm (Figure 5.8A) and this was augmented following stimulation with TGF-β1 (2 ng/mL) (Figure 5.8B). As the cells were not permeabilised it is presumed that the antibodies and subsequently the PLA probes only bound to cell surface protein. Target inhibition of the galectin-3 CRD with the galectin-3 inhibitor GB0139 inhibited this colocalisation in a concentration-dependent manner indicated by a decrease in red PLA signal. At the 0.1 µM concentration minimal colocalisation signal was detected in untreated cells (Figure 5.8A), this was increased to 1 µM with TGF-β1 treatment (Figure 5.8B).

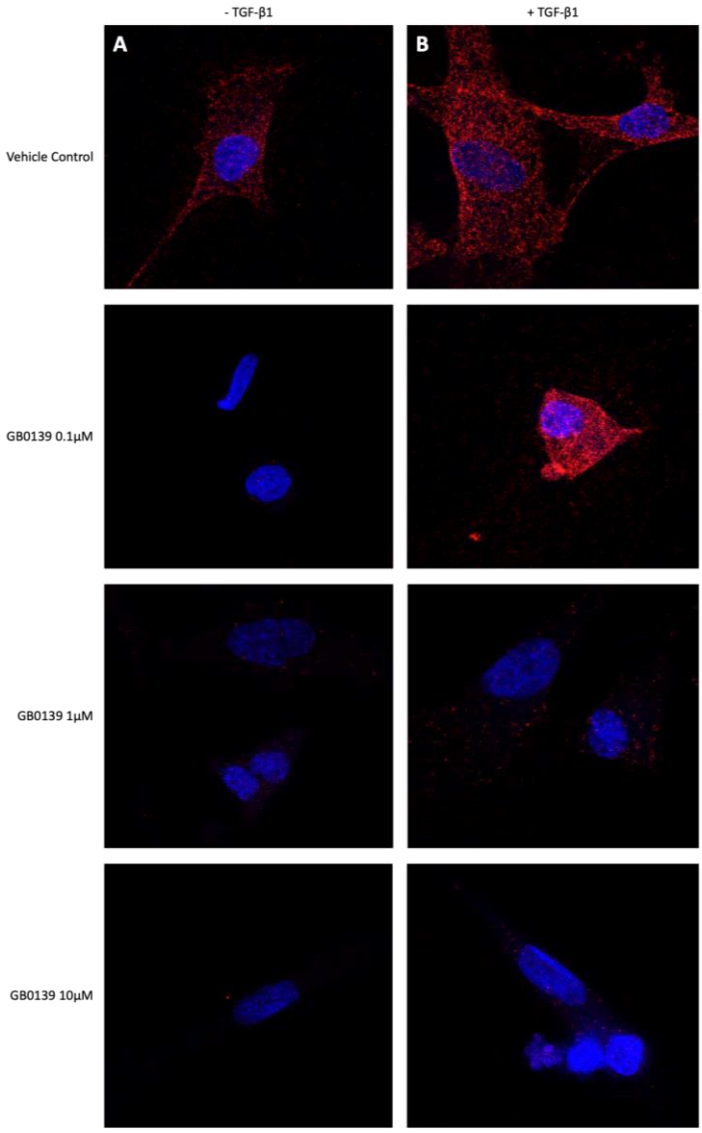


Figure 5.8: Colocalisation of galectin-3 and the β1 integrin in IPF HLFs by in situ PLA
*Representative confocal microscopy images (63x magnification) showing PLA of galectin-3 and the β1 integrin in IPF HLFs p3 (N=4) in the **(A)** absence or **(B)** presence of TGF-β1 stimulation (2 ng/mL TGF-β1 for 24 hours) and the galectin-3 inhibitor GB0139. Cells probed with a mouse anti-β1 integrin primary antibody (5 μg/mL) and a rabbit anti-galectin-3 primary antibody (5 μg/mL) followed by anti-rabbit PLUS and anti-mouse MINUS probes. Colocalisation of galectin-3 and the β1 integrin ≤40 nm indicated by red fluorescence with DAPI counterstaining (blue).*

5.5. Discussion

To confirm the key protein interactions previously detected by SPR (Chapter 4) the binding of galectin-3 to the αvβ1 integrin was investigated *in vitro* using a combination of protein interaction and colocalisation methods in primary lung fibroblasts. By Co-IP the binding of galectin-3 to the αvβ1 integrin was confirmed in HLFs, although numerous issues were initially encountered with the detection of the protein interaction. Firstly, in the standard Co-IP lysis buffer the β1 integrin was not soluble and therefore did not bind to the magnetic beads coupled to anti-β1 integrin antibody. Optimisation of the Co-IP lysis buffer was therefore required to increase the membrane protein solubility and optimal detection of galectin-3 and the β1 integrin was with samples lysed in buffer containing 1.5% detergent. Secondly, when performing the Co-IP with magnetic beads the galectin-3 band (27 kDa) was extremely close to that of the IgG L chain (25 kDa) subsequently further optimisation was required to remove/minimise the intensity of the IgG chains. By switching from magnetic beads to agarose resin (Ab coupled/crosslinked) for immunoprecipitation, antibody IgG chains were no longer detected by immunoblotting, specifically there was no IgG L chain that could obscure the protein of interest (galectin-3).

Overall, immunoprecipitation was more effective when using the agarose resin as evident from the target protein band intensity and the minimal non-specific binding (background) detected. There are, however, several known performance advantages of the agarose beads used which may explain the results observed. Firstly, although both types of beads have a high binding capacity, agarose resin is highly-porous and subsequently there is a larger surface area for the antigen-antibody-protein A/G

complex to bind (372). This enhances the capacity of agarose beads and the yields obtained, however a disadvantage is that more antibody is consequently required for surface coupling and to minimise non-specific binding (372). By comparison, magnetic beads are non-porous thus despite there being more beads per unit volume the theoretical binding capacity is lower (372). Secondly, as recombinant Protein A and G bind to the IgG subtypes from different species with varying affinities, immunoprecipitation beads have varying capacities to bind antibody depending on whether they are coupled to recombinant Protein A and/or G (373). The magnetic beads used here were coupled to Protein G alone, yet the agarose beads were coupled to both Protein A and G and therefore able to capture immunoglobulins from a wider range of species and antibody isotypes. This was advantageous given that the antibody used for β1 integrin pulldown was mouse IgG_1 and the antibody used for galectin-3 pulldown was rabbit IgG, which Protein G and Protein A have the highest affinity for, respectively (374). Considering the above, agarose resin was most appropriate for confirming the galectin-3 and β1 integrin interaction *in vitro*, although it should be noted that cell lysis prior to Co-IP can cause proteins that would never usually interact inside a cell to come into close proximity and associate.

Transient overexpression experiments were initially performed as proof of concept to assess the cellular localisation of galectin-3 and αv integrins, though it was not possible to use αvβ1 specifically for these experiments as cloning of the β1 full length GFP construct was unsuccessful (Section 3.4.1). Nevertheless, as galectin-3 also bound to αvβ5 and αvβ6 by SPR, preliminarily experiments were performed with both constructs to assess feasibility of the technique. As endogenous galectin-3 was

not detected extracellularly, the technique was concluded unsuitable for assessing colocalisation at the cell surface, even if the αvβ1 integrin construct was instead used. This was similarly shown in Madin-Darby canine kidney cells (MDCKs) whereby exogenous galectin-3 had to be added to assess the surface distribution of galectin-3 and the β1 integrin (375).

As a result, PLA experiments were performed in collaboration with Prof Bibek Gooptu (University of Leicester) for *in situ* detection of the endogenous proteins in IPF and non-IPF HLFs (Section 5.4.3). In these experiments there was variability of signal intensities by donor but the overall findings were consistent despite this. By PLA, colocalisation of galectin-3 and the β1 integrin was detected in untreated IPF HLFs but not non-IPF. This was unexpected considering that the successful Co-IP experiments were performed in untreated non-IPF HLFs. However, protein expression of both galectin-3 and the αvβ1 integrin is significantly elevated in IPF, thus perhaps the PLA signal was not observed in untreated non-IPF HLFs due to the method detection limit (226, 284). To address this possibility, non-IPF HLFs were stimulated with TGF-β1 for 24 hours and the PLA repeated. This timepoint was used as 24 hours of TGF-β1 stimulation has been shown to cause a marked increase in surface expression of β1 integrins (217). Following stimulation, a PLA signal was detected in these non-IPF cells but the PLA signal was still relatively low. As expected, TGF-β1 treatment also augmented the PLA signal detected in IPF HLFs. These observations are consistent with non-peer reviewed results from the University of Leicester showing that colocalisation of galectin-3 and the β1 integrin was only detectable in TGF-β1 treated *ex vivo* human lung tissue (non-IPF) and not in untreated

(287). Adding to this, they showed that TGF-β1 treatment (24 hours) significantly increased the protein expression of galectin-3 and the β1 integrin in A549 cells. Following success of the PLA, experiments were performed to assess the effect of galectin-3 inhibition on galectin-3 and β1 integrin colocalisation. The galectin-3 inhibitor GB0139 inhibited colocalisation of galectin-3 and the β1 integrin in IPF HLFs in a concentration-dependent manner. This suggests that the anti-fibrotic activity of GB0139 in lung fibrosis may be partly due to targeting galectin-3 in TGF-β1 signaling in lung fibroblasts. These data are also encouraging given that the concentration of compound used here covers the concentration range used in the phase 1/2a study and is therefore translatable to the clinic (NCT02257177) (291).

Summary: To summarise, galectin-3 bound to the β1 integrin in non-IPF HLFs by Co-IP irrespective of order of IP and western blot, therefore validating the previous SPR data whereby galectin-3 bound to the αvβ1 integrin. Colocalisation of galectin-3 and the β1 integrin was subsequently assessed in IPF and non-IPF HLFs by PLA. A higher level of colocalisation was detected in IPF HLFs when compared to non-IPF, although TGF-β1 treatment increased the PLA signal in both donor types. Pre-treatment of cells with the galectin-3 inhibitor GB0139 inhibited this colocalisation in a concentration-dependent manner, at clinically meaningful concentrations.

Limitations/ Future Work: Although these experiments were useful for investigating whether galectin-3 interacts with the β1 integrin in primary lung fibroblasts, several limitations do exist from the methodology. Firstly, by Co-IP and PLA it was not possible to determine if the protein interaction was direct or indirect via a third

unknown protein. Secondly, as an anti-β1 integrin primary antibody was used in both experiments it is unknown whether galectin-3 bound to both the αv and β1 subunits specifically. It would be possible to assess this by re-probing the western blots for the αv subunit, although the protein MW is very close to that of the β1 subunit. Finally, as both methods are highly dependent on the availability and quality of antibodies that recognise the target protein, similar experiments performed for galectin-3 and the TGFβRII subunit were unsuccessful (data not shown). In these Co-IP experiments multiple non-specific bands were detected in the IP reactions and this occurred with all anti-TGFβRII antibodies tested. Similarly, TGFβRII was not detectable by immunofluorescence or in western blot studies at the University of Leicester (287). However, if suitable antibodies were available then it would be important in future to perform a PLA for the β1 integrin and TGFβRII subunit, in the presence or absence of galectin-3 inhibition. This would address whether galectin-3 facilitates clustering of the αvβ1 integrin and the TGF-β1 receptor on their respective cell surfaces.

6. Investigating integrins as potential binding partners for SARS-CoV-2

6.1. Introduction

COVID-19 was first reported in Wuhan, China in December 2019 after its identification in hospitalised patients with severe pneumonia of unknown cause (376). The World Health Organisation (WHO) subsequently declared the outbreak a COVID-19 global pandemic following epidemiological assessment (377). By phylogenetic analysis it was revealed that the virus responsible for COVID-19 was a betacoronavirus showing clear similarities with the SARS-CoV virus that caused a worldwide outbreak of severe acute respiratory syndrome in 2003 (376). The enveloped, positive-sense, single stranded RNA betacoronavirus responsible for COVID-19 was subsequently named SARS-CoV-2 (previously known as "2019 novel coronavirus") (376). The sequence for the SARS-CoV-2 spike protein and receptor binding motif was shown to be similar to that of SARS-CoV with 76% homology and 73% homology, respectively (378). The primary functional receptor for SARS-CoV was identified as angiotensin converting enzyme 2 (ACE2) and the high degree of sequence homology led to speculation of a key role for ACE2 in cellular entry and transmission of the SARS-CoV-2 (379).

ACE2 is a single-pass type I membrane protein with an extracellular N-terminal peptidase domain and intracellular C-terminal collectrin-like domain (380). Using X-ray crystallography it was confirmed that SARS-CoV-2, like SARS-CoV, binds to the N-terminal ACE2 peptidase domain (381, 382). The SARS-CoV-2 genome encodes a nucleocapsid protein, membrane protein, envelope protein and spike protein and it

is the spike protein at the virion surface which recognises and binds to ACE2 (383) (Figure 6.1). The SARS-CoV-2 spike protein is comprised of two subunits namely S1 and S2 (384). The S1 domain of spike contains a receptor binding domain (RBD) used for attachment of the virus to ACE2 on host cells, and although not directly involved in receptor binding, the S2 subunit is important for membrane fusion (385).

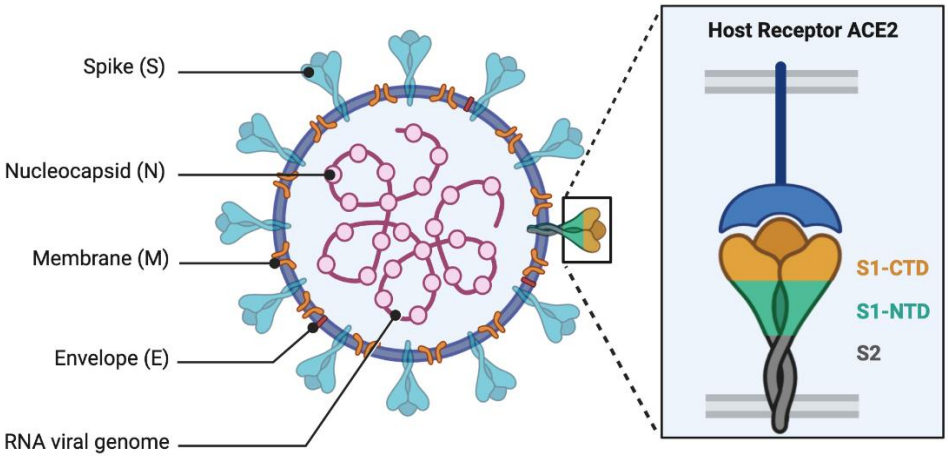


Figure 6.1: Structure of SARS-CoV-2
Diagram of the structure of SARS-CoV-2 and its attachment to host cell receptors. (Created in BioRender.com)

Early clinical studies suggested that severe COVID-19 symptoms were associated with SARS-CoV-2 infection of the lower respiratory tract (386). However analysis of the available scRNAseq data assessing the patterns of gene expression throughout the respiratory system suggested that there is low ACE2 gene expression and only rare ACE2 protein expression in the healthy airway epithelium and alveoli (387, 388). As ACE2 is an interferon-stimulated gene (ISG) it was hypothesised that SARS-CoV-2 infection of the upper respiratory tract may induce a type 1 interferon (IFN-I) response and upregulate ACE2 protein expression in the lower airways (389, 390). Additionally, it was suggested that the virus may use non-ACE2 mediated mechanisms of cell infection in the lung and a number of potential binding partners

investigated. Subsequently the virus has been shown to utilise several co-receptors: CD147/ basigin (BSG), neuropilin-1 (NRP1), HSPA5/ glucose-regulated protein 78 (GRP78) and proteases: transmembrane serine protease 2 (TMPRSS2), cathepsin L (CTSL), FURIN, a disintegrin and metalloprotease 17 (ADAM17) to facilitate infection (388, 391-397). As the spike protein S1 N-terminal domain has the same structural fold as human galectins, a feature that is universal to the betacoronavirus genus, it was also suggested that SARS-CoV-2 may also be able to mediate its effects by binding to galectin ligands (398). Supporting this, it has recently been shown that the SARS-CoV-2 RBD can induce pro-inflammatory responses (IL-6 and interleukin 8 (IL-8)) in a CD98-dependent manner (287). Interestingly, RGD-binding integrins (α5β1, α8β1, αvβ1, αvβ3, αvβ5, αvβ6, αvβ8, and αIIbβ3) were also proposed as a potential SARS-CoV-2 binding partner that may contribute to infection after identification of the RGD (403–405: Arg-Gly-Asp) motif in the RBD of SARS-CoV-2 (399). This is in agreement with previous studies that have shown RGD integrin binding to play a role in viral infection, although not all virus-integrin interactions are RGD-dependent (400).

Three different outcomes of the SARS-CoV-2-integrin interaction were initially proposed: 1) The virus is internalised, 2) SARS-CoV-2 binding stimulates integrin-mediated cellular responses without internalisation or 3) Their interaction is antagonistic by preventing ACE2 binding and downstream signaling (401). The SARS-CoV-2 spike protein has been shown to exist in two different confirmational states namely: “down” (receptor-inaccessible state) and “up” (receptor-accessible state) (384). When spike is in the “down” confirmation (receptor-inaccessible state) the RGD motif is not completely exposed (cryptic) as this would require removal of

residues 437-508 (399). However, these surrounding residues are sufficiently flexible to expose the RGD motif and facilitate the integrin interaction. When ACE2 binds to the S1 domain it induces a hinge-like conformational change in the virus RBD which then exposes the RGD motif (399) (Figure 6.2).

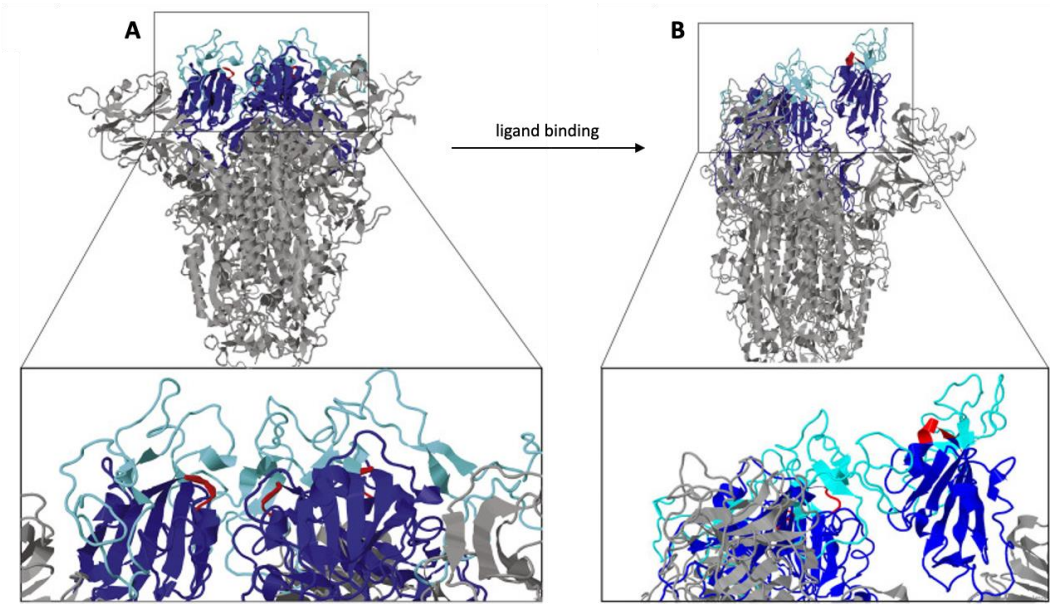


Figure 6.2: Confirmational change of SARS-CoV-2 spike protein
*Position of the RGD motif with spike in the **(A)** native confirmation or **(B)** following ligand (ACE2 receptor) binding. SARS-CoV-2 RBD is indicated in dark blue, the region of the RBD that ACE2 binds is indicated in light blue. The RGD motif is indicated in red. Figure adapted from article "A potential role for integrins in host cell entry by SARS-CoV-2" with permission (399).*

By understanding whether or not integrins and/or galectins mediate respiratory epithelial internalisation of the virus, drugs in development (integrin or galectin inhibitors) may be rapidly repurposed to treat COVID-19 by reducing viral entry within the lungs.

6.2. Aims

As both integrins and galectins could have a role in SARS-COV-2 infection, the aims of this chapter are as follows:

- To determine whether RGD integrins can bind directly to the SARS-CoV-2 spike protein
- To determine whether galectin-3 and/or galectin-1 interact directly with the SARS-CoV-2 spike protein

6.3. Methods

6.3.1. Solid-Phase Binding Assay

Principle of the method: Solid-phase binding assays are used to assess and quantify protein interactions. A protein of interest is maximally coated to the bottom surface of a 96-well plate and a serial dilution of the binding partner of interest then added across the different wells. An enzyme-labelled antibody (e.g. HRP-conjugated) specific for the binding partner of interest can then be used to detect whether or not an interaction has occurred with the immobilised protein. If an interaction has occurred this can be visualised by the addition of a colorimetric HRP substrate (TMB and hydrogen peroxide). Hydrogen peroxide becomes reduced and TMB oxidised by HRP which turns the reaction blue. Upon acidification, the reaction turns yellow and the absorbance of the stopped reaction measured at 450 nm. Absorbance readings alone are used to quantify the level of protein binding with comparison to a positive and negative control.

Protocol: Solid-phase binding assay experiments were performed to investigate if the SARS-CoV-2 spike protein S1 subunit binds to RGD integrins (αvβ1, αvβ5, αvβ6, αvβ8, αvβ3, αIIbβ3 and α5β1) or galectins (galectin-1 and galectin-3). Experiments were performed in either cation or cation free buffer conditions depending on the requirements of the assay (Table 6.1).

Table 6.1: Composition of solid-phase binding assay buffers

Buffer	Integrins	Galectins
Protein diluent	1X TBS pH 7.4 (+/- 1 mM $MnCl_2$)	1X PBS pH 7.4
Wash buffer	1X TBS pH 7.4, 1 mg/mL BSA (+/- 1 mM $MnCl_2$)	1X PBS pH 7.4, 1 mg/mL BSA
Blocking buffer	1X TBS pH 7.4, 5% BSA	1X PBS pH 7.4, 5% BSA
Spike protein diluent	1X TBS pH 7.4, 1 mg/mL BSA (+/- 1 mM $MnCl_2$ or 5 mM EDTA)	1X PBS pH 7.4, 1 mg/mL BSA
Antibody diluent	1X TBS pH 7.4, 0.5% BSA (+/- 1 mM $MnCl_2$)	1X PBS pH 7.4, 0.5% BSA

Binding partners of interest (2 µg/mL) were immobilised into wells of a 96-well plate (100 µL/well) by incubation overnight at 4°C. ACE2 protein (2 µg/mL) and BSA (2 µg/mL) were also immobilised onto the same plate as a positive and negative control, respectively. The next day, any unbound protein was removed from the plate by washing all wells once with wash buffer (300 µL/well). To prevent non-specific protein and antibody binding to the plate, wells were blocked for 2 hours at room temperature (300 µL/well) and then washed twice to remove excess block (300 µL/well). Binding of the S1 subunit to immobilised proteins of interest was assessed over a concentration range (experiment dependent) by serially diluting the S1 subunit protein. The S1 subunit was incubated on the protein coated plate (100 µL/well) for 1 hour at room temperature. Unbound S1 subunit was then removed by washing all wells 3 times with wash buffer (300 µL/well). A HRP-conjugated secondary antibody (Invitrogen, 31439) at 2 µg/mL final concentration was then added to all wells (100 µL/well) and incubated for 1 hour at room temperature to detect the protein bound S1 subunit. After 1 hour all wells were washed 3 times with wash buffer (300 µL/well). Bound HRP-conjugated antibody was detected following a 30-minute incubation with TMB substrate (200 µL/well) at room temperature. The plate was protected from the light during the incubation and the reaction stopped by the addition of 2 M H_2SO_4 (50

µL/well). Absorbance was measured at 450 nm (with correction at 570 nm) on a plate reader (FLUOstar Omega, BMG Labtech) with MARS software. Results were plotted in GraphPad Prism v9 and data presented as mean ± standard deviation (SD).

6.3.2. DNA Transfection

HEK293T cells were grown and maintained in culture as previously described (Section 5.3.1). When confluent, cells were seeded and transfected with FuGENE-HD transfection reagent (Promega, E2311) as previously described (Section 5.3.4). Briefly, cells were seeded into an 8-well chamber slide at a cell density of 25000 cells/well and returned to the incubator for 7 hours prior to transfection. After 7 hours, 200 ng of isolated plasmid DNA was mixed with 10 µL of serum-free DMEM media, 0.5 µL of FuGENE and incubated at room temperature for 10 minutes. The fuGENE-DNA mixture was then pipetted into the appropriate wells and the cells incubated at 37°C with 5% CO_2 for 36 hours before live-cell imaging. When required, spike protein (3 µM) was incubated with the co-transfected cells for 15 minutes at 37°C prior to imaging. All DNA transfection experiments were performed with Dr Siva Ramadurai (RFI, Harwell Campus) and confocal images taken on a DMi8 microscope (Leica) by Dr Siva Ramadurai due to COVID restrictions. Plasmid DNA used for experiments was generated by either myself or Dr Ramadurai as indicated below (Table 6.2).

Table 6.2: Plasmid-encoded proteins (αv integrins and ACE2) and recombinant proteins (spike) with fusion tags

	Target	Tag	Generated By
Plasmid DNA	αvβ5	GFP	Jessica Calver
	αvβ6	GFP	
	ACE2	GFP, mCherry	Dr Ramadurai
Protein	Spike	AF568, AF405	

6.4. Results

6.4.1. SARS-CoV-2 physically interacts with integrins αvβ6 and αvβ3 but not αIIbβ3

As previously described (Section 1.3.3) ligand-binding to integrins is dependent on the presence of divalent cations at their MIDAS motif with metal specificity for Ca^{2+}, Mg^{2+} and Mn^{2+} (198, 199). The three cations have markedly different effects on integrin-ligand binding capacity with Mn^{2+} promoting the highest levels of binding and subsequently solid-phase binding assays were performed in buffers containing 1 mM $MnCl_2$. Results showed that the SARS-CoV-2 S1 subunit bound to both αvβ6 and αvβ3 (2 µg/mL) in a concentration-dependent manner (Figure 6.3). Adding a chelating agent (EDTA at 5 mM) to the spike protein diluent buffer reduced the level of binding detected which therefore confirmed that both αvβ6 and αvβ3 integrin binding to the S1 subunit was cation-dependent. ACE2 protein (2 µg/mL) was immobilised onto the same plate as the integrins as a positive control and as expected there was a concentration-dependent increase in binding to the S1 subunit. Both integrin binding curves were shifted to the right of that for ACE2 which demonstrated that the S1 subunit has lower affinity for αv integrins than ACE2. Compared with αvβ6 and αvβ3 negligible binding was detected for the αIIbβ3 integrin, this was apparent from 0.75 µg/mL of spike protein and above. Minimal spike protein binding was detected in wells coated with BSA (2 µg/mL) as a negative control, however non-specific binding did increase as the concentration of the SARS-CoV-2 S1 subunit increased.

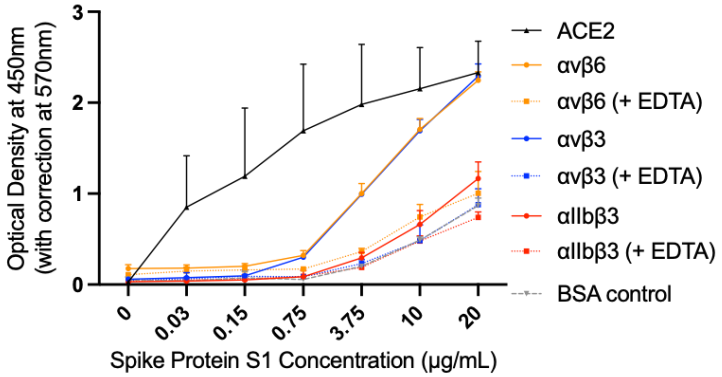


Figure 6.3: Cation-dependent binding of integrins to the spike protein S1 subunit
Solid-phase binding assay data of the SARS-CoV-2 S1 subunit binding to immobilised integrins αvβ6, αvβ3 and αIIbβ3 (2 µg/mL) in cation buffer conditions (1 mM $MnCl_2$), with or without EDTA present (N=3). ACE2 and BSA were immobilised at the same concentration and used as a positive and negative control, respectively. Data presented as mean ± SD with connecting line/curve shown.

6.4.2. SARS-CoV-2 does not physically interacts with integrins αvβ5, αvβ8 or α5β1

By solid-phase binding assay the SARS-CoV-2 S1 subunit did not bind to integrins αvβ5 (Figure 6.4A), αvβ8 (Figure 6.4B) or α5β1 (Figure 6.4C). Due to a shortage of ACE2 protein, binding of the spike protein to αvβ5 (2 µg/mL) was assessed with comparison to αvβ6 binding (Figure 6.4A). Results showed that the level of S1 subunit binding to αvβ5 was minimal when compared with αvβ6 and was instead similar to that detected in the BSA control well or in the presence of EDTA. Here it was not possible to assess the non-specific binding to BSA across the full protein concentration series due to shortage of S1 subunit, however the maximum level of non-specific was determined at the highest S1 subunit concentration (10 µg/mL). When assessing the binding of spike protein to αvβ8 and α5β1 there were no protein shortages thus binding was appropriately assessed with comparison to ACE2 and non-specific binding determined across the full S1 subunit concentration series. However, the SARS-CoV-2 S1 subunit did not bind to αvβ8 (Figure 6.4B) or α5β1 (Figure 6.4C) above the background level detected in BSA control wells.

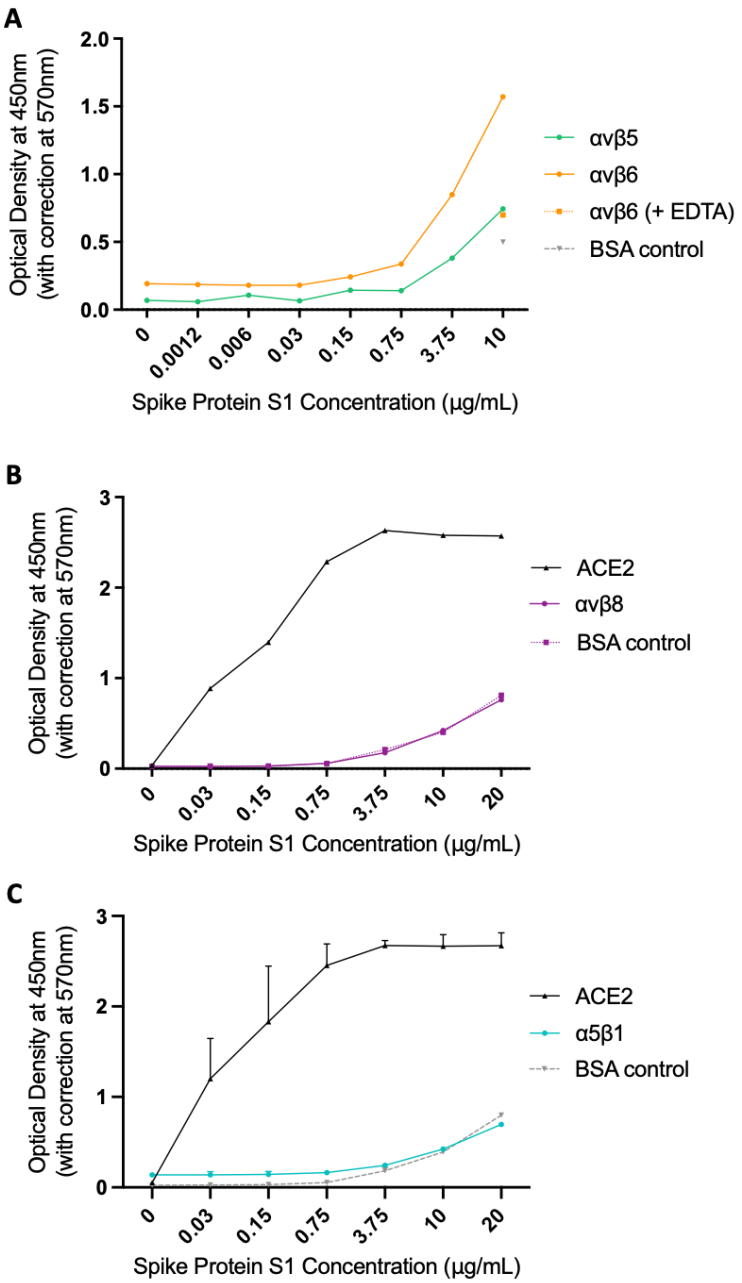


Figure 6.4: The spike protein S1 subunit does not bind to integrin αvβ5, αvβ8 or α5β1
Solid-phase binding assay data of the SARS-CoV-2 S1 subunit binding to immobilised integrins (2 µg/mL) ***(A)*** *αvβ5 (N=1),* ***(B)*** *αvβ8 (N=1) or* ***(C)*** *α5β1 (N=2) in cation buffer conditions (1 mM $MnCl_2$). ACE2 or αvβ6 were immobilised as a positive control and BSA used as a negative control. Data presented as mean ± SD with connecting line/curve shown.*

6.4.3. Preliminary data suggests cation-independent binding of SARS-CoV-2 to αv integrins

Initial solid-phase binding assay experiments were performed in cation free buffer conditions in error and consequently binding experiments had to be repeated in the presence of cations (data shown above). The preliminary data did however suggest

that the SARS-CoV-2 S1 subunit could also bind to αv integrins in the absence of cations (Figure 6.5). Across the spike protein concentration series, the level of binding was similar for integrins αvβ5, αvβ6 and αvβ8 but there was less binding to the αvβ1 integrin by comparison. Differences in integrin binding were apparent from the 0.75 µg/mL concentration of S1 subunit and above. As a negative control was not included in these early experiments it is not possible to confirm if the binding observed here was integrin specific or non-specific binding. These data do however demonstrate why αvβ1 was not taken forward for further binding experiments.

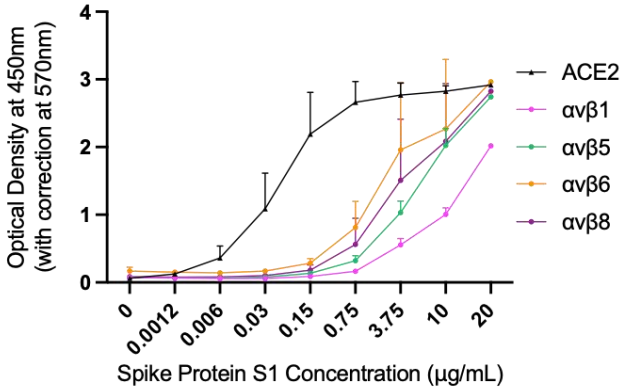


Figure 6.5: Cation-independent binding of αv integrins to the spike protein S1 subunit
Solid-phase binding assay data of the SARS-CoV-2 S1 subunit binding to immobilised integrins αvβ1, αvβ5, αvβ6 and αvβ8 (2 µg/mL) or ACE2 (2 µg/mL) in cation free buffer conditions (N=4). Data presented as mean ± SD with connecting line/curve shown.

6.4.4. SARS-CoV-2 does not physically interacts with galectin-3 or galectin-1

Solid-phase binding assays were also performed to assess binding of the SARS-CoV-2 S1 subunit to galectin-3 and galectin-1. Ligand-binding to galectin-3 and galectin-1 is not dependent on the presence of divalent cations and subsequently these experiments were performed in cation free buffer conditions. The level of S1 subunit binding detected for both galectin-3 and galectin-1 was above the background level detected with BSA, however with comparison to ACE2 this binding was negligible

(Figure 6.6). Though the S1 subunit binding was highest for galectin-1 at each concentration.

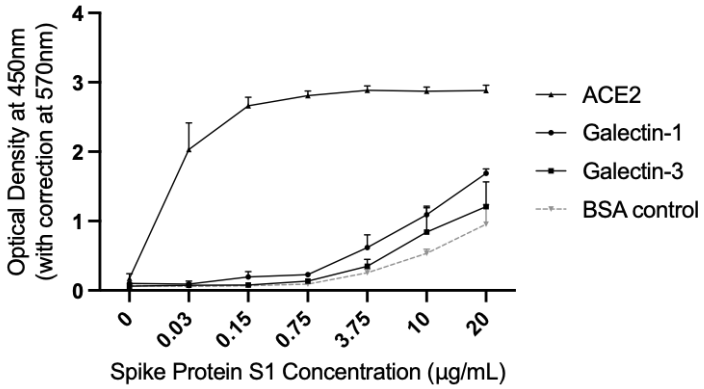


Figure 6.6: The spike protein S1 subunit does not bind to galectin-3 or galectin-1
Solid-phase binding assay data of the SARS-CoV-2 S1 subunit binding to immobilised galectin-3 and galectin-1 (2 µg/mL) in cation free buffer conditions (N=4). ACE2 and BSA were immobilised at the same concentration and used as a positive and negative control, respectively. Data presented as mean ± SD with connecting line/curve shown.

6.4.5. Spike protein, ACE2 and αvβ6 co-associate *in vitro*

To assess spike protein binding to ACE2 *in vitro,* HEK293T cells were transiently transfected with GFP-tagged ACE2 (200 ng) and then AF568-labeled spike protein (3 µM) subsequently added to the cells. At 36 hours post-transfection, ACE2 protein was predominantly expressed at the cell surface but was also distributed in the cytoplasm as indicated by green signal (GFP). When exogenous spike protein was added to the culture medium, it was found to co-localise with ACE2 at the cell surface after 15 minutes as illustrated in yellow (arrow head) (Figure 6.7). However, a proportion of cells did express ACE2 at the cell surface but no spike protein binding detected (asterisk).

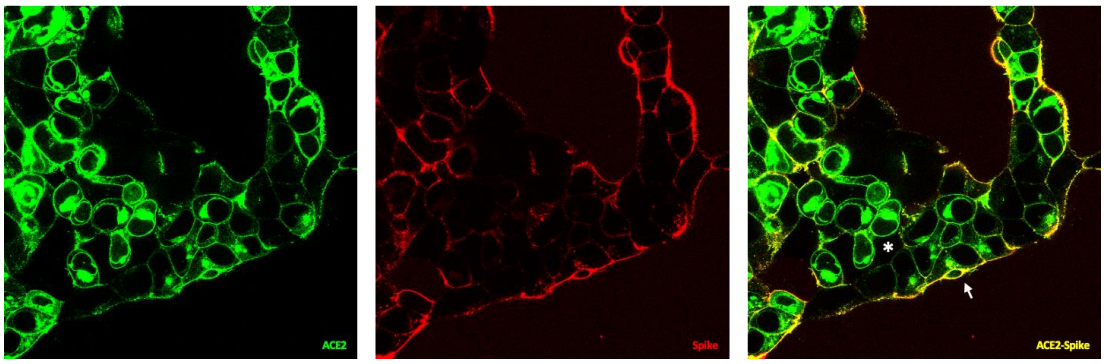


Figure 6.7: Colocalisation of ACE2 and SARS-CoV-2 spike protein (WT) by live cell imaging
Representative confocal microscopy images (63x magnification) showing that ACE2 co-associates with spike protein (yellow). HEK293T cells were transfected with GFP-tagged ACE2 (green) and then AF568-labeled spike protein (red) incubated with the transfected cells. Spike protein was added to the medium 36 hours post transfection and incubated for 15 minutes prior to imaging.

ACE2 and αvβ6 were overexpressed in HEK293T cells and confocal imaging determined that they were colocalised at the cell surface after 36 hours (Figure 6.8). The majority of the cells expressing αvβ6 also expressed ACE2, with areas of colocalisation illustrated in yellow (arrow head). Though a proportion of cells did only express ACE2 as illustrated by red signal (asterisk), low levels of ACE2 were also visible in the cell cytoplasm.

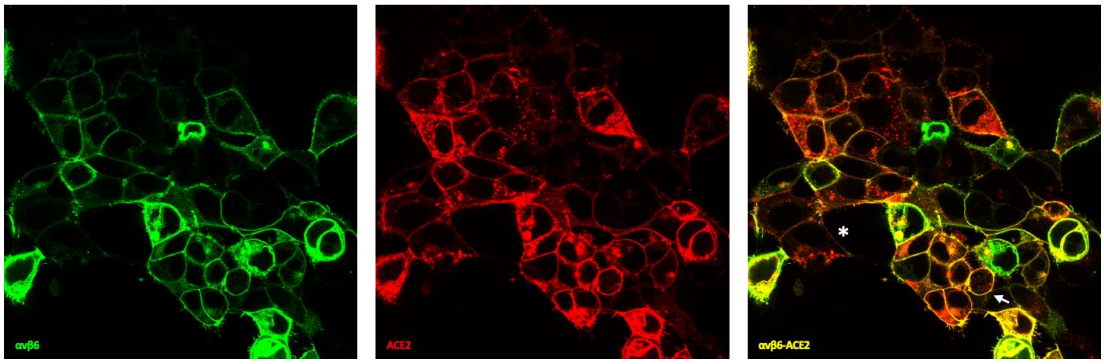


Figure 6.8: Colocalisation of integrin αvβ6 and ACE2 by live cell imaging
Representative confocal microscopy images (63x magnification) showing that αvβ6 co-associates with ACE2 (yellow). HEK293T cells were transfected with GFP-tagged αvβ6 (green) and mCherry-tagged ACE2 (red). Cells were incubated for 36 hours prior to imaging.

In cells co-transfected with αvβ6 and ACE2, exogenously applied spike protein bound to ACE2 at the cell surface as illustrated by the widespread distribution of purple colouration (arrow head). However, the highest level of spike protein binding appeared to correspond to areas where αvβ6 and ACE2 were colocalised as indicated in white (box region) (Figure 6.9). Images also indicated that spike can bind to αvβ6 in the absence of ACE2 as seen in cyan blue (asterisk).

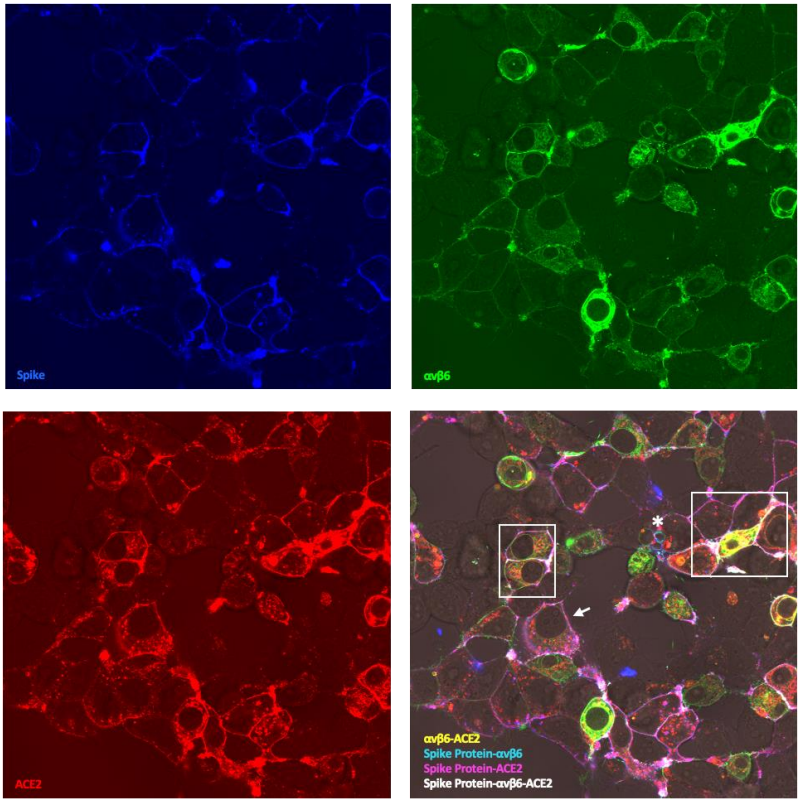


Figure 6.9: Colocalisation of αvβ6 with SARS-CoV-2 spike protein (WT) and ACE2 by live cell imaging
Representative confocal microscopy images (63x magnification) showing that spike protein binds to ACE2 (purple), ACE2 co-associates with αvβ6 (yellow), spike protein binds to αvβ6 (cyan) and triple-colocalization of αvβ6-spike-ACE2 (white). HEK293T cells were transfected with αvβ6 (GFP-tagged) and ACE2 (mCherry-tagged) with spike protein (AF405-tagged) then incubated with the co-transfected cells.

6.5. Discussion

Using scRNAseq data from publicly available datasets it was identified that only 1.5-3% of AT2 cells in the normal healthy lung are ACE2 positive (402). By RNAscope *in situ* hybridisation assay it was also recognised that in the SARS-CoV-2 infected lung there are large amounts of virally infected epithelium in the absence of ACE2 (403). As SARS-CoV-2 infection can lead to acute lung injury in a proportion of patients, it was hypothesised that the virus gains entry to the alveolar epithelium using a second receptor as well as, or instead of ACE2. A potential role for integrins in SARS-CoV-2 viral entry was subsequently proposed after identification of the acquired RGD motif in the SARS-CoV-2 RBD (K403R) (399).

Mn^{2+} has been shown to enhance SARS-CoV-2 viral binding and entry into VERO E6 cells (African green monkey kidney cells) (404). It was initially hypothesised that cation binding to the integrin MIDAS motif induces integrin extension and exposes the RGD recognition site to facilitate SARS-CoV-2 binding (404). Here both αvβ6 and αvβ3 bound to the S1 subunit by solid-phase binding assay, in a cation-dependent manner but with lower affinity than ACE2 (Section 6.4.1). It was concluded from these experiments that S1 binding to both αvβ6 and αvβ3 was RGD-dependent as divalent cation occupancy causes a change in the integrin confirmational state from bent (inactive) to extended (active) thereby exposing the RGD binding site (197). Although, it would be important in future to confirm this by performing a competitive solid-phase binding assay with an RGD-integrin antagonist. Validating this data, it has recently been demonstrated by SPR that both αvβ3 and αvβ6 bind to the SARS-CoV-2 S1-RBD in cation buffer conditions, with αvβ6 showing substantially higher affinity

for the S1-RBD than αvβ3 (405). By solid-phase binding assay no cation-dependent binding was detected for SARS-CoV-2 S1 with αvβ5, αvβ8 or α5β1 (Section 6.4.2) and minimal binding was detected with αIIbβ3 (Section 6.4.1). Perhaps this is due to the amino acid residues flanking the RGD sequence which could influence ligand-binding affinity. Additional data on S1 binding to these integrins appears limited although conflicting positive and negative binding results have previously been reported for SARS-CoV-2 binding to α5β1 (405-408).

The solid-phase binding assay data showing αvβ6 binding to the SARS-CoV-2 S1 subunit was of particular interest given that 'post-COVID-19 fibrosis' develops in a subset of patients after SARS-CoV-2 infection (155, 156). As previously described (Section 1.3.4) αvβ6 plays a key role in integrin-mediated TGF-β1 activation that is critical to the pathogenesis of fibrosis. Subsequently, the potential role of αvβ6 in SARS-CoV-2 infection was further investigated in a cell-based assay (Section 6.4.5). Exogenous SARS-CoV-2 spike protein (WT) was shown to rapidly bind and colocalise with ACE2 when overexpressed in HEK293T cells. Similarly, αvβ6 and ACE2 were shown to colocalise on the cell surface when co-transfected into HEK293T cells. The colocalisation of αvβ6 and ACE2 in the absence of spike protein is plausible given that ACE2 also contains an RGD motif (204-206: Arg-Gly-Asp) and has previously been shown capable of binding integrins (404, 409). When αvβ6 and ACE2 were co-transfected into cells and exogenous SARS-CoV-2 spike protein (WT) added, the highest level of spike protein binding corresponded to areas where αvβ6 and ACE2 were colocalised. This potentially suggests αvβ6 augmentation and is in agreement with previous data showing that αvβ6 enhances ACE2-mediated SARS-CoV-2

pseudovirus entry into HEK293T cells (403). Adding to this, αvβ6 has been shown to be upregulated in the alveolar epithelium following infection and RNAscope data shows colocalisation of the SARS-CoV-2 virus and αvβ6 in COVID-19 infected lung tissue (403).

Compared to the general population, IPF patients are more susceptible to SARS-CoV-2 infection and are at increased risk of developing severe COVID-19 (410). Expectedly αvβ6 protein expression is increased in IPF lung tissue, yet ACE2 is also upregulated in the distal lung compared to control samples (403). The regional increase in ACE2+ cells in the lower airways alongside an upregulation of αvβ6 in AT2 cells may potentially facilitate SARS-CoV-2 viral entry. Although, due to the RGD motif location it is most probable that integrins act as a SARS-CoV-2 co-receptor as opposed to a cell entry receptor independent of ACE2 (401). As previously described (Section 1.3.4) αvβ6-mediated TGF-β1 activation is a critical feed-forward loop which amplifies the fibrotic response in IPF. However, TGF-β1 signaling can suppress anti-viral IFN-I signaling activity by alveolar macrophages (95, 207, 216, 411). Consequently, TGF-β1 autoinduction and an upregulation of both ACE2 and αvβ6 may increase disease severity in IPF patients.

The major endothelial integrin αvβ3 is upregulated in the lung parenchyma and blood vessels of COVID-19 infected lung tissue (402). The importance of viral interactions with this integrin are still unclear but there is a wealth of evidence to support a key role for endothelial injury and vascular system dysfunction in COVID-19 pathogenesis. It has been hypothesised that the virus uses αvβ3 to promote the angiocentric

features of severe endothelial injury, widespread vascular thrombosis and significant new blood vessel growth observed in COVID-19 patients (412-415). Interestingly, SARS-CoV-2 binding to human endothelial cells is prevented by the αvβ3-antagonist cilengitide (an RGD-mimetic) in a concentration-dependent manner (413). Cilengitide also significantly reduces endothelial cell permeability and impaired barrier integrity associated with SARS-CoV-2 infection (413). Unpublished data from the Nottingham COVID-19 Research Group also shows colocalisation of the virus and αvβ3 in COVID-19 infected lung tissue by RNAscope *in situ* hybridisation assay (Joseph, C.) further supporting a potential role for this integrin in SARS-CoV-2 infection.

In contrast to the above, preliminary results showed that αvβ1, αvβ5, αvβ6 and αvβ8 bound to the S1 subunit in a cation-independent manner, although with a lower affinity than ACE2 (Section 6.4.3). However it should be emphasised that these preliminary experiments were performed in cation free buffer conditions in error and that the protocol was subsequently optimised with the help of Prof Martin Humphries and Dr Jonathan Humphries (The University of Manchester). Besides the RGD domain, several known integrin-binding motifs have recently been mapped to the SARS-CoV-2 RBD which could be involved in such interaction and may explain the cation-independent binding observed (416-418). Supporting this view, unpublished data from pseudovirus entry assays recently showed that ACE2 and integrin co-expression (specifically αvβ1, αvβ3 or αvβ6) enhances internalisation of the BA.2 Omicron variant despite the virus subtype lacking an RGD domain (D405N) (Monduzzi, F.). As integrin-binding motifs on the SARS-CoV-2 spike protein are more

widespread than initially thought, further work is required to establish the importance of binding independent of the RGD sequence.

Galectins have previously been suggested to regulate viral infection by directly interacting with viral glycoproteins (419). The plasma levels of both galectin-1 and galectin-3 are significantly higher in COVID-19 patients and given the potential role of galectin-3/-1 in fibrosis (Section 1.4.4.1 and Section 1.4.5.1), determining whether they directly bind to the SARS-CoV-2 virus was of particular interest (420). By solid-phase binding assay the binding of both galectin-1 and galectin-3 to the SARS-CoV-2 S1 was negligible over the BSA control. Despite this, galectin-3 has since been identified as a prognostic biomarker and therapeutic target in SARS-CoV-2 infection (421). The galectin-3 inhibitor GB0139 has been rapidly repurposed for a phase 1b/2a clinical trial in hospitalised patients with COVID-19 'DEFINE' (NCT04473053) in which the compound was hypothesised to reduce the severe effects of COVID-19 involving galectin-3 including post-viral fibrosis (155, 422). Additionally, it was proposed that the galectin inhibitors may possess dual-binding capabilities given that the SARS-CoV-2 S1 N-terminal domain has a highly similar structure to galectin-3 (423, 424). In the trial GB0139 (used in combination with standard of care) was well tolerated by patients hospitalised with COVID-19 pneumonitis and reduced plasma levels of biomarkers associated with inflammation, coagulopathy, major organ function and fibrosis (422). These preliminary findings are in support of a larger clinical trial of GB0139 in hospitalised patients with COVID-19.

Summary: To summarise, the SARS-CoV-2 S1 subunit bound to both αvβ6 and αvβ3 in a cation-dependent manner by solid-phase binding assay, although its affinity for integrins was lower than ACE2. In HEK293T cells the highest level of spike protein binding corresponded to areas where αvβ6 and ACE2 were colocalised. Taken together with the referenced literature, the data suggests that the virus may utilise RGD-binding integrins to facilitate entry into alveolar epithelial cells or endothelial cells and augment ACE2-dependent internalisation leading to the respiratory and vascular symptoms associated with COVID-19. Although the SARS-CoV-2 S1 subunit did not bind directly to galectin-3 or galectin-1, there is evidence to suggest that galectin-3 inhibition may reduce the severe effects of COVID-19 including post-viral fibrosis.

Limitations/ Future Work: Although these experiments were useful for investigating whether SARS-CoV-2 is capable of binding integrins, there are numerous limitations of the data. Firstly, the binding experiments were only performed by a single protein interaction method. As previously described (Section 4.5) it is important to consider the impact of immobilisation on protein interactions. Similar to immobilisation via amine-coupling in SPR, coating plates for solid-phase binding assay experiments is not orientated immobilisation and consequently the protein binding site may be masked or the protein inactivated. Ideally a combination of techniques would have been used to validate, characterise and confirm the protein interactions, however this was not possible at the start of the pandemic due to the short supply of recombinant proteins in high demand. More recently these data have been validated by an independent group showing SARS-CoV-2 binding to αvβ6 and αvβ3 by SPR

(405). For similar confidence in the negative binding data it would need to be reproducible by an alternative method, ideally an in-solution technique like ITC, although this would require a large quantity of protein. A second limitation to the binding data is the absence of ACE2 in these experiments. It would have been interesting to determine whether ACE2 had any effect on spike protein binding to integrins and vice versa, however it is important to note that this hypothesis was being addressed in parallel using a pseudovirus entry assay (403). Although the co-transfection data showed co-association of the spike protein with αvβ6 and ACE2, the proteins were overexpressed which limits its biological relevance. In order to fully establish the biological relevance of these findings, the experiments would need to be repeated in cells with native integrin expression and ideally performed using live virus.

7. Conclusions and Future Directions

7.1. Conclusions

IPF is a chronic and progressive ILD which falls under the major IIPs subgroup (2). It is the most common ILD subtype and the mortality rates are increasing globally (3-5). IPF has a poor prognosis with a median survival of 2.5 - 3.5 years from time of diagnosis (1, 6). It is defined by a pathological pattern of UIP and abnormal pulmonary function indicative of restrictive/impaired gas exchange (with no evidence supporting an alternative diagnosis) (1). Currently there is no cure for IPF, however pirfenidone and nintedanib are licensed for IPF management and slow disease progression (8-10). IPF has been proposed and accepted as a three-stage disease process: predisposition, activation and progression (41, 98). Integrin-mediated TGF-β1 activation plays a critical role in IPF pathogenesis and αvβ1 is suggested to be the principle integrin responsible for activation of latent TGF-β1 in fibroblasts (95, 223). Emerging evidence suggests that galectin-3 potentiates this TGF-β1 signaling to promote fibrogenesis, although the exact mechanism has not yet been defined.

Galectin-3 was initially identified as an essential mediator of TGF-β1-induced lung fibrosis in mouse models of fibrotic lung disease (284). Specifically, galectin-3 knockout mice were protected from the pro-fibrotic effects of bleomycin or TGF-β1 over expression (Ad-TGF-β1) evidenced by significantly reduced lung collagen content compared to WT mice. Following this study, *in vitro* experiments demonstrated that galectin-3 promotes TGF-β1 activation in primary lung fibroblasts (Parmar, N.). HLFs stimulated with galectin-3 had a significant increase in TGF-β1 SMAD signaling on par with cells stimulated with TGF-β1 (310, 311). The increase was

inhibited by the small molecule galectin inhibitor GB0139 which suggested that galectin-3 was directly involved in TGF-β1 signaling via its CRD. Furthermore, LPA-induced TGF-β1 SMAD signaling was blocked by GB0139 in HLFs and unpublished data similarly demonstrated concentration-dependent inhibition with the cell-impermeable galectin-3 inhibitor GB0149 (310, 313). This indicated that galectin-3, particularly extracellular galectin-3, was essential for integrin-mediated TGF-β1 activation in fibroblasts. Furthermore, GB0139 blocked TGF-β1-induced SMAD signaling in HLFs which additionally suggested that in HLFs galectin-3 may be involved in TGF-β1 signaling at the receptor level (310, 311).

Given the central role of galectin-3 in TGF-β1 activation and canonical signaling in fibroblasts, it was hypothesised that galectin-3 promotes fibrogenesis by directly interacting with key components of the TGF-β1 signaling cascade. Subsequently, the overall aim of this thesis was to investigate whether galectin-3 physically interacts with proteins responsible for integrin-mediated TGF-β1 activation and signaling (αv integrins, components of the large latent complex (LAP and LTBP1) and the TGF-β1 receptor). This work has encompassed cell, molecular and biophysical methods to define the mechanism of galectin-3-mediated TGF-β1 activation and to better understand the role of galectin-3 in development of IPF.

Using the biophysical technique SPR, both galectin-3 and galectin-1 were shown to bind to αv integrins as well as the TGFβRII subunit, key components of the TGF-β1 signaling cascade. Binding responses for both galectin-3 and galectin-1 were higher for αvβ1 and αvβ5 compared to αvβ6. Binding responses to all proteins were

however higher with galectin-3 than galectin-1, although the level of response suggested that galectin binding was heterogenous and not a 1:1 binding stoichiometry. This is consistent with the literature that has shown multivalency of carbohydrate-binding activity (260, 269-272). The protein interactions detected by SPR were glycan-dependent as enzymatic removal of cell surface glycans (on αv integrins or the TGFβRII subunit) blocked galectin binding. These findings are in support of the current literature which states that extracellular galectin interactions are carbohydrate-dependent (247, 249, 250). Additionally, small molecule galectin inhibitors (GB1107, GB1490 and GB0139) which target the galectin CRD inhibited galectin binding in a concentration-dependent manner. A difference in compound affinity was demonstrated by IC50 values obtained from the solution competition binding assay.

Current evidence supports that galectin-3 binding to the αvβ1 integrin on fibroblasts may be important for fibrogenesis (310, 311). Therefore, subsequent experiments focused on understanding the interaction of galectin-3 and the αvβ1 integrin in primary lung fibroblasts. By Co-IP galectin-3 bound to the β1 integrin in untreated non-IPF HLFs, this validated the SPR data in a cell-based system. PLA results showed colocalisation of galectin-3 and the β1 integrin in IPF HLFs which was augmented by TGF-β1 stimulation. In comparison, minimal colocalisation of galectin-3 and the β1 integrin was detected in non-IPF HLFs. The galectin inhibitor GB0139 inhibited colocalisation of galectin-3 and the β1 integrin in IPF HLFs at clinically meaningful concentrations (NCT02257177) (291). The anti-fibrotic activity of GB0139 in lung

fibrosis may therefore be partly due to targeting galectin-3 in TGF-β1 signaling in lung fibroblasts.

Despite some limitations (chapter specific), the data presented in chapters 3 - 5 supports the hypothesis that galectin-3 physically interacts with proteins responsible for integrin-mediated TGF-β1 activation and signaling. As the integrin, the latent TGF-β1 complex and TGF-β1 receptor are required to be in close physical association for integrin-mediated TGF-β1 activation this thesis proposes that galectin-3 promotes integrin-mediated TGF-β1 activation in lung fibroblasts by facilitating clustering of the αvβ1 integrin and TGF-β1 receptor on their respective cell surfaces (Figure 7.1).

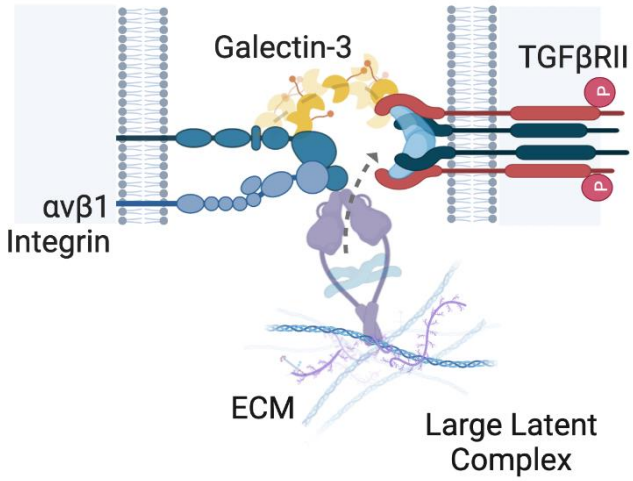


Figure 7.1: Visual summary
Galectin-3 binds to glycosylation sites on the αvβ1 integrin and the TGF-β1 receptor forming a galectin lattice at the cell surface to facilitate receptor clustering. This ensures that TGF-β1 can act on its receptor and potentiate TGF-β1 signaling. GB0139 binds to the galectin-3 carbohydrate recognition domain and blocks these protein-glycan interactions.

Further knowledge and understanding of the proteins involved in this signaling pathway may help to indirectly target TGF-β1 for IPF patient benefit. These findings may also be translatable to a subset of COVID-19 patients that have substantial fibrotic consequences following SARS-CoV-2 infection.

There is increasing evidence to suggest that TGF-β1 plays an important role in progressive lung fibrosis following SARS-CoV-2 infection (425). TGF-β1 is known to have two positive-feedback loops that amplify the fibrotic response 1) TGF-β1 autoinduction and 2) increased integrin expression (186, 187, 207). Given that SARS-CoV-2 has an RGD motif which may be recognised by RGD-binding integrins, TGF-β1-induced integrin expression in the distal lung may be of importance for virus infectivity. In this thesis a solid-phase binding assay was used to determine whether RGD integrins directly bind to the S1 subunit of the SARS-CoV-2 spike protein. Both αvβ6 and αvβ3 bound to the S1 subunit in a cation-dependent manner but with lower affinity than ACE2. Furthermore, In HEK293T cells the highest level of spike protein binding corresponded to areas where αvβ6 and ACE2 were colocalised. Pseudovirus assays performed in collaboration with Prof Wendy Barclay and Dr Tom Peacock (Imperial College London) also showed that the presence of αvβ6 on the cell surface enhances ACE2-mediated SARS-CoV-2 entry into HEK293T cells (403). Interestingly in these studies, expression of αvβ6 alone was not able to facilitate viral infection. The data presented in chapter 6 therefore supports the hypothesis that RGD-binding integrins are a co-receptor involved in SARS-CoV-2 infection and highlights the potential role for antifibrotic therapy in a subset of COVID-19 patients.

7.2. Future Directions

7.2.1. Fibroblast heterogeneity

Fibroblasts are heterogenous cells and recent advances in scRNAseq technology has facilitated the identification of lung fibroblast subtypes (426). Fibroblast classification has been explored in respect of localisation, molecular makers and function although there is currently no consensus on the subpopulations of cells (427). Fibroblasts were

initially classified according to their gene expression profiles which generally includes myofibroblasts, lipofibroblasts, mesenchymal progenitors and mesothelial cells (428, 429). However, they have since been reclassified based on their anatomical location in the peribronchial, adventitial and alveolar regions (430). Interestingly, there are also reports of a newly emerging fibroblast population in pulmonary fibrosis pathogenesis (428, 430, 431). This thesis has focused on defining the mechanism of galectin-3-mediated TGF-β1 activation and signaling in lung fibroblasts but in which subpopulation is unclear. This poses a challenge at present given that there are no definitive markers or transcriptomic signatures for each subtype and the impact of culture on these subpopulations is unknown. Future work would therefore explore in which fibroblast subtype galectin-3 binding is functionally relevant. This would provide further insight into the *in vitro* findings presented in this thesis and may help to determine which fibroblast subpopulation should be targeted therapeutically in pulmonary fibrosis.

7.2.2. Glycosylation in Disease

Glycobiology is an emerging field of research and there is increasing evidence that aberrant glycosylation is a feature of several diseases. Altered glycosylation has been most researched in congenital disorders of glycosylation (CDGs), cancer, autoimmunity and chronic inflammation (432). Although it is not currently known if glycosylation is atypical in pulmonary fibrosis. This is of particular interest given that galectins bind to β-galactoside carbohydrates and extracellular galectin interactions are commonly carbohydrate-dependent (231, 247, 249, 250). Future work would therefore investigate the glycosylation profile of fibrotic lung tissue from IPF patients compared to controls. Common methods used to assess glycosylation include

glycoprotein staining, lectin blotting and glycogene expression profiling (433). Mass spectrometry-based techniques can also be used for a detailed analysis of glycan structure (433). These experiments would determine whether glycosylation is elevated in IPF and if this subsequently impacts galectin activity and TGF-β1 signaling.

7.3. Concluding Remarks

Data presented in this thesis suggests that galectin-3 promotes TGF-β1 signaling in fibroblasts by directly interacting with components of the signaling pathway. Galectin-3 binds to the αvβ1 integrin and TGFβRII subunit and these interactions can be blocked by galectin-3 small molecule inhibitors. Furthering our knowledge and understanding of galectin-3-mediated TGF-β1 activation in IPF may highlight the importance of combination therapy in targeting multiple pro-fibrotic pathways involved in different stages of disease and/or on different cells (epithelial cells and fibroblasts). It will also aid the galectin drug discovery program and may help to successfully target this mechanism for IPF patient benefit.

8. References

1. American Thoracic Society; European Respiratory Society. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. Am J Respir Crit Care Med. 2002;165(2):277-304.
2. Travis WD, Costabel U, Hansell DM*, et al.* An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med. 2013;188(6):733-48.
3. Snell N, Strachan D, Hubbard R*, et al.* P272 Epidemiology of idiopathic pulmonary fibrosis in the uk: findings from the british lung foundation's 'respiratory health of the nation' project. Thorax. 2016;71(Suppl 3):A236.
4. Hutchinson JP, McKeever TM, Fogarty AW*, et al.* Increasing global mortality from idiopathic pulmonary fibrosis in the twenty-first century. Ann Am Thorac Soc. 2014;11(8):1176-85.
5. Navaratnam V, Hubbard RB. The Mortality Burden of Idiopathic Pulmonary Fibrosis in the United Kingdom. Am J Respir Crit Care Med. 2019;200(2):256-8.
6. Raghu G, Collard HR, Egan JJ*, et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788-824.
7. American Thoracic Society. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). Am J Respir Crit Care Med. 2000;161(2 Pt 1):646-64.
8. Richeldi L, du Bois RM, Raghu G*, et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2071-82.
9. King TE, Jr., Bradford WZ, Castro-Bernardini S*, et al.* A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2083-92.
10. Raghu G, Rochwerg B, Zhang Y*, et al.* An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. Am J Respir Crit Care Med. 2015;192(2):e3-19.
11. National Institute for Health and Care Excellence. Nintedanib for treating idiopathic pulmonary fibrosis. Technology appraisal guidance [TA379]. 2016.
12. National Institute for Health and Care Excellence. Pirfenidone for treating idiopathic pulmonary fibrosis. Technology appraisal guidance [TA504]. 2018.
13. National Institute for Health and Care Excellence. Nintedanib for treating idiopathic pulmonary fibrosis when forced vital capacity is above 80% predicted. Technology appraisal guidance [TA864]. 2023.
14. Iyer SN, Wild JS, Schiedt MJ*, et al.* Dietary intake of pirfenidone ameliorates bleomycin-induced lung fibrosis in hamsters. J Lab Clin Med. 1995;125(6):779-85.
15. Iyer SN, Gurujeyalakshmi G, Giri SN. Effects of pirfenidone on procollagen gene expression at the transcriptional level in bleomycin hamster model of lung fibrosis. J Pharmacol Exp Ther. 1999;289(1):211-8.
16. Iyer SN, Hyde DM, Giri SN. Anti-inflammatory effect of pirfenidone in the bleomycin-hamster model of lung inflammation. Inflammation. 2000;24(5):477-91.

17. Iyer SN, Gurujeyalakshmi G, Giri SN. Effects of pirfenidone on transforming growth factor-beta gene expression at the transcriptional level in bleomycin hamster model of lung fibrosis. J Pharmacol Exp Ther. 1999;291(1):367-73.
18. Conte E, Gili E, Fagone E*, et al.* Effect of pirfenidone on proliferation, TGF-β-induced myofibroblast differentiation and fibrogenic activity of primary human lung fibroblasts. Eur J Pharm Sci. 2014;58:13-9.
19. Cain WC, Stuart RW, Lefkowitz DL*, et al.* Inhibition of tumor necrosis factor and subsequent endotoxin shock by pirfenidone. Int J Immunopharmacol. 1998;20(12):685-95.
20. Oku H, Nakazato H, Horikawa T*, et al.* Pirfenidone suppresses tumor necrosis factor-alpha, enhances interleukin-10 and protects mice from endotoxic shock. Eur J Pharmacol. 2002;446(1-3):167-76.
21. Gurujeyalakshmi G, Hollinger MA, Giri SN. Pirfenidone inhibits PDGF isoforms in bleomycin hamster model of lung fibrosis at the translational level. Am J Physiol. 1999;276(2):L311-8.
22. Oku H, Shimizu T, Kawabata T*, et al.* Antifibrotic action of pirfenidone and prednisolone: different effects on pulmonary cytokines and growth factors in bleomycin-induced murine pulmonary fibrosis. Eur J Pharmacol. 2008;590(1-3):400-8.
23. Hilberg F, Roth GJ, Krssak M*, et al.* BIBF 1120: triple angiokinase inhibitor with sustained receptor blockade and good antitumor efficacy. Cancer Res. 2008;68(12):4774-82.
24. Wollin L, Maillet I, Quesniaux V*, et al.* Antifibrotic and anti-inflammatory activity of the tyrosine kinase inhibitor nintedanib in experimental models of lung fibrosis. J Pharmacol Exp Ther. 2014;349(2):209-20.
25. Hostettler KE, Zhong J, Papakonstantinou E*, et al.* Anti-fibrotic effects of nintedanib in lung fibroblasts derived from patients with idiopathic pulmonary fibrosis. Respir Res. 2014;15(1):157.
26. Wollin L, Wex E, Pautsch A*, et al.* Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. Eur Respir J. 2015;45(5):1434-45.
27. Ward HE, Nicholas TE. Alveolar type I and type II cells. Aust N Z J Med. 1984;14(5 Suppl 3):731-4.
28. Minutti CM, Knipper JA, Allen JE*, et al.* Tissue-specific contribution of macrophages to wound healing. Semin Cell Dev Biol. 2017;61:3-11.
29. Jorgensen L, Borchgrevink CF. THE HAEMOSTATIC MECHANISM IN PATIENTS WITH HAEMORRHAGIC DISEASES. A HISTOLOGICAL STUDY OF WOUNDS MADE FOR PRIMARY AND SECONDARY BLEEDING TIME TESTS. Acta Pathol Microbiol Scand. 1964;60:55-82.
30. Ross R. The fibroblast and wound repair. Biol Rev Camb Philos Soc. 1968;43(1):51-96.
31. Montesano R, Orci L. Transforming growth factor beta stimulates collagen-matrix contraction by fibroblasts: implications for wound healing. Proc Natl Acad Sci U S A. 1988;85(13):4894-7.
32. Clark RA, Folkvord JM, Hart CE*, et al.* Platelet isoforms of platelet-derived growth factor stimulate fibroblasts to contract collagen matrices. J Clin Invest. 1989;84(3):1036-40.

33. Finesmith TH, Broadley KN, Davidson JM. Fibroblasts from wounds of different stages of repair vary in their ability to contract a collagen gel in response to growth factors. J Cell Physiol. 1990;144(1):99-107.
34. Darby I, Skalli O, Gabbiani G. Alpha-smooth muscle actin is transiently expressed by myofibroblasts during experimental wound healing. Lab Invest. 1990;63(1):21-9.
35. Desmoulière A, Redard M, Darby I*, et al.* Apoptosis mediates the decrease in cellularity during the transition between granulation tissue and scar. Am J Pathol. 1995;146(1):56-66.
36. Desmoulière A. Factors influencing myofibroblast differentiation during wound healing and fibrosis. Cell Biol Int. 1995;19(5):471-6.
37. Haschek WM, Witschi H. Pulmonary fibrosis--a possible mechanism. Toxicol Appl Pharmacol. 1979;51(3):475-87.
38. Hagimoto N, Kuwano K, Nomoto Y*, et al.* Apoptosis and expression of Fas/Fas ligand mRNA in bleomycin-induced pulmonary fibrosis in mice. Am J Respir Cell Mol Biol. 1997;16(1):91-101.
39. Hagimoto N, Kuwano K, Miyazaki H*, et al.* Induction of apoptosis and pulmonary fibrosis in mice in response to ligation of Fas antigen. Am J Respir Cell Mol Biol. 1997;17(3):272-8.
40. Minagawa S, Araya J, Numata T*, et al.* Accelerated epithelial cell senescence in IPF and the inhibitory role of SIRT6 in TGF-β-induced senescence of human bronchial epithelial cells. Am J Physiol Lung Cell Mol Physiol. 2011;300(3):L391-401.
41. Wolters PJ, Blackwell TS, Eickelberg O*, et al.* Time for a change: is idiopathic pulmonary fibrosis still idiopathic and only fibrotic? Lancet Respir Med. 2018;6(2):154-60.
42. Myers JL, Katzenstein AL. Epithelial necrosis and alveolar collapse in the pathogenesis of usual interstitial pneumonia. Chest. 1988;94(6):1309-11.
43. Vaccaro CA, Brody JS, Snider GL. Alveolar wall basement membranes in bleomycin-induced pulmonary fibrosis. Am Rev Respir Dis. 1985;132(4):905-12.
44. Kuhn C, 3rd, Boldt J, King TE, Jr.*, et al.* An immunohistochemical study of architectural remodeling and connective tissue synthesis in pulmonary fibrosis. Am Rev Respir Dis. 1989;140(6):1693-703.
45. Kuhn C, McDonald JA. The roles of the myofibroblast in idiopathic pulmonary fibrosis. Ultrastructural and immunohistochemical features of sites of active extracellular matrix synthesis. Am J Pathol. 1991;138(5):1257-65.
46. Jockusch BM, Bubeck P, Giehl K*, et al.* The molecular architecture of focal adhesions. Annu Rev Cell Dev Biol. 1995;11:379-416.
47. Zamir E, Katz BZ, Aota S*, et al.* Molecular diversity of cell-matrix adhesions. J Cell Sci. 1999;112 (Pt 11):1655-69.
48. Cukierman E, Pankov R, Stevens DR*, et al.* Taking cell-matrix adhesions to the third dimension. Science. 2001;294(5547):1708-12.
49. White ES, Thannickal VJ, Carskadon SL*, et al.* Integrin alpha4beta1 regulates migration across basement membranes by lung fibroblasts: a role for phosphatase and tensin homologue deleted on chromosome 10. Am J Respir Crit Care Med. 2003;168(4):436-42.
50. Moodley YP, Caterina P, Scaffidi AK*, et al.* Comparison of the morphological and biochemical changes in normal human lung fibroblasts and fibroblasts derived from

lungs of patients with idiopathic pulmonary fibrosis during FasL-induced apoptosis. J Pathol. 2004;202(4):486-95.
51. Selman M, Pardo A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. Am J Respir Crit Care Med. 2014;189(10):1161-72.
52. Antoniades HN, Bravo MA, Avila RE*, et al.* Platelet-derived growth factor in idiopathic pulmonary fibrosis. J Clin Invest. 1990;86(4):1055-64.
53. Piguet PF, Ribaux C, Karpuz V*, et al.* Expression and localization of tumor necrosis factor-alpha and its mRNA in idiopathic pulmonary fibrosis. Am J Pathol. 1993;143(3):651-5.
54. Khalil N, O'Connor RN, Unruh HW*, et al.* Increased production and immunohistochemical localization of transforming growth factor-beta in idiopathic pulmonary fibrosis. Am J Respir Cell Mol Biol. 1991;5(2):155-62.
55. Pan LH, Yamauchi K, Uzuki M*, et al.* Type II alveolar epithelial cells and interstitial fibroblasts express connective tissue growth factor in IPF. Eur Respir J. 2001;17(6):1220-7.
56. Giaid A, Michel RP, Stewart DJ*, et al.* Expression of endothelin-1 in lungs of patients with cryptogenic fibrosing alveolitis. Lancet. 1993;341(8860):1550-4.
57. Pardo A, Gibson K, Cisneros J*, et al.* Up-regulation and profibrotic role of osteopontin in human idiopathic pulmonary fibrosis. PLoS Med. 2005;2(9):e251.
58. Abdul-Hafez A, Shu R, Uhal BD. JunD and HIF-1alpha mediate transcriptional activation of angiotensinogen by TGF-beta1 in human lung fibroblasts. Faseb j. 2009;23(6):1655-62.
59. Hinz B, Phan SH, Thannickal VJ*, et al.* The myofibroblast: one function, multiple origins. Am J Pathol. 2007;170(6):1807-16.
60. Abe R, Donnelly SC, Peng T*, et al.* Peripheral blood fibrocytes: differentiation pathway and migration to wound sites. J Immunol. 2001;166(12):7556-62.
61. Phillips RJ, Burdick MD, Hong K*, et al.* Circulating fibrocytes traffic to the lungs in response to CXCL12 and mediate fibrosis. J Clin Invest. 2004;114(3):438-46.
62. Willis BC, Liebler JM, Luby-Phelps K*, et al.* Induction of epithelial-mesenchymal transition in alveolar epithelial cells by transforming growth factor-beta1: potential role in idiopathic pulmonary fibrosis. Am J Pathol. 2005;166(5):1321-32.
63. Kim KK, Kugler MC, Wolters PJ*, et al.* Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. Proc Natl Acad Sci U S A. 2006;103(35):13180-5.
64. Desmoulière A, Rubbia-Brandt L, Abdiu A*, et al.* Alpha-smooth muscle actin is expressed in a subpopulation of cultured and cloned fibroblasts and is modulated by gamma-interferon. Exp Cell Res. 1992;201(1):64-73.
65. Tomasek JJ, Gabbiani G, Hinz B*, et al.* Myofibroblasts and mechano-regulation of connective tissue remodelling. Nat Rev Mol Cell Biol. 2002;3(5):349-63.
66. Kolodney MS, Wysolmerski RB. Isometric contraction by fibroblasts and endothelial cells in tissue culture: a quantitative study. J Cell Biol. 1992;117(1):73-82.
67. Gabbiani G, Ryan GB, Majne G. Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction. Experientia. 1971;27(5):549-50.
68. Majno G, Gabbiani G, Hirschel BJ*, et al.* Contraction of granulation tissue in vitro: similarity to smooth muscle. Science. 1971;173(3996):548-50.

69. Hinz B, Celetta G, Tomasek JJ*, et al.* Alpha-smooth muscle actin expression upregulates fibroblast contractile activity. Mol Biol Cell. 2001;12(9):2730-41.
70. Arora PD, McCulloch CA. Dependence of collagen remodelling on alpha-smooth muscle actin expression by fibroblasts. J Cell Physiol. 1994;159(1):161-75.
71. Desmoulière A, Geinoz A, Gabbiani F*, et al.* Transforming growth factor-beta 1 induces alpha-smooth muscle actin expression in granulation tissue myofibroblasts and in quiescent and growing cultured fibroblasts. J Cell Biol. 1993;122(1):103-11.
72. Scotton CJ, Chambers RC. Molecular targets in pulmonary fibrosis: the myofibroblast in focus. Chest. 2007;132(4):1311-21.
73. Uhal BD, Joshi I, True AL*, et al.* Fibroblasts isolated after fibrotic lung injury induce apoptosis of alveolar epithelial cells in vitro. Am J Physiol. 1995;269(6 Pt 1):L819-28.
74. Uhal BD, Joshi I, Hughes WF*, et al.* Alveolar epithelial cell death adjacent to underlying myofibroblasts in advanced fibrotic human lung. Am J Physiol. 1998;275(6):L1192-9.
75. White ES. Lung extracellular matrix and fibroblast function. Ann Am Thorac Soc. 2015;12 Suppl 1(Suppl 1):S30-3.
76. Timpl R, Brown JC. Supramolecular assembly of basement membranes. Bioessays. 1996;18(2):123-32.
77. Erickson AC, Couchman JR. Still more complexity in mammalian basement membranes. J Histochem Cytochem. 2000;48(10):1291-306.
78. Chung E, Miller EJ. Collagen polymorphism: characterization of molecules with the chain composition (alpha 1 (3)03 in human tissues. Science. 1974;183(130):1200-1.
79. Crystal RG, Fulmer JD, Baum BJ*, et al.* Cells, collagen and idiopathic pulmonary fibrosis. Lung. 1978;155(3):199-224.
80. Kadler KE, Holmes DF, Trotter JA*, et al.* Collagen fibril formation. Biochem J. 1996;316 (Pt 1)(Pt 1):1-11.
81. Nagase H, Woessner JF, Jr. Matrix metalloproteinases. J Biol Chem. 1999;274(31):21491-4.
82. Li Y, Jiang D, Liang J*, et al.* Severe lung fibrosis requires an invasive fibroblast phenotype regulated by hyaluronan and CD44. J Exp Med. 2011;208(7):1459-71.
83. Seyer JM, Hutcheson ET, Kang AH. Collagen polymorphism in idiopathic chronic pulmonary fibrosis. J Clin Invest. 1976;57(6):1498-507.
84. Raghu G, Striker LJ, Hudson LD*, et al.* Extracellular matrix in normal and fibrotic human lungs. Am Rev Respir Dis. 1985;131(2):281-9.
85. Booth AJ, Hadley R, Cornett AM*, et al.* Acellular normal and fibrotic human lung matrices as a culture system for in vitro investigation. Am J Respir Crit Care Med. 2012;186(9):866-76.
86. Chuliá-Peris L, Carreres-Rey C, Gabasa M*, et al.* Matrix Metalloproteinases and Their Inhibitors in Pulmonary Fibrosis: EMMPRIN/CD147 Comes into Play. Int J Mol Sci. 2022;23(13).
87. Selman M, Ruiz V, Cabrera S*, et al.* TIMP-1, -2, -3, and -4 in idiopathic pulmonary fibrosis. A prevailing nondegradative lung microenvironment? Am J Physiol Lung Cell Mol Physiol. 2000;279(3):L562-74.
88. Swiderski RE, Dencoff JE, Floerchinger CS*, et al.* Differential expression of extracellular matrix remodeling genes in a murine model of bleomycin-induced pulmonary fibrosis. Am J Pathol. 1998;152(3):821-8.

89. Madtes DK, Elston AL, Kaback LA*, et al.* Selective induction of tissue inhibitor of metalloproteinase-1 in bleomycin-induced pulmonary fibrosis. Am J Respir Cell Mol Biol. 2001;24(5):599-607.
90. Manoury B, Caulet-Maugendre S, Guénon I*, et al.* TIMP-1 is a key factor of fibrogenic response to bleomycin in mouse lung. Int J Immunopathol Pharmacol. 2006;19(3):471-87.
91. Selman M, Montaño M, Ramos C*, et al.* Concentration, biosynthesis and degradation of collagen in idiopathic pulmonary fibrosis. Thorax. 1986;41(5):355-9.
92. Philp CJ, Siebeke I, Clements D*, et al.* Extracellular Matrix Cross-Linking Enhances Fibroblast Growth and Protects against Matrix Proteolysis in Lung Fibrosis. Am J Respir Cell Mol Biol. 2018;58(5):594-603.
93. Barry-Hamilton V, Spangler R, Marshall D*, et al.* Allosteric inhibition of lysyl oxidase-like-2 impedes the development of a pathologic microenvironment. Nat Med. 2010;16(9):1009-17.
94. Olsen KC, Sapinoro RE, Kottmann RM*, et al.* Transglutaminase 2 and its role in pulmonary fibrosis. Am J Respir Crit Care Med. 2011;184(6):699-707.
95. Jenkins RG, Su X, Su G*, et al.* Ligation of protease-activated receptor 1 enhances alpha(v)beta6 integrin-dependent TGF-beta activation and promotes acute lung injury. J Clin Invest. 2006;116(6):1606-14.
96. Wipff PJ, Rifkin DB, Meister JJ*, et al.* Myofibroblast contraction activates latent TGF-beta1 from the extracellular matrix. J Cell Biol. 2007;179(6):1311-23.
97. Liu F, Mih JD, Shea BS*, et al.* Feedback amplification of fibrosis through matrix stiffening and COX-2 suppression. J Cell Biol. 2010;190(4):693-706.
98. Wolters PJ, Collard HR, Jones KD. Pathogenesis of idiopathic pulmonary fibrosis. Annu Rev Pathol. 2014;9:157-79.
99. Korfei M, Ruppert C, Mahavadi P*, et al.* Epithelial endoplasmic reticulum stress and apoptosis in sporadic idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2008;178(8):838-46.
100. Selman M, King TE, Pardo A. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. Ann Intern Med. 2001;134(2):136-51.
101. Mammoto A, Mammoto T, Kanapathipillai M*, et al.* Control of lung vascular permeability and endotoxin-induced pulmonary oedema by changes in extracellular matrix mechanics. Nat Commun. 2013;4:1759.
102. Steele MP, Speer MC, Loyd JE*, et al.* Clinical and pathologic features of familial interstitial pneumonia. Am J Respir Crit Care Med. 2005;172(9):1146-52.
103. López-Otín C, Blasco MA, Partridge L*, et al.* The hallmarks of aging. Cell. 2013;153(6):1194-217.
104. Raghu G, Weycker D, Edelsberg J*, et al.* Incidence and prevalence of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2006;174(7):810-6.
105. Gardner M, Bann D, Wiley L*, et al.* Gender and telomere length: systematic review and meta-analysis. Exp Gerontol. 2014;51:15-27.
106. Stuart BD, Lee JS, Kozlitina J*, et al.* Effect of telomere length on survival in patients with idiopathic pulmonary fibrosis: an observational cohort study with independent validation. Lancet Respir Med. 2014;2(7):557-65.

107. Duckworth A, Gibbons MA, Allen RJ*, et al.* Telomere length and risk of idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease: a mendelian randomisation study. Lancet Respir Med. 2021;9(3):285-94.
108. Anna D, Katherine SR, Julia KP*, et al.* Study of the associations between short telomeres, sex hormones and pulmonary fibrosis. medRxiv. 2022:2022.09.29.22280270.
109. Allen RJ, Stockwell A, Oldham JM*, et al.* Genome-wide association study across five cohorts identifies five novel loci associated with idiopathic pulmonary fibrosis. Thorax. 2022;77(8):829-33.
110. Partanen JJ, Häppölä P, Zhou W*, et al.* Leveraging global multi-ancestry meta-analysis in the study of idiopathic pulmonary fibrosis genetics. Cell Genom. 2022;2(10):100181.
111. Oldham JM, Allen RJ, Lorenzo-Salazar JM*, et al.* PCSK6 and Survival in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2023.
112. Allen RJ, Oldham JM, Jenkins DA*, et al.* Longitudinal lung function and gas transfer in individuals with idiopathic pulmonary fibrosis: a genome-wide association study. Lancet Respir Med. 2023;11(1):65-73.
113. Fingerlin TE, Murphy E, Zhang W*, et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet. 2013;45(6):613-20.
114. Allen RJ, Guillen-Guio B, Oldham JM*, et al.* Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2020;201(5):564-74.
115. Mushiroda T, Wattanapokayakit S, Takahashi A*, et al.* A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. J Med Genet. 2008;45(10):654-6.
116. Dhindsa RS, Mattsson J, Nag A*, et al.* Identification of a missense variant in SPDL1 associated with idiopathic pulmonary fibrosis. Commun Biol. 2021;4(1):392.
117. Seibold MA, Wise AL, Speer MC*, et al.* A common MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med. 2011;364(16):1503-12.
118. McCarthy MI, Abecasis GR, Cardon LR*, et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356-69.
119. Kropski JA, Blackwell TS, Loyd JE. The genetic basis of idiopathic pulmonary fibrosis. Eur Respir J. 2015;45(6):1717-27.
120. Evans CM, Fingerlin TE, Schwarz MI*, et al.* Idiopathic Pulmonary Fibrosis: A Genetic Disease That Involves Mucociliary Dysfunction of the Peripheral Airways. Physiol Rev. 2016;96(4):1567-91.
121. Cherniack RM, Colby TV, Flint A*, et al.* Correlation of structure and function in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 1995;151(4):1180-8.
122. Baumgartner KB, Samet JM, Stidley CA*, et al.* Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 1997;155(1):242-8.
123. King TE, Jr., Tooze JA, Schwarz MI*, et al.* Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. Am J Respir Crit Care Med. 2001;164(7):1171-81.
124. King TE, Jr., Schwarz MI, Brown K*, et al.* Idiopathic pulmonary fibrosis: relationship between histopathologic features and mortality. Am J Respir Crit Care Med. 2001;164(6):1025-32.

125. Antoniou KM, Hansell DM, Rubens MB*, et al.* Idiopathic pulmonary fibrosis: outcome in relation to smoking status. Am J Respir Crit Care Med. 2008;177(2):190-4.
126. Cisneros-Lira J, Gaxiola M, Ramos C*, et al.* Cigarette smoke exposure potentiates bleomycin-induced lung fibrosis in guinea pigs. Am J Physiol Lung Cell Mol Physiol. 2003;285(4):L949-56.
127. Scott J, Johnston I, Britton J. What causes cryptogenic fibrosing alveolitis? A case-control study of environmental exposure to dust. Bmj. 1990;301(6759):1015-7.
128. Hubbard R, Johnston I, Coultas DB*, et al.* Mortality rates from cryptogenic fibrosing alveolitis in seven countries. Thorax. 1996;51(7):711-6.
129. Mullen J, Hodgson MJ, DeGraff CA*, et al.* Case-control study of idiopathic pulmonary fibrosis and environmental exposures. J Occup Environ Med. 1998;40(4):363-7.
130. Baumgartner KB, Samet JM, Coultas DB*, et al.* Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case-control study. Collaborating Centers. Am J Epidemiol. 2000;152(4):307-15.
131. Hubbard R, Cooper M, Antoniak M*, et al.* Risk of cryptogenic fibrosing alveolitis in metal workers. Lancet. 2000;355(9202):466-7.
132. Miyake Y, Sasaki S, Yokoyama T*, et al.* Occupational and environmental factors and idiopathic pulmonary fibrosis in Japan. Ann Occup Hyg. 2005;49(3):259-65.
133. Gustafson T, Dahlman-Höglund A, Nilsson K*, et al.* Occupational exposure and severe pulmonary fibrosis. Respir Med. 2007;101(10):2207-12.
134. García-Sancho C, Buendía-Roldán I, Fernández-Plata MR*, et al.* Familial pulmonary fibrosis is the strongest risk factor for idiopathic pulmonary fibrosis. Respir Med. 2011;105(12):1902-7.
135. Awadalla NJ, Hegazy A, Elmetwally RA*, et al.* Occupational and environmental risk factors for idiopathic pulmonary fibrosis in Egypt: a multicenter case-control study. Int J Occup Environ Med. 2012;3(3):107-16.
136. Koo JW, Myong JP, Yoon HK*, et al.* Occupational exposure and idiopathic pulmonary fibrosis: a multicentre case-control study in Korea. Int J Tuberc Lung Dis. 2017;21(1):107-12.
137. Paolocci G, Folletti I, Torén K*, et al.* Occupational risk factors for idiopathic pulmonary fibrosis in Southern Europe: a case-control study. BMC Pulm Med. 2018;18(1):75.
138. Blanc PD, Annesi-Maesano I, Balmes JR*, et al.* The Occupational Burden of Nonmalignant Respiratory Diseases. An Official American Thoracic Society and European Respiratory Society Statement. Am J Respir Crit Care Med. 2019;199(11):1312-34.
139. Raghu G, Anstrom KJ, King TE, Jr.*, et al.* Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. N Engl J Med. 2012;366(21):1968-77.
140. Han MK, Zhou Y, Murray S*, et al.* Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. Lancet Respir Med. 2014;2(7):548-56.
141. Molyneaux PL, Cox MJ, Willis-Owen SA*, et al.* The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2014;190(8):906-13.

142. Invernizzi R, Barnett J, Rawal B*, et al.* Bacterial burden in the lower airways predicts disease progression in idiopathic pulmonary fibrosis and is independent of radiological disease extent. Eur Respir J. 2020;55(4).
143. Molyneaux PL, Cox MJ, Wells AU*, et al.* Changes in the respiratory microbiome during acute exacerbations of idiopathic pulmonary fibrosis. Respir Res. 2017;18(1):29.
144. Martinez FJ, Yow E, Flaherty KR*, et al.* Effect of Antimicrobial Therapy on Respiratory Hospitalization or Death in Adults With Idiopathic Pulmonary Fibrosis: The CleanUP-IPF Randomized Clinical Trial. Jama. 2021;325(18):1841-51.
145. Shulgina L, Cahn AP, Chilvers ER*, et al.* Treating idiopathic pulmonary fibrosis with the addition of co-trimoxazole: a randomised controlled trial. Thorax. 2013;68(2):155-62.
146. Sheng G, Chen P, Wei Y*, et al.* Viral Infection Increases the Risk of Idiopathic Pulmonary Fibrosis: A Meta-Analysis. Chest. 2020;157(5):1175-87.
147. Duckworth A, Longhurst HJ, Paxton JK*, et al.* The Role of Herpes Viruses in Pulmonary Fibrosis. Front Med (Lausanne). 2021;8:704222.
148. Lawson WE, Crossno PF, Polosukhin VV*, et al.* Endoplasmic reticulum stress in alveolar epithelial cells is prominent in IPF: association with altered surfactant protein processing and herpesvirus infection. Am J Physiol Lung Cell Mol Physiol. 2008;294(6):L1119-26.
149. Calabrese F, Kipar A, Lunardi F*, et al.* Herpes virus infection is associated with vascular remodeling and pulmonary hypertension in idiopathic pulmonary fibrosis. PLoS One. 2013;8(2):e55715.
150. Dowd JB, Bosch JA, Steptoe A*, et al.* Persistent Herpesvirus Infections and Telomere Attrition Over 3 Years in the Whitehall II Cohort. J Infect Dis. 2017;216(5):565-72.
151. Wong KT, Antonio GE, Hui DS*, et al.* Thin-section CT of severe acute respiratory syndrome: evaluation of 73 patients exposed to or with the disease. Radiology. 2003;228(2):395-400.
152. Das KM, Lee EY, Singh R*, et al.* Follow-up chest radiographic findings in patients with MERS-CoV after recovery. Indian J Radiol Imaging. 2017;27(3):342-9.
153. Wang D, Hu B, Hu C*, et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. Jama. 2020;323(11):1061-9.
154. Mo X, Jian W, Su Z*, et al.* Abnormal pulmonary function in COVID-19 patients at time of hospital discharge. Eur Respir J. 2020;55(6).
155. George PM, Wells AU, Jenkins RG. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. Lancet Respir Med. 2020;8(8):807-15.
156. Spagnolo P, Balestro E, Aliberti S*, et al.* Pulmonary fibrosis secondary to COVID-19: a call to arms? Lancet Respir Med. 2020;8(8):750-2.
157. Stewart I, Jacob J, George PM*, et al.* Residual Lung Abnormalities after COVID-19 Hospitalization: Interim Analysis of the UKILD Post-COVID-19 Study. Am J Respir Crit Care Med. 2023;207(6):693-703.
158. Sime PJ, Xing Z, Graham FL*, et al.* Adenovector-mediated gene transfer of active transforming growth factor-beta1 induces prolonged severe fibrosis in rat lung. J Clin Invest. 1997;100(4):768-76.

159. Derynck R, Jarrett JA, Chen EY*, et al.* Human transforming growth factor-beta complementary DNA sequence and expression in normal and transformed cells. Nature. 1985;316(6030):701-5.
160. Cheifetz S, Weatherbee JA, Tsang ML*, et al.* The transforming growth factor-beta system, a complex pattern of cross-reactive ligands and receptors. Cell. 1987;48(3):409-15.
161. Derynck R, Lindquist PB, Lee A*, et al.* A new type of transforming growth factor-beta, TGF-beta 3. Embo j. 1988;7(12):3737-43.
162. Herrera JA, Dingle L, Montero MA*, et al.* The UIP/IPF fibroblastic focus is a collagen biosynthesis factory embedded in a distinct extracellular matrix. JCI Insight. 2022;7(16).
163. Khalil N, O'Connor RN, Flanders KC*, et al.* TGF-beta 1, but not TGF-beta 2 or TGF-beta 3, is differentially present in epithelial cells of advanced pulmonary fibrosis: an immunohistochemical study. Am J Respir Cell Mol Biol. 1996;14(2):131-8.
164. Khalil N, Bereznay O, Sporn M*, et al.* Macrophage production of transforming growth factor beta and fibroblast collagen synthesis in chronic pulmonary inflammation. J Exp Med. 1989;170(3):727-37.
165. McCartney-Francis N, Mizel D, Wong H*, et al.* TGF-beta regulates production of growth factors and TGF-beta by human peripheral blood monocytes. Growth Factors. 1990;4(1):27-35.
166. Gentry LE, Webb NR, Lim GJ*, et al.* Type 1 transforming growth factor beta: amplified expression and secretion of mature and precursor polypeptides in Chinese hamster ovary cells. Mol Cell Biol. 1987;7(10):3418-27.
167. Dubois CM, Laprise MH, Blanchette F*, et al.* Processing of transforming growth factor beta 1 precursor by human furin convertase. J Biol Chem. 1995;270(18):10618-24.
168. Shi M, Zhu J, Wang R*, et al.* Latent TGF-β structure and activation. Nature. 2011;474(7351):343-9.
169. Miyazono K, Olofsson A, Colosetti P*, et al.* A role of the latent TGF-beta 1-binding protein in the assembly and secretion of TGF-beta 1. Embo j. 1991;10(5):1091-101.
170. Wrana JL, Attisano L, Cárcamo J*, et al.* TGF beta signals through a heteromeric protein kinase receptor complex. Cell. 1992;71(6):1003-14.
171. Radaev S, Zou Z, Huang T*, et al.* Ternary complex of transforming growth factor-beta1 reveals isoform-specific ligand recognition and receptor recruitment in the superfamily. J Biol Chem. 2010;285(19):14806-14.
172. Lin HY, Wang XF, Ng-Eaton E*, et al.* Expression cloning of the TGF-beta type II receptor, a functional transmembrane serine/threonine kinase. Cell. 1992;68(4):775-85.
173. Wrana JL, Attisano L, Wieser R*, et al.* Mechanism of activation of the TGF-beta receptor. Nature. 1994;370(6488):341-7.
174. Derynck R, Gelbart WM, Harland RM*, et al.* Nomenclature: vertebrate mediators of TGFbeta family signals. Cell. 1996;87(2):173.
175. Macías-Silva M, Abdollah S, Hoodless PA*, et al.* MADR2 is a substrate of the TGFbeta receptor and its phosphorylation is required for nuclear accumulation and signaling. Cell. 1996;87(7):1215-24.
176. Zhang Y, Feng X, We R*, et al.* Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. Nature. 1996;383(6596):168-72.

177. Tsukazaki T, Chiang TA, Davison AF*, et al.* SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor. Cell. 1998;95(6):779-91.
178. Lagna G, Hata A, Hemmati-Brivanlou A*, et al.* Partnership between DPC4 and SMAD proteins in TGF-beta signalling pathways. Nature. 1996;383(6603):832-6.
179. Nakao A, Imamura T, Souchelnytskyi S*, et al.* TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. Embo j. 1997;16(17):5353-62.
180. Heldin CH, Miyazono K, ten Dijke P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. Nature. 1997;390(6659):465-71.
181. Massagué J. How cells read TGF-beta signals. Nat Rev Mol Cell Biol. 2000;1(3):169-78.
182. Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature. 2003;425(6958):577-84.
183. Hayashi H, Abdollah S, Qiu Y*, et al.* The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. Cell. 1997;89(7):1165-73.
184. Imamura T, Takase M, Nishihara A*, et al.* Smad6 inhibits signalling by the TGF-beta superfamily. Nature. 1997;389(6651):622-6.
185. Kavsak P, Rasmussen RK, Causing CG*, et al.* Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. Mol Cell. 2000;6(6):1365-75.
186. Araya J, Cambier S, Markovics JA*, et al.* Squamous metaplasia amplifies pathologic epithelial-mesenchymal interactions in COPD patients. J Clin Invest. 2007;117(11):3551-62.
187. Tatler AL, Goodwin AT, Gbolahan O*, et al.* Amplification of TGFβ Induced ITGB6 Gene Transcription May Promote Pulmonary Fibrosis. PLoS One. 2016;11(8):e0158047.
188. Tamkun JW, DeSimone DW, Fonda D*, et al.* Structure of integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin. Cell. 1986;46(2):271-82.
189. Hynes RO. Integrins: a family of cell surface receptors. Cell. 1987;48(4):549-54.
190. Hynes RO. Integrins: versatility, modulation, and signaling in cell adhesion. Cell. 1992;69(1):11-25.
191. Plow EF, Haas TA, Zhang L*, et al.* Ligand binding to integrins. J Biol Chem. 2000;275(29):21785-8.
192. Erb EM, Tangemann K, Bohrmann B*, et al.* Integrin alphaIIb beta3 reconstituted into lipid bilayers is nonclustered in its activated state but clusters after fibrinogen binding. Biochemistry. 1997;36(24):7395-402.
193. Xiong JP, Stehle T, Diefenbach B*, et al.* Crystal structure of the extracellular segment of integrin alpha Vbeta3. Science. 2001;294(5541):339-45.
194. Hynes RO. Integrins: bidirectional, allosteric signaling machines. Cell. 2002;110(6):673-87.
195. Beglova N, Blacklow SC, Takagi J*, et al.* Cysteine-rich module structure reveals a fulcrum for integrin rearrangement upon activation. Nat Struct Biol. 2002;9(4):282-7.
196. Takagi J, Petre BM, Walz T*, et al.* Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. Cell. 2002;110(5):599-11.

197. Mould AP, Garratt AN, Askari JA*, et al.* Regulation of integrin alpha 5 beta 1 function by anti-integrin antibodies and divalent cations. Biochem Soc Trans. 1995;23(3):395s.
198. Gailit J, Ruoslahti E. Regulation of the fibronectin receptor affinity by divalent cations. J Biol Chem. 1988;263(26):12927-32.
199. Lee JO, Bankston LA, Arnaout MA*, et al.* Two conformations of the integrin A-domain (I-domain): a pathway for activation? Structure. 1995;3(12):1333-40.
200. Pierschbacher MD, Ruoslahti E. Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. Nature. 1984;309(5963):30-3.
201. Ruoslahti E, Pierschbacher MD. New perspectives in cell adhesion: RGD and integrins. Science. 1987;238(4826):491-7.
202. Xiong JP, Stehle T, Zhang R*, et al.* Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. Science. 2002;296(5565):151-5.
203. Horwitz A, Duggan K, Buck C*, et al.* Interaction of plasma membrane fibronectin receptor with talin--a transmembrane linkage. Nature. 1986;320(6062):531-3.
204. Lazarides E, Weber K. Actin antibody: the specific visualization of actin filaments in non-muscle cells. Proc Natl Acad Sci U S A. 1974;71(6):2268-72.
205. Weber K, Groeschel-Stewart U. Antibody to myosin: the specific visualization of myosin-containing filaments in nonmuscle cells. Proc Natl Acad Sci U S A. 1974;71(11):4561-4.
206. Goldman RD, Lazarides E, Pollack R*, et al.* The distribution of actin in non-muscle cells. The use of actin antibody in the localization of actin within the microfilament bundles of mouse 3T3 cells. Exp Cell Res. 1975;90(2):333-44.
207. Munger JS, Huang X, Kawakatsu H*, et al.* The integrin alpha v beta 6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis. Cell. 1999;96(3):319-28.
208. Sheppard D. The role of integrins in pulmonary fibrosis. European Respiratory Review. 2008;17(109):157.
209. Adams TS, Schupp JC, Poli S*, et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Sci Adv. 2020;6(28):eaba1983.
210. Neumark N. Idiopathic Pulmonary Fibrosis Cell Atlas. [Available from: http://www.ipfcellatlas.com/].
211. Reyfman PA, Walter JM, Joshi N*, et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. Am J Respir Crit Care Med. 2019;199(12):1517-36.
212. Sheppard D, Rozzo C, Starr L*, et al.* Complete amino acid sequence of a novel integrin beta subunit (beta 6) identified in epithelial cells using the polymerase chain reaction. J Biol Chem. 1990;265(20):11502-7.
213. Breuss JM, Gillett N, Lu L*, et al.* Restricted distribution of integrin beta 6 mRNA in primate epithelial tissues. J Histochem Cytochem. 1993;41(10):1521-7.
214. Horan GS, Wood S, Ona V*, et al.* Partial inhibition of integrin alpha(v)beta6 prevents pulmonary fibrosis without exacerbating inflammation. Am J Respir Crit Care Med. 2008;177(1):56-65.

215. Saini G, Porte J, Weinreb PH*, et al.* αvβ6 integrin may be a potential prognostic biomarker in interstitial lung disease. Eur Respir J. 2015;46(2):486-94.
216. Xu MY, Porte J, Knox AJ*, et al.* Lysophosphatidic acid induces alphavbeta6 integrin-mediated TGF-beta activation via the LPA2 receptor and the small G protein G alpha(q). Am J Pathol. 2009;174(4):1264-79.
217. Sheppard D, Cohen DS, Wang A*, et al.* Transforming growth factor beta differentially regulates expression of integrin subunits in guinea pig airway epithelial cells. J Biol Chem. 1992;267(24):17409-14.
218. Puthawala K, Hadjiangelis N, Jacoby SC*, et al.* Inhibition of integrin alpha(v)beta6, an activator of latent transforming growth factor-beta, prevents radiation-induced lung fibrosis. Am J Respir Crit Care Med. 2008;177(1):82-90.
219. Hahm K, Lukashev ME, Luo Y*, et al.* Alphav beta6 integrin regulates renal fibrosis and inflammation in Alport mouse. Am J Pathol. 2007;170(1):110-25.
220. Tager AM, LaCamera P, Shea BS*, et al.* The lysophosphatidic acid receptor LPA1 links pulmonary fibrosis to lung injury by mediating fibroblast recruitment and vascular leak. Nat Med. 2008;14(1):45-54.
221. Reed NI, Jo H, Chen C*, et al.* The αvβ1 integrin plays a critical in vivo role in tissue fibrosis. Sci Transl Med. 2015;7(288):288ra79.
222. Munger JS, Harpel JG, Giancotti FG*, et al.* Interactions between growth factors and integrins: latent forms of transforming growth factor-beta are ligands for the integrin alphavbeta1. Mol Biol Cell. 1998;9(9):2627-38.
223. Henderson NC, Arnold TD, Katamura Y*, et al.* Targeting of αv integrin identifies a core molecular pathway that regulates fibrosis in several organs. Nat Med. 2013;19(12):1617-24.
224. Scotton CJ, Krupiczojc MA, Königshoff M*, et al.* Increased local expression of coagulation factor X contributes to the fibrotic response in human and murine lung injury. J Clin Invest. 2009;119(9):2550-63.
225. Atabai K, Jame S, Azhar N*, et al.* Mfge8 diminishes the severity of tissue fibrosis in mice by binding and targeting collagen for uptake by macrophages. J Clin Invest. 2009;119(12):3713-22.
226. Decaris ML, Schaub JR, Chen C*, et al.* Dual inhibition of α(v)β(6) and α(v)β(1) reduces fibrogenesis in lung tissue explants from patients with IPF. Respir Res. 2021;22(1):265.
227. Heino J, Ignotz RA, Hemler ME*, et al.* Regulation of cell adhesion receptors by transforming growth factor-beta. Concomitant regulation of integrins that share a common beta 1 subunit. J Biol Chem. 1989;264(1):380-8.
228. Lancaster LH, Cottin V, Ramaswamy M*, et al.* PLN-74809 Shows Favorable Safety and Tolerability and Indicates Antifibrotic Activity in a Phase 2a Study for the Treatment of Idiopathic Pulmonary Fibrosis. B17 EMERGING DATA ON DISEASE AND SYMPTOM BASED THERAPEUTICS FOR PATIENTS WITH IPF. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2023. p. A2777-A.
229. Teichberg VI, Silman I, Beitsch DD*, et al.* A beta-D-galactoside binding protein from electric organ tissue of Electrophorus electricus. Proc Natl Acad Sci U S A. 1975;72(4):1383-7.
230. de Waard A, Hickman S, Kornfeld S. Isolation and properties of beta-galactoside binding lectins of calf heart and lung. J Biol Chem. 1976;251(23):7581-7.

231. Barondes SH, Castronovo V, Cooper DN*, et al.* Galectins: a family of animal beta-galactoside-binding lectins. Cell. 1994;76(4):597-8.
232. Leffler H, Carlsson S, Hedlund M*, et al.* Introduction to galectins. Glycoconj J. 2002;19(7-9):433-40.
233. Cummings RD, Liu FT. Galectins. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.
234. Lobsanov YD, Gitt MA, Leffler H*, et al.* X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-A resolution. J Biol Chem. 1993;268(36):27034-8.
235. Barondes SH, Cooper DN, Gitt MA*, et al.* Galectins. Structure and function of a large family of animal lectins. J Biol Chem. 1994;269(33):20807-10.
236. Hirabayashi J, Hashidate T, Arata Y*, et al.* Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. Biochim Biophys Acta. 2002;1572(2-3):232-54.
237. Johannes L, Jacob R, Leffler H. Galectins at a glance. J Cell Sci. 2018;131(9).
238. Drickamer K. Two distinct classes of carbohydrate-recognition domains in animal lectins. J Biol Chem. 1988;263(20):9557-60.
239. Hirabayashi J, Kasai K. The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution. Glycobiology. 1993;3(4):297-304.
240. Wilson TJ, Firth MN, Powell JT*, et al.* The sequence of the mouse 14 kDa beta-galactoside-binding lectin and evidence for its synthesis on free cytoplasmic ribosomes. Biochem J. 1989;261(3):847-52.
241. Cooper DN, Barondes SH. Evidence for export of a muscle lectin from cytosol to extracellular matrix and for a novel secretory mechanism. J Cell Biol. 1990;110(5):1681-91.
242. Lindstedt R, Apodaca G, Barondes SH*, et al.* Apical secretion of a cytosolic protein by Madin-Darby canine kidney cells. Evidence for polarized release of an endogenous lectin by a nonclassical secretory pathway. J Biol Chem. 1993;268(16):11750-7.
243. Sato S, Burdett I, Hughes RC. Secretion of the baby hamster kidney 30-kDa galactose-binding lectin from polarized and nonpolarized cells: a pathway independent of the endoplasmic reticulum-Golgi complex. Exp Cell Res. 1993;207(1):8-18.
244. Ohyama Y, Hirabayashi J, Oda Y*, et al.* Nucleotide sequence of chick 14K beta-galactoside-binding lectin mRNA. Biochem Biophys Res Commun. 1986;134(1):51-6.
245. Kasai K, Hirabayashi J. Galectins: a family of animal lectins that decipher glycocodes. J Biochem. 1996;119(1):1-8.
246. Hughes RC. Secretion of the galectin family of mammalian carbohydrate-binding proteins. Biochim Biophys Acta. 1999;1473(1):172-85.
247. Liu FT, Patterson RJ, Wang JL. Intracellular functions of galectins. Biochim Biophys Acta. 2002;1572(2-3):263-73.
248. Barondes SH. Soluble lectins: a new class of extracellular proteins. Science. 1984;223(4642):1259-64.
249. Haudek KC, Spronk KJ, Voss PG*, et al.* Dynamics of galectin-3 in the nucleus and cytoplasm. Biochim Biophys Acta. 2010;1800(2):181-9.

250. Joeh E, O'Leary T, Li W*, et al.* Mapping glycan-mediated galectin-3 interactions by live cell proximity labeling. Proc Natl Acad Sci U S A. 2020;117(44):27329-38.
251. Yang RY, Rabinovich GA, Liu FT. Galectins: structure, function and therapeutic potential. Expert Rev Mol Med. 2008;10:e17.
252. Smith DF, Song X, Cummings RD. Use of glycan microarrays to explore specificity of glycan-binding proteins. Methods Enzymol. 2010;480:417-44.
253. Hsieh TJ, Lin HY, Tu Z*, et al.* Structural Basis Underlying the Binding Preference of Human Galectins-1, -3 and -7 for Galβ1-3/4GlcNAc. PLoS One. 2015;10(5):e0125946.
254. Yamamoto F, Clausen H, White T*, et al.* Molecular genetic basis of the histo-blood group ABO system. Nature. 1990;345(6272):229-33.
255. Sharon N, Lis H. Lectins: cell-agglutinating and sugar-specific proteins. Science. 1972;177(4053):949-59.
256. Thiemann S, Baum LG. Galectins and Immune Responses-Just How Do They Do Those Things They Do? Annu Rev Immunol. 2016;34:243-64.
257. Brewer CF, Miceli MC, Baum LG. Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. Curr Opin Struct Biol. 2002;12(5):616-23.
258. Demetriou M, Granovsky M, Quaggin S*, et al.* Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. Nature. 2001;409(6821):733-9.
259. Partridge EA, Le Roy C, Di Guglielmo GM*, et al.* Regulation of cytokine receptors by Golgi N-glycan processing and endocytosis. Science. 2004;306(5693):120-4.
260. Massa SM, Cooper DN, Leffler H*, et al.* L-29, an endogenous lectin, binds to glycoconjugate ligands with positive cooperativity. Biochemistry. 1993;32(1):260-7.
261. Roff CF, Rosevear PR, Wang JL*, et al.* Identification of carbohydrate-binding proteins from mouse and human fibroblasts. Biochem J. 1983;211(3):625-9.
262. Robertson MW, Albrandt K, Keller D*, et al.* Human IgE-binding protein: a soluble lectin exhibiting a highly conserved interspecies sequence and differential recognition of IgE glycoforms. Biochemistry. 1990;29(35):8093-100.
263. Chiariotti L, Salvatore P, Frunzio R*, et al.* Galectin genes: regulation of expression. Glycoconj J. 2002;19(7-9):441-9.
264. Albrandt K, Orida NK, Liu FT. An IgE-binding protein with a distinctive repetitive sequence and homology with an IgG receptor. Proc Natl Acad Sci U S A. 1987;84(19):6859-63.
265. Kasper M, Hughes RC. Immunocytochemical evidence for a modulation of galectin 3 (Mac-2), a carbohydrate binding protein, in pulmonary fibrosis. J Pathol. 1996;179(3):309-16.
266. Cowles EA, Agrwal N, Anderson RL*, et al.* Carbohydrate-binding protein 35. Isoelectric points of the polypeptide and a phosphorylated derivative. J Biol Chem. 1990;265(29):17706-12.
267. Mazurek N, Conklin J, Byrd JC*, et al.* Phosphorylation of the beta-galactoside-binding protein galectin-3 modulates binding to its ligands. J Biol Chem. 2000;275(46):36311-5.
268. Menon RP, Hughes RC. Determinants in the N-terminal domains of galectin-3 for secretion by a novel pathway circumventing the endoplasmic reticulum-Golgi complex. Eur J Biochem. 1999;264(2):569-76.

269. Hsu DK, Zuberi RI, Liu FT. Biochemical and biophysical characterization of human recombinant IgE-binding protein, an S-type animal lectin. J Biol Chem. 1992;267(20):14167-74.
270. Mehul B, Bawumia S, Hughes RC. Cross-linking of galectin 3, a galactose-binding protein of mammalian cells, by tissue-type transglutaminase. FEBS Lett. 1995;360(2):160-4.
271. Yang RY, Hill PN, Hsu DK*, et al.* Role of the carboxyl-terminal lectin domain in self-association of galectin-3. Biochemistry. 1998;37(12):4086-92.
272. Ahmad N, Gabius HJ, André S*, et al.* Galectin-3 precipitates as a pentamer with synthetic multivalent carbohydrates and forms heterogeneous cross-linked complexes. J Biol Chem. 2004;279(12):10841-7.
273. Hsu DK, Dowling CA, Jeng KC*, et al.* Galectin-3 expression is induced in cirrhotic liver and hepatocellular carcinoma. Int J Cancer. 1999;81(4):519-26.
274. Wang L, Friess H, Zhu Z*, et al.* Galectin-1 and galectin-3 in chronic pancreatitis. Lab Invest. 2000;80(8):1233-41.
275. Henderson NC, Mackinnon AC, Farnworth SL*, et al.* Galectin-3 regulates myofibroblast activation and hepatic fibrosis. Proc Natl Acad Sci U S A. 2006;103(13):5060-5.
276. Henderson NC, Mackinnon AC, Farnworth SL*, et al.* Galectin-3 expression and secretion links macrophages to the promotion of renal fibrosis. Am J Pathol. 2008;172(2):288-98.
277. McCubbrey AL, Barthel L, Mohning MP*, et al.* Deletion of c-FLIP from CD11b(hi) Macrophages Prevents Development of Bleomycin-induced Lung Fibrosis. Am J Respir Cell Mol Biol. 2018;58(1):66-78.
278. Morse C, Tabib T, Sembrat J*, et al.* Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. Eur Respir J. 2019;54(2).
279. Joshi N, Watanabe S, Verma R*, et al.* A spatially restricted fibrotic niche in pulmonary fibrosis is sustained by M-CSF/M-CSFR signalling in monocyte-derived alveolar macrophages. Eur Respir J. 2020;55(1).
280. Serezani APM, Pascoalino BD, Bazzano JMR*, et al.* Multiplatform Single-Cell Analysis Identifies Immune Cell Types Enhanced in Pulmonary Fibrosis. Am J Respir Cell Mol Biol. 2022;67(1):50-60.
281. Fabre T, Barron AMS, Christensen SM*, et al.* Identification of a broadly fibrogenic macrophage subset induced by type 3 inflammation. Sci Immunol. 2023;8(82):eadd8945.
282. Gibbons MA, MacKinnon AC, Ramachandran P*, et al.* Ly6Chi monocytes direct alternatively activated profibrotic macrophage regulation of lung fibrosis. Am J Respir Crit Care Med. 2011;184(5):569-81.
283. MacKinnon AC, Farnworth SL, Hodkinson PS*, et al.* Regulation of alternative macrophage activation by galectin-3. J Immunol. 2008;180(4):2650-8.
284. Mackinnon AC, Gibbons MA, Farnworth SL*, et al.* Regulation of transforming growth factor-β1-driven lung fibrosis by galectin-3. Am J Respir Crit Care Med. 2012;185(5):537-46.
285. Inohara H, Akahani S, Raz A. Galectin-3 stimulates cell proliferation. Exp Cell Res. 1998;245(2):294-302.

286. Humphries DC, Mills R, Dobie R*, et al.* Selective Myeloid Depletion of Galectin-3 Offers Protection Against Acute and Chronic Lung Injury. Front Pharmacol. 2021;12:715986.
287. Panayiota S, Sara R, Wei W*, et al.* CD98 is critical for a conserved inflammatory response to diverse injury stimuli relevant to IPF exacerbations and COVID pneumonitis. bioRxiv. 2022:2022.08.12.503154.
288. Rock JR, Barkauskas CE, Cronce MJ*, et al.* Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. Proc Natl Acad Sci U S A. 2011;108(52):E1475-83.
289. Zhao J, Shi W, Wang YL*, et al.* Smad3 deficiency attenuates bleomycin-induced pulmonary fibrosis in mice. Am J Physiol Lung Cell Mol Physiol. 2002;282(3):L585-93.
290. Nishi Y, Sano H, Kawashima T*, et al.* Role of galectin-3 in human pulmonary fibrosis. Allergol Int. 2007;56(1):57-65.
291. Hirani N, MacKinnon AC, Nicol L*, et al.* Target inhibition of galectin-3 by inhaled TD139 in patients with idiopathic pulmonary fibrosis. Eur Respir J. 2021;57(5).
292. Hirabayashi J, Kawasaki H, Suzuki K*, et al.* Complete amino acid sequence of 14 kDa beta-galactoside-binding lectin of chick embryo. J Biochem. 1987;101(3):775-87.
293. Liao DI, Kapadia G, Ahmed H*, et al.* Structure of S-lectin, a developmentally regulated vertebrate beta-galactoside-binding protein. Proc Natl Acad Sci U S A. 1994;91(4):1428-32.
294. Cho M, Cummings RD. Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. I. Physical and chemical characterization. J Biol Chem. 1995;270(10):5198-206.
295. Bachem MG, Schneider E, Gross H*, et al.* Identification, culture, and characterization of pancreatic stellate cells in rats and humans. Gastroenterology. 1998;115(2):421-32.
296. Fitzner B, Walzel H, Sparmann G*, et al.* Galectin-1 is an inductor of pancreatic stellate cell activation. Cell Signal. 2005;17(10):1240-7.
297. Maeda N, Kawada N, Seki S*, et al.* Stimulation of proliferation of rat hepatic stellate cells by galectin-1 and galectin-3 through different intracellular signaling pathways. J Biol Chem. 2003;278(21):18938-44.
298. Friedman SL, Roll FJ, Boyles J*, et al.* Hepatic lipocytes: the principal collagen-producing cells of normal rat liver. Proc Natl Acad Sci U S A. 1985;82(24):8681-5.
299. Kristensen DB, Kawada N, Imamura K*, et al.* Proteome analysis of rat hepatic stellate cells. Hepatology. 2000;32(2):268-77.
300. Im HS, Kim HD, Song JY*, et al.* Overexpression of alpha1-protease inhibitor and galectin-1 in radiation-induced early phase of pulmonary fibrosis. Cancer Res Treat. 2006;38(2):92-8.
301. Lim MJ, Ahn J, Yi JY*, et al.* Induction of galectin-1 by TGF-β1 accelerates fibrosis through enhancing nuclear retention of Smad2. Exp Cell Res. 2014;326(1):125-35.
302. Kathiriya JJ, Nakra N, Nixon J*, et al.* Galectin-1 inhibition attenuates profibrotic signaling in hypoxia-induced pulmonary fibrosis. Cell Death Discov. 2017;3:17010.
303. Rajput VK, MacKinnon A, Mandal S*, et al.* A Selective Galactose-Coumarin-Derived Galectin-3 Inhibitor Demonstrates Involvement of Galectin-3-glycan Interactions in a Pulmonary Fibrosis Model. J Med Chem. 2016;59(17):8141-7.

304. Magi B, Bini L, Perari MG*, et al.* Bronchoalveolar lavage fluid protein composition in patients with sarcoidosis and idiopathic pulmonary fibrosis: a two-dimensional electrophoretic study. Electrophoresis. 2002;23(19):3434-44.
305. Rottoli P, Magi B, Perari MG*, et al.* Cytokine profile and proteome analysis in bronchoalveolar lavage of patients with sarcoidosis, pulmonary fibrosis associated with systemic sclerosis and idiopathic pulmonary fibrosis. Proteomics. 2005;5(5):1423-30.
306. Bennett D, Bargagli E, Bianchi N*, et al.* Elevated level of Galectin-1 in bronchoalveolar lavage of patients with idiopathic pulmonary fibrosis. Respir Physiol Neurobiol. 2020;273:103323.
307. d'Alessandro M, De Vita E, Bergantini L*, et al.* Galactin-1, 3 and 9: Potential biomarkers in idiopathic pulmonary fibrosis and other interstitial lung diseases. Respir Physiol Neurobiol. 2020;282:103546.
308. Dvořánková B, Szabo P, Lacina L*, et al.* Human galectins induce conversion of dermal fibroblasts into myofibroblasts and production of extracellular matrix: potential application in tissue engineering and wound repair. Cells Tissues Organs. 2011;194(6):469-80.
309. Chapman HA. Epithelial-mesenchymal interactions in pulmonary fibrosis. Annu Rev Physiol. 2011;73:413-35.
310. Parmar NR, Mackinnon AC, Zetterberg F*, et al.* Role of Galectin-3 in Idiopathic Pulmonary Fibrosis Development. C62 FIBROBLAST BIOLOGY. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2019. p. A5328-A.
311. Parmar NR, Tatler AL, MacKinnon A*, et al.* Galectin-3 Promotes Transforming Growth Factor b (TGFb) Activity in the Development of Idiopathic Pulmonary Fibrosis. A68 MOLECULAR DETERMINANTS OF REMODELING IN LUNG FIBROSIS. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2018. p. A2207-A.
312. Parmar N, Tatler A, Ford P*, et al.* S76 Role of galectin-3 in the development of idiopathic pulmonary fibrosis. Thorax. 2017;72(Suppl 3):A48.
313. Parmar N. The role of galectin-3 in the development of idiopathic pulmonary fibrosis. Nottingham eTheses: University of Nottingham; 2020.
314. Jones C, Patel A, Griffin S*, et al.* Current trends in molecular recognition and bioseparation. J Chromatogr A. 1995;707(1):3-22.
315. Dalton AC, Barton WA. Over-expression of secreted proteins from mammalian cell lines. Protein Sci. 2014;23(5):517-25.
316. Wingfield PT. Overview of the purification of recombinant proteins. Curr Protoc Protein Sci. 2015;80:6.1.-6.1.35.
317. Shi X, Jarvis DL. Protein N-glycosylation in the baculovirus-insect cell system. Curr Drug Targets. 2007;8(10):1116-25.
318. Kukuruzinska MA, Bergh ML, Jackson BJ. Protein glycosylation in yeast. Annu Rev Biochem. 1987;56:915-44.
319. Harbers M. Wheat germ systems for cell-free protein expression. FEBS Lett. 2014;588(17):2762-73.
320. Brondyk WH. Selecting an appropriate method for expressing a recombinant protein. Methods Enzymol. 2009;463:131-47.

321. Khan KH. Gene expression in Mammalian cells and its applications. Adv Pharm Bull. 2013;3(2):257-63.
322. Cooper S, Helmstetter CE. Chromosome replication and the division cycle of Escherichia coli B/r. J Mol Biol. 1968;31(3):519-40.
323. Sivashanmugam A, Murray V, Cui C*, et al.* Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. Protein Sci. 2009;18(5):936-48.
324. Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. 2007;35(Web Server issue):W43-6.
325. Frey UH, Bachmann HS, Peters J*, et al.* PCR-amplification of GC-rich regions: 'slowdown PCR'. Nat Protoc. 2008;3(8):1312-7.
326. Hilgarth RS, Lanigan TM. Optimization of overlap extension PCR for efficient transgene construction. MethodsX. 2020;7:100759.
327. Berrow NS, Alderton D, Sainsbury S*, et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res. 2007;35(6):e45.
328. Hasegawa H, Suzuki E, Maeda S. Horizontal Plasmid Transfer by Transformation in Escherichia coli: Environmental Factors and Possible Mechanisms. Front Microbiol. 2018;9:2365.
329. Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 1979;7(6):1513-23.
330. Boussif O, Lezoualc'h F, Zanta MA*, et al.* A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. Proc Natl Acad Sci U S A. 1995;92(16):7297-301.
331. Krasnoselska GO, Dumoux M, Gamage N*, et al.* Transient Transfection and Expression of Eukaryotic Membrane Proteins in Expi293F Cells and Their Screening on a Small Scale: Application for Structural Studies. Methods Mol Biol. 2021;2305:105-28.
332. Birch J, Cheruvara H, Gamage N*, et al.* Changes in Membrane Protein Structural Biology. Biology (Basel). 2020;9(11).
333. Birch J, Axford D, Foadi J*, et al.* The fine art of integral membrane protein crystallisation. Methods. 2018;147:150-62.
334. Golovanov AP, Hautbergue GM, Wilson SA*, et al.* A simple method for improving protein solubility and long-term stability. J Am Chem Soc. 2004;126(29):8933-9.
335. Gamage N, Cheruvara H, Harrison PJ*, et al.* High-Throughput Production and Optimization of Membrane Proteins After Expression in Mammalian Cells. Methods Mol Biol. 2023;2652:79-118.
336. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J. 2000;78(3):1606-19.
337. Gasteiger E, Hoogland C, Gattiker A*, et al.* Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005. p. 571-607.
338. Tripathi NK, Shrivastava A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. Front Bioeng Biotechnol. 2019;7:420.

339. Hussain H, Maldonado-Agurto R, Dickson AJ. The endoplasmic reticulum and unfolded protein response in the control of mammalian recombinant protein production. Biotechnol Lett. 2014;36(8):1581-93.
340. Kimple ME, Brill AL, Pasker RL. Overview of affinity tags for protein purification. Curr Protoc Protein Sci. 2013;73:9..1-9..23.
341. Zheng M, Fang H, Hakomori S. Functional role of N-glycosylation in alpha 5 beta 1 integrin receptor. De-N-glycosylation induces dissociation or altered association of alpha 5 and beta 1 subunits and concomitant loss of fibronectin binding activity. J Biol Chem. 1994;269(16):12325-31.
342. Isaji T, Sato Y, Zhao Y*, et al.* N-glycosylation of the beta-propeller domain of the integrin alpha5 subunit is essential for alpha5beta1 heterodimerization, expression on the cell surface, and its biological function. J Biol Chem. 2006;281(44):33258-67.
343. Lagana A, Goetz JG, Cheung P*, et al.* Galectin binding to Mgat5-modified N-glycans regulates fibronectin matrix remodeling in tumor cells. Mol Cell Biol. 2006;26(8):3181-93.
344. Goetz JG, Joshi B, Lajoie P*, et al.* Concerted regulation of focal adhesion dynamics by galectin-3 and tyrosine-phosphorylated caveolin-1. J Cell Biol. 2008;180(6):1261-75.
345. Saravanan C, Liu FT, Gipson IK*, et al.* Galectin-3 promotes lamellipodia formation in epithelial cells by interacting with complex N-glycans on alpha3beta1 integrin. J Cell Sci. 2009;122(Pt 20):3684-93.
346. Yang EH, Rode J, Howlader MA*, et al.* Galectin-3 alters the lateral mobility and clustering of β1-integrin receptors. PLoS One. 2017;12(10):e0184378.
347. Rossi B, Espeli M, Schiff C*, et al.* Clustering of pre-B cell integrins induces galectin-1-dependent pre-B cell receptor relocalization and activation. J Immunol. 2006;177(2):796-803.
348. Priglinger CS, Szober CM, Priglinger SG*, et al.* Galectin-3 induces clustering of CD147 and integrin-β1 transmembrane glycoprotein receptors on the RPE cell surface. PLoS One. 2013;8(7):e70011.
349. Janik ME, Lityńska A, Vereecken P. Cell migration-the role of integrin glycosylation. Biochim Biophys Acta. 2010;1800(6):545-55.
350. Isaji T, Sato Y, Fukuda T*, et al.* N-glycosylation of the I-like domain of beta1 integrin is essential for beta1 integrin expression and biological function: identification of the minimal N-glycosylation requirement for alpha5beta1. J Biol Chem. 2009;284(18):12207-16.
351. Cai X, Thinn AMM, Wang Z*, et al.* The importance of N-glycosylation on β(3) integrin ligand binding and conformational regulation. Sci Rep. 2017;7(1):4656.
352. Kim YW, Park J, Lee HJ*, et al.* TGF-β sensitivity is determined by N-linked glycosylation of the type II TGF-β receptor. Biochem J. 2012;445(3):403-11.
353. York WS, Mazumder R, Ranzinger R*, et al.* GlyGen: Computational and Informatics Resources for Glycoscience. Glycobiology. 2020;30(2):72-3.
354. O'Sullivan JM, Jenkins PV, Rawley O*, et al.* Galectin-1 and Galectin-3 Constitute Novel-Binding Partners for Factor VIII. Arterioscler Thromb Vasc Biol. 2016;36(5):855-63.
355. Dos Santos SN, Sheldon H, Pereira JX*, et al.* Galectin-3 acts as an angiogenic switch to induce tumor angiogenesis via Jagged-1/Notch activation. Oncotarget. 2017;8(30):49484-501.

356. Natarajamurthy SH, Sistla S, Dharmesh SM. Disruption of galectin-3 and galectin-3 binding protein (G3BP) interaction by dietary pectic polysaccharides (DPP) - Arrest of metastasis, inhibition of proliferation and induction of apoptosis. Int J Biol Macromol. 2019;139:486-99.
357. Pedersen K, Nielsen MA, Juul-Madsen K*, et al.* Galectin-3 interacts with PD-1 and counteracts the PD-1 pathway-driven regulation of T cell and osteoclast activity in Rheumatoid Arthritis. Scand J Immunol. 2023;97(2):e13245.
358. Gorshkova, II, Svitel J, Razjouyan F*, et al.* Bayesian analysis of heterogeneity in the distribution of binding properties of immobilized surface sites. Langmuir. 2008;24(20):11577-86.
359. Leppänen A, Stowell S, Blixt O*, et al.* Dimeric galectin-1 binds with high affinity to alpha2,3-sialylated and non-sialylated terminal N-acetyllactosamine units on surface-bound extended glycans. J Biol Chem. 2005;280(7):5549-62.
360. Kavanaugh D, Kane M, Joshi L*, et al.* Detection of galectin-3 interaction with commensal bacteria. Appl Environ Microbiol. 2013;79(11):3507-10.
361. Maljaars CE, André S, Halkes KM*, et al.* Assessing the inhibitory potency of galectin ligands identified from combinatorial (glyco)peptide libraries using surface plasmon resonance spectroscopy. Anal Biochem. 2008;378(2):190-6.
362. Xue J, Gao X, Fu C*, et al.* Regulation of galectin-3-induced apoptosis of Jurkat cells by both O-glycans and N-glycans on CD45. FEBS Lett. 2013;587(24):3986-94.
363. Lepur A, Salomonsson E, Nilsson UJ*, et al.* Ligand induced galectin-3 protein self-association. J Biol Chem. 2012;287(26):21751-6.
364. Vashist SK, Dixit CK, MacCraith BD*, et al.* Effect of antibody immobilization strategies on the analytical performance of a surface plasmon resonance-based immunoassay. Analyst. 2011;136(21):4431-6.
365. Shen M, Rusling J, Dixit CK. Site-selective orientated immobilization of antibodies and conjugates for immunodiagnostics development. Methods. 2017;116:95-111.
366. Fukushi J, Makagiansar IT, Stallcup WB. NG2 proteoglycan promotes endothelial cell motility and angiogenesis via engagement of galectin-3 and alpha3beta1 integrin. Mol Biol Cell. 2004;15(8):3580-90.
367. Ippel H, Miller MC, Vértesy S*, et al.* Intra- and intermolecular interactions of human galectin-3: assessment by full-assignment-based NMR. Glycobiology. 2016;26(8):888-903.
368. Lin YH, Qiu DC, Chang WH*, et al.* The intrinsically disordered N-terminal domain of galectin-3 dynamically mediates multisite self-association of the protein through fuzzy interactions. J Biol Chem. 2017;292(43):17845-56.
369. Kariya Y, Kawamura C, Tabei T*, et al.* Bisecting GlcNAc residues on laminin-332 down-regulate galectin-3-dependent keratinocyte motility. J Biol Chem. 2010;285(5):3330-40.
370. Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. Microbiol Rev. 1995;59(1):94-123.
371. Dunn KW, Kamocka MM, McDonald JH. A practical guide to evaluating colocalization in biological microscopy. Am J Physiol Cell Physiol. 2011;300(4):C723-42.
372. ThermoFisher Scientific. Overview of the Immunoprecipitation (IP) Technique. [Available from: www.thermofisher.com/uk/en/home/life-science/protein-

biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/immunoprecipitation-ip].
373. Eliasson M, Olsson A, Palmcrantz E*, et al.* Chimeric IgG-binding receptors engineered from staphylococcal protein A and streptococcal protein G. J Biol Chem. 1988;263(9):4323-7.
374. Abcam. Immunoprecipitation (IP) protocol. [Available from: www.abcam.com/protocols/immunoprecipitation-protocol].
375. Hönig E, Ringer K, Dewes J*, et al.* Galectin-3 modulates the polarized surface delivery of β1-integrin in epithelial cells. J Cell Sci. 2018;131(11).
376. Zhu N, Zhang D, Wang W*, et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33.
377. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed. 2020;91(1):157-60.
378. Wan Y, Shang J, Graham R*, et al.* Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol. 2020;94(7).
379. Li W, Moore MJ, Vasilieva N*, et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003;426(6965):450-4.
380. Yan R, Zhang Y, Li Y*, et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020;367(6485):1444-8.
381. Lan J, Ge J, Yu J*, et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581(7807):215-20.
382. Li F, Li W, Farzan M*, et al.* Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science. 2005;309(5742):1864-8.
383. Wu F, Zhao S, Yu B*, et al.* A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265-9.
384. Wrapp D, Wang N, Corbett KS*, et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483):1260-3.
385. Walls AC, Park YJ, Tortorici MA*, et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020;181(2):281-92.e6.
386. Huang C, Wang Y, Li X*, et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497-506.
387. Aguiar JA, Tremblay BJ, Mansfield MJ*, et al.* Gene expression and in situ protein profiling of candidate SARS-CoV-2 receptors in human airway epithelial cells and lung tissue. Eur Respir J. 2020;56(3).
388. Bui LT, Winters NI, Chung MI*, et al.* Chronic lung diseases are associated with gene expression programs favoring SARS-CoV-2 entry and severity. Nat Commun. 2021;12(1):4314.
389. Ziegler CGK, Allon SJ, Nyquist SK*, et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. Cell. 2020;181(5):1016-35.e19.
390. Nawijn MC, Timens W. Can ACE2 expression explain SARS-CoV-2 infection of the respiratory epithelia in COVID-19? Mol Syst Biol. 2020;16(7):e9841.
391. Wang K, Chen W, Zhang Z*, et al.* CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells. Signal Transduct Target Ther. 2020;5(1):283.
392. Cantuti-Castelvetri L, Ojha R, Pedro LD*, et al.* Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. Science. 2020;370(6518):856-60.

393. Ibrahim IM, Abdelmalek DH, Elshahat ME*, et al.* COVID-19 spike-host cell receptor GRP78 binding site prediction. J Infect. 2020;80(5):554-62.
394. Carlos AJ, Ha DP, Yeh DW*, et al.* The chaperone GRP78 is a host auxiliary factor for SARS-CoV-2 and GRP78 depleting antibody blocks viral entry and infection. J Biol Chem. 2021;296:100759.
395. Hoffmann M, Kleine-Weber H, Schroeder S*, et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020;181(2):271-80.e8.
396. Ou X, Liu Y, Lei X*, et al.* Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 2020;11(1):1620.
397. Zhao MM, Yang WL, Yang FY*, et al.* Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. Signal Transduct Target Ther. 2021;6(1):134.
398. Li F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. J Virol. 2015;89(4):1954-64.
399. Sigrist CJ, Bridge A, Le Mercier P. A potential role for integrins in host cell entry by SARS-CoV-2. Antiviral Res. 2020;177:104759.
400. Hussein HA, Walker LR, Abdel-Raouf UM*, et al.* Beyond RGD: virus interactions with integrins. Arch Virol. 2015;160(11):2669-81.
401. Makowski L, Olson-Sidford W, J WW. Biological and Clinical Consequences of Integrin Binding via a Rogue RGD Motif in the SARS CoV-2 Spike Protein. Viruses. 2021;13(2).
402. Calver J, Joseph C, John AE*, et al.* S31 The novel coronavirus SARS-CoV-2 binds RGD integrins and upregulates avb3 integrins in Covid-19 infected lungs. Thorax. 2021;76(Suppl 1):A22.
403. Joseph C, Peacock T, Calver J*, et al.* Binding of SARS-CoV-2 to the avb6 Integrins May Promote Severe COVID in Patients with IPF. TP105 TP105 BASIC MECHANISMS OF LUNG INFECTIONS: FROM SARS-COV-2 TO INFLUENZA. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2021. p. A4170-A.
404. Simons P, Rinaldi DA, Bondu V*, et al.* Integrin activation is an essential component of SARS-CoV-2 infection. Sci Rep. 2021;11(1):20398.
405. Norris EG, Pan XS, Hocking DC. Receptor-binding domain of SARS-CoV-2 is a functional αv-integrin agonist. J Biol Chem. 2023;299(3):102922.
406. Beddingfield BJ, Iwanaga N, Chapagain PP*, et al.* The Integrin Binding Peptide, ATN-161, as a Novel Therapy for SARS-CoV-2 Infection. JACC Basic Transl Sci. 2021;6(1):1-8.
407. Robles JP, Zamora M, Adan-Castro E*, et al.* The spike protein of SARS-CoV-2 induces endothelial inflammation through integrin α5β1 and NF-κB signaling. J Biol Chem. 2022;298(3):101695.
408. Liu J, Lu F, Chen Y*, et al.* Integrin mediates cell entry of the SARS-CoV-2 virus independent of cellular receptor ACE2. J Biol Chem. 2022;298(3):101710.
409. Clarke NE, Fisher MJ, Porter KE*, et al.* Angiotensin converting enzyme (ACE) and ACE2 bind integrins and ACE2 regulates integrin signalling. PLoS One. 2012;7(4):e34747.
410. Lee H, Choi H, Yang B*, et al.* Interstitial lung disease increases susceptibility to and severity of COVID-19. Eur Respir J. 2021;58(6).

411. Meliopoulos VA, Van de Velde LA, Van de Velde NC*, et al.* An Epithelial Integrin Regulates the Amplitude of Protective Lung Interferon Responses against Multiple Respiratory Pathogens. PLoS Pathog. 2016;12(8):e1005804.
412. Ackermann M, Verleden SE, Kuehnel M*, et al.* Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. N Engl J Med. 2020;383(2):120-8.
413. Nader D, Fletcher N, Curley GF*, et al.* SARS-CoV-2 uses major endothelial integrin αvβ3 to cause vascular dysregulation in-vitro during COVID-19. PLoS One. 2021;16(6):e0253347.
414. Caccuri F, Bugatti A, Zani A*, et al.* SARS-CoV-2 Infection Remodels the Phenotype and Promotes Angiogenesis of Primary Human Lung Endothelial Cells. Microorganisms. 2021;9(7).
415. Bugatti A, Filippini F, Bardelli M*, et al.* SARS-CoV-2 Infects Human ACE2-Negative Endothelial Cells through an α(v)β(3) Integrin-Mediated Endocytosis Even in the Presence of Vaccine-Elicited Neutralizing Antibodies. Viruses. 2022;14(4).
416. Tresoldi I, Sangiuolo CF, Manzari V*, et al.* SARS-COV-2 and infectivity: Possible increase in infectivity associated to integrin motif expression. J Med Virol. 2020;92(10):1741-2.
417. Beaudoin CA, Hamaia SW, Huang CL*, et al.* Can the SARS-CoV-2 Spike Protein Bind Integrins Independent of the RGD Sequence? Front Cell Infect Microbiol. 2021;11:765300.
418. Othman H, Messaoud HB, Khamessi O*, et al.* SARS-CoV-2 Spike Protein Unlikely to Bind to Integrins via the Arg-Gly-Asp (RGD) Motif of the Receptor Binding Domain: Evidence From Structural Analysis and Microscale Accelerated Molecular Dynamics. Front Mol Biosci. 2022;9:834857.
419. Machala EA, McSharry BP, Rouse BT*, et al.* Gal power: the diverse roles of galectins in regulating viral infections. J Gen Virol. 2019;100(3):333-49.
420. De Biasi S, Meschiari M, Gibellini L*, et al.* Marked T cell activation, senescence, exhaustion and skewing towards TH17 in patients with COVID-19 pneumonia. Nat Commun. 2020;11(1):3434.
421. Portacci A, Diaferia F, Santomasi C*, et al.* Galectin-3 as prognostic biomarker in patients with COVID-19 acute respiratory failure. Respir Med. 2021;187:106556.
422. Gaughan EE, Quinn TM, Mills A*, et al.* An Inhaled Galectin-3 Inhibitor in COVID-19 Pneumonitis: A Phase Ib/IIa Randomized Controlled Clinical Trial (DEFINE). Am J Respir Crit Care Med. 2023;207(2):138-49.
423. Caniglia JL, Guda MR, Asuthkar S*, et al.* A potential role for Galectin-3 inhibitors in the treatment of COVID-19. PeerJ. 2020;8:e9392.
424. Caniglia JL, Asuthkar S, Tsung AJ*, et al.* Immunopathology of galectin-3: an increasingly promising target in COVID-19. F1000Res. 2020;9:1078.
425. Carvacho I, Piesche M. RGD-binding integrins and TGF-β in SARS-CoV-2 infections - novel targets to treat COVID-19 patients? Clin Transl Immunology. 2021;10(3):e1240.
426. Jordana M, Schulman J, McSharry C*, et al.* Heterogeneous proliferative characteristics of human adult lung fibroblast lines and clonally derived fibroblasts from control and fibrotic tissue. Am Rev Respir Dis. 1988;137(3):579-84.
427. Liu X, Dai K, Zhang X*, et al.* Multiple Fibroblast Subtypes Contribute to Matrix Deposition in Pulmonary Fibrosis. Am J Respir Cell Mol Biol. 2023.

428. Xie T, Wang Y, Deng N*, et al.* Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary Fibrosis. Cell Rep. 2018;22(13):3625-40.
429. Liu X, Rowan SC, Liang J*, et al.* Categorization of lung mesenchymal cells in development and fibrosis. iScience. 2021;24(6):102551.
430. Tsukui T, Sun KH, Wetter JB*, et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. Nat Commun. 2020;11(1):1920.
431. Habermann AC, Gutierrez AJ, Bui LT*, et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. Sci Adv. 2020;6(28):eaba1972.
432. Reily C, Stewart TJ, Renfrow MB*, et al.* Glycosylation in health and disease. Nat Rev Nephrol. 2019;15(6):346-66.
433. de Haas P, Hendriks W, Lefeber DJ*, et al.* Biological and Technical Challenges in Unraveling the Role of N-Glycans in Immune Receptor Regulation. Front Chem. 2020;8:55.

9. Appendix

9.1. Tables referenced in the main text

Table 9.1: Galectin-1 and Galectin-3 K_d values and cell permeability for Galecto Biotech compounds

Compound	Galectin-1 K_d (µM)	Galectin-3 K_d (µM)	CACO-2 (A > B/B > A) Papp (10−6 cm/s)
GB0139	0.012	0.0023	0.07/0.05
GB1107	3.7	0.037	15/16
GB1490	0.41	2.7	11/31
GB0149	0.230	0.099	<0.03/<0.07

Table 9.2: MW estimation, (with percentage uncertainty) and polydispersity for galectin-3 and galectin-1 peaks by SEC-MALS

Galectin	Peak	Mn (kDa)	Uncertainty	Mw (kDa)	Uncertainty	Polydispersity (Mw/Mn)	Uncertainty
Galectin-3	Peak 1	23.2	0.80%	23.2	0.80%	1	1.09%
Galectin-1	Peak 1	13.4	26.50%	14	25.70%	1.047	36.93%
	Peak 2	24	1.20%	24	1.20%	1.001	1.65%
	Peak 3	16	48.10%	16.4	45.00%	1.024	65.86%

9.2. Figures referenced in the main text

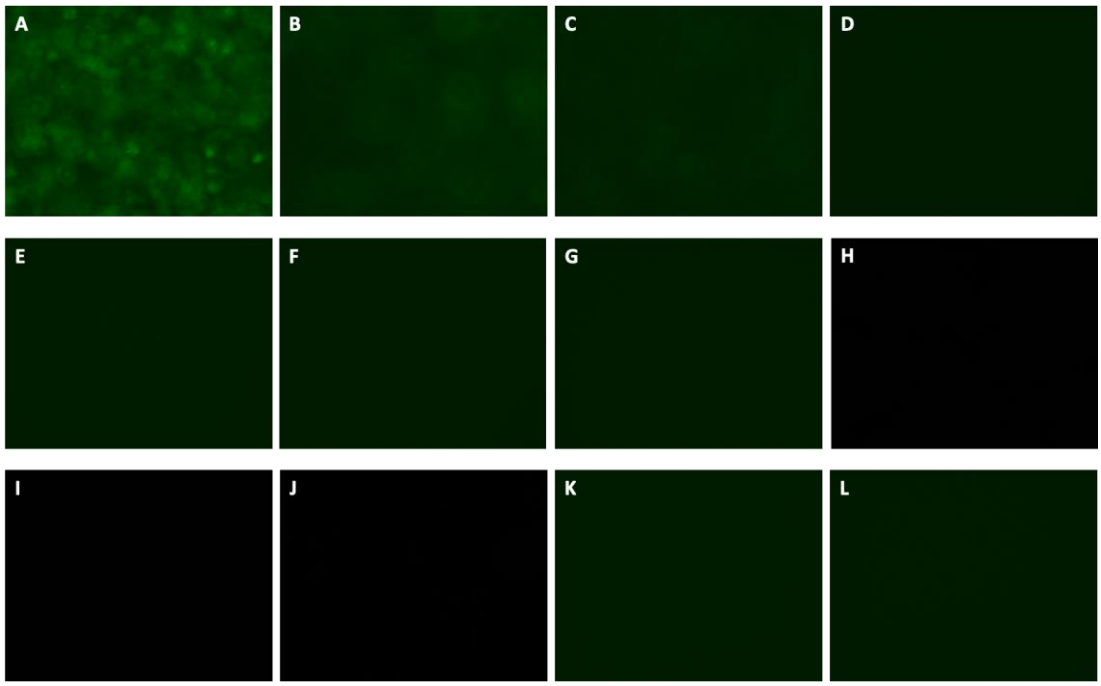


Figure 9.1: GFP fluorescence in Expi293 cells (3 mL expression) 48 hours post-transfection GFP fluorescence for truncated constructs: ***(A)*** *β1,* ***(B)*** *β5,* ***(C)*** *TGFβRI,* ***(D)*** *αv* ***(E)*** *LAP (L30-R278)* ***(F)*** *TGFβRII* ***(G)*** *β6 and full length constructs:* ***(H)*** *TGFβRI* ***(I)*** *β5* ***(J)*** *αv* ***(K)*** *LAP* ***(L)*** *β6.*

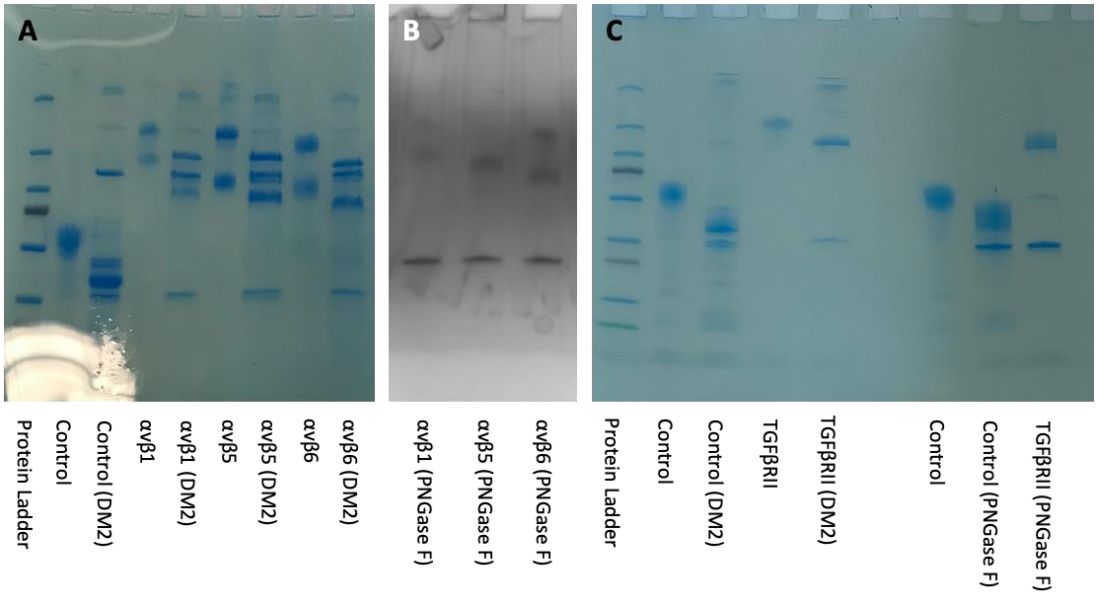


Figure 9.2: Shift in protein mobility by SDS-PAGE following deglycosylation
Protein migration pattern following ***(A)*** *full (N- and O- linked glycan removal) and* ***(B)*** *partial (N-linked glycan removal) deglycosylation of integrin proteins (αvβ1, αvβ5, αvβ6) and* ***(C)*** *the TGFβRII protein. Fetuin protein included as a positive control for deglycosylation under non-denaturing reaction conditions.*